Universidade de Aveiro
2012
Departamento de Electrónica, Telecomunicações e Informática

Alfredo Miguel Melo Matos

Privacidade em Redes de Próxima Geração

Privacy in Next Generation Networks

Universidade de Aveiro
2012
Departamento de Electrónica, Telecomunicações e Informática

Alfredo Miguel Melo Matos

Privacidade em Redes de Próxima Geração

Privacy in Next Generation Networks

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Informática, realizada sob a orientação científica do Prof. Doutor Rui Luís Andrade Aguiar, Professor Associado com Agregação do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro, e sob a co-orientação da Prof. Doutora Susana Barreto Sargento, Professora Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

Este trabalho foi realizado com o apoio de uma bolsa de Doutoramento, com referência SFRH/BD/35903/2007, financiada ao abrigo do programa POPH - QREN, - Formação Avançada,comparticipada pelo Fundo Social Europeu (FSE) e por fundos do Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) através da Fundação para a Ciência e Tecnologia (FCT).

Este trabalho é dedicado à minha família, pelo apoio incondicional e por acreditarem (sempre) em mim, e à Claudia, por ser o meu abrigo e por me ter dado o amor, a força, a coragem e o tempo que me faltavam mesmo quando eu não sabia que precisava...

Obrigado.

o júri

presidente

Prof. Doutor Fernando Manuel Bico Marques
Professor Catedrático da Universidade de Aveiro

Prof. Doutor Ian Brown
Senior Research Fellow, Oxford Internet Institute, University of Oxford

Prof. Doutor Paulo Veríssimo
Professor Catedrático da Faculdade de Ciências da Universdade de Lisboa

Prof. Doutor André Zúquete
Professor Auxiliar da Universidade de Aveiro

Prof. Doutor Rui Luís Andrade Aguiar
Professor Associado com Agregação da Universidade de Aveiro

Prof. Doutora Susana Barreto Sargento
Professora Auxiliar da Universidade de Aveiro

acknowledgments

I would like to start by thanking my supervisors, Rui Aguiar and Susana Sargento, for the inspiration, freedom and counseling, especially at those times where guidance was more on life than actual work.

I would like to thank João Girão (from NEC) whose friendship, motivation, support, and ideas never allowed me to stand still, and always encouraged me to do more and better.

I would also like to thank all of the wonderful people who participated in many of the ideas in this Thesis, especially the co-authors of several of the papers and ideas that are contained in this document, such as Rui Ferreira (IT), Rodolphe Marques (IT), Marc Barisch (University of Stuttgart), Ricardo Azevedo (Portugal Telecom Inovação), Amardeo Sarma (NEC), Frederik Armnecht (NEC), and all of those that are not mentioned that either through papers, projects or discussion had a directly influence on the ideas presented.

Finally, I would like to acknowledge all of my friends and colleagues at IT that over the years contributed with many ideas and endured through many discussions, which ultimately had an impact on the work.

Thank you all.

keywords privacy, networks, identifiers, protocols, layers, identity, architectures, threats.

abstract

In the modern society, communications and digital transactions are becoming the norm rather than the exception. As we allow networked computing devices into our every-day actions, we build a digital lifestyle where networks and devices enrich our interactions. However, as we move our information towards a connected digital environment, privacy becomes extremely important as most of our personal information can be found in the network. This is especially relevant as we design and adopt next generation networks that provide ubiquitous access to services and content, increasing the impact and pervasiveness of existing networks.

The environments that provide widespread connectivity and services usually rely on network protocols that have few privacy considerations, compromising user privacy. The presented work focuses on the network aspects of privacy, considering how network protocols threaten user privacy, especially on next generation networks scenarios. We target the identifiers that are present in each network protocol and support its designed function. By studying how the network identifiers can compromise user privacy, we explore how these threats can stem from the identifier itself and from relationships established between several protocol identifiers.

Following the study focused on identifiers, we show that privacy in the network can be explored along two dimensions: a vertical dimension that establishes privacy relationships across several layers and protocols, reaching the user, and a horizontal dimension that highlights the threats exposed by individual protocols, usually confined to a single layer. With these concepts, we outline an integrated perspective on privacy in the network, embracing both vertical and horizontal interactions of privacy. This approach enables the discussion of several mechanisms to address privacy threats on individual layers, leading to architectural instantiations focused on user privacy. We also show how the different dimensions of privacy can provide insight into the relationships that exist in a layered network stack, providing a potential path towards designing and implementing future privacy-aware network architectures

palavras-chave

privacidade, redes, identificadores, protocolos, camadas, identidade, arquitecturas, ameaças.

resumo

Na sociedade moderna, as comunicações e transacções digitais estão a tornar-se a regra e não a excepção. À medida que permitimos a intromissão de dispositivos electrónicos de rede no nosso quotidiano, vamos construíndo um estilo de vida digital onde redes e dispositivos enrirquecem as nossas interacções. Contudo, ao caminharmos para um ambiente digital em rede, a nossa privacidade vai-se revestindo de maior importância, pois a nossa informação pessoal passa a encontrar-se cada vez mais na rede. Isto torna-se particularmente relevante ao adoptarmos redes de próxima geração, que permitem acesso ubíquo a redes, serviços e conteúdos, aumentando o impacte e pervasividade das redes actuais.

Os ambientes onde a conectividade e os serviços se tornam uma constante, assentam em protocolos de rede que normalmente contemplam poucas considerações sobre privacidade, comprometendo desta forma o utlizador. O presente trabalho centra-se nos aspectos de privacidade que dizem respeito à rede devido à forma como os protocolos são utilizados nas diferentes camadas, e que resultando em ameaças à privacidade do utilizador. Abordamos especificamente os identificadores presentes nos protocolos de rede, e que são essenciais à sua função. Neste contexto exploramos a possibilidade destes identificadores comprometerem a privacidade do utilizador através da informação neles contida, bem como das relações que podem ser estabelecidas entre identificadores de diferentes protocolos.

Após este estudo centrado nos identificadores, mostramos como a privacidade em redes pode ser explorada ao longo de duas dimensões: uma dimensão que acentua as relações verticais de privacidade, cruzando vários protocolos até chegar ao utilizador, e uma dimensão horizontal que destaca as ameaças causadas por cada protocolo, de forma individual, normalmente limitadas a uma única camada. Através destes conceitos, mostramos uma visão integrada de privacidade em redes, abrangendo tanto as interacçoes de privacidade verticais como as horizontais. Esta visão permite discutir vários mecanismos para mitigar ameaças específicas a cada camada de rede, resultando em instânciações arquitecturais orientadas à privacidade do utilizador. Finalmente, mostramos como as diferentes dimensões de privacidade podem fornecer uma visão diferente sobre as relações estabelecidas na pilha protocolar que assenta em camadas, mostrando um caminho possível para o desenvolvimento de futuras arquitecturas de rede com suporte para privacidade.

Contents

List of Figures

List of Tables

Acronyms

A4C Authorization, Authentication, Accounting, Auditing and Charging

AAA Authentication, Authorization and Accounting

AID Association Identifier

AP Access Point

AttS Attribute Server

AuthNP Authentication Provider

BE Base Exchange

BU Binding Update

CN Correspondent Node

CoA Care-of Address

Daidalos Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimized personal Services

EPP Entity Profile Part

EPVH Entity Profile View Handler

EPV Entity Profile View

EP Entity Profile

FEPV Filtered Entity Profile View

FIDIS Future of Identity in the Information Society

FPD Freiburg Privacy Diamond

FQDN Fully Qualified Domain Name

GEL Generic Execution Layer

HA Home Agent

HIP Host Identity Protocol

HIT Host Identity Tag

HI Host Identity

HMIPv6 Hierarchical Mobile IPv6

HMN HIP Mobile Node

HoA Home Address

IaaS Infrastructure-as-a-Service

ID-FF Identiy Federation Framework

ID-Pointer Identity Pointer

IdAgg Identity Aggregator

IDBroker Identity Broker

IDManager Identity Manager

IdM Identity Management

IdP Identity Provider
IETF Internet Engineering Task Force
IS Information Set
LMA Local Mobility Anchor
MAG Mobile Access Gateway
MDP Mobility Decision Point
MEP Mobility Enforcement Point
MInP Mobility Information Point
MIPv6 Mobile IPv6
NAT Network Address Translation
NAV Network Allocation Vector
NDP Neighbor Discovery Protocol
NGN Next Generation Network
OECD Organization for Economic Co-operation and Development
PaaS Platform-as-a-Service
PANA Protocol for Carrying Authentication for Network Access
PET Privacy Enhancing Technology
PMIPv6 Proxy Mobile IPv6
PRIME Privacy and Identity Management for Europe
PRIVED Privacy Event Driven Model
QoS Quality of Service
RTT Round Trip Time
RVA Rendezvous Agent
RVS Rendezvous Server
SaaS Software-as-a-Service
SAML Security Assertion Markup Language
SIM Subscriber Identity Module
SIP Session Initiation Protocol
SLAAC Stateless Address Auto-Configuration
SP Service Provider
SSO Single Sign-On
SWIFT Secure Widespread Identities for Federated Telecommunications
TIM Traffic Indication Map
URI Uniform Resource Identifier
URL Uniform Resource Locator
VDM Virtual Device Manager
VID Virtual Identity
VIDID VID IDentifier
VIF Virtual Interface
VNS Virtual Network Stacks
WEP Wired Equivalent Privacy

Chapter 1

Introduction

> The true voyage of discovery lies not in seeking new landscapes, but in having new eyes.
>
> **Marcel Proust**

As the first chapter in the Thesis, the introduction sets the background for the topic of Privacy in Next Generation Networks, along with the motivation for pursuing a PhD on network privacy. It presents the hypothesis along with the goals that guided the work evolution, followed by the contributions that resulted from the concepts explored. Finally, it presents the overall structure of the document.

1.1 Background

In the past years I have been engaged in privacy related matters, dealing with networking issues. This is a very hard topic to synthesize. Most people cannot clearly state what privacy is. That is not uncommon, given that even the well-know scholars in different research fields cannot agree on what privacy actually is [136]. This confusion spills into areas in which privacy matters, such as computer networks, where the lack of effective privacy suggests that a deeper understanding is required. But, the absence of a common agreement on what privacy is should not halt research efforts in computer networks, given that users still feel threatened and violated when confronted with privacy compromising situations, regardless of scholarly definitions. Coming from a network background, it is easy to understand why considering privacy, as a user or engineer, can be a daunting task. The confusion around privacy concepts, along with their lack of systemization on the network, make them hard to address.

Privacy has been engraved as a fundamental Human Right into the European Convention on Human Rights [50], making it part of every country's law, carved into the foundations of our society. In this context, it is only natural that we expect privacy to be preserved in most aspects of our life. But, what if suddenly privacy begins to sink into a quicksand of different technologies? Over the past years we have witnessed the appearance of pervasive technologies, from mobile phones to micro-sized digital cameras, creating an ever present Internet. And so far, we have welcomed such intrusion in our daily life, as part of an enhancement of our daily activities, allowing us to do more, better and faster. We have been slowly and steadily becoming more dependent on technologies that we do not fully understand, up to a point that goes beyond what was originally intended for most of them. Only recently we became aware that letting digital technologies into our inner circle will come at a cost to our privacy. As we navigate towards future networks, technology adoption is rapidly increasing. What does this mean for privacy? Can we counteract the adverse effects of technology on our individual privacy? These questions have driven the concerns outlined in this Thesis.

The presented work aims at improving user privacy in a network environment. It does so by trying to identify network specific mechanisms that jeopardize privacy, and by questioning the assumptions upon which they work. Most protocols on the network resort to identification mechanisms that are associated with the layer on which they work. This leads to several identified entities (e.g. users or hosts) within the network space. These identifiers are fundamental to the network operations, specially in packet-based networks where every packet or frame must be properly identified. This introduces a set of privacy threats around the identification of the user, and associated properties (e.g. location), that can stem from observing ordinary network operations. While there are privacy enhancing features for specific protocols, there is a lack of a common vision on the privacy problems that can grow from normal protocol operation, especially as we step into an ever growing and pervasive digital world. The presented Thesis focuses on the assumption that next generation heterogeneous networks will amplify the privacy threats already present on the network, by reaching more users through ubiquitous and mobile technologies. By attempting to define these privacy threats, we propose several network oriented solutions that try to clarify meaning of privacy on the network, and how it can be addressed. To achieve this, we look at several privacy models, and try to propose meaningful concepts that fit the network aspects of privacy. Starting from these models, we tackle network specific problems that can stem from network protocol properties.

Beyond the technical aspects, the guiding principle behind the presented contributions is

to generally improve the privacy conditions for the end user, attempting to isolate and fix systematic problems. This frames the underlying reasoning that drives our proposals, where we strive to be as pragmatic as possible, identifying and clarifying privacy threats, showing their relationships to network mechanisms, and providing adequate solutions where possible.

1.2 Motivation

Privacy has always been a concern in many ways for the modern society. Appearing in Article 8 [143] of the Human Rights Declaration, it has been carried over to the communications and digital world, through Organization for Economic Co-operation and Development (OECD) Directives [47] and European Union directives [29, 30, 31, 32]. As we border the digital frontier, we are presented with a new array of interactions that are yet to be fully understood. This venture into new environments makes the study of privacy a relevant topic in networking. Society has already demonstrated that it does not possess the tools to understand the impacts of pervasive technologies [81, 118, 27], and has clearly shown not to grasp the full impact of privacy. Technology tries to keep up with a growing list of requirements that appear by adding hot-fixes and extending existing solutions. The common denominator is that most people do not understand the impacts on privacy, nor to what extent is their privacy already in jeopardy, given that there is no to clear way of measuring the consequences of adopted habits and technologies. This lack of means for assessing privacy, coupled with the fact that privacy requires preemptive measures that are only observed when they are negated, creates a delayed effect of privacy violations: the user usually discovers that there is a privacy issue only after his privacy has been breached.

A quick-paced technological evolution is bringing forth the much announced digital revolution that impacts not only networking but life, bringing overwhelming socio-economical changes. It is undeniable that these phenomena are tightly coupled with the offspring of future and next generation networks, accompanied by a changing life style, ever more digital. It is important to discuss in this context what are the major evolutions that contribute to a new paradigm in computing and networking, and more importantly how they relate to our privacy. With micro-electronics generally advancing at Moore's law, devices have become small, portable, and at times near invisible. These devices (first laptops, then mobile phones, and now smart phones, MP3 players, gaming consoles, sensors and all sorts of electronic gadgets) are small enough to become part of our daily apparel. We carry these devices everywhere, for a myriad of purposes, that go well beyond what was initially expected [138].

As we discuss these devices, we cannot neglect to mention the accompanying networking technologies that converted this trend into a connected lifestyle. Mobile telephony has become essential to our daily functions, powered by technologies like 2G and 3G. Most people have one or more cell phones, and it possible to observe a widespread dependency on such devices that have become part of daily routine. Moreover, we are extending this dependency to several wireless technologies, especially WiFi. We expect to have a WiFi connection (or any combination of 3G, LTE or WiMax) that brings us on-demand and on-the-go digital content. Such wireless technologies are becoming omni-present, and we expect nothing less.

However, it seems that we are taking (too) large steps given the fact we do not quite understand most of the implications of this connected lifestyle. The commonly neglected counterpart of this growing scenario is that, in order to function, the network uses several protocols that relate to users and devices. These protocols carry identifiers that can identify

a user or terminal, as well as information that can compromise the privacy (e.g. location aware applications are becoming commonplace). These different network protocols and their layered interactions with the network can provide mechanisms to breach user privacy, allowing user recognition, tracking or geographical pinpointing, among other threats. And while this is true for existing network paradigms, the pervasiveness of new electronic devices (all them with network connectivity) connected to a Next Generation Network (NGN), raises issues that have no straightforward solution.

By allowing digital technology to permeate our life, we are potentially enabling a Big Brother society that can reach beyond its intended purpose, violating our fundamental privacy rights. Privacy concerns in network environments can stem from the widespread incorporation of computing and mobile devices into the lifestyle of the average person. By linking these devices into wireless technologies, we enable several privacy threats, which are not being dealt with properly. The privacy aspects of such pervasive networks were well predicted by an early paper by David Chaum, where he speculates about a future Big Brother scenario surrounding digital environments [21]. What is expected over the coming years is that networks and devices become even more pervasive, and embedded in ordinary things, increasing the privacy threats as we move to an Internet of Things [106].

The most immediate threats are reinforced by the evolving user requirements of wanting to be "always best connected", a paradigm that is gaining traction. To support this paradigm, we have seen the appearance of Next Generation Networks, which not only define a set of entities, but also a blueprint for future networks that enable ubiquitous environments. We must understand what an NGN is, and what aspects need to be retained from this keyword along the lines of the privacy threats.

1.2.1 Next Generation Networks

There is a strong emphasis on NGN in this work. However, this acronym is becoming a synonym for several concepts, and not all of them are aligned. It is interesting to explore both the colloquial and formal notions of the term NGN.

NGN, in its original form, is a standard proposed by th ITU-T, following recommendation Y.2001 [73]. According to the ITU-T, and paraphrasing the overview document, an NGN is "a packet-based network able to provide Telecommunication Services to users and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent of the underlying transport-related technologies. It enables unfettered access for users to networks and to competing service providers and services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users". This definition provides a reference model, supported by several recommendations. Of those, the following principles better characterize an NGN:

- A packet-based network, IP in nature;
- A means to access services over multiple technologies, therefore considering a heterogeneous environment, as stated by the support for multiple last mile technologies;
- Lastly, a network which supports generalized mobility, and presents ubiquitous availability,

These three aspects are, in most generalized approaches, the terms that define an NGN, along with the common ground when discussing NGN concepts. Instead of dwelling on the

architectural details of the Y.2001 recommendations, in this Thesis we draw upon the larger definition of NGN, to lay ground to the default application scenario of privacy considerations. While the two first points, enumerated above, outline a technology scenario or base expectations, the last point, where we claim ubiquitous network access for mobile devices, can be seen as the fundamental aspect that turns privacy into a vital network feature. However, there are key principles in the NGN definition that provide guidance for future solutions and are worth discussing with greater detail. Of particular importance is the NGN reference model, outlined in Y.2001, which introduces a clear separation between the management, control and user plane. From this structure stems the functional split between control and bearer functions. If properly and consistently applied, this property alone can lead to new paradigms in networking, drifting away from current pitfalls in existing solutions.

It is also important to single out the features that can provide value for both security and privacy, from the plethora of functions and requirements imposed by the ITU-T "formal" NGN architecture. The need for identification and authentication mechanisms are among the most noteworthy requirements to build networks where access control and authorization functions are common. It is clear that NGN networks bring forth several new interactions, specially considering its ubiquitous intent and mobility support.

An important conclusion to extract from NGN recommendations is the scenarios that are presented in terms of network interactions: they show the pervasive scenarios that can amplify already existing privacy problems on the network. ITU-T specifications such as Y.2001 [73], should be considered only as a guiding framework for heterogeneous mobile networks. Applying the proposals mentioned in this document (or similar) to specific NGN interfaces and standards should be considered in future work.

We focus on the broader concept of NGN, sometimes referred to as 4G network scenarios, or even Heterogeneous Networks. These 4G network scenarios [1, 3, 4] typically encompass several administrative (federated) domains, handled by different providers. They provide a heterogeneous environment powered by different technologies, such as WiFi, WiMax, UMTS or DVB, seamlessly integrated on the architecture. These networks, from where we derive a simplified architecture (represented in Fig. 1.1), normally provide a common controlled environment regarding resources and authorization. Functional boxes such as Bandwidth Brokers (in NGN term, the Policy and Charging Rules Function or PCRF) control the network's resources and distribution, facilitating the optimum distribution of resources among the registered users. To sustain the controlled environment, Authentication, Authorization and Accounting (AAA) servers, or their augmented version also providing AAA plus Auditing and and Charging (A4C), take care of terminal authentication and authorization, providing a secure environment for network usage. In the NGN architecture these functions are represented by the Network Attachment Control Functions (NACF) and Resource and Admission Control Functions (RACF). Session Initiation Protocol (SIP) [125] proxies are in place for supporting multimedia applications (in NGN terms the Call Session Control Function - CSCF), while mobility anchors enable the support of mobility between the several types of networks (which can be roughly mapped to the Subscriber Location Function - SLF). A service pool, which defines the application layer, is composed by a set of application servers that can either be located locally or remotely (e.g. Internet).

From this model we highlight the support for mobile devices and session continuity using either Mobile IPv6 (MIPv6) [77] or SIP [125], which enable the "always best connected" paradigm over packet-switched networks. It is this ubiquitous support that magnifies the network based privacy threats towards the user.

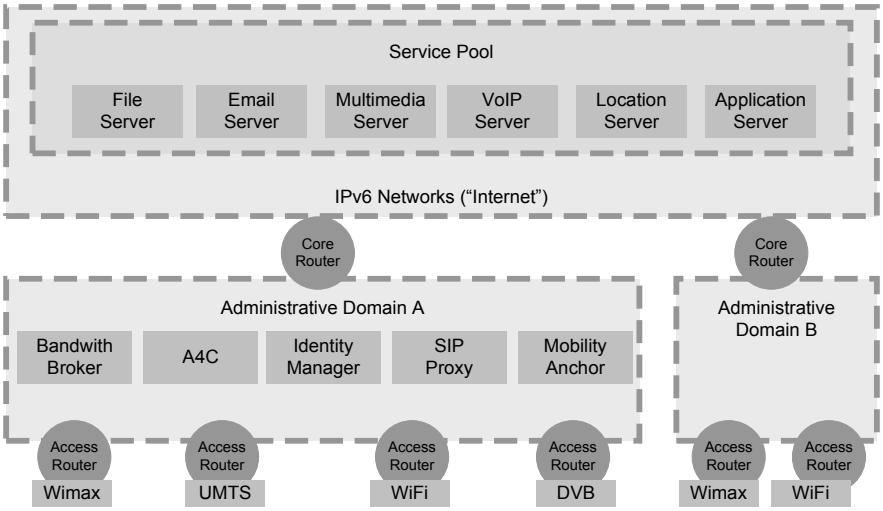


Figure 1.1: Heterogeneous Network Architecture Model.

1.2.2 Privacy in NGN

Due to the foreseen (and to some extent already existing) ubiquitous environments, it is safe to acknowledge that NGN can increase privacy threats [119], magnifying already existent security threats. In the proposed environment, we can imagine the user resorting to several devices and services. More importantly, it is a fact that we now see so many services appearing with a web-based component (a cloud service), which leads us to believe that the user's privacy is more endangered: by storing (even more) personal information on remote services, the user is opening the door to more privacy threats. When this model is combined with an NGN, it is possible to harvest levels of information which were not previously plausible. Taking NGN as a general-purpose mobility aware heterogeneous IP network, we can define the boundaries within which the user's privacy needs to be analyzed, understanding the threats that linger over the user when he interacts with his devices in this type of environment.

Several threats stem from the fact that the network mechanisms must always identify the user in some way. Either through addresses, a user ID, or any similar identification mechanisms, the user becomes traceable through his digital fingerprint[1]. It becomes possible to know what the user does, and where he does it from. This becomes a major peril when we tend to do almost everything on-line. Through network traffic analysis it is possible to determine user actions. It is also possible to determine the user's location, by converting his IP addresses to geographical information. Combining these two issues leads to another threat only possible in NGN environments: the provided mobility support allows tracking the user as he moves through different networks, converting his present location to an accurate geographical position. Most of these topics are covered in-depth in the following chapters.

We have already been confronted with several privacy risks in such environments, especially if we consider the individual threats on the each layer of the network stack. But we lack a *de facto* approach to protect the user against the multitude of privacy attackers that already exist today. To define such an approach, we must carefully understand the existing

[1]This refers to the identifiers that uniquely reference a particular user in the network or in a service, therefore serving as a unique proof that a particular user or device was involved in a communication or transaction.

privacy threats, in the network as a whole or in each individual layer, and how these can be mitigated. This must be done considering that NGNs are becoming the heart of the future communication infrastructure. By serving as the primary realization of the aforementioned digital lifestyle, it is clear that we must understand how privacy will function in this environment, and more importantly what can be done to improve it. Furthermore, we must see how the problems pertaining to identifiers on the different protocol layers, which already affect existing networks, play a role in NGN privacy thus showing how the privacy threats increase in ubiquitous networking environments.

1.3 Hypothesis and Objectives

It has become clear that the user faces an array of privacy threats, stemming from different situations and adversaries. In an NGN environment, the network contains several privacy jeopardizing mechanisms. Several of these threats are a direct consequence of the original design requirements of the different protocols: when created, most protocols focused on simplicity and pragmatism to accomplish their task, focusing on connecting hosts on the network rather than being the foundation of an all-purpose network, ubiquitous to our modern lifestyle.

Therefore, a question we must ask gradually becomes apparent, in the light of the outlined motivation and network scenario: How do network mechanisms, especially those inherent to heterogeneous mobile networks, compromise user privacy? This also entails a follow up question regarding the solution space: is it possible to mitigate the problems caused by network mechanisms on individual user privacy, while retaining the expected behavior of heterogeneous mobile networks?

The answer to these questions takes the form of the hypothesis of this thesis: the network identifiers that are at the core of network protocols, are the cause of several threats on user privacy. Such identifiers, which are used either for identification, routing or any other purpose, introduce mechanisms that enable an attacker to identify and track a user, as well as collecting (personal) information about the user and his activities. We believe this to be especially true when considering the interactions between different network layers and protocols: when linked together, and related to the user, identifiers can provide unique ways of undermining privacy, resulting in systematic privacy breaches.

To complement the hypothesis concerning identifiers as a major threat, we postulate how this can be mitigated: to handle the identifier related threats it is necessary to break the relationships between identifiers and the user. In the same context, it is also important to break the relationships between the different identifiers on the network stack, because their correlation can lead to privacy threats towards the user.

In summary, the hypothesis defines that identifiers are one of the main causes for privacy threats in the network, and that these threats can only be thwarted by breaking the relationships between user and network identifiers.

The formalization of this hypothesis can unveil several problems when subjected to a deeper analysis. The first question concerns the definition of user privacy on the network, which is a kernel issue to all network related discussions: it is important to define individual user privacy on the network, in order to understand the threats and how to mitigate them. Accordingly, we must realize what are the specific mechanisms that threaten privacy, and how they operate. This is crucial to understand, given that we must understand how these threats can be mitigated, specially the context of the postulated hypothesis. Finally, and as

a guiding principle, all these questions must be evaluated under the light of heterogeneous mobile networks, considering the expected ubiquitous network and service access.

These initial questions imply that there are many dimensions to privacy in a network environment, that cannot be properly handled unless considering a holistic approach to network privacy. We should carefully examine if the network, through its protocols and mechanisms, can contribute to the overall user privacy, or just simply undermines it. This suggest a duality: using privacy enhancing technologies on the network actually increases user privacy, but simultaneously, these solutions can be compromised through the layered composition of the network, especially considering the characteristics of each individual protocol. This flaw, which could be the result of privacy breach at any layer in the protocol stack, must be handled with a combined view of the network, encompassing user information and privacy features.

The stated hypothesis can trigger several questions that must be answered, in order to understand how privacy is upheld or forfeited in the network. If we analyze these different issues, it is possible to draw a set of objectives that will enable us to prove or disprove the hypothesis. The starting point should be to determine what models exist that accurately describe privacy in a network context, and how effective they are in exposing threats on next generation networks. This will inevitably lead to understanding the relationships between the network stack and privacy. The idea behind such a course of action is ultimately to determine how privacy can be assured, or even disrupted. The importance of this step lies in the fact that it is only possible to propose novel solutions that protect user privacy, in existing or future scenarios, if the relationships between network and privacy are well understood. This implies that we must clearly identify the threats against user privacy, originating from the current network stack or from next generation networks, which should only be possible through an in-depth study of network mechanisms.

After the privacy risks stemming from the network are clearly identified, we must strive to propose solutions that handle the privacy threats at different layers. We propose to achieve this with an integrated vision of the network stack and architecture, while still keeping in mind that privacy is not the only vector that exists in the network: we must make sure that proposed solutions do not compromise network flexibility, user flexibility and the overall pervasive environment of NGN while still retaining a high degree of privacy. Ideally, the proposed solutions should articulate a user-centric privacy environment that respects the network's layered architecture and uses it in favor of privacy. Therefore, it is part of our hypothesis that the proposed models and solutions can provide a framework where the layered design of the network is properly explored. In this approach, it is especially important to consider how identifier relationships can lead to privacy solutions, starting from the idea that the identifier relationships that threaten privacy must be broken.

However, as we evaluate and propose new models and solutions, we should try to understand what are the guiding principles of such threats and solutions, so that we can propose concepts that go beyond NGN, and carry us onto future Internet designs, thus improving privacy in the long run.

1.4 Contributions

At first glance, user privacy on the network appears to be a contained topic. However, this Thesis shows that, in fact, it involves a multidisciplinary approach, even on the network side, to provide effective privacy to the end user. Therefore, the scientific advances presented in this

Thesis are not confined to a single issue or topic, but rather span across different proposals, layers, and even architectures and paradigms. One of the most important contributions presented here is an integrated view of privacy on the network, supported by the definition of vertical and horizontal dimensions of privacy in the network.

Beyond the privacy overview presented in Chap. 2, the contributions begin to take form in Chap. 3. In this chapter we propose a privacy model, published in the PriMo Privacy Workshop [104], that draws on existing literature, outlining a new way to address privacy in the network. The model is built upon network-oriented privacy definitions, also presented in the chapter, and its key contribution is the articulation of the vertical and horizontal privacy threats on the network. The network application of the model resulted in a network-based threat analysis (Sec. 3.4) and the conceptual guidelines towards addressing the identified threats (Sec. 3.5). The potential mitigation approaches set the pace for the following chapters which led to two complementary approaches: exploring the vertical relationships of privacy across layers, and addressing the threats stemming from individual network protocols, shifting towards a horizontal perspective on privacy, i.e. within the same layer.

The vertical approach to privacy in the network is mostly discussed in Chap. 4 where we proposed a Virtual Identity (VID) Framework, resulting in a journal contribution entitled "VID Framework for Telecom Operators" [129], published in Springer Wireless Personal Communications, which presents an Identity based framework modeling user interactions with the network, perceived as virtual identities, that affect different levels in the network. The VID solution suggests the use of pseudonymity on the network, to support different personae towards the network, leading to a network based pseudonym solution, "Preserving privacy in mobile environments" [97] presented at *IEEE Globecom'07*. This work further suggested that we needed greater insight on network pseudonymity, which was later presented in a journal article entitled "Virtual Network Stacks: From Theory to Practice" [95], accepted in the Wiley Security and Network Communications Magazine.

The VID approach work also transpired into the EU project IST "Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimized personal Services (Daidalos) II" [71] and later IST "Secure Widespread Identities for Federated Telecommunications (SWIFT)" [72]. Also, the tight involvement in the Daidalos II project demonstrator, which used an instantiation of the Virtual Network Stacks (VNS) concepts, resulted in a contribution regarding testbeds on the Open NGN and IMS Testbeds Workshop hosted at *ISCT Tridentcom'09*, entitled "Deploying and testing an NGN testbed: IST Daidalos testbed".

To explore the horizontal dynamics of privacy, focusing on each individual layer mostly relating to the work in Chap. 5, we started by presenting a link layer solution presented at *IEEE Infocom'07*, entitled "Who said that? Privacy at link layer" [9]. The results were applied in a 802.11 environment, and consolidated in a journal publication on the IEEE Wireless Communications Magazine, as "Towards dependable networking: Secure location and privacy at link layer" [96]. The work presented in these papers was subjected to intellectual property evaluation, resulting in two distinct patents: the first, "Method for Protecting Location Information in Wireless Communication Networks" [8], and the second "Method for establishing a secret key between two nodes in a communication network" [7], both under the network security topic. On the network layer we proposed "Waypoint Routing: Privacy as a Service" [105], presented at *IEEE Globecom 2011*, dealing with location and identification issues using a Chaum Mix [22] based approach.

From the bi-dimensional approach to privacy, several architectural issues were identified

and addressed. The relationship between privacy and Identity Management (IdM) led to several complementary proposals that use identity concepts as a tool towards solving privacy and network issues, related to the concepts presented in Chap. 6. We proposed integrating a Security Assertion Markup Language (SAML) based IdM system with the VNS approach in "Cross Layer Privacy Support for Identity Management" [93]. The continuous exploration of the vertical role assumed by identity, led to "Identity Driven Mobility Architecture" [98] that shows how identity can be used to drive mobility aspects. Contributions developed in the scope of the IST SWIFT [72] project, led to a joint publication entitled "Security and privacy enablers for future identity management systems" [15]. These three mentioned publication efforts, with close relationship to the SWIFT project, were publish in the Future Network and Mobile Summit in 2010.

Additionally, using a locator/identifier split on the network layer to explore the privacy approaches outlined before, we proposed a solution using the Host Identity Protocol, published at the Mobiarch'06 Workshop, held at IEEE Globecom'06, under the title "HIP Location Privacy Framework" [102]. This work encouraged a standard track submission resulting in contributions to the Internet Engineering Task Force (IETF), more concretely the Internet Research Task Force which focuses on research and novel approaches, with two versions of a IETF draft entitled "HIP Privacy Extensions" [100, 101]. Also, the above efforts were consolidated in a national journal publication, "Location Privacy Extensions for the Host Identity Protocol" [99]. Further research into the Host Identity Protocol (HIP) protocol led to a tight integration with a SAML IdM architecture, resulting in a publication at *IEEE ISCC'09* with the title "Integrating user identity management systems with the host identity protocol" [14]. Finally, we proposed "Embedding identity in mobile environments" [103], published in the Mobiarch'07 workshop held at *IEEE Sigcomm'07*, which promotes tight network integration of identity and network protocols. Also, based on the exploration of how local mobility solutions can affect privacy, and in the scope of localized mobility aspects explored within the Daidalos project, an IETF draft contribution was presented covering a "Problem Statement on Common Interfaces for Local Mobility Management"[28].

As part of the evolutions originating from the privacy principles applied in the aforementioned solutions, in Chap. 7 we proposed solutions that use identity concepts and apply them in future scenarios and architecture. "Mobility aware paths: The identity connection" [94], published on the International Symposium on Wireless Personal Multimedia Communications (WPMC), outlines a novel way of integrating identity into future communication paradigms such as the concept of a path. We also published a journal article entitled "User Centric Community Clouds" [16], on the Springer Wireless Personal Communications (WPC) Magazine, that deals with privacy and identity issues in upcoming cloud computing environments.

For convenience, we summarize the contributions in Table 1.1, organized by type and date, so that outputs can be better understood.

1.5 Structure

The main body of work presented in this Thesis is structured around the concepts of the vertical and horizontal dimensions of privacy in the network, followed by the application of those same concepts on the network. We present a structure consisting of an introduction and conclusion, a related work chapter, and four central chapters incorporating the most novel contributions.

Type	Year	Title	Venue
Workshop	2006	HIP location privacy framework	MobiArch @ IEEE Globecom
Workshop	2007	Embedding identity in mobile environments	MobiArch @ IEEE Sigcomm
Workshop	2009	Deploying and testing a NGN testbed	ONIT @ TRIDENTCOM
Workshop	2011	A Privacy Model for Heterogeneous Mobile Networks	PriMo
Conference	2007	Who said that? Privacy at Link Layer	IEEE Infocom
Conference	2007	Preserving privacy in mobile environments	IEEE Globecom
Conference	2008	Mobility aware paths: The identity connection	WPMC
Conference	2009	Integrating User IdM Systems with HIP	IEEE ISCC
Conference	2010	Security and Privacy enablers for future IdM systems	FN'MS
Conference	2010	Cross layer privacy support for identity management	FN'MS
Conference	2010	Identity driven mobility architecture	FN'MS
Conference	2011	Waypoint Routing: Privacy as a Service	Globecom
Journal	2008	Towards dependable networking: Secure location and privacy at link layer	IEEE Wireless Communications Magazine
Journal	2008	Virtual identity framework for telecom infrastructures	Wireless Personal Communications (Springer)
Journal	2007	Location privacy extensions for HIP	Revista DETUA
Journal	2011	User Centric Community Clouds	Springer Wireless Personal Communication
Journal	2011	Virtual Network Stacks: From Theory to Practice	Wiley Security and Communication Networks
Draft	2007	P.S. on Common Interfaces for Local Mob. Mgmt.	IETF
Draft	2006	Hip privacy extensions - version 01 (revised)	IETF
Draft	2005	Hip privacy extensions - version 00	IETF
Patent	2008	Method for establishing a secret key between two nodes in a communication network	European Patent Office
Patent	2008	Method for protecting Location Information in Wireless Communication Networks	European Patent Office

Table 1.1: Publication contributions summary of the work presented in the Thesis.

The two initial chapters, Chap. 1 and Chap. 2 frame the content and topics presented in the Thesis. The Introduction presents the background and motivation, setting the stage for the hypothesis and goals of the Thesis. It also highlights the contributions stemming from the presented proposals. In Chap. 2 we present an overview on privacy, showing the current trends on privacy topics that relate to the network, and a survey of the most important solutions that either relate to network privacy, or provided key ideas for the work discussed in the following chapters. It presents a generic privacy discussion as well as a brief overview of the the legal boundaries defined within the European Union, establishing what is the current play-field for any and every privacy considerations. It is followed by the most well-known privacy models that handle network privacy, or that relate to it. It also shows important trends on user-centric privacy issues embodied by IdM, comprising projects, initiatives and architectures, along with the most important privacy solutions on different layers of the network stack, especially Link and network layer oriented.

Chap. 3 provides an overview of the difficulties that compose privacy on the network stack, starting with lexicon issues. It proposes clear definitions, building a consistent privacy terminology. It then goes on to provide a novel model for privacy, focusing on network issues. Feeding on the problem description, this chapter presents the underlying theoretical cornerstones that serve as the base attacker model for the privacy concepts on the network. The chapter closes with the application of such model on the network stack, highlighting both threats and possible solutions. It focuses on linking and correlation with a strong emphasis on a pragmatic view of the outlined privacy threats, along with possible solution spaces for them, mainly scoped around network identifiers. This chapter also provides the first introduction

to the vertical and horizontal dimensions of privacy in the network.

Closely following defined model, Chap. 4 presents a vertical approach to privacy. It introduces the concept of Virtual Identity, and promptly inserts it in a usable and tangible framework. The framework abides by the notion of partial user identities, stressing the notion of pseudonyms towards the network. With a clear focus on privacy, it attempts to provide personalization towards services, without compromising the user and pioneering the use of cross-layer IdM. The chapter finishes by evaluating network pseudonymity, instantiating it within the Virtual Network Stacks proposal, using the metaphor of user-centric virtual devices.

Exploring the second (horizontal) dimension of privacy, and according to the presented model, we dive into protocol specific aspects of privacy. Because the vertical integration provides a starting point for horizontal separation, through identity and pseudonymity, we can focus on each individual layer in Chap. 5, following a structured approach to privacy: we first investigate and propose new link layer privacy solutions, followed by the same approach towards network layer privacy issues, and finally focusing on targeted application layer issues. The proposed link and network layer solutions focus mostly on identification and location issues. The link layer section proposes a novel communication model that defines the recipients of messages based on cryptographic properties rather then on source and destination addresses. On the network layer, we reuse Chaum Mix concepts to provide a light hop-by-hop routing solution resorting to encrypted addresses. We finish the chapter by evaluating the feasibility of an application layer solution space that supports the privacy concepts presented up until this point, focusing on SAML based IdM technologies.

During chapters 4 and 5, several concepts frequently reappeared, particularly identity, hinting at more generic approaches for privacy-aware solutions. In Chap. 6 we expand identity concepts to become more than a vertical information layer: we propose identity as an architectural driver for privacy. We expose the different identity-centric architectural drivers, and provide instantiation architectures, using identity as the driver in different scopes. We propose a tight integration with SAML and the vertical concepts of virtual identities, particular pseudonymity and Virtual Network Stacks. We continue by proposing an architecture that splits actuation and control, with a heavy focus on mobility in NGN and IdM integration. This is followed by a privacy proposal using a locator identifier split concept, focusing on identity-centric protocols and location. Finally, we present an architecture that incorporates identity into the different layers, using identity references, closing the architectural instantiations, coming full circle to the initial proposals revolving around the vertical aspects of identity enabled privacy concepts.

As the final chapter, Chap. 7 summarizes the entire findings of the Thesis in the context of user privacy and presents future impacts of the proposed solutions and concepts through a discussion on some of the insights gathered throughout the research process detailed in this document. Beyond the results and achievements, it takes us on a journey: first through the immediate future work on privacy that can directly complement the Thesis; and second, it takes us out of the privacy conform zone and into future and evolving paradigms, highlighting how some of the presented concepts can be generalized towards Future Internet or even Cloud Computing, thus closing the Thesis.

Chapter 2

An Overview on Privacy

> Privacy is a concept in disarray.
>
> **Daniel J. Solove** *in Understanding Privacy*

This chapter presents an overview on privacy concepts, as context to understand privacy in the network. Starting on privacy definitions presented by several authors, we present a three-fold overview based on social, legal and technical aspects of privacy. Emphasizing the last point, to enable a pragmatic approach to privacy solutions on the network, we analyze several privacy models that are relevant to the work presented in later chapters, as well as technical solutions that handle several identified network threats.

2.1 Introduction

As the epigraph suggests, privacy can be a very confusing topic. This is especially true if we attempt to cover different privacy vectors. To properly study privacy without being completely overwhelmed, it is important to follow a structured approach; it is even more important to be prepared to absorb different definitions for seemingly similar concepts. We choose to first look at privacy as a whole, through its societal definitions, in order to understand what is intended to be. This involves looking momentarily at the most philosophical and legal aspects, which usually tend to go hand-in-hand, showing what privacy really means for a common user. However, we deal with digital environments, which have their own set of rules, expectations and possibilities. We briefly show the legal requirements and framework for data protection directives, particular in a European context, delimiting the network issues that must be addressed.

To start dealing with privacy on the network, we provide an overview of privacy models that deal with network related issues or provide interesting features, foundations or concepts that we can draw upon. While this allows us to better frame privacy in a network environment, we must then jump onto specific technologies, starting with Identity Management. IdM is currently one of the most promising trends in privacy related technology towards user information and service interaction. This is however, an application layer solution in most cases, and we must deal with lower network aspects of privacy. To do this, we survey the most interesting privacy solutions that deal with link layer privacy, network privacy, and later, mobility related aspects, all of which are relevant in an NGN scenario. These technologies will provide the starting point upon which to enhance the privacy landscape.

To tackle the above proposed concepts, we look into privacy in a social and legal context in Sec. 2.2, followed by a review of existing privacy models in Sec. 2.3. Upon dealing with the most theoretical-related aspects of privacy, we look at IdM concepts in Sec. 2.4, discussing recent trends and framing user privacy in connected environments. We later provide a structured overview of network related privacy technologies, focusing on link and network layer issues, in Sec. 2.5, and even on mobility aspects. We finish in Sec. 2.6 by providing a summary of what was concluded from the discussed trends and technologies.

2.2 Privacy

Privacy is a confusing concept, with multiple implications, often social and legal. The first place we look at when we talk about privacy is ourselves and others, which can be considered the social component of privacy. It is undeniable that privacy is a social phenomenon, only relevant within, and between, social groups. It also varies from person to person [2], meaning that there is a personal value of privacy. Right from the start, we become confused with the different personal aspects of privacy. And this is not uncommon. Quite often, we find privacy texts starting with a section entitled on "understanding privacy", or "defining privacy". It has been the tittle of books [137], chapters or sections, and has a place in almost every document on the subject, and this is no different. This is the typical path, followed by most privacy-centric documents, starting by term clarification, understanding, and later on, an acknowledgment of multidisciplinary concepts. This shows that we must understand privacy as a multi-disciplinary concept. In "Privacy and Freedom" [147], one of the first postwar themed privacy books and a reference document on privacy as a social and constitutional

right, Alan Westin ignites the discussion on privacy in modern times. He too acknowledges the privacy confusion by stating that *"few values so fundamental to society have been left so undefined in social theory or have been the subject of such vague and confused writing by social scientists"*. This alone shows how privacy is often misunderstood, ill-defined, and flat out confusing. Lacking clear social definitions of privacy, we cannot presume to provide a clear definition of what privacy is, especially in applied sciences, like computer science. However, Westin's work on privacy is extensive and provides many of the guiding principles on what privacy is and what implications it has. Westin goes on to provide a very brief definition of his understanding of privacy: *"Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."*

This shows that by nature, privacy deals with private information, as would be expected, and tends to be broad scoped, revolving around information disclosure about individuals, groups, and even institutions to others. Understandably, the definition is generic, given that indeed privacy is multidisciplinary and has a strong dependency on principles that are qualitative and relative to society. This leads to the four basic principles presented by Westin, which will not be discussed in depth, since their scope is sociological in nature: anonymity, reserve, solitude and intimacy. From such fruitful argumentation, Westin outlined the first legislative material on constitutional and common law responses to privacy threats and private surveillance, making remarkable contributions to privacy protection.

The complex nature of privacy is also acknowledged by Daniel J. Solove. His work on "Understanding Privacy"[137][1] provides several foundational concepts for privacy in this Thesis, along with some of more sensible approaches towards privacy in general. Solove argues that privacy is "a plurality of things" and that "questing for an overarching definition is a dead end". In his book, he states that privacy is a confusing and muddled concept, with multiple dimensions. The author presents a detailed overview on privacy definitions, legal, social and philosophical classified around six topics[2]: i) Right to be alone; ii) Limited access to self (shield one self from unwanted access); iii) Secrecy (conceal information); iv) Control over personal information; v) Personhood (identity, protection of personality, individuality and dignity) ; and vi) Intimacy - control over one's intimate relationships or aspect of life. Solove finally argues that existing definitions are "too narrow, too broad and too vague", a statement which perfectly scopes our view on privacy, and specially network privacy.

The most important conclusion, which can be implicitly found in our approach, is that privacy must be a focused effort contained to specific contexts. Even though a top-down approach should always be required, starting from the person or user, privacy can only by properly analyzed when the necessary contextual boundaries are defined. This is particularly inspiring for our network related work, by catalyzing the idea that we must avoid generic and metaphysical concepts, and jump into network privacy, clearly defining our context.

Adhering to the ideas stated above, Solove proposed a conceptualization of privacy, rather than a generic definition. Focusing on four distinct aspects, this conceptualization offers guidelines instead of definitions, on how to approach privacy problems, which are synthesized next:

[1]For both the casual reader or privacy expert, this book provides many vital insights on privacy, and can guide privacy discussions in different areas of expertise, and therefore is thoroughly recommended by the author of this Thesis.

[2]A detailed analysis of such related concepts is out of scope of this work, which must refrain from digressing into philosophical discussions, and focus on network related issues.

Method Privacy is pluralistic with no common denominator, and should be treated as such;

Generality Too generic concepts offers no guidance, whereas privacy should be worked out contextually and with pragmatic approaches, but still using a conceptual model;

Variability Leave room for significant variability in norms, attitudes and conditions about privacy;

Focus Attitudes about privacy must have a focus, and deal with concrete problems, rather than abstract scenarios.

This conceptualization indicates that we can benefit from understanding a focused version of privacy, and should rely on the contextual definitions of the problem at hand. In fact, this entirely supports our vision of a contextualized top-down approach, as well put in the "Generality" aspect, which states that privacy must be worked out contextual but still using a conceptual model. This in turn should yield contextualized and pragmatic solutions, where our focus, following closely the "Focus" aspect defined above, aspect should be the network context and scenarios.

To finalize the evaluation of the work proposed by Solove, we must look towards "A Taxonomy of Privacy" [136], where he outlines several different activities that affect privacy, and that can frame our work.

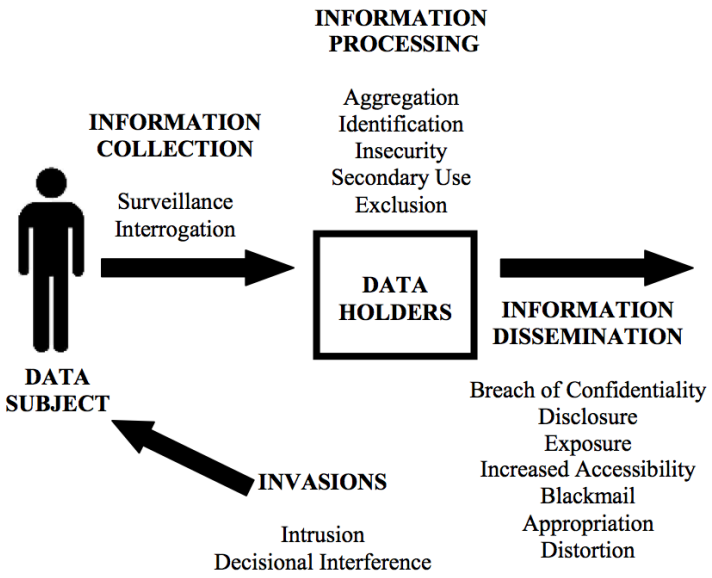


Figure 2.1: A taxonomy of different privacy affecting activities *in [136]*.

The key concepts outlined in Fig. 2.1 revolve around collection, process, dissemination and invasion. Collecting means watching, observing or probing. This is immediately relevant to the network, which is an environment where collecting information can be a simple and straightforward procedure. Processing deals with aggregation, the combination of information, identifying or linking information to individual. It is presented in [136] as directly

related to unique identifiers (of which the network has no shortage) focusing on operations and relationships built on top of collected information. Finally it deals with dissemination and invasion, the direct negative consequences that can stem from privacy attacks or security violations.

There has been sufficient studies and research to fill several pages worth of social definitions of privacy. However, the aforementioned ideas already expose the most important concepts, which in our opinion must be followed to achieve any relevant solution towards the network. First, we must acknowledge that privacy is an elusive concept. Consequently, the initial barrier to overcome is to understand and properly define privacy in the network context. This is also an important step, because this definition can become pivotal towards providing guidance and perspective over privacy solutions. The different faces of privacy assure that no single theory or context can grab all the different, legal, social and technical aspects of privacy. This three-fold approach is promoted within a privacy study by the ITU-T [119]. The study, as frequently occurs, contains an "understanding privacy" section, focusing on the importance and definitions of privacy, with special attention to term clarification. It presents privacy in a three domain perspective: technical, regulatory and sociological.

So far, we have focused on the social aspects of privacy, and how it can be understood. This has led to the reinforced idea that privacy is contextual, where generic and overreaching definitions are often insufficient. Next, we complement this view with legal and regulatory aspects of privacy, along with more recent data protection laws, guidelines and directives. While the previously discussed concepts and definitions frame the ideology on privacy, legal and regulatory frameworks also play an important role in focused objects of study, such as privacy in NGN.

2.2.1 Legal and Regulatory Frameworks

When analyzing privacy, it is important to look at the legal and regulatory frameworks that govern human and civil rights. Only by understanding this component can we develop solutions that simultaneously meet the privacy requirements mandated by law and respect the need of revoking privacy when required, e.g. lawful-interception by judicial institutions. This is a two-edged knife, where on one hand we must protect the user, and on the other, respect lawful authorities by providing means to access private information when lawfully required.

Privacy has been an acquired right for many generations, and is consecrated, as mentioned, in several documents. These documents can summarize the existence of privacy within a legal and regulatory context, providing boundaries for the above mentioned competing concepts. We analyze the European Convention on Human Rights to determine how privacy is defined as an acquired right, and then uncover how it is applied in the data-centric work of telecommunications, by analyzing the OECD guidelines, and derived European Union Directives on data protection and privacy matters.

2.2.1.1 European Convention on Human Rights

The European Convention on Human Rights in one of the first documents, within the European space, to engrave privacy as a fundamental right. In this document, privacy appears in article 8, stating that everyone has the *Right to respect for private and family life.* This right is broken up in two aspects: 1) the right to respect for private and family life, at home and

for one's correspondence; and 2) any interference with this right must be made in accordance with the law, and under defined exceptions. While this right was written in the late 1950's and did not foresee an information society, the parallels are obvious: the rights to private life still must be respected, and where previously was stated (mail) correspondence, now sits any digital communication. This is however better captured in the OECD guidelines on privacy and data protection.

2.2.1.2 OECD Guidelines

It is important to understand that privacy is an acquired right, consecrated in Human Rights declarations. But, for the purpose of our work, and for any work on telecommunications and networking, it is more important still to understand the rights distinguished within the scope of data protection guidelines.

The foundations of EU law, and individual EU members' law as well, originate from the OECD guidelines on data protection [47], dating back to 1980. These documents explicitly deal with digital environments, and set the pace for almost all privacy related matters, addressing issues on privacy and personal data protection. The guidelines can be sorted in two groups, composed of: first, pre and during data collection; and second, after data collection has occurred. For the first groups, major driving forces are *Notice*, *Purpose* and *Consent*. Generically, they describe that the user, or data subject, should be notified of what data is being collected and whom is collecting it, along with the purpose of such data. In certain conditions, that collection should even require the user's consent. After the data is collected, it should be governed by *Security*, *Disclosure*, *Access* and *Accountability*. This means that the data collector should ensure the security of the data, do not disclose it under any conditions except for the original purpose of the collection, provide access to the data by the data subject, be accountable for any violation of the principles. The principles are defined as follows:

Collection Limitation Principle There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

Data Quality Principle Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

Purpose Specification Principle The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

Use Limitation Principle Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with the Purpose Specification Principle, which are with the consent of the data subject or the authority of law.

Security Safeguards Principle Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.

Openness Principle There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identification and usual residence of the data controller.

Individual Participation Principle An individual should have the right to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has information relating to him. By holding information the data controller is mandated to report to the user, in a reasonable and readily intelligible manner, what information is being store and why is it occurring. Conversely, the individual has the right to challenge the data relating to him and to have that data erased, rectified, completed or amended, if the challenge is justified.

Accountability Principle A data controller should be accountable for complying with measures which give effect to the principles stated above.

The bottom line on all of the aforementioned principles is that the user should be informed, aware and protected. The user should know what is being collected, by whom, and certified that the information remains secure. It is also strongly stressed that the ability to collect data should be limited, justified, relevant and only carried out by lawful and authorized authorities. These are the guiding principles that are carried over to European Directives, regulating privacy within the European Community.

2.2.1.3 European Union Data Protection

As discussed, the OECD guidelines captured several of the regulatory principles that should govern privacy in environments which require data protection. This entails not only telephony, mobile phones, but also Internet (and networked) communications, and any of the new communications paradigms. With little surprise, these principles got adopted by the European Union in several Data protection Directions, which are now the baseline for privacy within Europe. From this transposition, stemmed several EU directives that are centered mostly around data protection, defining clearly the subjects and the information that is protected on telecommunications. We review the most relevant principles, scattered across several directives, from which the most important are Directives 95/46/EC [29], 97/66/EC [30], 2002/58/EC [31] and 2006/24/EC [32].

The objective of *Directive 95/46/EC* [29] is to define and regulate the processing of personal data. Processing of personal data means any operation, automatic or not, on personal data. Examples are provided on the directive text which include: collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction. In line with the OECD principles, this too defines a set of principles which should drive privacy guidelines for legislation, which are mostly discussed in the articles between 6 and 24, covering three areas: legitimacy of data collection and its associated conditions and transparency; confidentiality and security of the processing itself; and liability deriving from the stored data.

Collecting and processing personal data becomes legitimate if "the data subject has given his consent unambiguously". The user's consent can be obtained in several ways: by means of a contract, legal obligation, to protect the user's interest, public interest or official authority.

But even in the aforementioned cases, apart from positive and negative exceptions, the user must be generally informed of what data is collected, and the purpose of that collection, to maintain a transparency approach. The user should also be allowed to consult the collected data within reasonable time frames and cost, while reserving his right to object in case the data is collected under exceptional conditions (i.e. crime investigations).

Directive 95/46/EC also discusses the conditions under which collected data should be stored, such as retaining confidentiality, security and data protection in conformance with the OEDC guidelines. 95/46/EC also considers the liability, sanctions and remedies that derive from the collected data, and the enforcement of the previous guidelines.

Directive 97/66/EC [30], known as The Data Protection Telecommunications Directive, provides a complement to Directive 95/46/EC, with regard to Telecommunications, operators and users. It describes the guidelines for subscribers, users and their privacy protection in public telecommunications networks and telecommunications services. It mostly addresses the mandatory security principles that should be provided, stating that the providers of telecommunications services must adhere to providing a security infrastructure and technical measures to safeguard the security of their services with attention to the state-of-the-art in their domain and the associated implementation costs. The user must be informed of threats and breaches along with possible solutions and costs for such events.

The directive also defines that communications must be confidential, and that no listening, tapping or any type of storage can be done, with the exception of authorized legal entities. It also provisions traffic and billing data[3], stating that such information must be confined to the delegated entity of the telecommunications provider. Also, the data must be deleted or anonymized after processing or until the legal deadlines expire, which must be defined in accordance to regulatory law. The user must have the possibility of disabling such features as calling-line identification and automatic call forwarding.

Directive 2002/58/EC [31], known as Directive on Privacy and electronic communications, extends Directive 95/46/EC [29] and supersedes Directive 97/66/EC [30], because it only addresses a specific domain of Telecommunications, while 2002/58/EC is applicable to Communications in general. The importance of this directive relies on the fact that this document pertains to communications environments, which also include the internet and TCP/IP communications, therefore guiding any work to be done in the area of privacy in next generation networks. It focuses on five key points: Security, Confidentiality, Location Data, Spam and Consent. Regarding security, it states that the provider of electronic communication services must take appropriate technical and organizational measures to safeguard service security, while considering state-of-the-art and cost on their implementation. Also, the subscribers have the right to be informed of any breaches, along with the measures taken to remedy the occurrence and the costs involved. As far as confidentiality is concerned, it should be illegal to listen, tape, store or perform any other means of data or communication surveillance without the consent of the users, except for legally authorized entities. This directive also discusses how traffic data must be discarded or made anonymous after it has served its transmission purposes. Beyond that, traffic may be processed for billing purposes, and retained until the bill is payed. It first mentions marketing relating to electronic com-

[3]This data include: number or identification of the subscriber station; address of the subscriber and the type of station; total number of units to be charged for the accounting period; called subscriber number; type, starting time and duration of the calls made and/or the data volume transmitted; date of the call/service; other information concerning payments such as advance payment, payments by installments, disconnection and reminders.

munication means, where user information should be processed only with the consent of the user, which may withdraw at any time.

Directive 2002/58/EC also addresses Location Data issues by defining that, when available, such information may only be processed after being made anonymous, or with explicit user consent. The user must also be aware of what type of location information can be processed, and whether it will be relayed to a third party or not, prior to consenting to the data processing. As always, the user must have a simple and free of charge means to cancel such data processing. An increasingly important issue addressed in this document is the user's right to be informed whether his subscriber information, and what is that information, is stored in any directories, and who may access it under what terms, reserving the user's right to verify, correct or withdraw any personal data free of charge when required. Unsolicited Communications, also known as Spam, is described as only being possible with the direct and prior consent of the user. The aforementioned condition applies to a range of communication means such as voice or electronic mail. Beyond the previously mentioned articles, this directive also contains guidelines about the right to itemized billing, the presentation and restriction of calling and connected line identification and automatic call forwarding.

Focusing on Data Retention, and amending 2002/58/EC, *Directive 2006/24/EC* [32] sets the categories of data to be retained, the obligation to retain that data, and who accesses that data. It also defines the legal periods for retention, along with data protection, security and storage requirements. Service Providers are mandated to retain the user's data for a period no less than six months and not more than two years, where the access to it is granted based on lawful grounds, and according to national law. Article 5 determines the categories, or data types to be retained by the service provider. It makes the distinction between Internet communications and normal telephony situations. Focusing on Internet aspects, it states that providers must retain email addresses or user ID of the subscriber, complemented with name and address of the subscriber, if necessary. Similar information about the destination must be retained, along with duration, and type. Beyond these mentioned data types, Directive 20006/24/EC identifies the problem relating to location data on mobile communications. It seems originally intended for normal mobile telephony, but nowadays it will also apply to IP mobile devices. It states that the location label, or Cell ID, must be recorded at the start of the communication. Furthermore, data identifying the geographic location of the cells by reference to their location cells, must be retained for the duration of the communication.

2.3 Privacy Models

One way of contextualizing the privacy problem on network-related aspects is to explore existing models and technical solutions that deal directly with network related privacy. It is important to understand what are the existing contributions that relate to or enhance privacy on the network, before defining any new models or contributions. This analysis will enable a broader understanding of current privacy trends, threats and model, and contribute towards the objective of providing an in-depth study of privacy on the network. We focus especially on the relationship between user privacy and networked environments, departing from socio-legal aspects towards practical network-related models. We explore a wide array of privacy oriented models and systems, that can contribute to the understanding of what a network oriented privacy model must cover. Below, we present an overview of the relevant models, especially analyzing their main purpose and means towards achieving it.

As we focus on privacy related models, the first observation is that there is no single encompassing model that harnesses all different privacy aspects, relating to user or network privacy. Instead, most proposals focus on a specific subset of information or mechanism. Therefore, we present below several "pieces" of the privacy model puzzle, that, while lacking a consistent model that drives model overview, enables an insight on the overall aspects that govern privacy solutions: threats are discussed as stemming for specific aspects.

In the course of our study we have identified several areas that relate to network privacy, directly or not. We start with indirect concepts, regarding database privacy models, that have been extensively debated, and provide a bridge towards atomic user information, as reflected on the network. We later study anonymity models and their relationship to privacy. Only afterwards we focus on network specific privacy models, that are directly aligned with the course of our work. We finalize by studying interesting models that can provide further hints on how to handle privacy in the network, like Bayesian models, used for Spam classification or in intrusion detection systems, or orthogonal models, that are not directly related, but provide some insight on the problem at hand.

2.3.1 Database Privacy Models

The first relevant attempt to model privacy originated in the distributed database world, where the correlation of several data columns in distributed relational databases would yield privacy violations and allowed accurate identification of anonymous subjects. The most relevant model in this area is K-anonymity [141]. The model behind this proposal focuses almost entirely on data as a personal attribute on a table. Data is treated as a tuple of information, where an identifier is associated with a particular piece of data, siting in a database row. The model is built purely around identifier and data, upon which several anonymizing functions are discussed. The main idea is to ensure that the same information appears at least K-times, associated with different identifiers. The information is scanned and arranged to extract (almost) unique identifiers, dubbed Quasi-Identifiers [34]. These identifiers are used as the main correlation mechanism, establishing relationships between different seemingly anonymous pieces of information. This approach to relational databases has proved to be effective, and became the basis for a research branch that now sees further developments with derivative and improved schemes, such as K-unlinkability [92], l-Diversity [91] and T-Closeness [84]. These solutions all use different models to hide relationships established through the Quasi-Identifiers. While this does not seem directly connected to the network environment, it provides a powerful insight to the organization of information as tuples composed of identifier and data, as means to model information, and consequently threat user privacy. Also, the single fact that information nowadays is almost entirely becoming an entry in a relational database, as we converge to a computer-centric society, indicates that privacy on relational databases must be widely considered by any solution that wants to be taken seriously on any computational field.

2.3.2 Anonymity Models

While anonymity does not equal privacy, in recent years several approaches have tried to provided privacy through anonymity functions. The common assumption is that, when anonymity is provided, privacy is assured, even though this neglects several of the differences between privacy and anonymity. Regardless of the differences between them, it is important

to analyze the main trends that deal with anonymity, since it can be an important property towards providing privacy.

Most anonymity-related proposals rely on the concepts of anonymity sets, or anonymity by numbers. Anonymity usually deals with an omni-present attacker on the network and is usually discussed in terms of "knowledge" or exchanged messages. In most cases it requires that both senders and receivers remain anonymous in the communication process. This is exactly what is proposed by Chaum Mixes [22], which provides several important concepts for anonymizing messages, concealing the senders and receivers. In this approach, messages are distributed through a number of nodes, constituting a mix, where once a message crosses enough participants in the mix, it becomes impossible to determine its true origin, and is then forwarded to the final recipient. In this scenario, the mix constitutes an anonymity set, where any participant can be the original sender. The level of anonymity varies with the number of participants in the mix, where the larger the set, the more anonymous the message given that there are more possible senders. These concepts have been applied in several solutions. The most noteworthy solution derived from from Chaum Mixes, is Onion Routing [142], or TOR [37], which will be discussed in Chap. 5.

The value and success of Chaum Mixes [22] has lead to an attempt of building concrete privacy and threat models around it, explaining and measuring the amount of privacy provided. The initial effort of modeling and measuring the privacy associated with such anonymity ecosystems came from [133]. This paper presents the seminal work on using Shannon's Information Theory [134] to measure anonymity, instead of relying on the tried-and-tested technique of anonymity set sizes. It details the construction of an anonymity set where information is deposited, and later measuring it with entropy calculation. While such anonymity proposal seems to model a reality that is somewhat different from normal operations on the network, focusing on mixes, it provides an important contribution to a more generalized privacy model: information can be generally treated as a set, from which its anonymity value is extracted. Therefore, following the Information Theory rationalization, the more concrete and identifiable the information is, the smaller its anonymity set becomes, and the information begins to be clearly identifiable. The evolution of this research trend has lead towards measuring anonymity, rather than privacy models around anonymity.

A recent contribution [120] that builds on the above cited concepts focuses on hiding relationships, also resorting to information theory concepts. In the process, it provides definitions for anonymity, unlinkability and unrelatability. This shows how complex relationships can be observed for elements within a set, therefore focusing partially on set theory derived rules and theorems. However, the most interesting conclusion from [120] is that privacy indeed is about relationships, and these relationships must be properly acknowledged and accounted to truly understand the complicated privacy inter-connections that occur on the network space. But, this is not the first effort focusing on hiding relationships. Information hiding was made noticeable by [65]. This work presents a very high level framework for dealing with anonymity and privacy using an elaborate graph theory model. It is based on the same underlaying assumptions as Chaum Mixes, where senders and receivers should be indistinguishable (i.e. equiprobable), but focuses on an interesting concept: partial knowledge. The idea is that information can be broken up in different views, explored by the attacker to gain information about different observations. This is supported by graph theory, used to establish relationship gained by examining the knowledge put forth by different functions. However, while clearly showing the difference between privacy, anonymity and modeling difficult concepts such as pseudonymity, its pure mathematical and abstract nature makes the relationship with the

network, and network events, very elusive.

The above entire body of work focusing on anonymity, information theory and relationship concealment shows that, as recurrently stated, privacy is as complex as it is diverse. These are the most interesting trends in recent years, partially due to the increasing interest of anonymity and mix networks, that show the appearance of several concepts that we highlight: i) the appearance of a data or information set, that relates to private information; ii) the existence of multiple relationships between such sets of information pieces, guiding attackers (partial) view; and iii) the unification of information through the acknowledgment of information as data tuples. While these concepts were extracted from models that at times seem hardly related to network operation, i.e. packets on the network, they provide great insight into what a network oriented model can look like.

But, before moving on to network specific models, we must acknowledge that one of the most prevalent findings of the anonymity work is that each model tries to present first and foremost a definition of privacy, second a definition of anonymity, and thirdly how those concepts are addressed within the framework. Most solutions on the network, and discussed in several of the following sections, do not have any clear model, but rather make some underlying assumptions about the type of attacker and possible privacy attack. Due to the specificity of most discussed attacks, indeed it is not possible nor desirable to tackle a high level concept of privacy and privacy protection in each of these proposals. But, without understanding the model behind them, it is not possible to tie them together in common view of network privacy.

However, an important conclusion considering anonymity, is how it relates to privacy. As stated at the beginning of this section, it is important to clarify that privacy is not anonymity. A definition of privacy, that takes anonymity into account, comes from [65], where the authors state that privacy is simply relationship anonymity between two users or agents, whereas anonymity is the concealment of the user or agent identities. Another example comes from the outlined technical challenges of network anonymity [80]. This shows a clear differentiation of privacy and anonymity, and argues that anonymity is a property that can be used to achieve privacy.

2.3.3 Network Oriented Privacy Models

Most privacy models, and technologies, are usually centered on the user. The user is usually described as relating to information blocks, that can flow on the network or be statically stored in databases. This very abstract information model could fit anything from network, to databases, or even sociological models. However, the network is anything but abstract. On the network level, there are credible and immediate threats, relating to information disclosure, action disclosure, and correlation. Most threats stem either from particular identifiers (e.g IP or MAC addresses) and their properties (e.g location) or using specific mechanisms that jeopardize privacy (e.g, MIPv6). Given the specificity of the threats, and consequent proposals to address such gaps, the consequence is a very segmented approach to all threat that brings no real model behind network privacy. Therefore, solutions on different layers can originate from a diverse range of interactions, which are never captured.

The essence of privacy on the network eludes us. There is however, one model that focuses on network interactions, and indeed models several key participants in the network and how their relationships undermine privacy, although not covering the entire spectrum of privacy threats discussed previously: the Freiburg Privacy Diamond (FPD) [150]. FPD

uses relationships established between four key network entities and properties - user, device, location and action - to determine how privacy is breached (and protected). The proposed entities indeed cover a large spectrum of the network interactions, and can accurately model most network situations. Also, the establishment of relationships between the entities, as means of determining or breaching privacy, follows in the steps proposed by the before-mentioned abstract privacy models. They present a clear way to convey how the associations between the entities lead to unwilling disclosure of private information, but focusing entirely on the network level. This makes it interesting to explore this model more closely than others.

While many schemes exist to enhance user privacy in some level or particular aspect (c.f. Sec. 2.5), a very neglected issue is how to actually measure or compare the privacy level provided by each solution. FPD is a conceptual model that tries to, according to the authors words, "classify, analyze, and construct anonymizing mechanisms in respect of the type of mobility that is required for this anonymity mechanism". And while it focuses on anonymity, it can also capture the nature of privacy, which we analyze below.

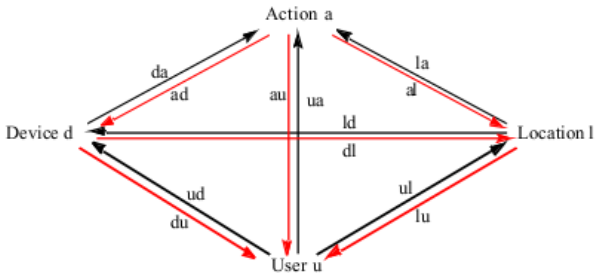


Figure 2.2: The Freiburg Privacy Diamond *in [150]*.

The FPD attempts to model privacy around a completely connected graph, diamond shaped with 4 vertexes, as shown in Fig. 2.2: User (U), Location (L), Device (D) and Action (A). The idea behind the four vertexes shows that an attacker can try to reach to map a user to an action by observing in which terminal was the action performed and who is the user associated with the terminal, or on the other hand, observing the location where the action happened and the matching user. Since the graph is completely connected, any variation of these steps is allowed, and every path that connects an action to a user has successfully broken the user's privacy. It is worth mentioning that FPD does not provide a solution that preserves privacy, but rather a model to evaluate if the user's privacy is being breached or not, which is exactly what we were striving for initially, and should precede any real solution. Also, the model assumes that time is an implicit dimension, and that no time correlation events occurs. They are assumed as mostly absolute events, that either yield information or not. This simplifies the graph in terms of edge weights, since it is assumed that all actions provide the same amount of information, i.e. once the link is established between a vertex, it is irrelevant the degree of information that was applied, since the link was established, making it binary in nature.

Perhaps the single most important contribution proposed by the FPD is the notion of using relations to connect the different edges. By looking at Fig. 2.2, we see that each edge has a particular identification, which denotes the relationship established between each of the vertexes. In total, the diamond is composed of 12 relationships that model the author's notion of privacy. As explained in [150], two elements (x, y) are in a relationship $R_{x,y}$, if the attacker

possesses knowledge that ties the elements together. This notion of relationship makes all the difference, because for the first time, we are able to derive the notion that something in the network may cause a dependency that ultimately leads to privacy violations.

Using these tools, the author proposes a set of definitions that are worth exploring. First, the anonymity of an action (A) towards a user (U) exists whenever there are two elements (U, A), which share no relationships $R_{U,A}$. In this situation, if there is no information that ties the user to an action (directly or otherwise), then not only is the anonymity of the action preserved, but also the privacy of the user concerning that particular action. The following rule defines transitivity between relationships, so that if $R_{a,b}$ and $R_{b,c}$ exist, then $R_{a,c}$ is also valid. This denotes the first inference reasoning on network privacy. Applying these rules to the entire set of relationships yield the attacker's view of the user, stated as a closure view over the information. As these views come together, the attacker forms a composed view, through the "composition of elements". This composed view is subjected to recognition, linking and intersection attacks. Recognition attacks means recognition relationships in two different composed views, and thus combining them into a richer view of the user. The linking attack means that the elements can be attributed to a specific user through contextual information, usually using exclusion. Lastly, the intersection attack means looking for similar traces in different views, also through contextual information.

These rules provide the first insight at a network privacy threat model, which yields relationships between user and the network, in such a way that it can be systematized. The diamond can be simplified to have a set of well defined, loop free, paths that lead to the positive identification of the user. These are summarized by the author as:

1. user to action directly
2. user via location to action
3. user via device to action
4. user via location and then device to action
5. user via device and then location to action

Whenever these observations are possible, the user's anonymity and privacy are forfeited. This means that to provide privacy, none of these paths can be detectable, an important conclusion that is worth extracting such a privacy model. A smaller observation is that the model defines itself as trying to capture the attacker's knowledge over the system, which could be modified with appropriate edge weights to provide a measure of anonymity rather than providing absolute linkage vectors.

While being the most promising attempt at a network counterpart of a privacy model, it neglects that privacy is built around identifiers (which are omni-present in the network), a prevalent discussion in K-Anonymity [141] with Quasi-Identifier [34]. Because identifier information is not a central element of the model, it is impossible to see how different identifiers play a role in different locations, since, depending on the location of the attacker, the information extractable from packets and actions is different. The result is that how and why relationships are established is not evident in the model, leaving a large gap, as made obvious by the discussions on databases and anonymity (and further supported by the evidence model presented in the following section). It also fails to acknowledge that in most scenarios, users desire privacy and not anonymity (e.g. service customization), making the ideas of minimal

disclosure or private data sets inescapable, something which is not captured by the model, which deals only with relationships.

Another discussion concerns edge weight. The model assumes no edge weight as a simplification, failing to acknowledge that different information can yield different value, for instance, different topological information. While at first this may seam a reasonable compromise, the rules presented for recognition, linking and intersection indicate that they work based on contextual information. These are at best, probabilistic inferences from network observations. And because these are important, they should play a role in the overall privacy model.

Despite these shortcomings, FPD still provides the most interesting overview of what are the key relationships in the network. It presents one of the most illustrative discussions of what privacy on the network implies or requires. The described model is a good fit for most network interactions, based on actions, locations, devices and user are actually suited in most current environments. This is both a blessing and a curse, because the specificity of the actors can hinder the abstractions that could be introduced into the model. Lastly, the modeling of an action, while interesting on a purely network level, forfeits the information plane that is widely discussed in the previous sections. Actions can pose a threat due to two different information planes: the action itself and the exchanged information.

2.3.4 Bayesian Models

An interesting field in privacy, which is not commonly considered, relies on Bayesian inference. Bayesian probability is based on the observation of multiple evidences, and building a probability according to the observations. In concrete terms, the Bayesian probabilistic approach assigns a starting (prior) probability, and continuously updates it (posterior) in the face of new data, thus evaluating the probability of a given hypothesis. As such, Bayesian probability is considered to treat a probability as measure of knowledge, that evolves over time (observation). This is the guiding intuition for Bayesian inference, where a collection of evidences are used to calculate a probability upon which the likelihood of an event or hypothesis is stated. This approach to evidence gathering, with appropriate probabilities, that increase or decrease based on the gained knowledge, has many applications in fields that do not directly relate to privacy.

Bayesian inference is the common approach [78] for Intrusion Detection Systems [35], where the different events or alarms are assigned a probability and related to the remaining events to determine whether any abnormal behavior is taking place in the network. This evidence collection mechanism is used because each event (positive or negative) takes part in a continuous network detection mechanism, gathering evidences until a decision is taken. A similar approach is taken for email spam filtering [128], with Bayesian Filters [58]. Some of the most popular filters today are based on Bayesian probabilities, or Bayesian networks. The normal procedure is to scan emails with multiple iterative rules, and build on the probability of considering an email as spam. The result of each rule is added to a Bayesian probability, that is adjusted based on evidences.

This introduction to Bayesian based mechanisms is interesting because it shows how several systems work based on evidence gathering. Instead of single and uncorrelated information evaluation, these systems perform a continuous evaluation, where different events or evidences contribute to an overall probability. In fact, this represents a new model for privacy that only recently gathered attention. What we have seen with Intrusion Detection Systems and spam filtering is that information can be extracted from several events. However, the Bayesian

approach can also be used to handle information in data mining architectures [25], allowing the extraction of large amounts of information from distributed databases, unquestionably breaching user privacy. Using Bayesian network to mine concealed information has been show in [23], which shows a new framework based on information-theoretic analysis using Bayesian networks to learn about private data. The framework proposes several algorithms for knowledge acquisition from data. Even though not directly about privacy, it shows how Bayesian networks can lead to learning information about users. Using these database techniques to invade user privacy has been show in the previously discussed l-Diversity [91] model. In l-diversity, a notion of privacy using Bayesian inference deducts the prior and posterior beliefs of an attacker in the face of a newly acquired piece of information. They argue that this information can be aggregated, building sufficient knowledge to defeat K-anonymity.

Dealing directly with the user, it is possible to treat a user privacy as a system subjected to intrusion. An interesting analogy would be to define user privacy in the same terms as an IDS, as we can view the user's privacy domain as an intrusion prone system. In this scenario, gathering information can lead to a breach, as in IDS systems gathering information can uncover an attack. This approach is loosely taken in [6], which shows how dynamic Bayesian networks can be used to collect evidence to assert whether privacy has been violated or not. An extension to this approach, which could be used to threaten user privacy, would be to use such an IDS-like approach as a means of evidence gathering and information inference about identifiers, data and personal identifiable information, classifying information until a user is positively identified, as opposed to detected privacy breaches.

All of the previous discussed approaches and techniques reinforce the notion that, by resorting to continuous analysis of private information through Bayesian inference, it is possible to compromise privacy when enough evidences have been gathered. The commonality of all these systems is that they all gather evidences, and properly mark them, to extract patterns and hidden information. However, the relationship to a network privacy model can seem foggy: there is no clear or consistent privacy model that can help in the definition of network driven privacy. But, there is a clear conclusion: privacy can be dealt with in terms of evidence collection, and not just as singular observations. This conclusion opens the door to a more complex ecosystem in the network, where multiple iterations can be harnessed by such systems, and turned against the user, similar to data mining mechanisms. And in turn, this needs to be properly accounted in any network privacy model.

2.4 Identity and Privacy

As we depart from privacy models towards technical solutions that deal with privacy, we acknowledge that, when discussing privacy in current digital systems, the focus is on the user. As in real life, the primary object of privacy is the user, a manifestation of a real person on the system. The metaphor for these user-centric concepts is identity, focusing entirely on the user. Identity is directly related to privacy because identity fosters notion of self, which is the subject of privacy. This was made clear by the definitions of privacy, and more importantly by the presented legal arguments. The relationship between identity and privacy is obvious and made more relevant in digital systems [54].

Therefore, it is not surprising that on the network, identity, in the form of the user and user-centric technologies, appears tightly coupled to privacy. Research on identity in the information society [62] shows that identity, embodied by IdM technologies, is being hailed as

a beacon for privacy and security in the Information Society. We must study these concepts in the light of privacy protecting mechanisms, that deal directly with user sensitive data. To do this, we present the most important aspects of the identity trend, especially on the application layer plane. We also review how identity is currently being used on the network, in different application scenarios that resort to user-centric technologies.

2.4.1 Identity Management

The evolution of Internet paradigms has lead to an explosion of services available over the Internet. This followed closely the trend of recent years, where Web 2.0 opened the door to the social web, making users signup and register with different services, blogs, forums, and a wide range of utility applications. As the number of accounts and corresponding passwords grew, a problem emerged facing the authentication and access control of end-users, which now faced multiple accounts and login forms, locked in information silos, resulting in a large amount of replicated information. This ecosystem provided unique conditions for the growth of IdM concepts as means to relieve the burden of user management, both on the user side (by handling multiple providers and service through a single account, requiring a single login action), and on the service side (by reusing authentication and authorization systems, and enabling federative properties).

It is arguable to actually state when the IdM trend began to gain momentum, but it is unquestionable that one of the driving forces behind was the work proposed by Kim Cameron, with Microsoft Research, that came to be known as the "The Laws of Identity" [19]. Rather then proposing yet another user-centric platform or architecture, this work proposed an identity meta-system defining the requirements for IdM operation and adoption. Below, we summarize these laws and their meaning:

- **User Control and Consent:** Only reveal information that identifies the user with the user's consent;
- **Minimal Disclosure for a constrained use:** Release the minimal amount of identifiable information;
- **Justifiable Parties:** Limit information release to justified parties that express a real necessity for the information;
- **Directed Identity:** Limit the identifiable information towards a specific goal, and do not reuse it, making it unidirectional;
- **Pluralism of Operators and Technologies:** Promote interworking of multiple identity technologies and the federation of multiple providers.
- **Human Integration:** Make the user (human) part of the system design.
- **Consistent Experience Across Contexts:** Guarantee a simple and tangible user experience throughout multiple operators, technologies or devices.

These apparently simple and straightforward rules, make up a very solid foundation for the discussion of user-centric technologies that build on identity concepts. However, this only defines a meta-system, where several architectures can be used to fulfill several of these laws.

Below we explore the technologies that can be used to fill in the gaps promoted by the various laws as means to provide privacy.

One of the primary objectives of many IdM solutions is to solve the multiple account and authorization issues, by providing a Single Sign-On (SSO) experience. With this goal, towards enterprise authentication models, the Liberty Alliance proposed an architecture, Identiy Federation Framework (ID-FF) [5] that enables an SSO platform with federative capabilities, using SAML [20] assertions. By using an Identity Provider (IdP), which provides a central point for authentication, that can be federated with other IdP system, it is possible to provide an SSO experience across many providers. Therefore, the ID-FF platform provides authentication and federation capabilities that can be complemented by attribute exchange procedures and web service interfaces, thus composing the entire Liberty Alliance Framework. Interestingly, the ID-FF defines the usage of pseudomyity towards the providers using the federation system, creating a unique pseudonym association between user and provider, that is uncorrelatable across different providers. The entire specifications of ID-FF and family protocols were later contributed to the Kantara Initiative [79] that now aims at steering identity related discussions. It brings together several of the Liberty alliance members, along with many of the proponents of Internet based solutions, discussed below, providing a new forum for discussing identity, privacy and the network.

Using similar concepts to ID-FF, applied in the scope of educational environments, Shibboleth [70], a unilateral proposal from the Internet2 Initiative [69], focused on a user-centric approach using identity providers. It further defines the concept of Service Provider (SP), that is able to discover the user's IdP, for facilitated federation mechanisms. Both these initiatives, ID-FF and Shibboleth, led to the creation of the SAML 2.0 [20] standard, which covers a wide range of scenarios, such as SSO, Federation and attribute exchange, all with a strong emphasis on end-user privacy. It makes use of options discussed earlier to create a combined specification that is being widely adopted as the standard for enterprise SSO. It uses the same basic components, and reflects the same basic operations as its originating technologies.

The adoption of SAML 2.0 is also reflected on Microsoft Cardspace [109], a Microsoft effort to create a recognizable paradigm or user identity. Cardspace has its root on an Internet oriented scheme known as Microsoft Passport. Passport aimed at providing an SSO environment for internet based applications. It later evolved into Windows Live ID [110], which provided a consistent user account across different services along with a single authentication mechanism used across multiple services, whoever limited within the Microsoft portfolio. This has become somewhat stagnant due to the shift towards Cardspace, which is foremost a metaphor that provides information cards as a tangible identity management concept to the user. Cardspace builds on web service technology, but aims at being a meta-system for identity concepts, and can be integrated with other IdM solutions, complementing them. Using Internet-based services, it already reflects some of the concepts present in IdM systems, such as an IdP and a service provider entity, named Relying Party, but empowers them with a much needed user relationship through the cards - something the user can understand - that can be presented at different websites.

With the emergence of the Web 2.0 phenomena, OpenID [122] has gained momentum as widespread user adoption grows. It is also an Internet based IdM system, where users are identified by a Uniform Resource Identifier (URI). Using the concept of Users, Relying parties and Identity providers, it enables SSO, secure authentication and attribute exchange. It provides the expected and necessary features for the basis of an identity system, but only

tackles application (web) based concepts, extremely biased towards Internet-based services. OpenID has seen adoption by providers like Google, converting all their user accounts into OpenID identities, usable at any enabled website. Another open protocol that enables SSO in a web environment is OAuth [116]. Major social websites like Twitter and Facebook provide OAuth support, enabling their users to login on other services using their Facebook or Twitter credentials, through the website's API, without compromising privacy. The login credentials provide a limited access to the user's profile, thus enabling SSO, and even some user attributes, without providing full access to the user's profile at the corresponding provider. OAuth has been standardized within the IETF [40], and is picking up momentum in the cloud environment.

On a completely different front, european projects have been experimenting with identity as a privacy enabling technology, of which the most relevant are Privacy and Identity Management for Europe (PRIME) [63] and Future of Identity in the Information Society (FIDIS) [46].

PRIME is a European research project that aims to develop and promote Privacy Enhancing Technologies, based on Identity Management Systems. Its primary goal is to enhance user control over the overwhelming amount of data scattered through different networks. The approach relies on the concept of "partial identities", supported by an IdM framework that takes into account legal, technical, social and economical requirements. Using several technologies and concepts, the PRIME project defines an IdM ontology to serve as the interoperability foundations. Also, a key concept is using Identity Mixes for privacy support. It is primarily a web oriented project, taking advantage of several HTTP based technologies and web services. The PRIME architecture is based around generic concepts, such as Entities, Data and data exchanged between entities, in the form of Claims. The entities are differentiated as *Users*, which refers to natural persons as defined in the European Union directives and *Organizations*, representing generic legal persons. Data exchange between entities is a very important part of the PRIME definitions, which is accomplished by claims. There is a key exchange protocol executed beforehand to establish trust between entities, in such a way that the subsequent claim exchange is trusted. All of the aforementioned components define a PRIME system, which can be summarized as a a framework for secure data exchange, with a very strong regard for privacy. However, it uses only application level protocols to exchange application level information, showing little or no regard for network aspects and concepts.

The other interesting initiative is FIDIS, which is a Network of Excellence centered on privacy and identity management sponsored by the European Community. It aims at supporting interoperability of identity with prevailing security. Much like PRIME, FIDIS focuses on IdM solutions for the European Union. The general approach is somewhat common for both projects, focusing first on a set of socio-economic and legal requirements for identity management systems, which include privacy constraints. This project's primary goal is to push forward on identity solutions, with a high interoperability capability therefore unifying scattered existing systems. Also, it focuses on security and privacy along with profiling and forensic implications, which are mandatory topics when discussing IdM systems. FIDIS aims at becoming a driving force for future IdM solutions, establishing a widespread base of documentation that describes the implication of several socio-economic factors and identity management systems. FIDIS devotes particular attention to the high level integration of identity management systems, rather than to protocol specific or too detailed information bits. It considers an important topic of Identity and Mobility, but only scratches the surface by laying the groundwork of requirements along with a few interesting directions. The practical results of this approach can be seen more as models, as opposed to deployable architectures,

and it is not restricted to computer networks and digital identity, as it grasps real identity and computer aided systems like Radio-Frequency Identification (RFID) enabled passports.

Most of the situations where (user) identity has been considered are at the application layer, which deals more directly with user and user related information. As an application level feature, focusing mainly on Internet-based scenarios, identity has played a part in several services and architectures, especially in the new web 2.0 landscape. Several architectures and protocols follow this approach to user identity, using concepts such as Identity providers and service providers, relying on HTTP/XML based protocols and web services.

Regardless of the previous trend, the paradigm shift towards IdM is not happening solely on the Internet. The user is becoming increasingly defined as an identity, and identity driven architectures and services are becoming a reality. As part of the research expressed in this Thesis, as mentioned before, the IST Daidalos [33] and IST SWIFT [72] projects, have promoted the use of identity in the network, bringing it to all levels of the network stack and not only the service layer. This led to a user centric network where identity is the driving force behind network provided services, preserving privacy while providing an unprecedented level of pervasiveness and personalization. However, in different places, the network has shown traces of using user-centric information and identity related paradigms. Below we try to uncover such traces, even if implicit in certain efforts.

2.4.2 Identity in the Network

The network has long resorted to the use of identification of the user or user related resources. This has been done implicitly in most cases, and results in a disguised use of identity related technology that must be understood. Several technologies have always needed to either identify the user, or the device under the control of the user, creating a unique coupling between user and device. Therefore, it is interesting to observe that identity has already made brief appearances in different network contexts.

All of the technologies mentioned in the previous section are agnostic to lower layers and fail to address the "network side of things". At the same time, there is already a large amount of identity information that resides scattered across the network stack. But, on the lower layers, such identities are usually tied to devices and protocols rather than to the user and his digital identity, as observed in GSM and UMTS. The Subscriber Identity Module (SIM) and the Universal SIM (USIM) are in fact cards that store the International Mobile Subscriber Identity (IMSI), which is a unique number that identifies the user and acts as a key for network access. The SIM (or the USIM) can be perceived as the "user", or a representation of the user. Similarly, a growing trend is, for example, mobile payments supported by Near Field Communications (NFC), where the division between user and device is further diluted, and yet another real-life aspect is moved towards digital interactions, with all the privacy associations it implies. There is a tightly coupled view of the device and the user in this model. If we consider the TCP/IP network stack, we also find identity information. Unique protocol and network identifiers provide simple handles to recognize or identify a specific user or device, a phenomena that is not limited to a single layer. At the link layer, the terminal is identified by its unique MAC addresses. At the network layer, this is achieved by a public and usually dynamic IPv4 or IPv6 addresses. Mobility bound identifiers, such as the MIPv6 [77] Home Address (HoA), are also unique. These addresses identify both the user and terminal, since they are tightly coupled. Furthermore, some addresses are tied to public keys, such as the proposals for Cryptographically Generated Addresses [10]. The key can be bound to

the user identity or to the terminal, leading to more information based on identity acting as singleton information pieces, with no real link to the upper layer identity concept.

An interesting case of identity (or identification) in the network stack is Host Identity Protocol [113]. HIP is, in its essence, a key exchange protocol based on a new cryptographic namespace. It accomplishes a clean separation of locator and identifier, between network and transport layers, by introducing a new namespace through the Host Identity (HI). A HI is a statistically global unique name for hosts with an IP stack; it represents the identity of a host on the network, which can assume multiple identities, some "well known" and others unpublished or "anonymous". The HI is the public part of a cryptographic key, where the private part is stored in the host, or kept by the user. In order to represent the Host Identity in other protocols, either a Host Identity Tag (HIT), a hash of the HI or a Local Scope Identifier, a legacy identifier that fits in IPv4 fields, are used. HIP introduces the concept of an identity namespace, where the identity object is the host. As with the other discussed identifiers, this can be easily mistaken by the user, given the close relationship that can exist between user/device (host).

On other layers, dealing with the use of identity in the network usually involves the representation of the user through handles (e.g. usernames) or application identifiers. At application layer protocols and services, this usually takes the form of usernames or even URIs. This brings us once again to the IdM systems, discussed before, and how identity is represented in such systems, given that application layer protocols and services already fall under the category covered by IdM systems.

Every piece of information mentioned in the preceding paragraphs identifies the user, or at least some part of the user identity, regardless of its intent to disclose this information. Each of these protocol or network identifiers either uniquely map to the user or to the device being used by the user. The lower layers are thus a gray zone, where the user "is" the device, and the device is the user. And no concerns on user privacy, or identity protection, exist. The network stack currently does not support the semantics of several digital identities representing the same user, which may be used for different purposes or actions, or for different roles. While this is simple to address at the application level, the task becomes complex at the network level, where we can have two "personae" representing the same or different users simultaneously in the same device.

2.5 Network Aware Privacy

Apart from the previous section, which covered identity paradigms in the network stack, most of the discussed work discussed so far has no particular network oriented focus. In fact, most approaches are conceptual in nature. However, as previously discussed, to understand privacy on the network, we must reduce the applicability scope. Accordingly, we try to further narrow the studied solutions towards network related approaches. This focus results in a review of privacy enabled solutions that aim at protecting different aspects of network operation. Usually, these solutions are only applicable in the layer on which they operate.

The first part of the discussed work focuses on link layer threats and solutions. We mainly intend to explore solutions focusing on the network access. In the scope of link layer solutions, we are faced with individual protocol threats, identification and tracking through unique link layer identifiers. We then focus on solutions that target the network layer, where the most important issues are the location and identification of end points, due to network properties.

In this scope we also explore some mobility related solutions that focus on privacy. Finally, we explore a trend that primarily resorts to using different pseudonym identifiers i.e. alias identifier on different network levels, for privacy protection. This comes at the end because solutions using pseudonymity tend to have relationships between several layers or concepts.

2.5.1 Protecting Network Access

The network access corresponds to the link layer domain, to which users connect their devices. These technologies are usually broadcast mediums, such as WiFi or Ethernet. Most link layer security solutions aim at reducing the security impact of a shared broadcast medium, where users can observe all the traffic flowing in the network.

Beyond this shared-medium security problem, there are other security and privacy threats stemming from the link layer, which are discussed in depth in Sec. 3.4. To evaluate the security protocols for the link layer it is worth mentioning that the most pressing threats stem from the usage of a unique identifier, the MAC Address, leading to identification and tracking of the user. Through a MAC address, it is possible to uniquely identify users through their devices, especially as they roam through different networks. These threats are particularly observable in environments that enable roaming and ubiquitous access, such as wireless technologies, where an attacker does not even have to participate in the communication to monitor all the data. These properties require a deeper discussion of such technologies.

Wireless technologies usually resort to broadcast transmission, translating into inherently insecure communication: all nodes located in the broadcast cloud can effectively monitor every packet traveling through the air. This is the default behavior of IEEE 802.11 [66] technologies, offering no guaranties on either privacy or confidentiality. The initial proposal to protect WiFi environments, conveyed by the 802.11 standard [66] aims at providing Wired Equivalent Privacy (WEP). WEP is based on shared key between station and Access Point (AP), relying on the fact that an unassociated station cannot listen to the ongoing communications, without access to the shared secret. Besides being a weak mechanism [140], WEP does not provide any protection from the already associated nodes, forfeiting node privacy and confidentiality. It is easy for an attacker to deploy illegitimate probes to monitor traffic and location of a roaming STA. A later specification amendment, IEEE 802.11i (which is now a part of the base 802.11 standard [66]), commonly known as WPA2, was designed to replace WEP, providing a more adequate security solution. It defines different ways of generating a Session Key from a Master Key, which can be based on a pre-shared secret (shared key mode), or server authentication (enterprise mode). The session key mechanism does not suffer from WEP's insecurities, and provides better confidentiality, since the secret is only shared between one STA and the AP. Neither legitimate nor illegitimate stations can listen to each other's traffic. But 802.11i only protects the payload of the 802.11 messages, leaving the protocol headers susceptible to information collection from any eavesdropper. Under such traffic inspection, stations can be easily tracked, forfeiting their location privacy, along with link layer data, which can be used to derive privacy breaching information. Also, none of the previous solutions conceal the STA global identifiers, namely their MAC address, which uniquely identifies them in any situation, further simplifying tracking by an attacker.

The link layer is also afected by location privacy issues, given that each AP has a fixed geographical position, and the nodes can be placed inside its antenna range, leading to geographical tracking.

It is worth mentioning that several network layer architectures provide location and

identification privacy features [53, 24, 130, 139, 114], as discussed in the following section (Sec. 2.5.2). But despite their effectiveness on the network layer, they are still vulnerable to link layer attacks. Therefore these proposals will needs to be complemented with adequate link layer privacy protection, tackling both location and identifiers.

2.5.2 Protecting Location and Identification

The threats identified on the network layer, considered by most of the presented related work, relate to tracking a user's location by uniquely identifying the user through unique addresses. While these threats are discussed extensively in Sec. 3.4 in the network layer, we can advance that they are associated with the fact that a user is uniquely identified through the IP address on the network. Nevertheless, the address also yields topological location: the hierarchical assignment of IP addresses can be surveyed and mapped, tracing addresses to specific cells with a fixed geographical position. The increase of private and commercial services surveying network locations has led to surprising accuracy for such IP to location services [148].

Location privacy has been mostly addressed in absolute geographic location environments, particularly with GPS technologies [59, 17, 39]. These solutions show that introducing confusion [17] or imprecise location [59] benefits location privacy. While not directly applicable, such solutions provide a better understanding of location privacy issues, especially considering that IP addresses can yield surprisingly accurate geographical position, available from public services [148] (and even more accurate with private services). Considering absolute location issues, an approach to attain privacy introduces imprecise measures (less accurate or relative) in the location metrics, omitting any accurate positioning. Gruetser and Grundwald [59] attempt to reduce the accurateness of location queries both in time and space, to provide location privacy. They show, by measuring location accuracy deviation, that confusion can increase location privacy, thus enhancing the overall user's privacy. After showing how anonymous traces can lead to identification through information correlation (i.e an office location trace can be matched to desk sitting and worker identification), Beresford and Stajano [17] propose an entropy quantifier to show how correlation probability can decrease by performing unexpected actions. The disruption of movement and location patterns shows that entropy increases user privacy. In a similar approach, Duckham and Kulik [39] propose an obfuscation mechanism that introduces inaccurate and imprecise locations in the coordinates sent by a user responding to a location query. They then use "levels of privacy" to measure privacy. Location privacy is achieved by creating an anonymity set that yields multiple locations. Such an approach shows that location privacy does not depend only on the amount of samples available but also on their precision.

Another approach is to conceal the node's IP address altogether, removing location information, thus protecting user privacy, as proposed by Chaum Mixes [22]. In this approach, the message receiver is not able to determine the message sender because the message is anonymized by the mix, through store-and-forward cryptographic mechanisms. This concept has been applied in Onion Routing [142] and TOR [37] using the notion of encrypted virtual circuits. The circuit progressively decrypts the routed packets, according to a layered construct: the packet is encrypted many times over by the sender, and is successively decrypted towards the destination, providing anonymity, privacy and resistance to traffic analysis. However, the TOR architecture favors user enrollment as TOR routers, making traffic reach the network edges, which results in reduced privacy and suboptimal routing. It also redefines several transport mechanisms which were solved in other places of the network stack

(e.g. segmentation). In all, TOR mechanisms are very secure and flexible - route telescoping allows reusable circuit building - but the required full packet encryption along with the drawback of the users becoming the routers, leads to severe performance issues. But despite the performance drawbacks, TOR provides location protection because it conceals any user information concerning addresses.

Blind [149] introduces a security framework that allows endpoints to identify themselves without revealing their identity to eavesdroppers. This is accomplished by an initial cryptographic handshake, and through the usage of forwarding agents to provide location privacy. It is worth mentioning that Blind does not propose any elaborate threat analysis, and focus directly on providing anonymity. However, it requires an entire new mechanism, similar to HIP, that intrudes in the current layered configuration.

It is worth noticing that the negative impacts of location privacy and user identification on the network are expanded when coupled with mobility aspects. Mobility protocols, such as MIPv6 [77] and HIP [112], need to resort to global identifiers for user identification. These identifiers can be used to track the user and his current location. By tracking the mobility updates performed by the user, either a Correspondent Node (CN) or an illegitimate eavesdropper can learn the user's current location. In MIPv6 this is achievable by tracking the Binding Update (BU) messages [44, 60]. Also, given that it is a global identifier that does not change regardless of point of attachment, it suffers from the same threats as the link layer address, which were discussed in the previous section. Therefore, we can only conclude that mobility further endangers the user. These mobility related aspects were addressed by Escudero-Pascual in the scope of a PhD thesis, entitled "Privacy in the next generation Internet: Data Protection in the context of the European Union Policy" [45], baring a title similar to the current document, but with a different focus and format. The focus is on Mobility aspects, first with a solution [42] adapted to the Freedom Network (a pseudonymity oriented network), where the user resorts to pseudonym based tunnels for mobility construction. Second, it proposes a solution that conceals the users location through the usage of proxies [43]. This means that the user is protected from remote peers, as well as from location services in the network, by interposing a location proxy that shields the mobile node.

Paradoxically, several location privacy solutions stem from mobility solutions. Hierarchical mobility solutions, such as Hierarchical Mobile IPv6 (HMIPv6) [135], or localized schemes like Proxy Mobile IPv6 (PMIPv6) [41], reduce the topological information carried by the IP address while increasing location privacy. But, as privacy requirements grow, there is a need to incorporate it in different architectures, and more importantly, future network designs can be built from the ground up having such requirements. Disruptive solutions that break today's network assumptions can provide location privacy, or even identification masking. New architectures or mechanisms are able to cope with location privacy as a side effect of privacy or architectural requirements. IP^2 [117], Turfnet [130], I3 [139] and Blind [149], while not focusing on location privacy, address some of the issues.

IP^2 [117] is able to hide the user location through the use of anchor points in the network which also deal with mobility. This resembles what happens in HMIPv6 [135] and in an architectural instantiation proposed in Sec. 6.5.2), but faces a large deployment overhead mostly caused by features not related to privacy, making location privacy a small-side effect of a complex architecture not properly adaptable to existing network technologies. Overlay networks provide also good approaches to hide location information. In Turfnet [130], location privacy is achieved implicitly mainly due to an innovative method of routing and the use of Turfnet Gateways connecting each Turf. However, it is difficult to achieve optimal routing. In

I3 [139], a new realm for routing is defined based on names. Using a rendezvous point between communicating peers, it is possible to achieve some degree of location privacy, but it is still an overlay network and we argue that location privacy should be achieved through architectural support. Blind [149] describes a complete identity protection framework for endpoints. It proposes a Diffie-Hellman authenticated agreement for identity exchange. Regarding location privacy, a solution based on identity aware Network Address Translation (NAT) is proposed: when an endpoint tries to initiate communication with a node in the network, it uses a Forwarding Agent that selects a virtual IP address for it. The peers are able to see only the virtual address, not the real address of the endpoint. However, it does not contemplate security between endpoint and Forwarding Agent, or has support for mobility. The Layered Naming Architecture [11] aggregates several existing solutions, providing a unified and integrated system. The LNA introduces two layers of names: for service identifiers (SIDs) and for unique endpoint identifiers (EIDs). Both of them are independent of IP addresses. Since LNA is partially based on HIP, it also incorporates the notion of identity, although diluted. The two introduced layers have the objective of providing a decoupled view of different layers, and do not aim at providing a tight integration with identity, even though LNA presents a first step in abstracting sets of identifiers that are not affected by mobility.

2.5.3 Pseudonymity based solutions

Several proposals exist in the literature on the usage of pseudonymity across different layers to ensure privacy. In all of these proposals it is important to consider: i) how pseudonymity systems are controlled; ii) to what granularity they are applied; and iii) if any sort of evaluation or justification is presented on such approaches. It is also important to understand if they are used in conjunction with any other technique, given that privacy is a cross-layer issue.

"True Anonymity without Mixes" [111] is one the first proposals of using pseudonymity in the network. It argues against using mixes, since most mix based approaches do not protect from the mix itself, when nodes are colluding. It argues towards using a non-personal, temporary, random identifier IP address. It also tackles the identifier problem of link layer protocols that rely on 48 bit MAC addresses, and finally resorting to random MAC addresses with vertical integration for service consumption. Without going into lengthy architectural analysis, the authors propose a few pragmatic solutions to handle network related privacy. There is an implicit notion of a vertical threat to privacy expressed in the proposal of simultaneous link, network and service solutions. However, such vertical awareness is neither acknowledged nor systematized.

In the context of network layer pseudonyms, i.e. using multiple concurrent IP addresses on the same host for different purposes, Flasche [151] stands out by using the already discussed Freiburg Privacy Diamond. It puts forth a complete system based on location dependent addresses, associated with a control layer that manages addressing. Flasche is rather flexible, enabling the creation of virtual interfaces when the need for anonymity appears, along with location based addresses tied with an application model. However, it does not focus on how to control the management of such devices, and hence, on controlling the multiple privacy dimensions that the users require. Although FPD [150] presents a potential implementation roadmap, it does not detail actual behavior or performance. Addressing impacts are discussed, based on collision probability of adding more addresses to the network, which is a first step in the direction proposed in Sec. 4.4.

On a different track, virtualization software is accidentally providing network pseudo-

nymity. Software such as VMWare [146], VirtualBox [145] or Xen [12], provide network configurations that enable a virtual machine to appear decoupled from the host. This could possibly allow running one virtual machine per identity or application, but it is hardly scalable, especially considering mobile devices such as smartphones. A recent addition to the Linux Kernel (Network Namespaces [86]), part of virtualization enhancements, could be used for pseudonymity purposes. However, it has a strict application of namespaces, and is not connected to the concept of identity in upper layers.

Until now, there was no complete study that demonstrates the impact, feasibility or performance of pseudonymity based systems. Without a well-defined analysis process, including practical and theoretical aspects and a clear set of requirements, it is hard to assess the advantages or drawbacks of such (or any) solution. This leaves a wide theoretical and practical gap that is explored in Chap. 4.

2.6 Summary

As we start lifting the veil on privacy, we come to the conclusion that we must define it to provide much needed boundaries for our work. But, the most important lesson to retain from the different views on privacy is that providing overreaching concepts can lead to failure. We must understand that privacy is a multidisciplinary subject, and often, it involves a plethora of considerations, stemming from different environments. From this, we need to scope and contextualize our privacy landscape, in order to achieve any meaningful definitions, something that will be carried out in the next chapter.

The followed approach, as discussed, was to study the social, legal and technical aspects of privacy, so that they can serve as boundaries for the topics on network privacy. And while these studied concepts may not provide the indented focused definitions towards the network, they can definitely provide guidelines that enable us to reach those definitions. And while we saw that, from the social aspects, privacy is a concept in disarray, we can study the legal and technical aspects without prejudice from this confusion.

The legal guidelines, or directives, studied in the scope of network privacy focused more on data protection and privacy as a fundamental right. After analyzing all of these directives, it is important to retain that they deal in abstract terms, establishing regulatory requirements for communications, as well as privacy. It becomes clear that objective of such directives and regulations is to define how data should be handled by the different parties involved in communications. Nevertheless, the focus seems to be on what is lawful and how it should be carried out, especially dealing with user related information. It does not set many requirements for network mechanisms. But, it does present a corollary requirement: if only legal, authorized or mandated entities should be able to inspect, collect or analyze user (private) information, every other entity that does not fit this mold should be promptly denied access to such information. This means that the current mechanisms yielding information fitting the aforementioned description, to either eavesdroppers or legitimate network peers, are violating user privacy and should be treated as a flaw that must be fixed. Beyond this corollary, the provided information also provides a wealth of definitions, conditions, and limitations around privacy, data protection, and how to deal with it in the modern world. The focus now is put on privacy in the technical aspects, that must comply with all the discussed guidelines, and more importantly, to frame them in models that abide by these limitations.

In our privacy analysis we reviewed several models, such as database, anonymity, Bayesian

and even network models, coming from different privacy fields. There are common aspects to them, where the first is that the definitions they offer are usually vague, and not fitting a pragmatic network privacy model. Most of them deal with larger contexts for privacy, focusing on the entire user dimensions, and not just on the network. So, at a first glance, we do not find the network focus we required. The database models provide an interesting view of information correlation than can be helpful when looking at the network, given that relationships can be established similarly and provided we consider the network as a dynamic and evolving database, thus contributing in several ways towards a solution space. The presented anonymity models do not try to model the same thing. They aim at providing anonymity towards the end user, allowing us to understand clearly that anonymity is not privacy. As for the network oriented privacy models, while very interesting, we come to the conclusion that we still do not have the required tools to reach the goals of this Thesis.

The presented models have a consistent problem of providing a level of detachment from the network, and can become abstract and hard to transpose to practical mechanisms. Specifically focusing on the most notorious solution, FPD, it seems that while it captures generically the essence of the relationship between user and network, it does not provide any insight towards the actual means of providing those relationships. This is the missing step, that makes a large difference, and can become the most important part, because it leads to the materialization of the models. Also, what seems to be a common approach is that the presented models tend to alleviate the process of information capturing by determining that only absolute deductions from information can be processed. By neglecting complex or event probable relationships, most models can be simplified. This view is incoherent with todays advanced data mining infrastructures and processes, which tend to go beyond simple information inspection. This is a lesson extracted from the database models, that expose liabilities present in other models.

As a general remark regarding the privacy models, we can only conclude that none of the presented models provide a perfect environment to tackle network privacy aspects. While they can serve as guidelines on how to build a new model, they need adjustments, such as a more focused or pragmatic approach.

Another place that we can look at for privacy guidelines are the existing technical solutions for privacy. In the search for technical solutions for privacy, we started by exploring the relationship between privacy and user, in the form of identity. The conclusion is that, by centering on the user, IdM is acquiring the necessary tools to provide a consistent approach to user privacy. While the solutions discussed are all similar in nature, they all advocate the same principals, guided in part by the conceptual notions presented. The conclusion we gather is that the user, and its identity, are becoming an omnipresent part of communications. Despite the solutions presented are limited to the application layer, i.e. interacting with services, it seems that the mantra they advocate of protecting user privacy and information can be applied generically to all vectors of user interactions. Because of this, we tried to understand to what degree identity information is already present in the network. Our conclusion was that, while only implicitly, the user identity is becoming reflected in the different protocols through unique identifiers that lead to a user. This seems both a hazard and a benefit, because while improper use can breach privacy, proper use, as proposed by IdM technologies, can lead to privacy protection.

There is no consistent threat model behind these solutions, rather than protecting user privacy. In many cases it does not directly match the network aspects or the privacy models discussed before, which only touch when discussing database privacy models, given that user

information can be stored in databases. An interesting view that stems from these observations is that a model that captures the network essential threats in a user-centric view, as is the case with IdM, would be a very beneficial tool. This would enable the identification of network processes and their relationships towards the user, filling in the gap of existing models, and providing the tools for understanding network privacy.

The definite step towards understanding network privacy is to study privacy solutions. For this, we studied three different solution spaces that aim at providing privacy in the network. First, we focused on link layer privacy, where we observed that current technologies are more suited towards providing security in link layer interactions, than actually providing privacy towards the end-user. When studying technologies which protect the network access, we saw that they protect the network aspects, but ignore the privacy issues that make users trackable and identifiable. Similarly, they leave these efforts to network solutions, that can be compromised by link layer mechanisms. We realized that there is a lack of tools to determine the threats, and also to understand them on a very pragmatic level. It seems that identifiers on lower layers do indeed compromise upper layers, such as link layer addresses compromising the network layer, something which is frequently ignored.

On the network layer, we observed that most solutions aim at protecting location and identification, without much consideration for lower layers. The solutions regarding location tend to use confusion as their main tool, and those regarding identification privacy tend to use encryption as the preferred mechanism. While these are the two primary threats, they are often not consistent from solution to solution, considering different aspects of multiple protocols. There seems to be, however, the notion that privacy affects all layers, and that while they solve part of the issue, there is a bigger issue regarding the entire concept of user privacy, spawning across all layers. It was also interesting to see that some of these threats stem from the close relationship between layers, e.g. network and transport, and solutions that decoupled them are providing solutions that can improve user privacy as a byproduct of futuristic architecture design without a clear focus on privacy. We also observed that mobility aspects, common in NGN, generate more privacy threats that enable simpler tracking and identification, thus accentuating the network layer threats. While solutions exist, they do not follow any particular model, and have almost no consideration on how they affect either the network, or the remaining threats.

One trend that can handle several different layers in the network, discussed in Sec. 2.5.3, is pseudonymity. The ideas coming from the notion that we can generate multiple identifiers for the same user showed that it can be applied to more than one layer. We saw it emerging on IdM solutions, as well as in some models that try to propose joint link and network layer solutions. This seems to be a very reasonable approach, but most solutions analyzed do not have any guiding concepts in terms of privacy, rather than generating multiple identifiers, which makes the feel incomplete. While we acknowledge that this may be part of a solution, it is not determined how it can work or how it integrates into a larger view of user privacy.

The conclusion is that, while privacy definitions exist, there is no concrete model for network privacy. This becomes evident as we explore, first, the models that govern privacy, and second, the solutions that currently exist. Most of them address different threats, with different solutions, but without overall commanding principles, making them niche solutions. There is a consistent pattern that shows that different threats require different solutions, but all deal with the user. This suggests that there is a two-vector characteristic to privacy that deals with the overall privacy of the user, and with the particular aspects of each network mechanism. This is discussed in-depth throughout the following chapter, exploring what is

referred as the vertical and horizontal approaches to privacy. However, we first need a model that guides our privacy approach.

Chapter 3

Modeling Privacy in Network Environments

> Great things are done by a series of small things brought together.
>
> **Vicent Van Gogh**

It is important to have clear definitions, as to avoid confusion when dealing with the different privacy solutions discussed throughout this document. For this purpose we present our own definition of privacy and associated concepts. Using these definitions, we propose a privacy model that tackles the conceptual nature of privacy in the network, and how it can be synthesized into a clear and consistent framework. Using this model, we evaluate the privacy threats that exist in current network protocols and operations, structured around network threats and information relevance of the highlighted mechanisms.

The presented definitions, model, and pragmatic network analysis, result in a high level systematization of privacy protection concepts. We distinguish between protecting privacy and protecting identifiers, and the role they play in extracting generalized approaches towards privacy, which will be followed in subsequent chapters.

3.1 Introduction

As highlighted in the previous sections, privacy is a complex and multi-disciplinary concept that causes technological, legal, social and philosophical debates. It becomes clear that, to properly handle network-related privacy, we must divide and conquer, a strategy that leads to the partitioning of the problem to cover the interactions of users and communication networks.

When considering individual user privacy in modern networks, we must split the concept into different views, each with its own set of both threats and sensitive information, turning privacy at different layers into a tractable problem. To address the subset of issues that relate to network privacy, a logical solution can be to conceive a privacy threat model that covers the necessary assets on the network. There is a need to identify what is the information (and information flows) that jeopardizes privacy, and how it can be properly modeled, specially concerning threats, so that we can later propose a clear set of measures that mitigate such threats.

After analyzing different models that can contribute to user privacy (Chap. 2), it is possible to understand the basic needs of a network privacy model. Some models provide an interesting framework to address privacy in networking environments. Models like the Freiburg Privacy Diamond [150] (FPD) already provide abstractions for evaluating privacy threats, and potentially can cover network based attacks. The 4-way model proposed by the FPD system covers several network interactions: by individualizing the user, the device used for network and service access, the location of both user and device, and the actions performed by the user at the device. This provides a generic model for network interactions and identifies key objects of a privacy model.

But, a recurring problem of the discussed models is the exceedingly generic approach to privacy. This can undermine the applicability of the different models, as they are instantiated into network scenarios, given that we deal with concrete privacy breaches, that translate into bits and bytes on the wire. It can be very hard to go from generic relationships to tangible concepts in the network. To accurately address privacy concerns, we must understand the nature of data flowing in the network and provide a model that correctly describes and isolates the network threats, guiding the potential threat mitigation approaches. Consequently, we need a model that, in practice, addresses the gaps of privacy models, mapping directly into network concepts.

To tackle some of the shortcomings of the discussed models, we propose an approach that covers the diverse interactions on the network, and provide a theoretical and practical set of rules and assumptions that lead to a correct privacy evaluation. In particular, we need a model that focus on users, handling both data and identification, which can grow to identify how their user is compromised. To achieve these goals we must understand how we can relate high level concepts to the specificity of different network protocols. We propose resorting to the notion of events as a bridge between conceptual approaches and practical network applications. This demands that we recognize that privacy (or the loss of it) is subjected to the observation of network related events, which will exist in the form of data packets exchanged between different entities.

Not only can these events yield information on their own, but they can also be correlated to harvest more information about the participants in the communication. The corollary of the aforementioned concepts is a twofold analysis that must be accounted for in privacy oriented models: 1) the immediate information an eavesdropper or peer can extract from the information flowing on the network, and 2) the derivation of knowledge from a continued

analysis of network events. Basing the model on such apparently simple concepts allows a straightforward comprehension of the core principles that can compose a broader model, which can be applied to different levels of the network. The presented definitions allow establishing a clear relationship between the way we perceive privacy (as it relates to information) and how protocols operate.

Supported by the previous concepts, in this chapter we propose a network privacy model that has its foundations on clear definitions, and especially, on a simple information model that adheres to the principles of the above analysis: event information extraction and event correlation. This provides a mechanism for threat assessment that is flexible, yet simple, and permits the re-usage of existing literature and proposals towards creating a network oriented attacker model that shows how privacy can be protected in the network. This broader model is a framework that should be instantiated into a more detailed attacker model, thus capturing the essence of different attack zones along with the varying importance of attacks confined to specific topological areas.

In the process of solving existing privacy model constraints there are several requirements that must be met by a new approach to a privacy model. First, the model must handle identifier based privacy threats. This means that it should be possible to determine if privacy is threatened by an identifier or event observed on the network, since they are the primary starting point for recognizing Personal Identifiable Information. Because each identifier has variable purpose and relevance, the identifier scope must be properly defined and formalized, within the model or the instantiation of that said model, to access their threat potential. This implies that we define that threat space, which can be achieved by defining an attacker model. In this attacker model[1] we delimit the network conditions that are relevant towards applying the privacy model. It is in this space that we must formalize Linkage and Correlation, coupled with the proper definitions providing a clear understanding of what each concept means and how it translates into network observations. Lastly, it is important to consider the practical aspects of the model, such as the network applicability and protection. We should understand how to apply the formalized models onto the network stack, identifying how the stack can be improved, ultimately leading to the protection measures that deal with privacy and identifiers.

We thus divide the modeling of network privacy threats into three separate steps. We provide clear definitions to delimit the scope and comprehension of the defined concepts, within Sec. 3.2. Afterwards, in Sec. 3.3, we propose a definition of the network privacy model that is mostly built around the presented information model, which contains Events and Information Sets, showing the purpose behind these mechanisms and how they can operate. The final step of the model is to extend the conceptual definitions onto concrete network applications, done in Sec. 3.4. Applying the model in the network requires that we first determine the threat space, through an attacker model, but more importantly, we define the threats along with the relevance of the information present on the network. This enables scoping the problem in a well defined context, but more importantly, allows defining privacy protection solutions. This is outlined in Sec. 3.5, where we show how to protect privacy, identifiers and the relationship between them. We then summarize the results of the chapter in Sec. 3.6.

[1] An attacker or attack model determines the amount and scope of information that is available for an attacker to carry out an attack. In our scope, this directly concerns privacy related information available to an attacker.

3.2 Privacy Definitions

The first thing we must do to enhance privacy on the network is to define it. The definition must be as clear as possible, with the goal of dissipating possible doubts with respect to what is privacy in the scope of this Thesis and of the proposed solutions in the following chapters.

To support the proposed privacy definition, we also define concepts and terms that contribute to privacy discussions, both as arguments and contextual boundaries that frame the different proposed approaches. These terms can sometimes be the source of doubt, and experience shows that the presence of ambiguities can undermine privacy discussions. Therefore, when discussing privacy, at any layer or context, it is always necessary to agree on a common lexicon, since several terms in the privacy terminology can have different meanings, specially when traversing contexts. Acknowledging these limitations justifies our effort to provide clear and simple definitions that fit our privacy model and assist in its comprehension.

To understand our approach, we must first explain that we propose identifier based definitions for privacy (and associated concepts), contextualized in network operations, which we think is the most adequate strategy towards addressing the problem. For us, an identifier is any single piece of information that can uniquely map towards a subject (e.g. a person or device) within a specific scope, thus containing or indexing identifiable information. This will be better understood as we present the model, but for now, it will serve as the base for our privacy definition. We present two different sections, one dealing directly with privacy, and another dealing with two privacy related terms, which are often overused in privacy discussions with different meanings: linkage and correlation.

3.2.1 A Privacy Definition

We propose a definition based on two different aspects: first, the disclosure of information, and second, the content of the disclosed information, as relating to a subject (user). In the scope of our work, our privacy definition is the following:

Privacy is the property of retaining control over the disclosure of identifiable information on the network.

The most important part of the definition relates to the disclosure of information, focusing on privacy loss or unwilling disclosure of private information. K-anonymity [141] states that "a disclosure means that explicit or inferable information about a person was released that was not intended." We agree with this definition, which is a ground assumption for privacy, and add that the information release is done through the network (thus voiding any other contextual scopes). Therefore, whenever an unwilling disclosure occurs, i.e. that cannot be controlled by the user, there is a privacy loss to some extent. It is also important to consider that our definition is almost a reverse definition, since privacy is highlighted as the property of preventing or concealing disclosures (either involuntary or forced by the network). The underlying assumption is that we can only lose privacy, as privacy should be considered the initial state (which will be implicit in most of the proposed solutions), if we are to protect or even trade it.

The second part of the proposed privacy definition, highlighted above, deals directly with identifiable information. Here, we turn to identifiers as the main driver of privacy definitions as highlighted before, and that will be handled in detail within the privacy model (Sec. 3.3). However, it is important to notice that we do not include "personal" on purpose, as in Personal

Identifiable Information, because a large portion of the targeted information might not even relate to the user, but rather to network properties and assets that are indirectly related to the user, thus better matching the concepts in the network, and following the disclosure definition on explicit or inferable information.

A corollary of the proposed definition is that privacy must hold in face of network properties, implying both immediate or continuous observations. The objective is to include information that might be observed atomically on the network or inferred over several observations, leading to private information that can either be atomic or extracted (mined). The user consequences of this is that information can directly link or implicate the user, or alternatively be correlated with other information to establish the link to the user. Such distinct actions bring us to the definition of the terms linkage and correlation.

3.2.2 Linkage and Correlation

Establishing relationships between observed information, what it means and to whom it pertains, can lead the unwilling disclosure of private information. In this context, linkage and correlation become important concepts as they define the means by which relationships can be determined, a key aspect of the privacy model presented in the following section (3.3). These concepts have been used before in privacy related models [88, 150], either using aggregation [88] or inference [150] as means of establishing relationships, but with no clear definition.

Given their central role in the privacy model that follows next, it is important to provide clear definitions of linkage and correlation. Stemming from their use in different contexts, these concepts often have competing definitions, colloquial and mathematical. We must analyze both types in order to define these concepts in the scope of (network) privacy.

Colloquially, linkage is often defined as the act of linking (*Mirriam-Webster* [108]), or as the relationship that connects (or ties) one thing to another. The link itself is presented as the connecting element association, correlation, or even a causal, parallel or reciprocal relationship. Also in a colloquial sense, correlation is defined as the mutual relation of two or more things, as well as the act of correlating or state of being correlated. From a mathematical point of view [82], only correlation seems relevant, in the field of statistical analysis and probability theory, as it is a measure between two random variables, denoting the strength and direction of their (linear) relationship.

These definitions enable us to assess that the usage of linkage and correlation in privacy mostly relates to their colloquial meanings. Linkage relates to establishing a link between observed information, which can be a relationship or shared property, creating potential privacy threats. Similarly, correlation most often relates to its colloquial definition of establishing a relationship, but its mathematical bias often points at a complex nature, as opposed to the simple and direct nature of linking.

In a privacy context, linkage is often perceived as the observation of links, whereas correlation is presented as the act of establishing a (non-obvious) link as an inference[2] of observed information, as shown by Lunt [89]. This is supported by the K-anonymity [141] work on relational databases, stating that "to draw an inference is to come to believe a new fact on the basis of another information."

[2]Inference is assumed as the process of deriving strict logical consequences of assumed premise [108]. Mathematically it is very similar, and stated as the act or process of deriving a conclusion from premises [82], resorting to either deductive or inductive logic.

These definitions view on linkage and correlation, stemming from different sources, all contribute to the definitions used in the scope of our work:

Linkage is the process of extracting relationships from information directly observed on the network. Linking can be achieved in several ways, but always built upon single and obvious observations.

Correlation is the process of extracting new relationships based upon preexisting relationships. While correlation can be carried out in different ways, it always defines creating non-obvious links based on information that was previously observed on the network.

The key difference between linkage and correlation is that, while linkage refers to deriving links from simply observing network information, correlation deals with creating new links by analyzing preexisting ones. This is further highlighted by the different natures of linkage and correlation, reactive and proactive relationship establishment, respectively. In the case of linkage, the relationships are established by reacting to network observations i.e. factual observations on the network. Conversely, correlation means a proactive search for relationships based on already observed link, where we are not simply reacting to new information observations, but actively searching for relationships based on the facts we have already collected. In this light, a non-obvious link can be defined as a link based on preexisting links rather than network observations.

These definitions outline two major approaches that can threaten user privacy. One is simply observing network information, and extracting the observed relationships e.g. transported identifiers, using simple yet effective means to void user privacy. The second approach, stemming from correlation, creates a more complex privacy threat and involves relating existing information to threaten user privacy e.g. building dependency graphs to establish a probability relationship between different pieces of information, where linkage and correlation are tied to inference[3].

As we establish relationships, obvious or otherwise, between different pieces of information, we are drawing new conclusions based on the observed information. These links are privacy threats that originate from network observation, and it is in this scenario that we highlight the definition of linkage and correlation. These key definitions help to scope the privacy problem, and bring the attention to the network stack, where privacy can be compromised just by careful observation and through simple (linkage and correlation) techniques, which are modeled in Sec. 3.3.

3.3 Modeling Privacy

Digital environments are mostly characterized by information blocks, exchanged and stored across the network. As the basic information units that can take any shape or form, they depend on context and purpose. This is particularly true in packet based networks, where information flows in well defined finite sized datagrams. Whenever information can be extracted from those blocks, or even from the relationship between different blocks, privacy is threatened. In this scenario, private information stems from attributing generic data onto particular and identifiable subjects, which are the users or references to the users. When this

[3]For more information on mathematical induction, and other abstract mathematics topics, the interested reader should refer to *"Foundations of Abstract Mathematics"* [82]

connection is possible, we are in the presence of potentially private information that must be protected. There are different models for privacy, as noted in Sec. 2.3 where we discussed databases, anonymity, network and bayesian models. Most of these do not deal with the network, and even those that do, only focus on high level interactions. This exposes a large gap when it comes to direct applicability on the network, both for protocol understanding and vertical network interactions, which must be bridged to take full advantage of the proposed approaches. The requirement for network privacy mandates a network oriented model, that enables easy identification of privacy threats on the network level. This implies that the network semantics must be understood, along with potential implications of the network operation models on user privacy.

To provide an effective model for privacy, concerning network aspects, we must look at packet based exchanges and realize that they are event driven (a packet occurrence on the network can be thought of as an event), and assume that all the observed information is, in different ways, indexed. This is achieved by associating a data block with an identifier, either by explicit relationship or by context[4]. This approach generically covers data oriented systems that deal with blocks of data. In network protocols this paradigm occurs on several levels: we match the information that flows across the network to different identifiers conveying origin, destination, data type, and the data itself. But a common shortcoming is that these information systems (and networks) have no concrete representation of a user. Instead, most just rely on references to the user (e.g. a handle, an email), usually pointing to a user profile that might have or not a legal entity behind it. From this observation we can infer that, in digital systems, a user can be represented as a set of references aggregated at some point in the system, usually referring to a person, group or entity.

Combining the two aforementioned complementary views can the basic tools from which we can outline an information model: a contained view of information (data objects) and an identifier-based view of users in the network (references to information). This approach can provide the tools to understand the nature of information relationships on the network, in a concise and reproducible model. It handles information as indexed and related objects, with relationships between them, independently of how they were created, i.e. the means used to establish links between observed information.

Based on the idea that it is possible to aggregate information into particular sets, through relationships between them, we propose the Privacy Event Driven Model (PRIVED), that uses the concepts of events, information sets and the relationships between them as the threefold drivers for privacy breaches on the network. However, we must further understand how the proposed representations (events, sets and relationships) can actually be used to model user privacy.

3.3.1 Information Model

The privacy related models explored in the previous chapter cover many privacy dimensions, either considering threats or preservation relating the end user, in different context and scenarios. However, in the light of the privacy, linkage and correlation definitions presented before, we conclude that most of them do not deal with the potential network instantiations that we expect in our approach. While they present a rather amorphous view of user and information, useful for defining theories and conceptual work on privacy, the direct network

[4]Implicit or contextual relationships are defined by the data itself conveying unique and identifiable information, such as references to unique data blocks.

applicability of these models can still elude us, given that they do not focus on the practical events that occur on the network.

We need a model that focuses on privacy in the network, relying on the presented definitions. We propose a data model and concepts that enable an instantiated network view of privacy, in order to specify both attacks and protection mechanisms around network information. In the effort to define the model, it is important to understand how to actually represent information blocks within it. We split the representation of information into data and identifiers for that data, as described below:

Data: The data, or data block, is a generic piece of information that by itself has no particular meaning towards specific subjects. This piece of information, when coupled with a subject (directly or through a reference), can constitute private information, and therefore can threaten the privacy of the subject, but without a subject it can be considered innocuous in terms of privacy.

Identifiers: The identifiers are references that represent a subject[5]. These references can be perceived as properties unique to the subject, allowing their identification, and when associated to a data block, provide information that relates to the subject consequently threatening their privacy.

We come to the conclusion that information on the network is usually a two part observation, consisting of data and identifier. This tuple, composed by related identifiers and data, defines an important support concept for the model: the **information block**.

These data and identifier constructions, combined into information blocks, allow the definition of an information model specifically geared toward capturing the essence of this bilateral existence. While the contents conveyed by the data block might be relevant for privacy, we argue that, privacy-wise, the important part of the information directly relates with its identification. In computer networks and most information systems, information must be indexed. This alone usually conveys a uniqueness to the handled information, which can be tracked back to the owner, threatening his privacy. To be relevant, private information must always be associated with a particular subject or user. This view of identifiers is further supported by the proven use of *Quasi-Identifiers* [34], which shows that identifiers can be extracted from information in database tables that can later be used to establish (and destroy) relationships between different tables as used for K-anonymity [141]. K-anonymity further supports this view over information, given it is entirely based on information tuples that can be approximated to *Id, Data*. In our work, we propose to leak these definitions to the network view of privacy, properly harnessed in the information model.

This leads to a clear separation of the privacy problem into identifier based threats, and data mining privacy threats. Such partitioning enables us to focus on the identifiers based issues, reducing the complexity associated with information analysis. Nevertheless, when such analysis is performed, through pattern recognition, advanced syntactic analysis or any other data mining strategy, we assume that the outcome is a properly tagged information block, with an identifier for the collected information that can later be used for retrieval or identification purposes. The result is a recursive approach that again lead to an identifier and

[5]The subject, either a user, a group or a (legal) entity on the network can be abstracted as an identifier or a set of identifiers. At this point in the model, we make no effort to differentiate between subjects and identifiers that reference subjects, given that in the network we mostly observe identifiers and not complete subject information.

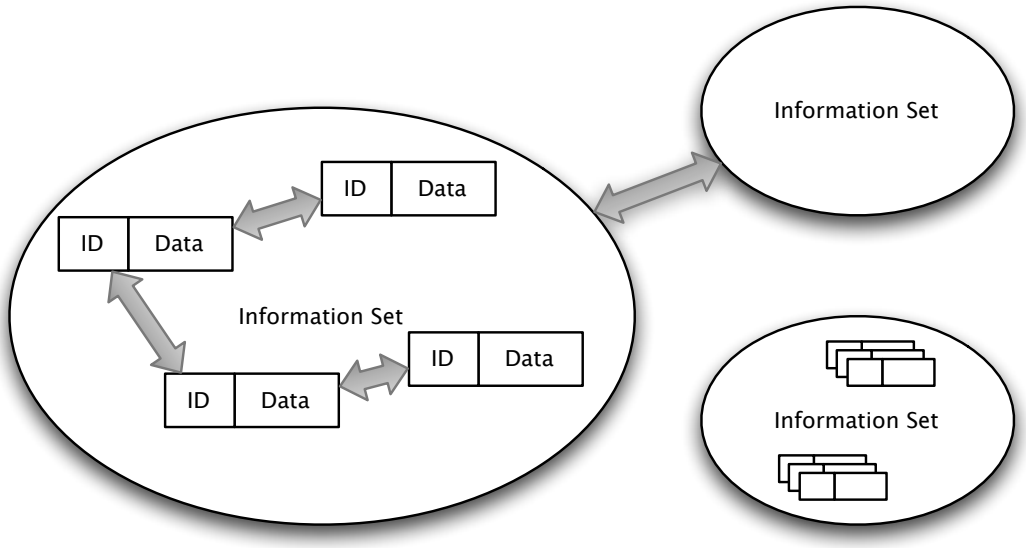


Figure 3.1: Abstract information model

a data block, restating the privacy threat as an identifier based issue. This allows focusing on identifier based threats, without disregarding more complex techniques for privacy threat. Such an approach to information is the cornerstone of the privacy model, which can now focus on describing the relationships between identifiers and subjects (end users), and even between different information blocks.

While there is usually no concrete representation of a user in digital systems, for privacy this should exist to allow modeling the information around the user, given that the user is the primary stakeholder in the privacy discussion. We propose that, digitally, the user can be regarded as a set of identifiers (along with their respective data), i.e. a set of related information, either expressed by the user itself or by the communication systems with references towards identifiable information. In this scenario, the user becomes a set of references, composed by identifiers (and corresponding data) that relate to the same subject, building the representation of a user in the information system. Such an approach is shown in Fig. 3.1, where we present the most basic building blocks: the *information block* organized as the tuple *identifier* and *data*. This tuple can then be aggregated into a set, which represents the knowledge gained about a particular user, outlining potential privacy threats.

Each individual information block can be related to others by establishing a relationship between the identifiers. Once this relationship is established - which is where the privacy threat resides - a larger set is built, resulting in aggregated information. This aggregate view provides information on a user, and more importantly, represents the knowledge gained by an attacker around a certain subject, undermining his privacy.

Therefore, we must understand how those relationships are established, from which information they stem from and how they can be maintained over time, resulting in the concept of information sets. To answer these questions we propose a three-fold approach that attempts to mimic the natural behavior of the network. We resort to finite and discrete *Events* that allow creating *Relationships* thus building the *Information Set (IS)*. Below we present a succinct description of each building block:

Events: An event is a finite and discrete occurrence (e.g. on the network) that conveys an information block, i.e. identifiers and data, containing potential information for relationship establishment.

Information Sets: The aggregate set of information that is composed of different information blocks, observed from the aforementioned discrete events.

Relationships: Relationships are the links established between events or information sets, through their common information blocks and properties (e.g. identifiers). Observing the relationships between information blocks defines the threats that undermine privacy lays as well as the assumptions for building the information set.

While these concepts are explored in the following sections, it is important to notice that they are based on the operations observed in packet based networks. Events matter because they can model the occurrence of information in the network. As finite discrete events, packets convey observable information for attackers, associating private information back to the user, and thus defining the baseline for any attack. Through the information conveyed in a single packet (each can carry multiple basic information blocks), it is possible to establish relationships that breach user privacy. By concatenating several packets (or events) through the relationships between them, it is possible to build a larger information set that will surely include private information about the user, that was not intended for public release. And yet, through simple assumptions and concepts, user privacy is voided in many dimensions, especially on the network.

To better understand the model, we must discuss the events, information sets and relationships between them in greater detail, allowing us to bridge these concepts to the network.

3.3.2 Events

Packet-based networks are defined by the exchanged of small, finite, data packets, the most atomic unit in such technologies. In the proposed privacy model, we adopt a similar strategy by introducing an atomic unit: the event. The most simple definitions is to classify an event as the occurrence or observation of an information block. This information usually appears fragmented, and when properly identified, can be used to build a steady stream of information about users, devices or even networks, closely coupled with packet network concepts. The identifiable information blocks allow modeling events as providing the most basis for establishing relationships between different pieces of information. As a result, an event can be described as a tuple of an identifier, an optional subject that can be omitted when unknown or when similar to the identifier, and the associated information.

The potential threats stem from observing the information block, which can convey all sorts of personal identifiable, and even private information. The identifier acts as an index of such information, serving as a straightforward correlation mechanism, regardless even of whether the subject is explicit or not. In most cases, the type of identifier used as index conveys the subject itself. The defined tuple is presented in Eq. 3.1, which shows the formal definition of events in a privacy scope.

$$e_x = (identifier, [subject,]information) \tag{3.1}$$

Because events can by typified, the model is not restricted to any particular instantiation. Therefore, it can be applied to different contexts, entities or scopes, enabling an extensible

environment. Other models only consider a well defined set of subjects, thus suffering from limited applicability. In fact, events can relate to any identifiable entity in the network, such as devices, addresses, protocols or any other type of network information. In other words, an event is a privacy threat because it conveys sensitive and potentially able to be cross-related, information on the network. This is a key step that allows applying the PRIVED model not only to network events, but also to more complex information correlation scenarios, about user behavior.

We defined two possible scenarios for event observation, based on the nature and content of the event. Either it yields privacy threatening information when observed individually, or it can be correlated with other events, maximizing the amount of available information. As a corollary of our event definition, we can determine that information which is not identifiable or recognizable has no meaning and therefore cannot yield private information about a subject.

3.3.2.1 Correlating Events

As mentioned before, events alone can contain interesting information, from a privacy perspective. But, for an attacker, more information is collected when different events are linked together. This can be achieved by establishing relationships between events that can be related to the same subject or user. In many cases, the described event driven mechanism can be compared to stream communication employed by many protocols, such as TCP, where the stream is rebuilt based on the address, protocol and port, and every piece of atomic information is correlated using the protocol identifiers. Consequently, some events appear unavoidably linked, due to the nature of the system where they appear e.g. capturing a network packet at the link layer can yield identifiers for all layers. However, in some scenarios that link is not obvious, acting as probabilistic evidences that can be collected to prove a relationship. These two complementary cases define the basis of event correlation. Depending on how the events are linked, the connection between them might be inferred, rather than explicitly deduced, leading to a probability of correlation rather than a certainty. Even in the most seemingly obvious cases, as in networks, there are factors that erode the event relationship. The analysis of the network factors is deferred to Sec. 3.4.3, where the different network protocols, topologies and namespaces are equated into privacy observations.

In theory, we can determine the existence of a relationship between two events, e_1 and e_2, when the relationship probability, P_r between them is higher than the a predefined relationship factor r_f.

$$P_r \geq r_f \Rightarrow \varphi(e_x, e_y) \tag{3.2}$$

We define a relationship between two events x and y as $\varphi(e_x, e_y)$. The relationship factor can be tuned according to the model applicability expectations, as outlined by Eq. 3.2, with values in the interval of]0..1]. Conceptually, we can define that when a relationship exists, i.e. P_r is equal or larger than r_f, then we can assert a relationship between the two events. Accordingly, when $P_r < r_f$, there is no relationship between two events. However, the definition of r_f has some impact on the perception of the model. When $r_f = 1$, then we only accept "certain" events, leading to a simpler instantiation of the model, where condition matching and correlation requires less effort. All other cases require a probabilistic analysis and evidence collection, typically depending on several observations and factors to build a decimal value for P_r, as discussed below. The advantage of such definition is that the model fits generic requirements and several types of analysis, whether working in absolute

terms (unitary relationships) or with probabilistic relationships (require complex correlation models). It is worth noting that we assume the condition $r_f > 0$ is always true, meaning that we not do consider any unrelated events, which would otherwise pollute the existing relationships.

$$(ID_x = ID_y) \Rightarrow P_{xy} = 1 \Rightarrow \varphi(e_x, e_y), ID_x \in e_x, ID_y \in e_y \tag{3.3}$$

In the scope of this network based approach, it is important to define a corollary to the definition presented in Eq. 3.2, for certain events (with $P_r = 1$) with the relationship determined through identifiers. When matching identifiers that relate to the same subject (e.g. observing the same network addresses on two different events), it is possible to use a simplified event composition. Eq. 3.3 demonstrates that regardless of the value of r_f, which is omitted, whenever there a relationship is defined by two matching identifiers (assumed to carry meaningful information), we can deduct a relationship probability of 1, leading to a relationship between the two events. This will prove particularly useful for network observations, based on identifiers that presumably belong to the user. Interestingly enough, the simplicity of the proposed model enables the definition of correlation as an event itself. This is a recursive definition that state that the correlation of two events yield a third event that can be handled similarly, as presented in Eq. 3.4.

$$\varphi_{(e_1,e_2)} \geq r_f \Rightarrow e_3(e_1, e_2, \varphi_{e_1,e_2}) \tag{3.4}$$

The missing link in event correlation is the definition of both P_r and r_f. The relationship factor, representing the probability value upon which a relationship to be true, is left intentionally open[6]. As for P_r, we have already discussed the situation where $P_r = 1$ ("certain" events), as a true condition regardless of the value of r_f. But, not all relationships will have an obvious nature, especially when stepping out of the realm of deterministic information (e.g going from network identifiers to user preferences). Whenever $P_r \neq 1$, the relationship between two events should be considered a weighted evidence. In these cases the model must foresee that non-trivial relationships can be established. For that, we propose a weighted variation of individual events, treated like evidences[7]. Each individual event will possess a certain contribution w_i to the overall correlation, weighted in a defined probability of p_i for a number n of correlatable events. Whenever n is sufficiently large for the deduced model's purpose, it is possible to check whether the summation in Eq. 3.5 is equal or larger than p_r, leading to a relationship. This approach, using composed or weighted events is highlighted by Eq. 3.5.

$$\varphi_{(e_x,e_y)} = \begin{cases} 1 & \exists P_i = 1, i = 1..k \\ \sum_{i=1}^{n} \frac{w_i}{n} * p_i & \sum_{i=1}^{n} w_i \leq n \end{cases} \tag{3.5}$$

The most important conclusion from Eq. 3.5 is that evidences can contribute with a weight factor, bounded by the observation probability, leading to a probability, that can be matched against r_f. For each of the n evidences, it is possible to determine different w_i and p_i values.

[6]It should be the focus of future work, to determine what values of r_f should be used, in light of the theories discussed in Sec. 3.3.3.

[7]The usage of the term "evidence" immediately suggests the consideration of the Bayesian model, which is used in many circumstances such as Spam detection and on Intrusion Detection Systems.

For an collected evidence i, w_i represents the weight of the evidence type. This is important in different instantiations e.g. on the network, observing a link layer identifier has different semantic value of observing an IP address. The second variable p_i, determines the trusted conveyed by the observation, which can also vary which the model instantiation, where on the network, observing a link layer identifier on a core network might not yield a trustable relationship. The same applies in scenarios where the user could be employing some sort of privacy countermeasures.

In this scenario, if a relationship is not immediately established, it is possible to collect more evidences i, with a corresponding weight factor of w_i, and a trust factor of p_i, resulting in either discarding the collect information, or to establishing a relationship between e_x and e_y. This is shown by the second branch in Eq. 3.5, which states that for n collected evidences, each evidence i, contributes with weight w_i and probability p_i to find the relationship between e_x and e_y. How to actually obtain the w_i evidences should be a result from the discussion presented in Sec. 3.3.3, because it refers to using contextual event information, such as belonging to a set of captured user information, to actually establish the link between events. Nevertheless, the most relevant definitions are in Sec. 3.4.2 were we show how to actually take the network related events and use them to build the information set of them, considering mostly network information. It should be noted that the presented model allows a variable number of event correlation, at any particular weight or correlation factor. For the purpose of this Thesis, these concrete values can be omitted, given that most of the work will revolve around certain relationships, established with network identifiers (as discussed in Sec. 3.4). But, the model must accommodate a more generic approach that depends on several types of information correlation. It is up to specific applications of the model to determine the concrete values of n, p_r or even w_i, depending on the type of information, observations and individual event contents.

The conclusion that can be drawn from the presented definition of events is that such an atomic component provides simple mechanisms for relationship establishment regardless of the level at which the model is being applied. Establishing relationships between events is the focal point of the proposed rules concerning events. In fact, the nature of events allows the definition of several correlation properties used to establish relationships. Events can be correlated by matching the identifier, or subject, or by linking the contained information (or even pieces of it). This indicates a two-fold system, designed to tackle the establishment of relationships through either identifiers or subjects, as is covered from the network point of view in Sec. 3.4. Alternatively, we have a semantic or contextual evaluation that depends on several factors (this is outside of the scope of this work). But, when we focus on specific domains, with well defined roles, these correlation can be greatly simplified. Nevertheless, once a relationship is created, we can apply several mathematical properties to extend it to other events, as is described in the following section.

3.3.2.2 Formal Rules and Relationships

The concreteness of the event-based approach provides straightforward means of establishing relationships between distinct events. Once the relationship is defined, it is possible to apply rules that determine how events relate to the user and to other events. This provide a formal set of rules that can be applied to relate events to each other, based on a previous relationship.

Such an approach has been followed before in a privacy context by the Freiburg Privacy Diamond, which defines a set of rules for relationships and relationship extension. In our case,

the defined set of rules is fairly similar but scoped around events, providing a more focused application on the network. Nevertheless, by drawing upon the rules defined by FPD, we can define a model that takes into account events and relationships, establishing a bridge between *PRIVED* and FPD. This is possible because events, or sets of events, can be used to describe the tools that provide the relationships between User, Device, Location and Action in FPD.

To reuse FPD relationships, it is necessary to describe them in terms of events. Given that the notion of event is not present in FPD, we can use it to bridge the gap of how relationships exist within FPD. Taking FPD nomenclature, a relationship $R_{x,y}$ can be expressed in terms of events, providing a clear understanding of how the relationship is built (through events), and how events can be represented in FPD, thus bridging the two models. First, we can describe the relationship actors in FPD as events which convey the appropriate information. We can represent the $User$ with an event that conveys an identifier that identifies the user e_u. Similarly, an action is an event that denotes a particular action (or a set of events that illustrates a high level action), e_a. The same is true for an event that carries a location, e_l, which can be a packet with an IP address or a GPS coordinate, and an event that identifies a device (e.g. a MAC address), e_d. Now, representing FPD relationships becomes straightforward. As discussed in Sec. 2.3.3, FPD defines a set of relationships that undermine privacy, such as relationship between action and location. The example presented in Eq. 3.6 shows how the relationship between action and location (R_{al}) can be represented by using events, in this case linking an action event (e_a), with a location event (E_l) thus reusing the important relationship privacy work presented by FPD, but eliminating the gaps that do not show how relationships appear.

$$R_{al} = \varphi_{(e_a, e_l)} \tag{3.6}$$

With the above example, it becomes clear that it is possible to use events to represent all of the FPD relationships. This leads to the conclusion that relationships, including those of the FPD mode, can have a pragmatic network application. It is also important to establish that a relationship in FPD terms is a relationship between well-defined event types, that carry either, location, device, cation or user information. Accordingly, we can generalize FPD relationships and formal rules to use generic events. From the set of FPD rules, we conclude that the most important one towards events is transitivity, which we elaborate in Eq. 3.7, stating that if a relationship exists between e_x and e_y, and another between e_y and e_z, then a relationship exists between E_x and E_z, thus showing a linkage between events.

$$R_{\{e_x, e_y\}} \wedge R_{\{e_y, e_z\}} \Rightarrow R_{\{e_x, e_z\}} \tag{3.7}$$

This rule provides the basis for establishing multiple relationships across different sets of information, based only on events. This indicates that the information conveyed by those events can be assimilated into a larger set, discussed in detail in the next section, as a means of exposing user information.

3.3.3 Information Set

We have defined the information flowing through the network as a tuple, composed by identifier and data, where the identifier refers to a subject or user, allowing the construction of an Information Set (IS) around it. From a privacy perspective, the IS provides a representation of the information that can be gathered by an attacker. It is this feature that, privacy-wise,

makes the IS concept very relevant, since by building it, we can define what is the potential view of an attacker on the network, drawing attention to the private information that is being exchanged.

Once the information collection process is underway, if two identifier relate to the same subject, we expand the information set. This is the basic correlation process, where identifiers become the common ground to threaten the user's privacy. In these circumstances, identifiers become a concern, closely related to the uniqueness of the identifier's scope. Each identifier is associated with a particular information set relating to a user.

The entire set of associated information is defined as the IS. It relates to the user, or any relevant subject (depending on the attacker's intent, this can be a host, a service or any other entity), and compounds the whole range of information that an observer, eavesdropper or attacker can obtain. The combined information is anything that can be observed about the subject: it can be anything from network specific items, such as addresses, location, consumed bandwidth, services, to anything that is user specific, such as eye color or height. Nevertheless, we impose the requirement of an associated identifier, because every piece of data must be properly indexed or labeled as in database systems.

The IS description opens the door at an informal definition, which is that of a user. Given that, as argued, computational systems usually have no perception of the user, we can treat the information set as the representation of the user on the system. We can assume this representation as a global identifier, a common handle for the user, with associated information, ranging from usernames to network addresses. One of the advantages of this generic representation is that it leaves space for a user which can be the manifestation of a legal entity, an not just a real person.

But, there is also space for a formal description stemming from events and the relationship between them, leading to a mathematical concept of IS. Formally, we can see the information set as a bundle of related events: whenever the subject of the information observed in different events is (considered) the same, we are building an information set associated to that subject. This relates to the rules defined in the previous section, which establish links or relationships between events. A good representation can come from Set Theory [61], where we can state that an Information Set is the set formed by all the events which share a relationship, as shown in Eq. 3.8, where S is the set of all n events which share a relationship between them.

$$S = \{e_1, e_2, e_3, \ldots, e_n\}, \varphi(e_x, e_y) \geq t \tag{3.8}$$

Therefore, we can model the information relating to the user as the information set built around a particular set of identifiers, which represent the user. The sensitive and private information is the data that can be accumulated into the information set built around recognizable identifiers. This is the information that must be modeled around identifiers and that any privacy oriented solution must acknowledge and protect. With this definition, we can define the threats on information gathering through Set Theory, where the attacks on privacy are built on expanding a knowledge set, using correlation of events.

3.3.3.1 Building the Information Set

To prevent linking and correlation, we have to understand how it can occur within the proposed model. From this, we are in a position to evaluate the best means to prevent or even conceal the relationship that may threaten privacy on the network.

As mentioned above, a good approach towards building the IS is through Set Theory [61]. According to Eq. 3.8, every information set is built upon events that can share relationships with events external to the set. Assuming that S_k represents an IS built of k identifiers, we can outline the rules necessary to expand the IS. Assuming an event e_x that belongs to S_k, and a new event e_y, which does not belong to any set, if there is a relationship between e_x and e_y, then e_y also belongs to S_k, as demonstrated by Eq. 3.9.

$$e_x \in S_k, e_y, \varphi(e_x, e_y) \Rightarrow e_y \in S_k \tag{3.9}$$

Assuming that multiple sets exist, such as S_i and S_j, we can state that, if they do not share any events, they are unrelated (Eq. 3.10). But, if an event e_i from S_i exists and relates to an event e_j in S_j, then the expanded information set S_k is the union of both sets, as denoted in Eq. 3.11, representing the expanded knowledge. This is not only present in [97], but it is also expressed as a recognition attack in FPD [150].

$$S_i \cap S_j = \emptyset \tag{3.10}$$

$$e_i \in S_i, e_j \in S_j, \varphi(e_i, e_j) \Rightarrow S_k = S_i \cup S_j \tag{3.11}$$

These expansion rules, based on Set Theory, define the basic operations to build an information set. They assume that the relationship between events is binary, leading to one of two outcomes: an event either belongs to a set, or it does not. However, one of the outlined values of the model is the possibility of probabilistic relationships. In these scenarios, where relationships are not necessarily binary, the value of Set Theory becomes less apparent, as it does not take into account weighted relationships.

Depending on how relationships are defined, the φ factor can be less than 1, forcing us to assume a non deterministic relationship. A more adequate model to represent such weighted relationships is Graph Theory. With Graph Theory each event can be represented as a vertex, and each relationship is an edge between two events. As events occur they are mapped onto a graph, with an associated edge and a respective weight. Then, the IS that represents the information collected about a user, corresponds to the vertexes that are connected through edges which have a value greater than a predefined threshold. This graph representation of the IS can be simplified, as it is possible to traverse the graph and coalesce the vertexes which share an assumed relationship, depending on the defined values for φ.

In practice we can assume that the representation is a mix of both graph and set theory, because the coalesced vertexes form a set of their own: when a coalescence occurs, several vertexes are combined into an information set, subjected to the same rules defined for set theory. This occurs with minimal information loss, as the assumption for coalescing implies that the relationships between information blocks in the same vertex are certain (or at least above the coalescence threshold).

While both approaches, Set and Graph Theory, are possible and complementary, our focus on certain relationships, observed through network events, places more importance on set theory. Therefore, it is in this context that we observe how privacy in the network can be defined, along with what relationships can be uncovered on the network, that lead to a certain link.

3.4 Privacy in the Network

One of the key aspects of the proposed model is that it shares a close relationship with the network. Transforming privacy into events, information sets and relationships, mappable to network concepts, provides the basic tools to provide privacy in the network. Whoever, the three-fold approach is not yet completely palpable in network terms. The still existing abstraction gap must be bridged by converting the model concepts into occurrences in the network. The remaining challenge for the model is to complement the abstract definitions with network terms, something that can be done through a practical analysis of the network, given that the presented definitions already have their roots in network behavior.

In network terms, packets can be seen as events: each one has unique identifiers or circumstances, containing data (or payload) which can be linked or mined for more information. However, given that each packet can contain several information blocks, which define the contents of an event, we define on the network a "Packet Event", as the occurrence of a network packet that can contain one or several Events as defined by the information model. It then quickly follows that Relationships are established between packet identifiers, leading to linkage between independent (discrete) packet observations in the network. Basic relationships can be extracted through observation, by looking at the plethora of identifiers carried in one single packet.

From these packet events, more information can be inferred, either from the mentioned data mining on the payload, or by establishing complex relationships between the pattern of events, inferring more information about the user. Finally, the collected information is aggregated into the Information Set, built around the observed identifiers, and forming the notion of a user behind the gathered information. One important parallel that can be drawn is that the identifiers present on data packets can mostly be considered Quasi-identifiers [34], as done for the K-anonymity [141] and derivative solutions [84, 91, 92].

By approximating the model to the network, it is possible to understand the nature of network interactions. This includes understanding the fullness of what relationships can be established, and how, by understanding the composition of the model in a networking environment. Therefore, it becomes possible to witness the consequences of introducing an identifier into the Information set, and its impact on network related privacy, considering that different identifiers have different meanings and bear different weights on end user privacy. We highlight the most meaningful relationships and analyze existing network identifiers, along with the information they carry. Understanding the nature of network based threats will be increasingly important as we study and propose solutions that aim to increase user privacy on different levels. Only in the light of a well structured model can we make such assertions and proposals. But this knowledge must first start on the assumptions provided by an attacker model, which defines the amount of information available to an attacker, threatening user privacy.

3.4.1 Attacker Model

Before diving into the application of the model and potential threats on the network, it is important to establish the attacker model behind the proposed discussions. As we instantiate the model onto the network, part of the realization assumes that an attacker is obtaining information on some point of the network, which will either soften or strengthen the consequences of the attacks depending on where and how they occur. The attacker model assumes

any type of passive listeners: this means any eavesdroppers on the network, at any point, as well as legitimate correspondents, such as services, that can have access to user and network information. This means that the location from where the attack is carried out is not limited, nor is the eavesdropper: it can be either on the access network, either by legitimate peers or eavesdroppers, or it can be accomplished throughout the network, either by the network operator, wire-tapping (illegal or not), and by remote peers, typically services, both malicious or normal.

This defines an attack scenario where the user can be taped on different levels in the network, requiring the evaluation of several network assumptions, identifiers, and consequently multiple threats. However, we can summarize the attacks along the following main lines:

- Passive local eavesdropping, typically on the wireless or even wired first hop.
- Remote colluding peers that can establish information relationships.
- Aggregated local and remote eavesdroppers or legitimate providers, such as the network operator, resulting in an omnipresent-attacker[8] threat scenario.

In fact, the network operator is the entity that harnesses the most means to void user privacy. It has cross-layer information access, ranging from low level network information to location, and even to high level privacy on the services it provides and associated legal contracts. The remaining attackers can be a derivation of the omni-present attacker embodied by the network operator. However, for the proposed work we do not aim to distinguish between malicious or non-malicious providers or passive listeners. In the privacy spotlight, every listener, legitimate or not, can assert or gather information about the user, and by defining privacy as unwilling disclosure of information, all listeners are potential threats.

As we focus on possible threats and solutions, this attacker model frames the potential information threats, as well as the type of attacker that can access and benefit from the collected (private) information.

3.4.2 Applied Network Threats

The primary focus of any applied threat in the proposed model, must now turn to network packets. Special attention should be applied to the identifiers they convey, how they are used, and what information they carry. The underlying assumption is that observing a packet on the network is observing at least one event that conveys meaningful information. This is how data flows across the network, with purpose and intent. The information conveyed by packets is not devoid of meaning, and can relate to the user, either directly or indirectly, contributing to the construction of the IS. This further supports the notion that the information set can be perceived as a set of references to the user, especially concerning the network, where several seemingly dispersed identifier actually create a detail snapshot of the user.

Establishing these relationships between events, based on the conveyed identifiers, provides the missing link on the conceptual model, leading to the notion of packets as events in the model. The transmission of a packet will lead to a piece of information flowing on the network. As network protocols demand, these events are properly tagged with identifiers and types that define specific network interactions. It is this uniqueness that is also observed

[8]The omnipresent attacker is assumed to be the most harmful for privacy, because it can listen and gather information on all points of the network.

in databases [141], that enables the Quasi-Identifier [34] properties of identifiers conveyed in network packets.

Based on these properties, we can assert several conditions that lead to relationships between apparently unrelated identifiers. Being state driven, the network provides intrinsic mechanisms for correlating identifiers and events. Many protocols hold internal state machines based on identifiers. This alone enables relating multiple events. Furthermore, the layered structure of the network creates a co-dependency of identifiers that can be explored for privacy threatening purposes. Because protocols must be transported over other layers, identifiers get reused and relationships are purposely created. We can safely state that everything in the network is designed to be identifiable, given the discrete nature of protocols and (frequent and finite) packets. For that reason, we have unveiled only the tip of the iceberg for network correlation, as we will see throughout this Thesis. But, this clearly shows that there are straightforward mechanisms that provide linking between events, assuming that there are Quasi-Identifiers present in them.

In the network, we can highlight several means of identifier driven correlation, such as time-based, contextual or layered. While these different types of correlation will be explored in greater detail on the following sections, they already allows us to establish guidelines for privacy threats: the relationships between events are based almost entirely on identifiers contained inside the network packets, such as link, network layer or transport layer identifiers, causing several threats. Such relationships can be classified according to two distinct approaches, that result from the nature of observations: relationships can be established based on a single event that conveys distinct information (identifiers), or multiple events related together through network techniques or inference. Further examining this conclusion shows that in the former case there is a vertical relationship between identifiers, while the latter exhibits an horizontal predisposition, when discussed in terms of network layers. This leads to the following categorization of network based linkage:

Vertical Linkage: Correlation of identifiers present in a single packet event, typically by observing multiple layer identifiers, corresponding to several information blocks, in a single packet.

Horizontal Linkage: Establishing relationships between identifiers of the same nature (belonging to the same layer, horizontally on the network), leading to relationships between multiple (packet) events.

These definitions make a concrete difference in the network, due to the nature of protocols and the internal structure of network packets. Therefore, both are described in greater detail below, properly explained and discussed.

3.4.2.1 Vertical Linkage

The proposed definition for vertical linkage indicates that it relies mostly on single packet event observations that contain multiple information blocks. In practice, a vertical linkage or relationship is established by deriving a correlation between a lower layer identifier and a higher layer one. The simplest way to achieve this is through packet inspection, discovering which (inner) identifier is being transported by the outer identifiers. This is shown clearly in Fig. 3.2, where a link layer packet transports a network layer packet. If captured in the access network, this relationship is unique to a user, associating all identifiers (and contextual

information) to that user. Depending on the location of the observation, because different identifiers have different relevance, we can harness information that uniquely binds together several identifiers. This sort of relationship does not require further information to enable its aggregation under a common information set pertaining to a user.

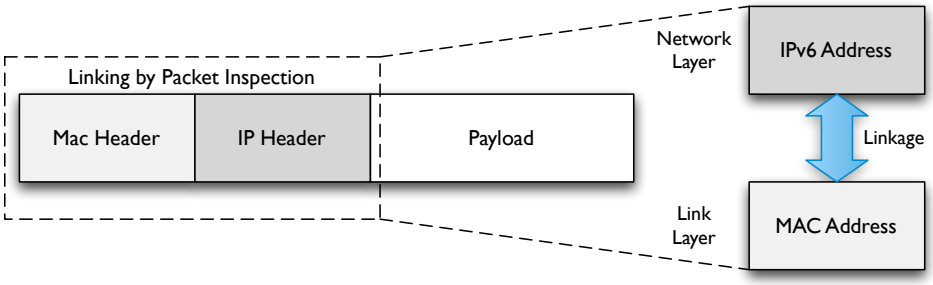


Figure 3.2: Linking between L2 and L3 Identifiers by packet inspection.

In the explained example, two identifiers appear together, creating a lasting relationship between them. In this case, resorting to the proposed model, we can observe that each packet event carries multiple information blocks, or payloads, which sustain the relationship. It is also possible to conclude that it is a recursive property, innate to packet structure, that allows us to establish the line to the conceptual model: because packet events can contain several information blocks, corresponding to a simple event in the model, they can convey multiple correlated events, establishing certain links between them. An example of this property is linking between link layer identifiers and IP addresses shown in Fig. 3.2 where, by packet inspection, a link between the L2 and the L3 identifiers could be established. This can be expressed by Eq. 3.12, which defines a practical aggregation through identifiers, one of the most fundamental linkage rules observed throughout the proposed work.

$$E_x(E_{L2}, E_{L3}) \tag{3.12}$$

Once the relationship is established, it is straightforward to aggregate it to the information set, a process which simply requires finding a set with matching identifiers. In this case, using set theory is sufficient, given that the relationship is considered certain. While yielding powerful and simple results, it is still worth mentioning that vertical correlation is highly dependent on where it is performed. Identifiers have different semantic values depending on their location in the network, and their direct relationship towards the end user is not always relevant. As packets flow through the network, several identifiers can change, specially on lower layers, confusing the relationship between user and identifier. This is where identifier properties become important to consider: the scope of the identifier (i.e. the grasp of its namespace) leads to different privacy-related conclusions depending on the network path it is observed. As an example, consider a Link Layer identifier captured on the core network. The relationship highlighted in Fig. 3.2 would mix a user IP address with an ISP-related MAC address (probably belonging to a core router), yielding no positive relationship between identifiers and with no consequence towards privacy. These properties are further discussed in Sec. 3.4.3, where each individual layer is properly analyzed.

3.4.2.2 Horizontal Linkage

Establishing horizontal relationships means that identifiers in the same layer or protocol are related together, building on the previously collected knowledge. Because they operate on the same layer, identifiers do not appear on the same packet events, thus avoiding the threats discussed as vertical linkage. To establish such horizontal relationships, several collected events must share a property, e.g. a common identifier, thus creating a link between them. The relationship is not as obvious as those which appear in the same packet event, but can still be performed.

It should be possible to establish a link between seemingly unrelated events, by either deducing a relationship resorting to other identifiers (that appear on other layers), or by inferring it due to the context of the layer they operate on (e.g. time based). The most typical case is using a simple deduction through lower layer identifiers: whenever two identifiers on the same layer appear in packet events (that yield through vertical linkage) along with a common lower layer identifier, we can deduce the horizontal relationship. This fits the exact definition of correlation, presented earlier, since it uses already collected knowledge (observed events) as the basis for inferring new relationships. This is useful to correlate higher layer identifiers such as usernames (through IP addresses) or even IP addresses (through common MAC addresses), as presented in Fig. 3.3, which shows an example of linking two network identifiers, B and C, by inference. This eventually leads to a link between the other previously linked identifiers (achieved by packet inspection), A and D. From this process we can extract a single set of identifiers that belong to the same user, or a fully connected graph, as explained before. This is what usually happens with MAC address the network interface card usually utilizes the "original" address, inserted into the firmware of the card by the manufacturer, resulting in the fact that any communication using that device will use the same MAC address. As a consequence, every IP address used on that device can be tracked back to the device, whenever a link can be made between IP and MAC. More elaborate methods for performing inference can of course be applied.

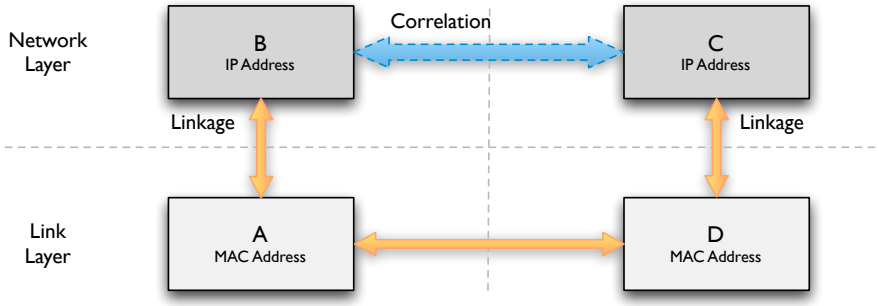


Figure 3.3: Linking the same layer identifiers by horizontal inference.

This is expressed in the model through the functions that enable building the information set: two different events yield a link between them (vertical, through identifiers) when observed in the same packet event; later, if a relationship is built between two similar layer identifiers present in different packet events, the previous relationships can be extended, forming the information set. The importance of these horizontal relationships forces us to understand how such a correlation can occur on the network.

As discrete events, packets appear on the network at different times. This suggests a temporal quality to events and identifiers. The interesting case for horizontal correlation, in the terms discussed above, is that identifiers have different life-spans, which will lead to a time based correlation.

Short lived identifiers (e.g. dynamic IP addresses) only correlate upper layer identifiers during a defined time frame (e.g. the DHCP lease time). Therefore, inferring a correlation between those identifiers is only valid within a well defined period. Beyond that, there is no conclusion that can be made by using the vertical relationship between a lower layer identifier and many upper layer ones. Once the period expires, the assumption is no longer valid, and the relationship is dropped, most likely not to repeated itself in a near future. In the DHCP example, it is only possible to establish such relationships while the acquired dynamic address is in use, because once the lease expires, that address may be reassigned to another user.

However, depending on the identifier type, the lifespan may vary. Some long-lived identifiers can persist throughout longer periods. By lasting for a considerable amount of time (e.g. lifetime of a device), some identifiers are used in an larger amount of communications. The privacy threats that stem from the constant reutilization of the same identifier tied to the same user or device, can facilitate the occurrence of horizontal correlations, followed by larger information sets. Therefore, we can conclude that Long lived identifiers are prone to more correlation threats of (transported) upper layer identifiers. This behavior is exemplified in Fig. 3.4 using network and application identifiers, where the link between A and B occurs in $t_{(0,x)}$, and link between A and C only occurs during $t_{(x+y,t)}$. Through these observations, it is possible to conclude that a link between B and C exists. A practical use case for this is contacting a service, using the same username, through different IP addresses, which can be obtained dynamically through DHCP leases.

Coupling the identifier life-span property with the fact that the network is naturally structured around layers and protocols, we have provided simple means for correlation across several identifiers and layers. This conclusion is further supported by the analysis done later in Sec. 3.4.3, highlighting the individual properties of several relevant identifiers, where lifespan is just one of many interesting properties privacy-wise.

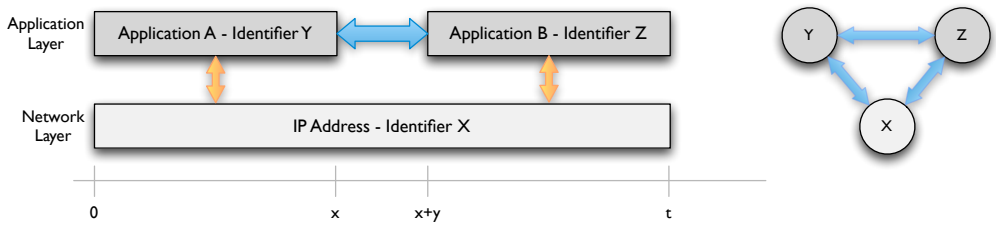


Figure 3.4: Temporal linkage of identifiers.

The so far discussed means for attacking privacy are only valid in the scope of a long lived analysis. If the mechanisms employed by an attacker are not adequate for evidence collection, possibly complemented by off-band tools, the effectiveness of the discussed threats can be almost non-existent. Because these threats rely on temporal properties, they also depend on the type of attack executed, either short lived attacks or longer periods of information collection, used to build the IS. We make an effort not to distinguish between them due to the frequency and simplicity involving the storage and exchange of user information in the network. When

coupled with the fact that there is a long-term "memory" in the network, due to book-keeping, logs or even legal mechanisms, this over-time correlation is not only possible, but probable. In which case, either by direct live analysis, or post-mortem database information mining, such attacks will be easy to carry out. This is how the discussed database attacks are performed, but now scoped to network based Quasi-Identifiers, rather than Personal Identifiable Information.

Horizontal correlation introduces the concept of probable events, since the observed events and identifiers can be unrelated, leading to potential false positive relationships. This is part of the theory described in Sec. 3.3.3, which includes the possibility of erroneous relationships, depending on the strength of the collected evidences. This is possible in any evidence based model, where weaker evidences may lead to wrong conclusions, and stronger evidences better support a correct decision. In our approach, we consider that most results stemming from horizontal correlation are so likely that usually leave almost no margin for error. As an example, the probability of a device using random identifiers and generating a similar identifier as another user (e.g. MAC addresses) is almost zero. These are identifiers that can cause great losses in terms of privacy, where they are applicable. This type of probabilistic correlation is classified as very likely and treated with $P = 1$, in our scenarios, simplifying any analysis based on the PRIVED model.

The temporal correlation process establishes relationships between identifiers by resorting to property that remains static over a determined time-frame (e.g. identifiers). But, it is possible to establish additional relationships by chaining of vertical and horizontal correlation mechanisms on the network, contributing to the information set. Relying on protocol and identifier context, obtained from events and surrounding conditions, it is possible to infer relationships between identifiers (and events). The key issue here is data mining on network information, or on the event payload, to extract relevant private information. In this scenario, the probability of the relationship depends on the mined information and the means to extract it, leading to probabilities than will most likely not be certain. In this case, it is possible to benefit from the PRIVED approach, as it can use conditional and probably relationships, resulting from the evidence based system. And while we focus on deterministic relationships, using probable relationships could prove to be an interesting privacy research path. Network relationships are deemed probable, handled as evidences, and correlated against each other to determined information leaks, and potential threats, as discussed in Sec. 2.3.4. Furthermore, in one way or another, this type of approach is already surfacing on complex policy based privacy negotiation schemes [87], performing privacy inspection on changed information.

3.4.3 Network Information Relevance

With the network correlation mechanisms described before, and the theory behind them in place, we now turn our attention towards the identifiers seen on the network. As stated, we treat network identifiers as Quasi-Identifiers centering much of the privacy discussion on identifier observation and their relevant properties. However, we are yet to determine the individual nature of identifiers that make up for different threats. Unless we isolate what information can be gathered from each identifier, and how that information is relevant in an overall privacy landscape, we cannot fully understand the consequences of capturing or observing a particular identifier on the network.

To provide a privacy overview of identifiers, we look at each individual layer, analyzing the unique privacy and network contributions asserted by each identifier exchanged on the

network. So far, we have looked at network identifiers simply as unique numbers or strings that have associated information. But network identifiers have more dimensions that need to be explored. In fact, the different properties, specific to different identifier types, influence the scope and relevance of identifiers.

In a nutshell, the relevance of the identifiers varies accordingly to the network segment they appear on, leading to the idea that not all identifiers are interesting everywhere. As discussed in Sec. 3.4.2.2, depending on where horizontal correlation is performed, observing an identifier might yield private information or relationships: a cited example is binding a core router link layer (L2) address with an end-user IP address (L3). Based on the previously postulated ideas, we arrive at the definition of a *Locality* property of identifiers, detailed in Sec. 3.4.3.1, which does not exist in the discussed literature. It can be defined as the scope of the identifier, in privacy terms, that is connected to the purpose and namespace of identifiers. Also related to the namespace is the uniqueness of identifiers, a direct consequence of the namespace definition, that together with the locality, defines the important privacy analysis vectors for different identifiers.

An indirect consequence of a namespace is the fact that identifiers themselves have meaning, depending on their network functions. As an example, an IP address not only provides a unique identifier for a host, but can also convey its topological position in the network. The implicit meaning of identifiers must not go unchallenged, at a cost to user privacy, and must be subjected to careful analysis, as carried out in the following sections.

The discussed properties indicate that, for the practical application of the privacy model, the location of the events is indeed important. The places where events are observed can define the scope of a relationship, making it important to define the purpose of identifiers, because they can convey other meanings that influence privacy (e.g. IP can yield geo-location). In practice, we must consider a combination of both locality and uniqueness to define the threats posed by identifiers, specially considering packet observation, without neglecting their intrinsic meaning. That is why some relationships can only be set within a defined spatial bound, as discussed below. As an example, in a core router it is possible to link L2 and L3 identifiers. The result is that every single IP served by that router will be linked to its L2 address, forming a set of all IP addresses that cross that particular router. It is obvious that this does not yield any knowledge gain on an attacker's part, since it bares no meaningful information concerning the users.

3.4.3.1 Spatial Relevance

We have been discussing that identifiers are usually confined to a particular space of action where they are relevant. This spatial relevance, denoted as *Locality*, defines the network scope in which an identifier has privacy meaning. This means that, when observed in an event, identifiers are only relevant in a certain network context, or location, beyond which they can become devoid of meaning, concerning the user that originated the message.

The locality of an identifier is a direct consequence of namespace and purpose. The namespace defines the properties of the identifier, especially considering its uniqueness. Regardless of how identifiers are used, whenever we observe them, depending on the reach of the namespace, there are direct privacy consequences. Therefore, we can classify namespaces according to their reachability: local or global. Identifiers are required to be either globally unique, yielding a one-to-one relationship with a user, or local, where the uniqueness of the relationship is only assured in a confined scope. Despite the namespace characteristics, we

must also consider what is the actual purpose of the identifier in terms of its properties: if it is required to uniquely identify an object within a global namespace, or within a restricted location. While these two properties may seem similar, they are actually very different, and this can be made very clear with an example: typically a link layer identifier is global in nature, due to its namespace, but the purpose of the identifier is just to single out a device on the access link, and is not used beyond that.

According to identifier purpose, it is important to understand how unique an identifier must be, in what scope it operates, and what are the associated privacy consequences. Depending on the locality of an identifier class, it can yield more or less information regarding a possible correlation, especially concerning the location where the identifier was observed. This shows that the occurrence of specific identifiers cannot be decoupled from the location of the event that yielded the identifier. Consequently, to understand locality and its usefulness for linkage and correlation, we must define several network areas where different identifiers bare different significance.

We tackle this issue by resorting to the earlier presented view of Next Generation Networks. The typical NGN approach shows that the network is composed of i) the access, where an access technology is used, e.g. 802.11 or UMTS, ii) the provider domain or local domain, and lastly iii) the global domain, or in a more mundane view, the Internet. These three areas are presented in Fig. 3.5, clearly showing a separation that is not only physical, but logical with consequences on the identifiers. We must analyze each of these network sections separately to understand what information they convey.

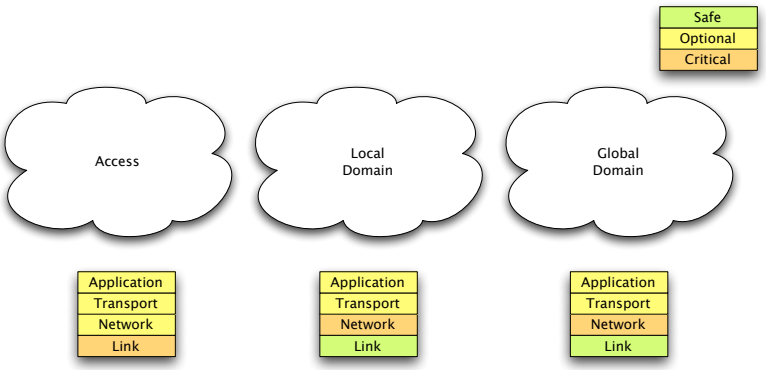


Figure 3.5: Spatial Relevance of identifiers.

We review each layer (link, network, transport and application) on the access, local domain, and global domain. For each of these combinations, we analyze the subject identified, the observability of the identifier, and (as a result of the two previous) the relevance of such an identification. In this analysis, the subject represents the network entity to which the identifier relates; the observability classifies the identifier according to the need to be seen i.e. if is it used by the network entities on that layer observe the identifier, and if not, if it can be concealable (e.g. through encryption); lastly, the relevance, a qualitative classification based on our understanding of the subject and observability, matched against the function of the layer.

3.4.3.2 Access Network

When referring to the access network, we are really considering the network's last hop, sometimes referred to as the last mile. This is where an attacker can obtain the most information, since packet inspection reveals the largest set of linked identifiers. The L2 Identifier is crucial since it is represents the peer (and the user's device) that is using the upper layer communication, i.e. the communicating peer. This means that the packet recipient and transmitter is in fact the real owner of any identifiers contained inside packets, like the IP addresses, endpoint identifiers, URI's or usernames. Also, this is where it is usually easier for an attacker to obtain packets for inspection, since these are usually broadcast mediums, e.g WiFi or Ethernet.

The most important identifier at the access network is the link layer identifier, which in most technologies must be visible. It must uniquely identify a node in the access network, making it very important for both network operation and user privacy. Beyond that, every other identifier can be concealed (e.g. using 802.11i [67]) between the node and the first IP router. But, their importance stems from the fact that they identify the end user, thus gaining an important dimension when not hidden. This is summarized int Table 3.1, and represented in the leftmost part of Fig. 3.5, showing the access part of the network and the important identifiers. This figure indicates that the link layer privacy is critical, while the others are of optional importance, as they can be concealed.

Layer	**Subject**	**Observability**	**Relevance**
Link	Endpoints	Observable	Crucial
Network	Endpoints	Concealable	Important
Transport	Endpoints	Concealable	Important
Application	Applications	Concealable	Important

Table 3.1: Access Network

3.4.3.3 Local Domain

The local domain corresponds to the local part of the network where the topological information on IP addresses is most relevant. This can also be seen as the Network Operator or ISP domain. Here, the L2 identifiers bare almost no relevance since they pertain to network routers and switches, given that this predominantly an L3 network segment where most packets are routed or switched. Therefore, the addresses present on packets have no particular privacy importance to the end-user. Regardless, the addresses used at L3 have a significant topological mapping and cannot be encrypted since routers need to properly process them. In this case, an attacker is also able to inspect them and link the contained information.

The IP address is the most relevant identifier within the local domain. Even when using different IP transport solutions, such as localized mobility, the IP is relevant and pinpoints the node on the network. The remaining addresses are either not relevant or can be protected, as summarized by Table 3.2. This is also reflected in Fig. 3.5, where the user is "safe" from any link layer privacy threat, but the network layer is critical for privacy.

Layer	Subject	Observability	Relevance
Link	Routers and Switches	Irrelevant	Irrelevant
Network	Endpoints	Observable	Crucial
Transport	Endpoints	Concealable	Important
Application	Applications	Concealable	Important

Table 3.2: Local Domain

3.4.3.4 Global Domain

The third part of the proposed division consists on the global domain, which is usually perceived as the network backbone (e.g. the Internet backbone). While for the most part, it can be considered an L3 domain, it does not necessarily imply the same characteristics as the Local Domain. These are usually high speed (optical) links that resort to label switching for fast packet forwarding. Multiprotocol Label Switching (MPLS) tunnels are usually in place (or other protocols for that matter), operating between link and network layers. But privacy wise, it is not assured that the upper layer identifiers are concealed within these tunnels, and could be observed. That means that the identifiers, or addresses, flowing inside the communication channels can contain meaningful information. The L3 addresses seen on packets should be the ones used in the local domain, threatening user privacy, as well as all the upper layer addresses, which could be concealed, as synthesized by Table 3.3. Remarkably, on Fig. 3.5 we can see that the global domain is aligned with the local domain properties, given that the identifiers used are the same, and refer to the same subjects.

Layer	Subject	Observability	Relevance
Link	Routers	Irrelevant	Irrelevant
Network	Endpoints	Observable	Crucial
Transport	Endpoints	Concealable	Important
Application	Applications	Concealable	Important

Table 3.3: Global Domain

This similarity between local and global domain only dissipates in special conditions. On of such exceptions is the when a locator-identifier split architecture is used. In this case, there is the possibility that the conveyed (network) addresses only have a generic topological significance, and might not even always map to the same upper layer identifiers, similar to the L2 identifiers in the local domain. The direct translation of this into privacy terms is that an attacker might not always gather correct information using the L3 identifiers, and needs to dig deeper (end-point identifier, username, etc) in order to obtain relevant linkable information.

3.4.4 Identifiers and Privacy

Exploring the relevant network spaces where identifiers exist showed that there is a discrepancy between the identifier namespace and the areas relevant for privacy information, and most importantly, between the scope of identifier and its purpose: identifiers (or addresses)

which are only used locally, are unique, and therefore establish a one-to-one relationship with the user, revealing a stronger than required relationship.

Therefore, we must analyze what each identifier carries as its meaning, especially considering the properties of the associated namespace, and the contextual information provided by the identifier. This is because, the identifier is usually not used simply as subject identification, but also conveys information about the associated subject (e.g location).

We take a layered approach, and evaluate the most important identifiers on each layer in the network, to better pinpoint the primary threats stemming for network identifiers, and to better understand how these threats can be solved, presenting the user with a privacy preserving environment. By following the TCP/IP model, we focus on four different layers, the ones most used at that convey more information: Link, Network, Transport and Application. We also introduce layers that attempt to offer separation or decoupling between different layers, such as layers often referred as 2.5 or 3.5.

3.4.4.1 Link Layer

The most common identifier used at link layer is the MAC address, which uniquely identifies a network interface, and is usually 48 bits long. Even thought it provides two bits expressing scope and administrative control, these are easily ignored at no cost to the purpose of the identifier. In different technologies, e.g. UMTS, it can be a 64 bit identifier but the meaning is similar. The MAC address could be generated as a random number given that it has no imposed structure beyond its size (apart from the bits mentioned above). But for commodity and uniqueness, a per-manufacturer prefix is usually imprinted in the device, making it easier to track cards by manufacturer and model.

A MAC address has a global scope, but has only a narrow usage. It only has meaning in the link reachable through the network device. In both wired and wireless technologies, the scope of the identifier is limited to the broadcast domain, in which the device must respond to direct queries. Beyond that, even though it is a globally scoped identifier - it should be unique for each device, and consequently user - the identifier gets replaced by the target device, which will route packets based on upper layer information. The conclusion is that the identifier corresponds to a user, which is coupled to a unique device identifier, threatening privacy when captured within the link scope, i.e. before getting replaced at the router.

Related to the link layer, there is an identifier class that is used only for transporting benefits of upper layer packets. These identifiers are used between the link and network layers, forming a wedge layer for special operating purposes. For simplicity, we assume that they belong to the Link Layer because they share the same fundamental properties. These identifiers do not define links with a real connection between them (e.g. an Ethernet cable), but rather a login link layer connection run atop the link layer. Examples of these scenarios are Virtual LAN (802.1Q) identifiers, which are used to emulate different physical networks over the same switching devices. Another example is MPLS, which is used to create switching environments through the backbone networks, serving as a fast switching technology that simplifies the routing mechanisms by using label matching rather than complex longest-match algorithms used on the network layer. These identifiers do not actually identify a device, but rather a network (VLAN) or a logical tunnel (MPLS), and therefore do not constitute a direct user privacy threat.

Considering link layer technologies, we conclude that in the access network, the link layer identifier uniquely references an interface belonging to a terminal, along with the user

associated with that device, binding also all the upper layer identifiers. This association disappears when packets are routed at the IP layer and bridge different link layer domains, where the MAC addresses are replaced, since they are only meaningful within a local scope. But when a relationship with the L2 identifier is made, it will probably remain valid, since the device identifier usually defaults to the card's original MAC address (present in the firmware). This is true until the card is replaced, or the MAC address is voluntarily modified by user or operating system.

3.4.4.2 Network Layer

The primary identifier at the network layer is the IP Address. Either in version 4 or 6 format, its purpose is to identify and locate a host on the network. It is a hierarchical addressing structure divided into two different parts: the network prefix and the node identifier. The network prefix is used to topologically locate the network to which the node belongs. Additionally, the node identifier uniquely identifies a particular node within its subnet. Together, both parts point to a unique peer in the network at a specific topological location. The IP address was originally designed to be a global identifier, and that is the case of both IPv6 address and public IPv4 addresses. But there are exceptions, both due to nature and applicability of addresses. By using NAT mechanisms in IPv4, private addresses can be used to extend the size of the identifier base: one public address can have multiple users behind it. This technique also conceals the topological network information behind the public address, rendering private addresses meaningless outside of their applicable domain. There are also some specific reserved ranges that do not carry any associated meaning, (e.g Multicast addresses), which we do not consider for the purpose of this discussion.

On IPv6 there are special purpose addresses that also have singular properties. The IPv6 Link Local address is used for messages that should not be routed, i.e. only relevant within the link layer domain, used mostly in the scope of the Neighbor Discovery Protocol [144]. These addresses are built using the MAC address (EUI 64 bit expansion), establishing a one-to-one relationship with the device, reusing identifiers.

The conclusion of the previous analysis is that, at any point in the network, an IP address (v4 or v6) yields the topological position of the node, along with the identifier that distinguishes the node from another in the same network. This holds true unless special purpose mechanisms are used (like NAT). This conveys the identification and location of host and user, resulting in both tracing and tracking threats.

3.4.4.3 Network Layer Mobility and Separation

The wedge layer created between network and transport is important enough to deserve a separate discussion. As means of providing mobility, along with other properties that depend on the applied scenario, there are special purpose identifiers that create an indirection over the network layer, enabling IP based mobility, retaining on-going connections as the user roams through different networks.

While this is discussed in detail on Chap. 5, it is worth mentioning that to enable mobility, solutions like MIPv6 [77] and HIP [112] have been proposed. MIPv6 [77] creates a separation layer based on the same namespace with overlapping functionality to the network layer. While the identifiers have a similar structure, they are trackable regardless of location, since multiple locations of the same device (and user) can be aggregated through a common IP address. This

is also true for IPv4. HIP [112] proposes a different approach by creating a completely new namespace. It decouples the node identity from location, but with similar results as MIPv6 regarding mobility. Due to continuously updating location to peers, i.e. IP address, it is possible to easily track the location of associated devices and users.

The two presented use cases show that, even though this wedge layer may seem beneficial at first, privacy-wise it amplifies the network layer problems. Identifiers still suffer from the same privacy concerns that affect IP addresses, but with simplified means to track the location of endpoints (for mobility support). Since MIPv6 relies also special IP addresses, not only can we identify the current location of a node, but also its primary origin, which is given by the Home Address. HIP's Host Identity (Tag) is a globally unique identifier that singles out a node, and is mappable to the appropriate location. Even though HIP acts mostly on the control plane, it unequivocally binds to the location of network addresses, so that it functions in today's networks.

However, this separation can allow the creation of solutions that tackle the privacy problems at those layers, because they enable underlying identifier replacement without breaking connections, like we currently have on most L2 technologies.

3.4.4.4 Transport Layer

As complex as transport protocols may be, this is actually one of the most straightforward layers to analyze. In existing IP architectures, the transport layer identifiers are usually the network layer identifiers. This means that the transport layer suffers from the exact same problems mentioned in the network layer, either when binding to the normal mechanisms or to a wedge layer. Even in the face of these wedge layers, where the transport identifiers used become the wedge identifiers, along with the port information, there is no actual semantic change from one layer to another. In short, it suffers from the exact same issues as the network layer, but with the added nuance of port numbers: port numbers allow inferring exactly what application is being used, since there are well defined ports for well-known applications. This along with topology and user pinpointing, can not only tell who is the user, but what is the nature of its interaction in the network, resulting in an extraction of action through the collection of observed events.

3.4.4.5 Application Layer

The main problem of addressing the application layer has to due primarily with the plethora of available protocols, all of which deal with identifiers and identification in some shape or form. Still, it is possible to single out trends and directives that deal with identification on the application layer. Most Internet related protocols usually deal with HTTP based interactions, resorting to Fully Qualified Domain Name (FQDN), supported by DNS, for most (if not all) types of interactions. Therefore, on the Internet, which has a strong HTTP component, it is common to use Uniform Resource Identifiers (URI) [18] and Uniform Resource Locators (URL). These identifiers imply using an FQDN that requires conversion to a Transport Layer identifier (e.g. IP addresses resolved through DNS) and a service related component. Also, in some cases it is possible to include a user reference in the URI, resulting in the explicit identification of the user.

In order to deal with growing requirements on the naming layer, a new class of identifiers was introduced, the eXtensible Resource Identifiers (XRI), that are an enhanced and flexible

extension of the URI/URL standards. With the added flexibility comes the potential for stating more information about user and action on the identifier, raising privacy concerns.

As mentioned, the application layer is too large to come up with one approach or solution that handles user privacy. As it stands, providing privacy on the application layer is out of the scope of this Thesis, unless it concerns a conceptual privacy approach. Beyond that, it is only treated as a generic identifier threat mechanism on a URI based scheme that must be properly assimilated. Application layer identifiers are usually either application or domain bound, making them uniquely identifiable within a domain or an application namespace. This means that an email address identifies a user within the mail related protocols such as SMTP, POP and IMAP among others or a username at a particular site. The degree of privacy information that is leaked through these types of identifiers relates to the size of the domain to which they apply. However, they tend to be global, since the unique user identification part is usually coupled with its domain counterpart to properly scope the conveyed information, such as an email address like *user-id@domain-name.com*.

3.4.4.6 Identifier Summary

The main conclusion of the analysis on the most significant identifiers in the network stack shows that they are as diverse as their meanings. Table 3.4 provides a necessary summary about the scope and meaning of identifiers.

By inspecting the MAC address we can see that, regardless of having only local meaning, it is globally scoped identifier, as it uniquely addresses a device in the world. Its flat structure shows that there is no particular need for extra meaning beyond an identifier, also falling short on the justification of using a global scope. Similarly, IPv4 is global both in scope and in meaning, as it uniquely identifies a host in the a globally connected network. Moreover, there are variations to this approach based on specific deployment scenarios and protocols, which reduce the scope to local. Nevertheless, the main usage scenario is global. For IPv6, these scenarios are still uncertain, and the current application is global, not considering exceptions or link local addresses. These consideration on IP addresses take care of both network and transport, because ports only add local multiplexing, and do not impact addressing structures.

Finally, if we consider the application layer as merely URI/XRI based identifiers, we can see that their size is variable, as well as its applicability. However, the nature of the identifiers, built using a global FQDN and an identifier unique within that FQDN, leads to the notion of a globally scoped identifier.

Generally, what we observe is that each identifier or namespace possesses different properties that adequate its use to their specified layer. These properties, suited to particular application of network protocols, lead to different privacy related information conveyed by identifiers, as already discussed. But, the most interesting conclusion is that there is a clear

Identifier	**Size**	**Scope**	**Meaning**
MAC Address	48	Global	Link Domain
IPv4 Address	32 bits	Global	Global
IPv6 Address	128 bits	Global	Global
URI/XRI	variable	Global	variable

Table 3.4: Summary of most relevant identifiers and their scope.

overreach in scope. From the discussed identifier, all have a global scope, even if some can be applied in limited domains. The most critical example concerns link layer addresses, which are only used in the local physical network, but uniquely identify a device in the entire world. In fact, many identifiers have a global meaning and only a scoped meaning, out serving their purpose. This shows that there is much that can be done for network privacy, considering identifier structure and the applicability domains of identifiers.

3.5 Network Privacy Protection

The previous sections addressed the process of how to undermine user privacy using a systematic model. In most cases, this is only interesting when used as means to defining countermeasures against the identified threats. In our case, we showed that knowledge can be constructed by observing events and establishing relationships between them, to form information sets. This points at two different steps that undermine privacy: first, the establishment of relationships between events and, second, the aggregation of those events into an information set.

This leads to the conclusion that, in order to protect privacy, we should either conceal the relationships between events or prevent them from being aggregated into an information set (or even combine both approaches for greater privacy). In practice, we can hide the relationships between identifiers by protecting them, and thus avoiding any privacy leakage. Alternatively, we can try to protect the user's privacy as a whole, by protecting information from being aggregated towards the information set. This translates into roughly two segmented approaches with distinct characteristics. On one hand, to protect user privacy we must look at the network as a whole, guided with a consistent vertical view over all protocols. On the other hand, we must focus on a designated layer with specific properties, in order to conceal eventual relationship between identifiers, either through their interactions with above, below between layers.

The only way to achieve this is by understanding what protecting privacy means, what protecting identifiers means and how both can be achieved.

3.5.1 Protecting Privacy

Protecting privacy directly relates to the user, and not to identifiers and protocols, which are merely tools employed by user and network to carry out information exchange. In our approach, we imply that protecting privacy means the protection of the information set associated with the user, preventing it from increasing due to network observed relationships.

While using identifier protection mechanisms necessarily improves privacy on the basis that the identifiers based relationship are protected, the reverse is not necessarily true due to several conditions: i) the correlation can be desired, under certain conditions (e.g. personalization); ii) when communicating with peers or services, some identifiers must be released to identify the user; iii) correlation free environments cannot be guaranteed, in most cases. In an optimal privacy environment, establishing relationships would be impossible, eliminating the need to to define and build an information set based on the proposed model. But, the enumerated conditions lead to the conclusion that first, its is necessary to provide an approach that guides privacy solutions, and only then concern ourselves with identifier protection. As such, identifier protection solutions can be important, as they provide mechanisms that can aid in reducing the information set, but are only one approach. The privacy goal should focus

on reducing the information set, which could be achieved by using the transient identifiers that loose relative importance, instead of focusing on protecting identifiers. The concept of reduced information sets is presented next.

3.5.1.1 Reduced Information Sets

Preserving user privacy must rely on reducing the information set associated with a particular identifier (or user). Given that the user information is aggregated into the information set, a larger set directly results in less privacy enjoyed by the user. Therefore, the first step towards increasing user privacy is to reduce the information set associated with the user. This can resort to any number of strategies, such as identifier protection, discussed below. But, conceptually it is necessary to establish barriers between events. The basic idea is to create vertical shields between information, and instead of one large information set, produce several smaller unrelated sets where the user privacy is maximized. The result is that information previously associated with the user is now broken across different aggregated identifiers, scattering the information. To create these shield, more importantly than concealing relationships, is controlling how they are aggregated. An approach to solve this would be to generate random identifiers that are only related to the intended IS.

Using any of the introduced techniques contributes to fewer relationships and smaller information sets, which can boost user privacy. The optimal situation in information reduction is to randomize identifiers per atomic operation, associating one operation with one untraceable identifier. Nevertheless, to control these relationships, it is necessary to protect identifiers, in most cases. There should be a balance between techniques that provide inner and inter layer privacy, but all under the scope of mechanisms to provide reduced information sets, that are the true threat to user privacy.

To assist the idea of creating vertical shields, that define several reduced information sets, we also use techniques that aim to protect identifiers (and conceal relationships), through vertical and horizontal separation, as discussed next.

3.5.2 Protecting Identifiers

Most threats on network identifiers stem from the vertical or horizontal threats discussed in Sec. 3.4.2. Such links can be avoided by protecting identifiers, thus breaking the relationships between them. But, no matter how commonly the expression is used, there is usually no clear understanding of what protecting identifiers means or what it entails. Protecting identifiers means resorting to special procedures or mechanisms to conceal a particular identifier, therefore protecting its existence. If an entity (e.g an attacker or eavesdropper) is unaware of a particular identifier, that identifier is protected in the sense that it does not yield any information about any user, be it real persons or protocols to which the identifier belongs.

To safeguard identifiers, we can provide specific layers of protection. By applying layer specific solutions that protect identifiers, used or conveyed on a given level of the network stack, we can segment privacy solutions according to what they provide. This can be converted to the approach that we undertook for threats, which is to create vertical and horizontal separation layers between identifiers:

Vertical Separation This describes a horizontal barrier that separates the identifiers on each layer. Any identifier conveyed by the layer in question is hidden, and unavailable

to any attacker on the network or to lower layers. This thwarts vertical correlation threats, by concealing its existence.

Horizontal Separation Horizontal separation means to create a shield between identifiers on the same layer, thus avoiding their correlation through observation on the same network layer. Identifiers should be shielded in such a way that it is impossible to draw a conclusion based on straightforward horizontal threats.

Both of these concepts, can be combined to reduce the amount of relationships observed on the network. When coupled with the reduced IS approach, also associated with randomized identifiers, can provide a strong set of tools to provide privacy. But first, it is necessary to understand how vertical and horizontal separation can be achieved.

3.5.2.1 Vertical Separation

To achieve vertical separation, it is important to create a barrier between identifiers. One example is to use encryption mechanisms on network packets. In fact, encryption at any layer hides any upper layer identifiers, that are conveyed inside the encrypted payload. With encryption techniques, like IPSec, it is possible to create an opaque tunnel that no entity is able to inspect, besides the endpoints. Opaque tunnels define secure communication channels that conceal any information flowing inside. Such an approach provides a few advantages such as tunnels are reusable for several communication sessions, and in some cases, can be fairly simple to deploy and establish. However, they fail to protect the endpoints, since they only provide a horizontal separation layer and do not conceal the currently used (endpoint) identifiers. Another drawback of using tunnels results from the increased communication overhead, reducing packet efficiency as control data, and processing, is required for the same amount of exchanged data. Also, a negative aspect often overlooked concerns the use of multiple encryption tunnels at different layers. It is not uncommon to use Transport Layer Security (TLS) or Secure Socket Layer (SSL) connections that overlap in functionality with mechanisms such as IPSec. When used together, we face a two-tunnel penalty that can negatively impact performance. Each of these tunnels is protecting the conveyed data payload, but the application data is protected by the TLS tunnel and by the IPSec tunnel, which can be waste of resources and a security redundancy.

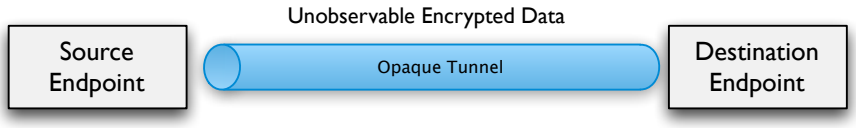


Figure 3.6: Identifier protection through encrypted tunnel, unobservable by eavesdroppers.

These drawbacks do not necessarily mean that opaque tunnels are a flawed solution. In fact, such technologies provide a quite effective privacy protection mechanism, for the sole purpose they were designed to achieve. Therefore, they should be extended towards a consistent use, taking advantage of other means to protect the transport identifiers, without having rely on lower layer tunnels. However, this type of solution only covers identifier-based threats on the network where there is identifier matching or aggregation. Even so, opaque

tunnels are just an example of how the vertical separation can be achieved. The underlying principle is to create a barrier between two adjacent layers in the layered model, which could be achieved through other means e.g. creating a shim layer with interchangeable identifiers.

3.5.2.2 Horizontal Separation

While vertical separation is discussed between layers, horizontal separation focuses on identifiers in the same layer. Breaking horizontal relationships must unequivocally resort to using different identifiers that no not provide obvious relationships within the same event space. This implies the use of solutions like pseudonymity, where multiple identifiers are used on the same layer to hide user interactions.

Using different identifiers on the same layer is only vulnerable to the use of band or contextual information to link two events together, forcing the attacker to resort to the discussed probabilistic scenarios that become harder to deploy. When necessary, these example horizontal shielding techniques should be complemented with vertical separation techniques for a more complete ecosystem in privacy protection. However, as discussed in Sec. 3.4.2.2, horizontal correlation threats that resort to side channels or contextual information requires more than changing identifiers or concealing them in opaque tunnels. They demand more evolved privacy considerations as will be discussed in the next chapters. Nevertheless, the concept is simple enough to expand to both situations and is summarized as a need to establish a barrier between identifiers or events in the same layer in such a way that they are not relatable. This can be done by different identifier mechanisms along with a privacy protecting mechanism.

3.6 Conclusion

In this chapter, we started by proposing a privacy definition, supported by previous definitions that rely on information disclosure and identifiers [141]. This paved the way for the proposed privacy model, established upon simple definitions that try to describe privacy interactions using concepts that can directly relate to the network: identifiers.

The PRIVED model, presented in Sec. 3.3 and subsequent sections, uses the proposed privacy, linkage and correlation definitions and introduces the concepts of Information Set, Event and Relationship. The IS creates the necessary comprehension on user information directed towards the network, while the events translate the repercussions of network interactions on the user. This is achieved through the notion of relationship, a key component of the PRIVED model as well as of other discussed proposals [150, 65].This can provide the right balance between abstract and pragmatic concepts for building the PRIVED model.

One important advantage of our approach is the effort to keep the model as close to the network as possible, allowing a clear integration path between model and network. By discussing how linking and correlation occurs on the network, we were able to show the usefulness of the model, and how it can describe network privacy. By extending the PRIVED definitions with identifier-based threats, established through horizontal and vertical relationships on the network, it is possible to show the resulting privacy loss.

Focusing on identifier-based threats, we discussed the relevance of each major identifier on the lower network substrate, which includes link, network, transport and even application. Here, the network privacy review (presented in Sec. 3.4) becomes important: it provides a basic understanding of the network, which is now subjected to our proposed model, thus resulting in a synergy between the theory and realization of the problems we face on the

network. By combining these two approaches, we were able to outline a clear vision of the identifier issues we face on the network, which are the core of privacy threats, along with the potential hazards of each individual protocol identifier.

One of the most important conclusions regarding the analysis of identifiers and the network as a whole is that protocols can single-handedly compromise above layers, thus jeopardizing privacy throughout the network stack. Not only can they compromise above layers, but also other identifiers on the same layer, further undermining user privacy. This conclusion emphasizes the need for a vision of vertical and horizontal privacy threats. Using a systematized view, the vertical and horizontal conceptualizations, we were able to explain the difference between protecting privacy, which is a vertical coordinated effort, as shown in Sec. 3.5, and protecting identifiers, relating to a horizontal view of privacy. These, in our opinion, are fundamental principles that show a solution path towards privacy preservation in the network: the vertical aspects are directly related to the user, and must be handled vertically, whereas the horizontal aspects deal with each individual protocol or layer, and can be handled orthogonally (as long as the vertical issues are addressed). These conclusions shape the contributions in the following chapters, which further explore this duality: in Chap. 4 we tackle the vertical problem, as opposed to Chap. 5 where we deal with the horizontal threats that appear in different layers. There is always a connection between vertical and horizontal, given that horizontal threats can undermine vertical approaches, requiring a balance which is also explored in the face of different solutions, and even architectural paradigms, in the following chapters.

Chapter 4

A Vertical Approach to Privacy

> The whole is more than the sum of its parts.
>
> **Aristotle** *in Metaphysica*

After modeling privacy, in the previous chapter, we dive into the different possibilities and approaches that increase user privacy in the network. In this chapter we aim for a vertical solution for privacy that encompasses the different dimensions of the user's interaction with the network, focusing on his information. We do this by proposing the concept of a Virtual Identity (VID), a generic approach that models users in NGN systems, describing an identity centric approach to network paradigms. On top of this concept, we present a framework composed by the necessary entities, a data model, an overall architecture and the necessary support components. Together they form the VID Framework.

The structure of these partial identities is oriented at limiting the information seen on the network. Therefore, the proposed solution segments user information, resorting to pseudonyms used towards network and services, and simultaneously, provides the means for concealing private information whenever possible, supported by the application layer. Moreover, because the proposed approach must reach the different layers of the network, we must handle how pseudonyms are translated into network mechanisms, leading to a network pseudonymity concept. By understanding what moves and differentiates different pseudonym solutions, as shown later in the chapter, we introduce a cross-layer pseudonymity approach to support virtual identities in the network.

4.1 Introduction

It has become apparent that privacy is not limited to a single layer or protocol. The threats on each layer along with the dependency between layers, define a clear landscape: individual layers have their own privacy issues, due to specific protocols, mechanisms, identifiers and conveyed information; but, different layers, especially those adjacent in the network stack, can compromise each other through linkage and correlation techniques. The resulting inter-layer dependency, in privacy terms, requires a consistent approach, where we must consider the network as a whole, before considering solving individual threats. To achieve this, we propose a vertical approach that translates into privacy considerations that spawn to every layer in the network, providing the basic instruments for preserving user privacy. In practice, this can be described as a vertical view over the network privacy issues, that relates to user privacy. In this context, the most pressing issues concern the presented identification problem, and every associated mechanism which builds on the user information set.

There must be a common framework to aggregate user related information, and therefore preserving user privacy by controlling how information gets associated to an information set. In this chapter we present a framework that handles user related information, tackling how the user is perceived in a cross-layer (network) view. This enables control over what information blocks are actually disclosed, which can be either user or network related. The vertical nature of the framework requires an aggregation point for all the user and network information. We propose using identity as the vertical driver that handles user information, through IdM concepts and technologies. We extend this feature to become the network privacy glue that holds the different layers together in the light of a VID concept, defining a new paradigm to approach network privacy problems.

Using the VID as a fundamental concept, alongside all the necessary support concepts, we present the VID framework in Sec. 4.3, which proposes a vertical identity-based aggregation layer, consolidating privacy in the network stack. An important part of the framework is the VID IDentifier (VIDID), which provides the basic building block for the vertical view over the network. We also cover the high-level mechanisms required by such a layer, that also serve as the privacy tools for different elements in the network stack. Given that the VID framework heavily relies on the concept of pseudonyms, we explore how network pseudonyms can be used for privacy, in Sec. 4.4. The results in the introduction of a cross-layer pseudonymity solution, Virtual Network Stacks, detailed in Sec. 4.5. The proposed solution is evaluated in Sec. 4.5.4 through a theoretical and practical analysis that provides conceptual boundaries, by highlighting design barriers and benefits, as well as performance expectations based on an experimental prototype. We close the chapter by discussing the benefits of the pseudonymity-based vertical identity solution in Sec. 4.6.

4.2 Virtual Identities

The privacy analysis done so far (including network models) has led to the conclusion that neither the user nor providers (service or network) know how to express privacy in a clear and tangible way. This stems in part from the fact that there is still no accurate or consistent representation of the user, making it hard to design or provide any privacy solution. But this problem goes beyond privacy: there is a generalized lack of understanding on how to represent the user in the communications ecosystem, undermining several network and service efforts.

In most cases, the user is a number or an identifier, whereas in other cases it is represented by personal information. However, this information can vary from interaction to interaction, leading to the complex problem of how we can represent the user in a digital system.

Apart from user representation and identification issues, there other aspects that reflect on the way users currently interact with communication systems. As users resort to different applications, seeking comfortable access conditions and services, they establish different trust relationships exchanging privacy-sensitive information. This user information, the digital representation of user identity as data, can end up distributed (and even duplicated) among different platforms, introducing several obstacles that both user and system need to overcome. The first and foremost problem is that user information can become inconsistent. As data modifies and "evolves", the user alone is incapable of maintaining his information properly updated, simply because it's an overwhelming task. Moreover, the user has multiple services with different sign-on procedures and credentials, which are hard enough to manage even without requiring the coordination of the information associated to those accounts. This shows that there is a crippling inability to make use of data across different platforms, even if that is in the best interest of the user. Consequently, it becomes difficult for the user to manage and control privacy information across several platforms, resulting in the loss of control over all his digital information.

As users move towards an connected lifestyle, the concept of *Digital Identities* can provide a meaningful representation of user data in digital formats, showing potential to be a breakthrough and convergence technology that addresses these concerns. The Virtual Identity concept, explored in the topic of this Thesis and also within the IST Daidalos II [71] European Project, adds another dimension of flexibility to digital identities, making privacy and the multiplicity of roles part and parcel of the solution. The VID approach not only tries to provide a framework that defines how the user is presented in networked systems, but also seems adequate to be deployed by telecommunication operators, exploring already existent trust relationships.

The main goals behind virtual identities are to overcome the current privacy shortcomings, targeting a unified approach to privacy, along with new communications paradigms. First and foremost, the objective is to link identification to the user, rather than to the device, reconverting the network and privacy discussion. This shift allows us to rethink many network processes, especially access across several devices and while on the move (user mobility rather than terminal mobility). In fact, this removes the limitation of owning a device that travels with the user, leading to a concept of seamless and ubiquitous network access. This simplification and transparency is extended to different aspects, such as billing, enabling the user to limit customer relationships to fewer trusted providers, and also simplified management and control of private information across platforms.

While these objectives may seem daunting, they can be feasible with a vertical approach that ties user identities to a cross-layer network concept. We focus on the privacy implications and proposals of putting the user, and not the terminal, at the center of the communications, preserving privacy control even towards the network provider.

Fig. 4.1 illustrates a user-centric view of the world that frames the VID concept. The user, in one of several roles (worker, boss, father), uses one of many access possibilities (e.g. DSL, car-to-car) to access the digital world, such as at home, in the office or while shopping. Each such use could correspond to a specific "Virtual Identity" of the user, and different providers will only perceive the information which the users so desire - regardless of the communication layer where they operate. The overall concept, aligned with the framework presented below,

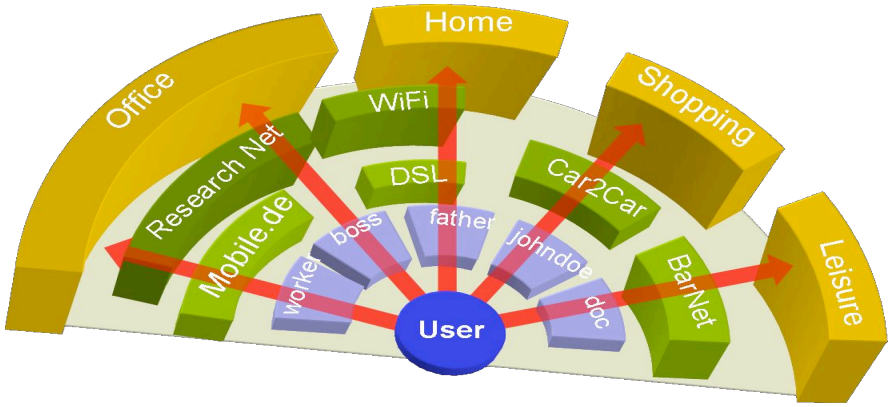


Figure 4.1: The virtual digital world for a user

links the real and the digital worlds, thereby considering various contractual and contract-like relationships that a user has. These relationships serve to vouch for some aspects of the user's attributes, such as nationality, driving capability or financial credibility. Nonetheless, the user should remain in control of personal sensitive data. In addition, various contractual partners should be able to exchange user data in a controlled way (federation), while respecting both the users' and the contractual partners' needs.

Fig. 4.2 illustrates the basic approach. To begin with, we note that the VID reflects a collection of data constructed by the user for a particular purpose. It contains an identifier, called VID IDentifier, but the VID itself is far more than an identifier. Though the VID concept is not restricted to persons, we use persons for explanation purposes as the starting point - the extension of this to other entities is afterwards immediate. The physical person is the real identity (or uniqueness over time) of the person with all its associated attributes, as shown in Fig. 4.2. Some of these attributes may change without affecting the identity. Such uniqueness over time may also be held by legal persons, such as companies or societies, though their uniqueness over time is more easily challenged. Often, in common language and in public debates, "identity" is also used to denote belonging to a specific group, reflected in terms such as "national identity" or "religious identity". This latter usage is not more than a user attribute or group membership. These "common language identities" are dealt in our model as attributes or properties of a real identity.

Many attributes or properties of a physical person are stored by contractual partners and governmental organizations. A birth certificate, a driver's license or a passport are examples that reflect data stored at an office, and may contain user data, such as date of birth, ability to drive, names of parents or address. These credentials - or issued identities - are then directly used for purposes such as travel to some other countries. A very different type of issued identity is the SIM card, which stores data related to mobile network access rights and a link to the person using it. These issued identities are not only inflexible, because they cannot easily federate across organizations, but they also do not allow the user to control what information is revealed. At the other end, for the web and for applications, we often have names and passwords as the issued identity. This incorporates what we note as the contractual or acquired attributes (Fig. 4.2) extending issued identities. However, it is worth taking a broader view of what this means in a more grasping communications scenario.

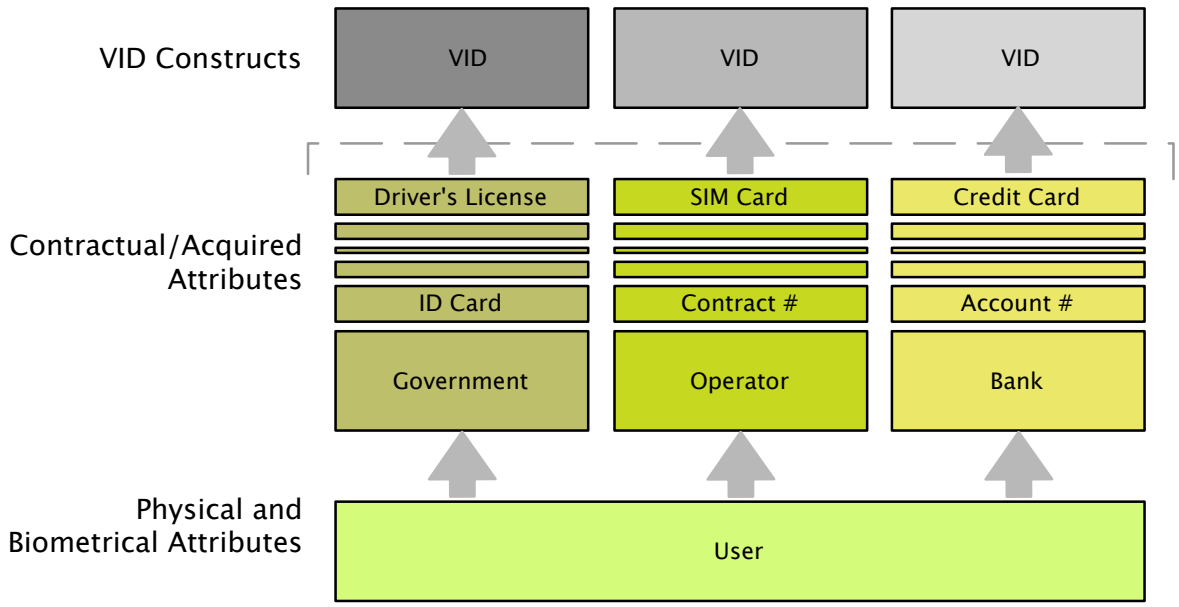


Figure 4.2: The construction of an encompassing Virtual Identity.

The VID concept builds on such real world schemes, adding a new dimension of flexibility by enabling the choice of revealing, or not, user information. Credentials that allow the user to make well-defined assertions, which are nothing more than user-related information blocks certified by a trusted party, can be used by the user to construct virtual identities, each of which reflects a specific scope of use, which may be linked to role assumed by the user (e.g. boss, teacher, client), or persona (the contextualized view of the user highlighted in Fig. 4.1). In this scope, the VID is equivalent to the persona embodied by the user on the network, and the terms become synonyms.

The overall data contained in a virtual identity should be under the user's control, and brings together an intended scope of use, such as reflecting a private person who does not want to reveal any personal information for use in a low cost sector only. It should be noted that the user does not even need to reveal all the data of the VID when accessing a service: the revealed VID data can be filtered to whatever is needed or whatever the user wants to reveal.

This presents a new approach in communication systems. When access to the Internet and other communication infrastructures and services is not a privilege of the few but a right of many, the VID opens the door to on-line digital identity. In this new scenario we must respect the use privacy, and the VID shields user information, protecting his personal life, e.g. relieving the user from the worry that the bank has access to information about services he paid for. The VID takes into account the different identities and roles we incarnate each time we turn on our computer, mobile phone or PDA. Only the a subset of the identity information, required for a given role and acceptable to the user, should be disclosed. Depending on the situation, anonymity and even concealment of attributes such as name, profession, or age, should be possible. The information needed to provide a service varies greatly and seldom requires detailed personal information. In some cases, only the ability to pay is required, in other cases proof is needed about being above some age limit or about being citizen of some

country. Nevertheless, legal tracking capabilities will need to be supported.

All of these attributes are part of the user's digital identity, therefore the user should have control over what is disclosed when accessing a service, as well as the choice on whether this information is kept private or shared with others. If the information required by a service is more than what the user is willing to provide, service access may fail. Furthermore, users are not the only ones with several (virtual or digital) identities. Any entity capable of establishing legal relations with other entities can benefit from this framework. We propose that users, groups, service providers, network operators and even banks all share this identity framework and utilize it to communicate with each other and establish their relations, not based solely on their "real" identity but also on their "virtual" one.

Given the ambitious reach of defining virtual identities, it must be contextualized in a support framework (Sec. 4.3), which, bound to the law, presents its users with a range of new and interesting possibilities in the fields of privacy, identity and federation. However, this can only be done in light of the privacy model, and with clear design guidelines (as discussed below).

4.2.1 VID in a Privacy Model

The virtual identity can be seen as a privacy aware information aggregator, leading to the assumption that all the information present within a VID relates to each other in some way. This view is aligned with the Information Set (IS) from the PRIVED model, where the VID is a set of user related information blocks. The parallel drawn between the VID and the IS, constructions that share a natural relationship, only differs from the observer's perspective, presenting a small difference concerning the purpose of the set. The IS is a privacy construct used within the model to detail the aggregated information through relationships extracted from events. With the VID, the paradigm is slightly different, because it is a user oriented construct where information is controlled and disclosed willfully by the user, according to privacy aware operations. But, regardless of who builds the set, the contents are similar.

From the attempt to model the IS under the specific conditions mentioned in Sec. 3.5.1.1, where we highlight that privacy comes from reducing the information contained in each set, comes the privacy focus of the VID. The aim of the VID is to provide this reduced set, a part of the user information set that is not related to any other IS. In this set we place the information required for a VID to access a service without any issues (e.g. information shortage). Obviously, it can contain more information, but that urge should be feigned off, to retain a better chance of privacy. Nevertheless, in some situations, as we will see below, the information set can be quite large, providing a highly customizable environment for the user. The nature of the partial identities, presented as a Virtual Identity, presents the initial building block of a more generic abstraction, which is to reduced the information associated with a Virtual Identity as much as possible.

We use the VID to tackle the vertical aspects presented in the model, splitting information across several sets through the partitioning principle. This makes it less vulnerable to privacy threats, minimizes the consequence of a privacy breach, and puts the disclosure of information (conceptually) in the hands of the user. And, while the IS is the result of privacy mining technologies, as outlined before, the VID provides strategies and structure to control information vertically. It must be noted that, to provide the vertical component based on partial identities that limit the information set size, the VID must respect the IS principles to avoid into linkage and correlation in the network. This is where the IS and VID walk

hand-in-hand.

While the PRIVED model and the mapping between VID and IS exist only to deal with privacy related concepts, the VID defines a paradigm that goes beyond privacy. This requires that we explore other dimensions in the light of the overall framework, as highlighted next.

4.2.2 Dimensions of a Virtual Identity

When undertaking the task of defining how users are perceived from the network point of view, and aggregating that with an information model that suits both user and network needs, it is important to define a clear set of guidelines that enable building a complex architecture. Specially considering that the Virtual Identity aims at providing a new paradigm for user representation and consideration in the networking systems, it requires considering several dimensions that affect the network ecosystem, with special emphasis on the privacy aspects. The VID concept can provide significant advantages in several network vectors. We define several dimensions that are part of the guiding principles present in the VID framework, that stem from the VID concept: *privacy*, *Unification and Uniformity*, *Contractual Information*, *Context Data* and *Access Control*.

Privacy, as discussed before and embodied by the PRIVED model, is one of the core features that a VID approach can provide, especially in an NGN environment. Users travel from and in networks with no concern for their privacy or even awareness of privacy threats. Wherever they go and connect their devices, they inform the world of their presence and movement. One of the main objectives of this model is that the user not only regains control of his data, but also that he can access the network via distinct untraceable "avatars". The network and service layers, including the operators owning them, cannot link two different avatars to the same user unless he allows it. In addition, the user will control which information is linked to which avatar and can create distinct virtual identities to access the network and services. We thus propose a mapping of the users' multiple identities into multiple network counterparts. This allows us to devise a mechanism that enhances user privacy by maintaining attributes for each of these network counterparts separately, which also separates individual service and network perspectives.

The VID, by not being limited to a single layer or protocol, provides unique means for *Unification and Uniformity* throughout the network, focusing on user-centric paradigms. Since the VID concept cuts across layers, it can be used to unify the handles by which users or services are represented by the VID, including access data across different layers. This imposes uniformity in protocol design, thus simplifying the process by which data is stored, archived, transported and used. The objective is that the VID maintains consistent semantics throughout all affected layers of next generation network, and can be accessed by different network components such as services, billing, accounting, authentication or authorization.

The relationship between the above mentioned components and the VID can often be based on *Contractual Information*, something that is handled within the concept. Although network access, as well as service access, is restricted to the information contained in the VID, most of the user's contractual data still relates to the real identity. Either directly or indirectly, by cross-certification, the entity that issues the contract must be assured that a valid person exists. In most cases, this means assuring that the user can pay for the contracted services. All contractual information should then be translated into a language that the network can understand, verify and account for. This information is linked to the VID, and can include Quality of Service (QoS), authentication or authorization credentials, among

the several sources that can relate to network information, as well as any other contractual binding data.

The contractual data is usually partnered with *Context Data* as it translates into the network. Context data or context information characterizes the current situation of an entity. This information can either be relevant for the entity itself or of broader interest. In consequence, two different kinds of context data have to be distinguished: Context data bound to a VID and context data independent of a VID. In the first case, context data is associated to the VID and represented only in relation to that same VID. The architecture itself has no notion of user. The mechanism by which entities access and provide context has to consider the VID concept. In the second case, context data is considered a commodity that can be traded and should be decoupled from the acquiring context source. Furthermore, this context data must not be linked to a VID of the providing entity. Context is also one of the most critical parameters, which can be used to link two ore more different VIDs of one entity. Therefore, an entity has to pay special attention to ensure that context data cannot be used to link two or more VIDs. One measure to achieve this is context obfuscation. Nevertheless, this is a gray area in which no solution covers all angles since the concepts of context and privacy are somewhat contradictory.

Finally, we present a traversal aspect to the VID and its possibilities that is *Access Control*. The user sets rules on VID attributes filter the information a service or operator can see. By controlling the access to such information, the user can present to a service only a small part of the information contained in a VID. Access control provides the notion of a "filtered" VID that limits the view on that identity to the amount needed by the service and desired by the user. This provides another partitioning layer upon the Information Set, something that will be detailed in the VID model (Sec. 4.3.1). However, this also triggers the concept of ownership around a VID and its related components, because a user may own the VID, but that does not mean that he has full control over its parts. Having attributes, such as age, does not imply that the user can change them, nor does it mean that he owns all the information contained in it. A VID is composed by information linked to the avatar of the user, which may come from many different places. The nature of the information in the VID also determines who owns it. Information resulting from a contractual right of the user is under the control of the user, to be used as deemed fit. On the other hand, information about contractual obligations of the user will need to be enforced by contracted business entity and thus cannot therefore be under full control of the user. For example, the maximum QoS level the user has a right to use is under his control while related metering information is not, since it represents a contractual obligation.

By understanding the different dimensions of the VID, we can take the proposed paradigm, along with the model in which it is inserted, and carve the basilar stones upon which to erect he VID framework.

4.3 Virtual Identity Framework

The previous section tried to clarify the concept Virtual Identities, where a user has different personae, each showing a different side of the user, associated with different information. When aligned with the PRIVED model, this approach fosters the notion of information sets, along with providing a way to properly partition them. However, there is a gap between the conceptual approach of partitioning the information set based on virtual identities, and

achieving that partitioning in the network. We turn to IdM to bridge that gap, which already deals with the notion of user identities or digital identities. This provides the mindset of the VID framework, which can be summarized as a vertical model to tackle privacy, using IdM as the corner stone of the infrastructure to support the cross-layer approach.

It is in this vertical approach that we can consider the complete array of objectives and guidelines that were considered earlier. But, as we consider the different layers, we are no longer restricting IdM to service providers. In fact, by using a cross-layer approach to satisfy such requirements, we are bringing the NGN requirements onto the drawing board, and considering both user and network operator interests in the process.

By building on all this entire vector of concurrent motivations, and taking into account the previously highlighted guidelines, we can move onto a model that portraits the user and operator reality, using a cross-layer VID approach.

4.3.1 VID Model

The VID model can be described as the integration of virtual (partial) identities with the PRIVED model, by resorting to identity management semantics. While succinct, this definition highlights the key values of the VID model: by being based on tangible concepts of how privacy works, heavily influenced by PRIVED, and on how we can shape information, according to the VID, we can define a model that brings us one step closer to an architecture (based on IdM). The identity semantics bring several definitions that compose the core of the logical model, which is later instantiated into proper network components. We provide a definition of a user, that is properly generalized to fit the virtual aspects of the user, described as generic "entities". By using logical "entities" it is possible not only to define a privacy preserving architecture, using virtual concepts, but also to define handy tools that enable architectural support, working on multiple levels. This means that services must be properly considered, along with different user-related aspects such as personalization, access control and other features that are convoluted into the presented model.

Therefore, we provide a privacy-aware communication model for NGN, and possibly Future Internet approaches. The VID Model, which is a fundamental component of this framework, is described below, along with an architecture to support the model (Sec. 4.3.2). One of its core properties is a cross-layer approach that supports aggregated visions over several network aspects. First, it supports uniform namespaces, relying on identity (one ID for all purposes). It is this bound to identity that makes it suitable for network identification, and promotes controlled exchange of user/service/group information. Still based on identity aspects, it enables the maintenance of pseudonymity at a higher level, without depending on a single protocol instance or application. This results in a top-down approach, that can be independent of application, service, interface and even terminal.

4.3.1.1 Model Overview

The VID model is a data model that aggregates user information into a comprehensible set of structures that allow the interaction between users, services, focusing on privacy-aware VID constructions. Fig. 4.3 shows an overview of the VID model. The model is best described as the relationships that can be established between an Entity, its profile, the parts that compose that profile, and the views of that profile. An "entity" can either be a user (person) or a group of users, generalizing its applicability. On the left hand side of Fig. 4.3 there are several parts

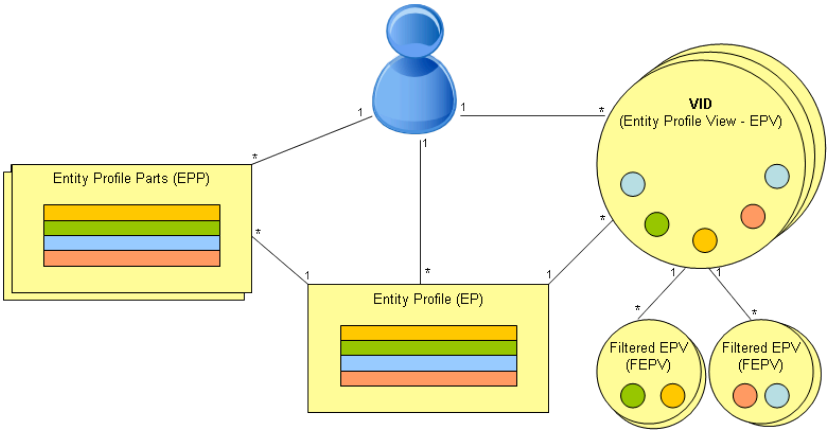


Figure 4.3: VID Data Model

of the profile about this entity. The Entity Profile Part (EPP) represent the fragmented pieces of information that stem from different contracts at different networks and/or services. In the middle, there is the aggregated logical profile about the entity consisting of all profile parts, called Entity Profile (EP). It is important to note that the EP is an abstract notion that is not stored anywhere as data, but represents an aggregation of all the user related information that can come from user, network, contracts or any other source. On the right hand side, several views on this overall profile are shown. They represent the Entity Profile View (EPV), which holds a subset of the information aggregated under the EP. The EPV represents a VID of the user, an contains all the relevant data for the user, that can be provided for service consumption or interaction. In any case, there is a single logical profile, the EP. The views are only pointers to the real entity profile parts and implicitly to entity attributes stored in a database (e.g. operated by a network provider).

An entity profile should be considered as a pool of entity profile parts and consequently as a set of attributes about the respective entity. From this profile, different pieces can be chosen, and together form a view on entity. However, each of these individual components will be further clarified, given that their functionality and definition can exceed this brief summary.

4.3.1.2 Entity, Profile and Parts

An entity is any actor or “real identity” that is able to establish legal, contractual bindings with other actors. The most common entities are users, service providers, network operators, banks or groups of any of the former. Although network or software components that act on behalf of an entity do not have a VID themselves, they can be associated to the identity that owns them. When discussing entities, we highlight two particular realizations: user and group. The user in the VID framework, is an abstract entity, which is not represented directly in the framework, but rather impacts the architecture through his VIDs that do have a network representation. A user is able to create legal bindings with other entities, which allow the user to create VIDs and to utilize the services underlying the architecture. Additionally, a group is a set of entities which also can be represented by a VID. A group can also be seen as an entity with a set of common attributes, and may be treated by the architecture as one.

Such a group, in contrast to a legal entity, does not necessarily exhibit the identity properties of being identifiable as the same group over time.

Each entity has its own profile, the Entity Profile. The EP is an abstract concept representing a group of EPPs that relate or belong to an entity, as described below, including the knowledge possessed only by the entity itself. By holding the information related to the entity, the EP becomes a controlled realization of the Information Set concept, stemming from the PRIVED model. There is only a slight difference between since the IS models the knowledge of the attacker, from a privacy point of view, whereas the EP models the knowledge of the entity over its data.

The compartmentalized view of the EP, which builds on several EPP, allows the user to define how to best build the VID, choosing what information is contained in a single identity. By using different blocks to build VIDs, we are in fact defining different (and hopefully unrelated) information sets that relate to a single entity.

It is important to understand that the EP cannot be stored in a single well-defined location because it represented the collective set of user related information dispersed through multiple databases. The EPP is a system representation of partial information on the entity profile. This means that the entity will have a set of attributes, which can be joined together to form an EPP. The rule for creating an EPP is that it should be the minimum consistent and coherent set of data, which can be extracted from the EP. The EPP attributes should be atomic (cannot be broken into smaller parts), holding the information that belongs to the EPP. We can also look at an EPP as a block of data, which exists in a database in the system. This block of data can be read, written into, copied and acted upon.

When establish an analogy between VID and PRIVED, we can map the EPP to an Event, provided that we make a few assumptions: if the EPP is transferable and observed on the network, it is an event from the PRIVED point of view. While this is valid for several observations, we draw this parallel due to the nature of the information: both the EPP and PRIVED events should be as atomic as possible, allowing in one case to build the VID by stacking information, and in the other case allowing evidence stacking that breaches privacy.

4.3.1.3 Entity Profile View - The Virtual Identity

The VID represents an avatar, or persona, of the user in the network. It is a subset of EPPs, which belong to a single EP. Because the VID is simply a scoped view of the EP, it can be formally described as an Entity Profile View (EPV). The EPV is an alias for the VID because it defines a partial view of the user's identity, forming a virtual persona. From a different perspective, the EPV is comparable to the Information Set, when properly reduced. The only true difference between the EPV and the IS occurs when there is a privacy leak, and the information traveling on the network is larger than that intended by the EPV owner. This happens because the VID model tries to build on the information known by the user, and the IS models the information actually seen on the network from a privacy perspective, known by an attacker, which should be different. The VID is used to reflect different personal information, as well as different network circumstances, and is purposely intended to difficult the task of an attacker linking several identities. It is expected for a user to have, in average, 5-7 different VIDs, depending on the roles he assumes. Managing more than this may prove difficult from both a user and network perspective. To add more flexibility to the EPV, or VID, we accompany it, first with an identifier, the Entity Profile View Handler (EPVH), and second with a filtering mechanism, to enable granular use of the information contained in the

VID, outlined as a Filtered Entity Profile View (FEPV).

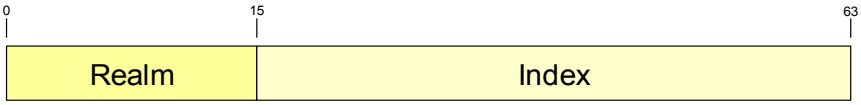


Figure 4.4: VIDID - The Virtual Identity Identifier

The EPVH provides a unique mechanism to identify and reference an EPV. Because it identifies the VID, we also term it VIDID. The VIDID, shown in Fig. 4.4, is the identifier that acts as the initial reference towards the VID, allowing a holder to reach the appropriate information of a "realm" identifier, which maps to the server or domain responsible for the VID, and a flat identifier, which uniquely identifies the VID within a certain realm (while not revealing any information about the owner). The length of the VIDID is in direct relation to the diameter of the Internet and the number of expected identities which use it. In fact, a VIDID should uniquely identify a VID in a certain realm (this format allows 248 such users per realm, and 216 realms).

The VIDID can be a useful tool for integrating identity references into the network stack, as we will discuss later in the Thesis. Its identification purposes can be combined with the network identifiers of existing protocols, creating an environment where common protocols carry identity references that lead back to the user identity and to the IdM plane. Because the identifier can be represented in different ways (e.g. hexadecimal, string), or can be derived from specific information pieces, i.e. it can be a hash of a well knwon value (e.g. a public key), it is suited for integration with different protocols, ranging from low level network to application layer identifiers, such as an IP address or a Uniform Resource Locator (URL).

The unique relationship that can be established between the VIDID and user identify can be explored in an entire new architectural dimension, building a playground where identity is integrated into "legacy" protocols, as explored in Sec. 6.6. Because the VIDID can serve as a unique cross-layer identity reference, it can also be used as the identifier by which an attacker tries to build an IS, providing a common denominator for user information, identifiers or references that relate to the same user identity. This property leads to the idea that the VIDID must be handled with care, to build privacy aware solutions, as discussed in Sec. 4.4.

The second structure that aids the VID is the Filtered Entity Profile View, which represents a second level of access control on the EPV, thus holding a subset of data of an EPV. This enables a better control over the information revealed to the outside world. While the use of distinct FEPVs as a subset of the same EPV may be identified as belonging to that particular EPV, two FEPVs representing different VIDs should be completely distinct in all revealed data to make it impossible to link different VIDs. Of course, the owner of the VID has the power to allow this cross-linkage.

4.3.2 Architecture

To support the conceptual aspects of the VID model, we define the core components that support the identity oriented operations and functionality. Unified and uniform namespaces enable the user to be reached despite a very heterogeneous environment and must rely on a common entry point for identity information, providing consistent access to privileged information based on VIDID resolution and credentials. Therefore, access control, which allows

the user and the provider to limit access by others to the user or to user data based on well-defined criteria or rules, must use architectural components. These components need to be tightly coupled with the billing and charging infrastructure, the AAA backend, enabling charged services to be offered in a flexible way, allowing the user to aggregate billing to one or a few billing entities. Finally, the base components must provide strong privacy policies, enabling the user to protect her data as desired, i.e. to prevent disclosing its own attributes and to limit reachability.

The common components of the aforementioned operations are two specific functional elements: The Identity Manager (IDManager) and the Identity Broker (IDBroker). With them, we can define the basic identity framework components and how they facilitate identity oriented operations.

Identity Manager The responsibility of the IDManager is to handle VID creation and maintenance. While providing the interface for creating, managing and destroying VIDs, it becomes the repository for context and personalization information, which is not stored on distributed EPPs. It is one of the trusted components on the architecture, since it handles on one side the VID information that the user chooses to store on the network and to the other, it may retain information that is considered sensitive, acting as an EPP repository for specific data, and applying several control policies on the data it stores. The IDManager has a strong relationship with the IDBroker.

Identity Broker The IDBroker component provides the location of the EPP or proxy to the requester, and forwards the request to the holder of the EPP (the place where it is stored). The IDBroker thus becomes the main contact point for identity information and the preferred location to set the initial privacy and content access policies. Depending on the operation mode, different amounts of trust are placed at the IDBroker. If the IDBroker only redirects requests, the burden of policy control is redirected to the EPP holders and possibly the IDManager. Moreover, when applying access control rules directly on the mapping and content access, pointing to different locations depending on the requesting entity, it becomes a more crucial architecture component. The IDBroker must therefore be fluent in handling VIDIDs, since this is the preferred notation to handle the unified namespaces. The process of resolving and accessing information through the identity broker are described below.

4.3.2.1 Identity Brokerage

The IDBroker is the component that is responsible for mapping the revealed FEPV to the EPPs, and possibly also applying some access control rules and actions. It is called a broker because it correctly redirects requests for EPPs belonging a certain VID. It is most likely the component identified by the realm part of the VIDID. The IDBroker is the first contact point in a realm, although it may be possible that the requests are forwarded to a different machine. This process is illustrated in Fig. 4.5, where Bob accesses the IDBroker with a specific handle for Alice (a VIDID), which should be subjected to access control, and lead to a filtered access onto specific network components (e.g. A4C, PKI).

With only the VIDID to resolve information on a certain VID, the requester uses the realm part to contact the right IDBroker, and the ID part of the VIDID to identify the VID within the realm. Additionally, the resolver has to identify to which EPP it requires access,

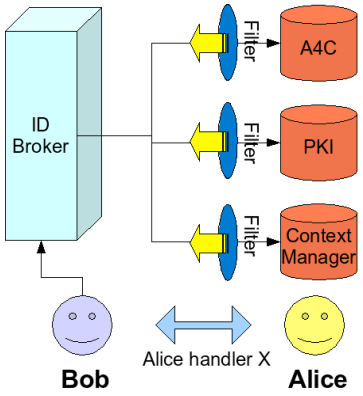


Figure 4.5: Using the Identity Broker.

possibly with accompanying credentials for access control. The IDBroker will keep a table of the VIDID, the associated EPPs and their types, as well as the place where they are stored.

One of the most important functions provided by the the IDBroker is access control on whom accesses information and what information is accessed. This results in several tiers of access control applied on user information, which enable a consistent and privacy preserving approach to Identity information.

EPP Access Control This is the simplest form of access control: directly on the data storage place. The broker will always direct the requesting entity to the right storage place (the storage place can already reveal information for linking VIDs, e.g. referring to the same attribute.). Access control is then performed directly as the requesting entity attempts to access the information. The drawback is that the requesting entity knows that this information exists for that VID. Also, on the storage place, the identity of the requester is revealed (at least partially) for matching access control rules.

Broker Access Control Before redirecting the request, the broker applies access control rules to tell, or not, the requester that the information exists and where. Optionally, the broker can "hide" the storage place by presenting credentials from the requester and obtaining the information encrypted for the requester in such a way that the requester does not know where the information is stored, and the broker does not know the information itself, only the type and the requester. Also, at the same time, the storage entity does not know who requested the information. This depends on the credentials system and whether this is also brokered by the identity broker. Furthermore, the storage place for certain attributes can also facilitate VID-linkage.

Terminal Access Control As a last resort, access control can be applied by the rules in the terminal itself. The owner of the VID requests the information and filters it before directly giving it to the entity that requested it. This actually means the IDBroker can redirect all requests to a terminal where the entity can then decide what information to give out and how to access it. Although this approach can impact performance, from a security perspective it is the one which offers most user control.

4.3.2.2 Resolving Information

To obtain information on a VID, an entity must first obtain a VIDID. With the proper VIDID, it should contact the IDBroker (defined by the realm) to either directly obtain the information or be redirected to the place of storage of this data. This defines a resolution process required to go from VIDID to actual information.

We outline this process with an example where the entity trying to reach a particular user starts with a FQDN. The first step is to resolve the other entity's FQDN, through DNS, obtaining the corresponding VIDID. It will then use the realm part of this value to contact the correct IDBroker. This connection may depend solely on the protocol being used, for example Diameter for AAA information. In some cases, the IDBroker is simply a box that uses the same protocol as the information requester, resorting to its private tables to determine how the requested information can be obtained.

To reach the information itself, it is possible to use two strategies that can be applied in different situations to enable a requester to resolved the require information. We define push and pull models for VID information access. In Fig. 4.6, we illustrate both push and pull mechanisms, involving two entities, A and B, which already have the necessary EPV (or VID) and request handler (VIDID). In Fig. 4.6, entity B uses a VIDID to access information about A, that want to access a service provided by B.

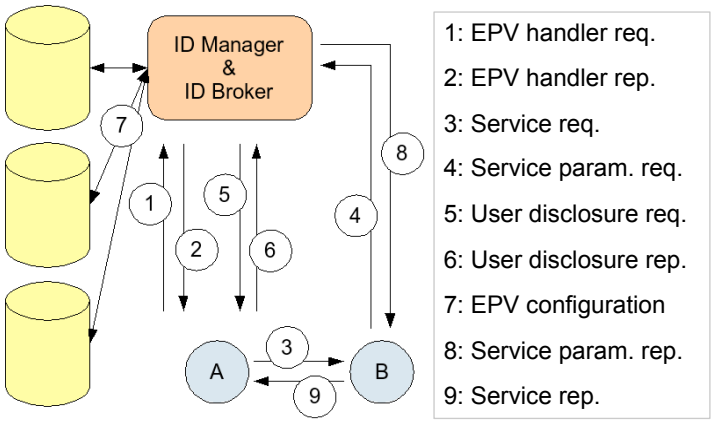


Figure 4.6: Identity Brokerage Push and Pull process

Phases 1 and 2 represent the initial establishment of the EPV (or VID) and the initial part of the push mechanism, where VIDID information on handling a VIDID is initially pushed to the IDBroker. Steps 5 and 6 represent the pull mechanism, where in order to respond to an information request from B, the IDBroker pulls information from the user. Following the initial setup at the IDBroker, the user, entity A, requests a service to B, step 3, triggering an information request (a service parameter) to the IDBroker in step 4. At this stage if its a pull situation, the appropriate steps is to consult the user (steps 5 and 6), or alternatively (push model), the IDBroker retrieves the information from a database (step 7), providing it to the service (step 8). This concludes the information resolution, and the service is provided to entity A in step 9.

Both these mechanisms refer to the way in which the VID information is populated and accessed, along with the applied access control rules.

4.3.3 VID in the Network

The vertical model focuses on the control part of the privacy solution, that resides atop the network stack, closely coupled to application level technologies (especially IdM). The vertical issues addressed are tightly coupled to information exchange and identifier resolution. But, the model needs to have network repercussions to fit the envisioned cross-layer architecture, which is a requirement for the privacy features. We must go from vertical virtual identities to tangible cross-layer instantiations, properly aligning the different communication layers with the proposed a model.

This integration can be done with a virtual terminal approach to support the virtual identities present within the VID framework. The concept relies on the fact the network should perceive each user identity as belonging to a different (virtual) terminal. The virtual terminal should then be instantiated on top of its physical counterpart, the real end-user device, following the same reasoning used for the real versus virtual user duality, stemming from the VID concept. For each information set associated with a VID, we must create an accompanying virtual terminal that handles communication for that specific identity. As we expand this notion to different layers, we see that there must be a support for pseudonymity (or anonymity) on lower network layers, which represents a challenge in itself.

While the VID defines mechanisms for vertical information containment, supported by the PRIVED approach and the IS, it is the IdM system that provides containment across different virtual identities. This is represented by a requirement extracted from PRIVED: if each VID is characterized by an an information set, there can be no shared identifiers between identities, regardless of layer.In order to preserve privacy, the defined containment requires that each Identity uses its own pseudonym at each layer. Using the identity model as a cross-layer solution, it can provide the architectural requirements to fulfill the multi-pseudonym approach: it enables a vertical control plane that interacts with applications, yielding the input for network stack management, while still providing control over different network stacks through their identity relationships. This defines a control mechanism for multiple pseudonyms where each VID can leverage its own contained network stack, with multiple uncorrelated pseudonyms. To support this, it is possible to create different identifiers at each layer. In this scenario, each VID would receive different link, network and transport pseudonyms, making it impossible to use these identifiers for correlation purposes. This mitigates such correlation across different identities, where identifiers are generated or used at the pace they are required to connect to the network at different layers. In this context, it is important to understand the use of pseudonymity as a solution for privacy in NGN, along with its impact on existing technologies.

4.4 Network Pseudonymity

Using multiple identifiers, or pseudonyms, for the same user as means for preserving privacy, is becoming an important approach that has not seen sufficient discussion. Before it can be considered a reliable privacy tool, pseudonymity requires a formalization that takes into account network stack impacts. Therefore, we attempt to demonstrate how pseudonyms can be used to preserve privacy, resulting in the key issues and requirements presented in this section, which enable the use, control and evaluation of pseudonym based solutions at the network level. The conclusions serve as the starting point for our pseudonym privacy proposal defines as Virtual Network Stacks (VNS) [97].

Identifier correlation (including network layers) can render useless several privacy-oriented solutions, regardless of how and where that correlation occurs. As network heterogeneity increases, the privacy defined by application-centric systems (from simple username and password approaches to complex Identity Management solutions) is further questionable. Upper layer information can be linked or aggregated through lower layer identifiers, hindering most application-level privacy efforts. In most of these scenarios, attackers reside in the access network (either the network operator itself or a malicious node) and is capable of scanning most (even all) network traffic. Service providers or endpoints can also be untrusted entities: in some scenarios they can even collude to increase the knowledge of private information about the user, and have access to the network traffic (at least the data intended for each peer). In these scenarios many conditions emerge that can lead to privacy loss, as discussed in Chap. 3.

Different approaches have been employed to mitigate such fundamental privacy threats, discussed exhaustively in Sec. 2.5. Anonymity based solutions [149, 55, 38] try to conceal the user involved in the communication or interaction. Albeit an improvement, this is insufficient: for an efficient privacy solution, the intent is not to completely prevent the disclosure of information, but rather to control the disclosure process that enforces privacy concerns at each moment. Pseudonymity solutions take a different approach: they use different identifiers, i.e. addresses across the network stack, to effectively hide selected interactions between users (or devices) and services, thus providing privacy. Whenever these identifiers are used for misdirection (of eavesdroppers or peers) we call them pseudonyms, given that they now serve a privacy protection purpose: introducing confusion.

Comparing different pseudonymity based solutions can be hard due to the lack of systematized concepts that can guide a fair comparison. This gap also results in the absence of concrete studies on the impact of such solutions, either conceptual or experimental. In fact, the privacy requirements for such solutions have not even been formally expressed, which are hard to accept or adopt without understanding their consequences. Therefore, a systematized solution that enables an evaluation and comparison of different solutions is needed. Consequently, there is a need for practical results that explore cross layer pseudonymity solutions. We propose to fill this gap with a set of formal and applied requirements that can provide clear guidelines for the solution presented in Sec. 4.5, which enables consistent cross-layer pseudonym usage. With theoretical and practical evaluation, we can determine the feasibility of pseudonymity approaches, and simultaneously extract the perceived "cost of privacy". Based on the PRIVED model, we are able to provide a framework for pseudonymity that enables the means for understanding how pseudonymity impacts the network. Key points reside in the theoretical and practical insight gained on pseudonyms, covered by the requirements and evaluation mechanisms presented. Moreover, the requirements and evaluation harness the knowledge of an experimental deployment of network pseudonyms, enabling the evaluation of potential impacts on the network layer. The main objective is to establish a knowledge base to understand and analyze the VNS solution (Sec. 4.5), both in theory and practice.

4.4.1 Privacy with Pseudonyms

In network and computing contexts, a pseudonym is a fictitious name or identifier used to hide the real information of a user, device or network entity. It is achieved by using several identifiers of the same type, at each layer, that differ from the default identifiers (such as the MAC

address). The alias identifiers, created for privacy purposes, are named pseudonyms[1]. Pseudonymity is a concept that has already permeated the computing disciplines. IdM concepts are in essence pseudonymity approaches, at the application layer.

Today, when accessing services, we choose a username and release a certain amount of information to that service. This builds our IS towards that service. Given that we do this for every service, we in fact have multiple identities towards different services, which build different information sets. The goal of pseudonymity solutions is to break the links between the different identifiers, to appear in the network as if different terminals or users are consuming the services, instead of a single, linkable entity.

An example of how a user can inadvertently create links between identifiers over time is illustrated in Fig. 4.7. It shows how identifiers relate to each other, to devices, and to the end user. A common use of pseudonyms is when the user chooses different usernames for different services. Typically, a pseudonym strategy would resort to multiple network IP addresses for a single interface, separating each username even at the network level.

Pseudonyms can be a fundamental tool to protect user information, and as such, must be characterized from an operational point of view. For that, we must rely on IS concepts first presented in Sec. 3.3.1, where user information is described as belonging to a set built around identifiers, such as an email (or a username) or network identifiers, from link to application layer, that share a common thread of knowledge.

4.4.1.1 Protecting Users with Pseudonyms

As already extensively discussed, in networking environments, information can be tracked back to the owner, effectively threatening his privacy. This happens on several levels and can be associated with a particular user, building an IS. As the information around the identifier increases, the privacy hazards also increase. It is important to remember that the simplest way to build an IS is precisely by collecting all possible information around a particular identifier. However, it is worth mentioning that, to build the set, it is irrelevant whether the correlated identifiers are pseudonyms or not.

Further examining the definition of IS allows us to understand how pseudonymity protects the user. Concealing identifiers from passive eavesdroppers, e.g through encryption techniques, hides certain interactions but is ineffective against endpoints and services. Therefore, protecting the identifiers cannot be the only privacy means applied in a network, because the IS can still be built associated with the endpoints. To preserve user privacy, the pseudonym strategy works by reducing the information set associated with an identifier, by generating alternative identifier, pseudonyms, that have little or no associated information. This is one of the major protection techniques, discussed in Sec. 3.5.1.1.

Information reduction also safeguards the top level user identity, which in our model is simply represented by its identifiers, by reducing the information associated with specific identifiers: instead of one revealing set, several smaller and unrelated sets ((different VIDs) are built for different purposes . The optimal situation in terms of information protection would be to randomize identifiers per atomic operation, associating each operation with one untraceable identifier. This is an extreme case that presents serious performance drawbacks as will be discussed in the following sections.

[1]Pseudonym and identifier are used interchangeably to denote an identifier, or set of identifiers, created to reduced the information associated to a particular user, and can belong to any layer in the network stack.

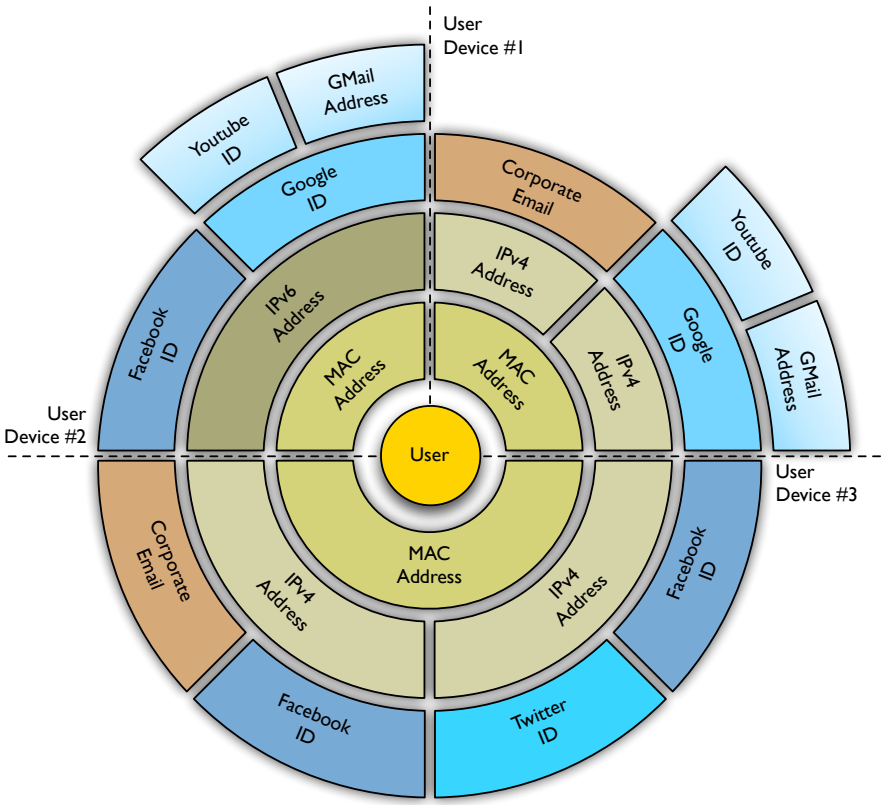


Figure 4.7: Identifier associations created, over time, by a user using several devices, resulting in a broad information set.

4.4.1.2 Threats

The analyzed threats in the PRIVED model allow concluding that application layer mechanisms are not sufficient to provide a safe privacy environment for the end-user. Even if a user presents different identities to different services, the network stack can disclose the linkage between identities, breaking the application layer privacy model. This is the essence of horizontal correlation mechanisms, discussed in Sec. 3.4.2. By establishing relationships between identifiers on the same layer, the privacy of upper layer identities is jeopardized. Lower network stack layers use identifiers that are common to the different virtual personae, rendering them useless in terms of privacy. At the transport layer, the used endpoint identifier, usually the IP address, allows linking different virtual personae because every application uses the same identifier. At the network layer, the same IP address is used for every transport and application connection, providing the cross-linkage of the virtual personae through the common identifier. In the same line, link layer addresses can bind higher layer addressing structures, whether they are IP addresses, transport identifiers or virtual personae.

Our objective is to leverage the concept of cross-layering virtual personae, embodied by the VID concept, as the starting point for pseudonymity, using different and non-linkable identifiers for different virtual users. From the entire formal threats presented in Chap. 3, we focus on a privacy-enabled network where an attacker cannot be able to:

- Correlate two identities by mapping them to the same endpoint identifier;
- Correlate two identities by mapping them to the same locator, or different endpoint identifiers to the same locator;
- Correlate two identities by mapping them to the same link layer identifiers, or mapping endpoint identifiers (or locators) to the same link layer identifiers.

These are the basic threats that the VNS approach attempts to address, relying on the VID concept to provide the guiding mechanisms for vertical containment layers, disallowing any shared information between them, and blocking horizontal correlation mechanisms, that work either by Set or Graph theory. We do not limit the attacker to a simple eavesdropper. This model, coupled with the VID, is targeted at protecting from any single network entity, like the network operator or the service provider, even in the face of collusion.

4.4.2 Controlling Pseudonyms

Each IS contains sensitive information that is complex to handle. As we attempt to partition information, we maintain pseudonyms associated with each set. Therefore, the control strategy becomes an important part of any pseudonym based solution, because it has to fulfill two major roles: control information linkability and scale the system usage.

Controlling information linkability includes the management of pseudonyms generation, and more importantly, the definition of an IS. Given all the exchanged information, managing these sets is crucial to any solution that seriously considers pseudonym usage. The IS, along with the associated identifier(s), must be properly managed, especially considering that information reaches all network layers; failing to do so may result in unforeseen information leakage. The control process depends on the applied paradigm: user-centric solutions try to provide identity pseudonyms [20, 122], whereas application and device centric solutions [56, 151] try to provide per application pseudonyms.

Most user centric solutions rely on IdM standards [20, 122], providing pseudonymity focused on the user, at the application level. In fact, these approaches can be described as striving to maintain an IS for each identity. Consequently, one of the major benefits of IdM is that it provides not only the means to control identifier exposure, but also the mechanisms to operate on the attributes that must be shared with services. The fact that IdM limits the required identifiers on a per-identity basis can guarantee that pseudonyms scale with identities: because the IdM layer throttles identity usage (limits the amount of identities used for service consumption), it can provide a set of tools (and boundaries) to tackle scalability issues caused by the increased number of identifiers.

The alternatives [56, 151] focus on using per-interaction pseudonyms, meaning that each network connection will use a pseudonym. This will require a strong effort from the control layer, to maintain the sets consistent across applications, and from the network layer to maintain a (prohibitive) large amount of pseudonyms on each layer. It also delegates the pseudonym control to mechanized solutions that focus on system interactions rather than on the user. Regardless of the control issues, the total number of pseudonyms must be limited, and it will be necessary to determine which application require pseudonymity and which applications can share information without compromising privacy, as means to limit the potentially large pseudonym number. The two aforementioned paradigms will share the same

requirements that will be common to all control solutions: scaling the system to reasonable performance boundaries and maintaining consistent information sets.

4.4.2.1 Controlling pseudonyms on the network

For effective privacy preservation, pseudonymity must be supported at the network level. This requires extending the information set semantics onto the network stack through a segmented view that enables the dynamic creation of pseudonyms for different sets. We must understand the impact of such segmentation mechanisms in current addressing models, particularly on pseudonym collisions (detection or avoidance), which determines not only the viability of any pseudonymity solution, but defines the scalability boundaries to which the control layer must abide.

The first stage for handling pseudonyms is the definition of a clear strategy regarding how pseudonyms are created and managed. Because an Information Set can be built based on any user identifier, the network stack must be treated as a whole. We need to avoid the establishment of unwanted correlations at any layer, so pseudonyms must be created and managed as sets of identifiers for the network stack. For each Information Set, a corresponding set of network identifiers, that act as pseudonyms, must exist to allow communication without compromising privacy. This creates a direct relationship between pseudonym set and information set, defining the usage boundaries, where two requirements appear: 1) always use the identifiers consistently as a whole set; 2) do not share identifiers across sets i.e. different user (pseudo) identities.

Beyond the two above mentioned requirements, an important network application criteria is pseudonym management, where clear policies shall determine the granularity of how pseudonym sets are applied. However, this responsibility rests on an IS management entity or function (eventually an extended IdM function), preserving an aligned view over the user information spectrum as discussed before.

4.4.2.2 Dynamic Pseudonyms and Collisions

The second stage for handling pseudonyms is to provide generation functions that ensure unlinkability between the generated identifiers. This is typically solved by a proper random number generator aligned with the identifier structure, which will depend on the target layer and obey the namespace ruleset (e.g. network addresses must have a prefix matching the target network to be properly routed).

However, dynamic pseudonyms imply that there will be an increased number of addresses and identifiers on the network, straining the addressing assumptions. The strain degree will depend on the required granularity at which sets are generated and consumed, and raises one important issue behind pseudonyms usage: the overload of the addressing space for each identifier type. Namespaces usually have a well defined addressing space, limited by the size, in bits, of the identifiers. With pseudonyms, the original design preconditions for each namespace are extended, and must be handled accordingly. The impacts on the address space must be predicted and measured, along with proper solutions for handling collisions. Usually, it is sufficient to determine the collision probability of a namespace, and from there, determine whether the identifier size and properties adequately fit the namespace requirements. The address collision probability can be modeled as a birthday paradigm [107] where the collision probability is that of any two addresses within the same space colliding. This proposition

results in Eq. 4.1, which determines the collision probability depending on identifier size (in bits), where n is the number of random bits in the identifier and k is the total number of addresses in the collision domain.

$$P_a = 1 - \frac{(2^n)!}{(2^n - k)! \times 2^{nk}} \tag{4.1}$$

Solving the presented equation for a given size is computationally heavy. A common solution is to deduce an approximation based on the expansion of the exponential function using the Taylor series. Solving the approximation to n leads to Eq. 4.2, as commonly used for hash collision probabilities, where n estimates the maximum allowed addresses that match a specific collision probability p, within an address space of 2^k. These mathematical principles lay the base to understand the impact of pseudonym usage.

$$n(p, 2^k) = \sqrt{2 \times 2^k \times ln(\frac{1}{1-p})} \tag{4.2}$$

4.4.2.3 Performance and Limitations

To consider the support of multiple pseudonyms in network stacks, one must also consider the implications for the stack itself. The stack is responsible for specific tasks, and how the ability to perform those tasks is affected should be determined. In Sec. 4.5.4.1, the details of enabling the use of pseudonyms in the network stack are considered; however, before tackling the low level details we need to understand the nature of pseudonyms and the operations involved in their use, so that we can approach the experimental results obtained later in Sec. 4.5.4.

Some of the addresses employed are subject to registration with the network - like IP addresses. Using pseudonymity means that multiple registrations will be required in situations where previously one was sufficient, increasing the amount of internal state necessary to accommodate these registrations. It can be considered that the resources necessary for the stack to register new addresses grow linearly with the number of generated pseudonyms, in terms of required time, computational effort and stored state. While the effort requirements cannot be easily overcome, the registration of new pseudonyms with the network should be able to take advantage of parallel operations.

Addressing the above mentioned concerns will introduce additional logic in the network stack, and consequently overhead. It is also necessary to understand what are the consequences for Key Performance Indicators (KPI). The impact that these changes have for metrics such as traffic delay and overall throughput must be taken into consideration. Such values offer a measure of the limitations at the network stack, and are a requirement to understand how far we can push the pseudonymity approach before incurring in unsustainable performance penalties.

4.4.3 Addressing Space

While address or identifier collision must be considered at all times, there is a broader impact on the addressing space itself that is not considered in such calculations or concepts. Therefore, while useful, having the collision probability as a function of only the number of devices and bit size falls short when designing pseudonymity solutions, because it is independent of user or IS. Therefore, we expand on a notion of a bit size calculation that relies only on

device consideration, introducing the concept of a metric that depends also on the number of sets expected per user/device. Only this paradigm shift will allow a proper evaluation of the true consequences to the address space, and consequently, to the network.

We model the existence of i pseudonyms (one per IS on each layer) per user, and from this, we get a maximum of s pseudonym sets per network, represented by Eq. 4.3, where n is the number of addresses per addressing domain. However, the collision probability is still dependent on the number of available addresses.

$$s = \frac{n}{i} \tag{4.3}$$

On the other side, reusing the notions of pseudonyms per network (Eq. 4.3 and of collision probabilities (Eq. 4.1)), we can update the estimation of the collision probabilities in terms of how many sets can be supported per network.

For this, we introduce the Virtual Address Space (VAS)[2], which represents the available addressing space based on the expected number of pseudonyms introduced per user/identifier. The virtual address space varies with the amount of pseudonyms per user, yielding a different perspective from the traditional collision probability, which depends solely on the bit size of the identifier. To create this metric, we replace s in Eq. 4.3 by 2^k, representing the pseudonyms per network as if it was a bit sized identifier, leading to Eq. 4.4.

$$2^k = \frac{2^a}{n} \tag{4.4}$$

We then rewrite the number of bits k in terms of the n identifier per set, leading up to Eq. 4.5, thus creating a user dependent view on the identifier size. This presents a virtual metric, where the size of the identifier will depend on the number of pseudonyms per user.

$$\begin{aligned} k &= \log_2 (\frac{2^a}{n}) = \log_2 (2^a) - \log_2 (n) \\ &= a - \log_2 n \end{aligned} \tag{4.5}$$

With this new formulation, we gain a better understanding of the address space in the presence of pseudonymity. By applying Eq. 4.2 on the virtual address space we obtain the different curves shown in Eq. 4.8. We observe here the maximum number of addresses for 64 bit identifiers and a variable number of pseudonym set size at an admitted collision probability of 0.1%. Note that the original address space is still represented when the addresses per set is 1. The goal is to estimate the loss of effectiveness of the address space through the usage of pseudonyms. This loss is represented by the curves of 2,4,8 and 16 addresses per pseudonymity set.

By taking the results of Fig. 4.8, we can establish a comparison between the original space and the reduced VAS. We can determine the proportion of space that is consumed by pseudonyms as the difference of addresses from the curves in Fig. 4.8. If T_1 is the total number of addresses given by the collision probability approximation for 1 address per user (the real address space), and T_n is the number of addresses as calculated for Fig. 4.8 (the virtual

[2]The definition presented is virtual since the network addressing space will always ultimately depend on the size in bits of the addressing structure and its eventual hierarchy, regardless of how many identifiers are used per user or device.

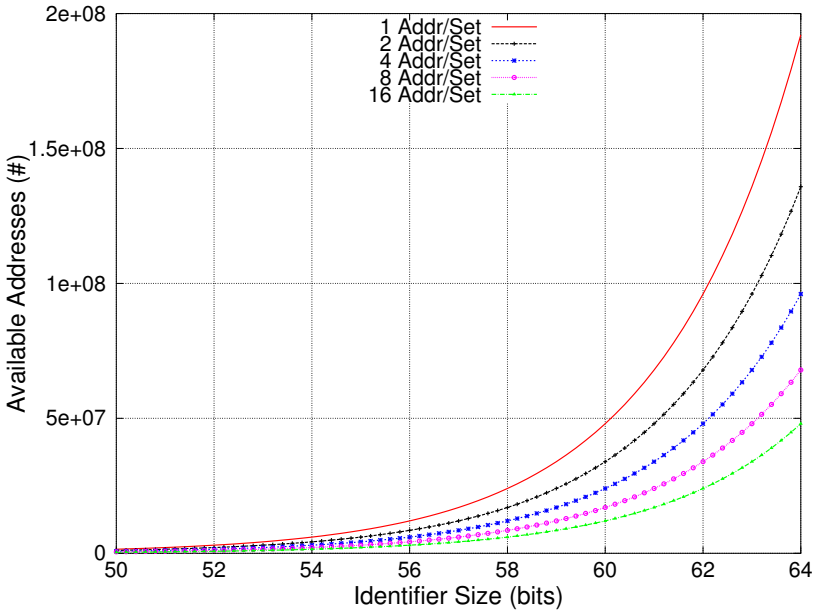


Figure 4.8: Available addresses for different pseudonym set sizes (1,2,4,8 and 16) and increasing identifier sizes (at 0.1% collision probability).

address space), it is possible to determine the wasted space (Fig. 4.6), shown in Fig. 4.9, for 64 bit addresses.

$$W_{space} = \frac{T_1 - T_n}{T_1} \tag{4.6}$$

The results clearly show the addressing space loss introduced by pseudonymity, that increases by each pseudonym introduced into the user portfolio. The important conclusion from these results is that any pseudonym solution must constrain the usage of pseudonyms to an acceptable level. Given current link layer and network technologies, this introduces a practical requirement of no more than 5 to 10 sets per user, at the cost of more than 70% of the address space, a price only affordable when the identifiers are long enough to sustain a vast address space even at 30% of their capacity.

This analysis provides us with the tools to understand whether a particular pseudonym solution is able to be effectively deployed at the network level. Such findings ultimately impose two overall requirements: any solution must provide a control plane that determines policies to guide and manage the IS and its associated network pseudonyms reinforcing the previous control requirements; and also, measures must be taken to determine and minimize the burden on the respective namespaces, ultimately pointing at design constraints.

4.4.4 Requirements

Throughout the previous sections we have highlighted major requirements that any pseudonymity solution must address. Here, we formalize such requirements to enable proper evaluation and discussion of any pseudonymity solution. We focus on privacy (*R1*), information control

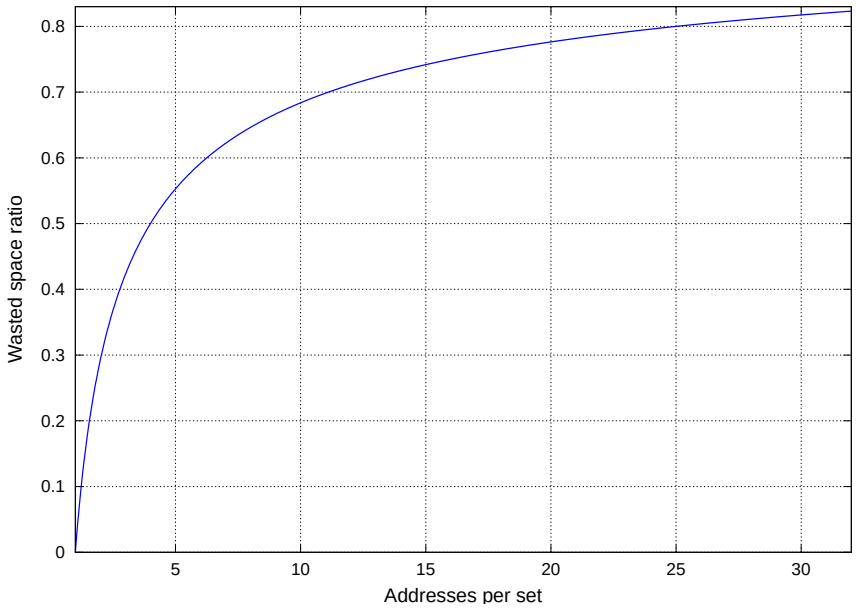


Figure 4.9: Space consumed by using pseudonymity sets.

(*R2*) and network (*R3*) requirements, along with corollary requirements (e.g R1.1-R1.2), as cornerstones of any pseudonymity solution for privacy. Failing to acknowledge any of these requirements may render any solution ineffective or unusable.

R1. Reduced Information Set The information contained in each set must be reduced, to minimize information leaks under a minimal disclosure approach.

R1.1. Identifiers cannot be shared between sets Sharing identifiers between sets is the fundamental principle for linkage and should be made impossible.

R1.2. Information Set scalability The sets must be made as large or as small as needed, with granular control. Large sets increase linkability, but also personalization, while smaller sets provide greater privacy.

R2 Control Information Linkability Each IS must be adequately managed, controlling how and what it links, because uncontrolled information can lead to linkage.

R2.1 Handle all network layers Identifiers must be handled on every network layer. Failing to handle one layer can lead to linking on other layers, forfeiting privacy.

R2.2 Manage network identifiers collectively: All generated pseudonyms must be handled jointly on all network layers, conforming to *R1* and *R2*.

R2.3 Dynamic Identifier Generation Functions Functions that dynamically generate identifiers, to act as pseudonyms, should exist and avoid easy correlation of the generated identifiers.

R3. Preserve Networking Functions Protocol function must be retained even in light of several pseudonym identifiers on the same communication layer.

R3.1 Preserve Address space Retain address space properties and assumptions, evaluating any impacts resulting from using several identifiers per user (pseudonyms).

R3.2 Handle identifier collision Dynamic identifiers, acting as pseudonyms on the network, can lead to collisions, which must be resolved at the cost of failed communication.

R3.3 Minimize Performance Impact Introducing multiple identifiers on different network layers must have a low overhead, preventing excess communication delay or inflict hefty bandwidth restrictions.

These requirements provide the foundations upon which we can build a scalable and encompassing solution towards network based privacy. By relying on cross-layer pseudonymity, we are in a position to instantiate the multiple requirements and limitations into a realization of the Virtual Identity approach onto the network. In this context, Virtual Network Stacks appears the natural realization of VIDs on the network.

4.5 Virtual Network Stacks

Instantiating the Virtual Identity model implies creating several layers of pseudonyms, wrapping the user in pseudonym identifiers, defining a virtual entity. To accommodate this assumption, we must follow a similar approach with the user's device, thus providing a virtual terminal that adapts to the new network reality. However, our notion of a virtual device pertains only to the network interfaces, where the most relevant metaphor for our solution becomes that of virtual network stacks, instantiated as required by each identity, towards the network. In summary, Virtual Network Stacks [97] (VNS) provide a privacy enhancing solution by applying a multiple pseudonym approach, where each pseudonym set constitutes defines a virtual network stack, bound to a single Information Set. These stacks are instantiated per user identity, avoiding any possible correlation between identities, partitioning the network information into a contained IS.

Sec. 4.4 outlined the methods required to apply pseudonyms on any network layer. It defined key principles and requirements that should appear in privacy-oriented pseudonym solutions. Those requirements are taken into account in our proposal, which aims at providing a full network pseudonymity solution, tightly coupled with Identity as a control layer, stemming from the VID layer and relying on IdM technologies. It provides a cross-layer approach, scaled through identity properties and tied with virtual identities.

The support and management of a virtual stack requires terminal architecture modifications, both in terms of control and data. The current legacy model is connection oriented: identifiers are used or generated at the pace they are required to connect to the network at different layers (e.g., an IP address is generated or assigned at the time the terminal connects to an access router, and is normally used by all upper layer protocols). The proposed approach turns the focus to identity, generating different identifiers when an identity wishes to connect to the point of service.

In practice, relying on virtual stacks to convey pseudonymity consists on assigning different Link, Network and Transport layer pseudonyms for each identity, making it impossible to correlate different identities. Thus, one device is transformed into several virtual terminals, And, as mentioned, to seamlessly support a virtualized network stack, we must provide modifications to the terminal network support, which we achieve trough the metaphor of virtual devices, used jointly with the virtual identities. always hand in hand with the identity

management system. This coupling allows retaining pervasiveness and personalization by linking the pseudonyms to the identity model, creating different identifier sets per VID. This is only possible with a clear separation between control and date plane: managing information sets, assigning connections and distributing actions across each set, are handled on the identity control plane; translating information concerning each virtual stack into packets in the network is handled on the data path.

However, to properly address the network environment, we must start off with a proper network model, where the (NGN) network architecture is properly outlined, so that pseudonyms are applied consistently throughout all the network required mechanisms. We present such an NGN model to clearly define the boundaries of how and where we should apply the control, through identity, and network mechanisms, through virtual devices, that enable the support for multiple identities.

4.5.1 Network Model

When discussing network pseudonyms, it is easy to arrive at the conclusion that generating pseudonyms for all network interactions will greatly depend on the employed network protocols and mechanisms. Therefore, a practical functional solution must be framed within a particular network model, which in our target cases revolves around a NGN network, focusing on all-IP (4G) scenarios. This is important, because ignoring more complex network interactions, such as mobility support in current and future network, may yield a too simplistic network instantiation, rather than a framework that enables future network evolutions in the light of a broad array of services, which are appearing in every dimension of the network. Furthermore, the type of privacy issues that we try to tackle are not only present in today's networks, but are amplified by NGN, or 4G, network models, by all the reasons mentioned in Chap. 1.

Therefore, the network model that serves as reference for our approach, shown in Fig. 4.10, is based on 4G scenarios. These environments contain heterogeneous access technologies, such as WiFi, Wimax or DVB, seamlessly integrated into the global architecture. Similarly, user terminals are evolving into multi-technology devices, capable of sustaining several connections across different interfaces. This multi-technology availability enables a better user experience, serving the 4G paradigm of "always on, always best connected"[3] and creates a real possibility for multi-homing and development of multi-homed based solutions. In heterogeneous multi-homed capable networks, mobility is no longer governed by signal availability, but by user preferences, possibly identity based - each persona governs its mobility patterns and selections. In this context, flow based mobility is a well suited candidate for the minimum granularity available to mobility mechanisms. The way flows are distributed through the available interfaces should depend on network availability, provider information, cost and more importantly, on preferences set by the user or, in our privacy model, by the virtual persona. This allows the user to take full advantage of the concepts of identity management, personae and multi-homing.

While the proposed solution to tackle privacy can be applied to any mobility scenario, it becomes particularly interesting in such volatile environments, where a user can distribute flows belonging to different identities across the same interfaces. It would lead, in the current

[3]This is also commonly referred to the "always best connected" paradigm, verbalizing the ever more pressing expectations of users to enjoy faster and more reliable connection even when on mobile platforms and nomadic scenarios.

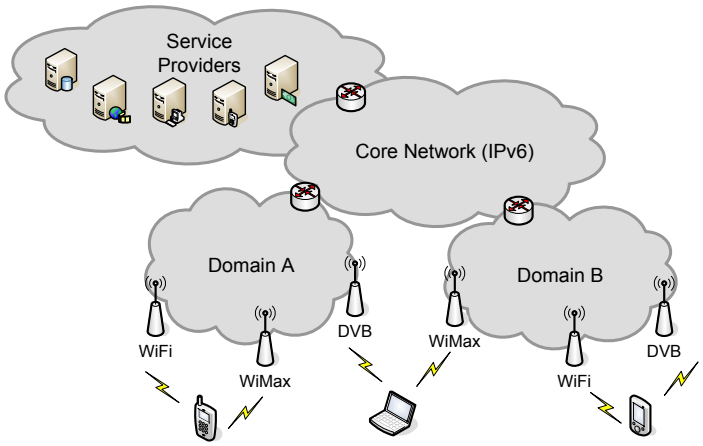


Figure 4.10: Network Architecture Overview.

models, to the sharing of similar network addresses, and consequently, to the linkage of identity personae, which would increase the threats previously presented. However, multi-homing capabilities can be used to diminish the effects caused by mobility on identity correlation. Considering an example of a terminal with only one interface, when a new network with better signal is sensed, the terminal performs handover and all flows, even belonging to different virtual personae, need to handover simultaneously as well, which gives information on the correlation of the different flows and different personae.

It is worth mentioning that this is the reference network used within the Daidalos II project platform [71], which aimed at providing a 4G all-IP network with multiple service support and heterogenous technologies. The proposed approach was included and explored within the scope the highlighted network model, and also within the Daidalos II architecture.

4.5.2 Control Through Identity

The first step towards supporting the VNS paradigm is to introduce a control layer, that directly instantiates identity layer functionality, pulling the VID concept further down the OSI stack. This control plane interacts with applications, which are used as input for network stack management. As can be seen in Fig. 4.11, in the terminal control plane, applications might be identity aware and provide specific inputs to the management plane, deciding independently on the usage of identity personae; or applications might be legacy, where the management decisions will be extrapolated by a legacy interface component, that analyses the application requirements and selects an appropriate persona.

The control plane dynamically handles the creation and removal of the pseudonym sets that become available to different identities. It also enables the selection of identities, i.e. enabling an application to set an identity, implicitly selecting all the associated pseudonyms. In VNS this is achieved by closely coupling the network events and identifiers to an IdM system. By introducing an IdM layer that interacts with applications (which are the real information producers and consumers), it is possible to decide when and how new information sets need to be introduced.

Binding applications to identities implies that applications are identity aware, and have

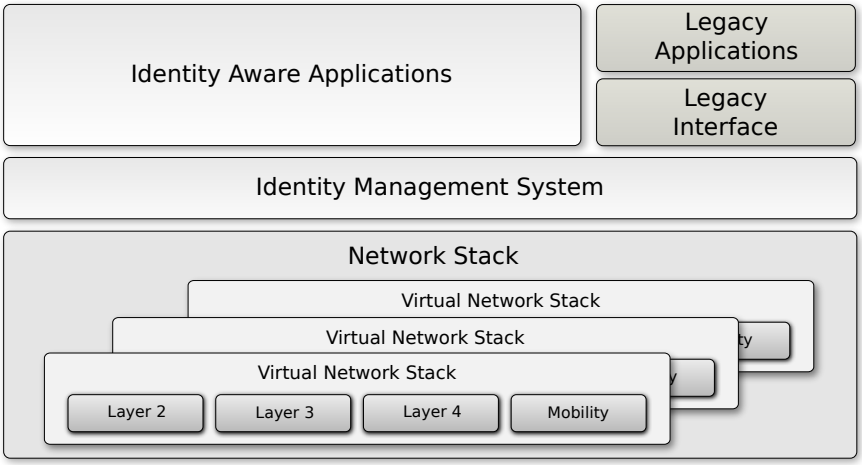


Figure 4.11: Terminal control plane and associated Virtuanl Network Stacks.

an algorithm for selecting an identity from the Identity Control Layer (Fig. 4.11). Otherwise, it is necessary to develop an abstraction layer that preserves the current operating system semantics, resulting results in the shown legacy interface. This legacy interface is required for identity-aware protocols and applications. As such, mechanisms in the terminal must enable the instantiation of multiple protocols, by means of a control application that adequately connects the correct input and output flows, tying them to identity information.

Addressing scalability issues identified in Sec. 4.4 are tackled by a tight integration with the identity plane. Scaling is provided through IdM properties, which inherently provide a way of throttling network usage versus provided privacy. Given the limited identities or personae assumed by the users when interacting with the digital world, it automatically restricts the total number of Information Sets, and thus the number of required network addresses. Given the highly customizable privacy framework provided by the IdM system, through policies, it is possible to support a fine-grain control to suit the requirements of the user, and still keep the creation of pseudonyms under control.

4.5.3 Cross Layer Pseudonyms

For a terminal to disguise itself under several layers of pseudonyms, it is necessary to support the concept of multi-layer pseudonyms on an Identity basis, leading to the concept of a VNS per identity (or VID). The VNS approach provides the instantiation that enables the required network support for the VID framework. As such, each VNS represented in Fig. 4.11 is associated with a VID. As such, the identifiers present in each VNS belong to a distinct IS, as defined in the VID framework. Each identity is assigned different L2, L3 and L4 pseudonyms, making it impossible to use these identifiers for VID correlation. We assume that, at the application layer, the identity scheme in use provides different pseudonyms for the user, presented at different services.

While the control has been discussed, several aspects of how to support network pseudonyms in the network have not yet been discussed. We must provide a mechanism by which pseudonyms are made persistent on each layer, obeying the network model. To gain control

of the network stack, identity needs to have direct influence on the control plane. However, the data plane is entirely sustained by a VNS, given the fact that identity operations are not required on data packets. The objective is to provide control over the network stack without linking identifiers that can be correlated. While the VNS concept is fairly simple, it needs to be extended to all mentioned layers, providing the means to instantiate pseudonyms at every layer, taking privacy into account. The next sub-sections address this instantiation over the network stack.

An additional layer, the Virtual Device Control Layer, is introduced in Fig. 4.11 to enable virtual identities towards the network and supporting the virtual device metaphor, as discussed in the following sections. With the proper abstractions and control path, the modifications to the data path in the terminal are minimal. The data path must be capable of handling packets directed at different identifiers, which are not directly related to the existent physical devices, but rather to the identifiers or pseudonyms associated with the different identities.

4.5.3.1 Virtual Devices

Integrating the VID concepts in the network, and especially into the terminal, requires more than just support mechanisms on the device. It requires a metaphor that adapts and adjusts to the requirements imposed by virtual identities, and especially virtual network stacks.

We turn to the concept of a Virtual Interface (VIF), or a virtual network device, to support the idea of multiple concurrent users attached to the same physical device. Emulating physical devices presents the desired metaphor for enabling different pseudonym sets, or VID, since at the operating system level the network addresses are associated with a device. A VIF is a normal device from the operating system perspective, but which does not represent a real device towards the network. Instead, it must be associated with a real network device, thus virtualizing a network device over its real network counterpart. This facilitates the existence of several virtual devices, over a single network interface card, where each VIF is associated with a new link layer address, a pseudonym, with different upper layer addresses for that particular device. This enables the support of disjoint identifier sets across different virtual devices, while preserving operating system semantics.

From a terminal architecture perspective, the virtual device abstraction presents the necessary bindings to cope with virtual network stacks and cross-layer pseudonymity. By instantiating multiple devices that are bound to a user identity, we are able to support different pseudonyms on different layers, aligned by the layer shown in Fig. 4.11 where the Virtual Device Control Layer interacts with the identity layer to provide several virtual interfaces that are associated with their own VNS. Once these concepts are in place, we can start instantiating virtual devices to provide pseudonyms, as shown next.

However, from the network perspective, every VIF should be independent of of its real associated device, and all other virtual devices. This creates the illusion that multiple terminals exist in the network, one for each VID.

4.5.3.2 Link Layer Pseudonyms

On the Link layer, independent addresses need to be generated for each VNS. These addresses do not correspond to the physical address present in every network interface card. This is in fact a virtual address that is created for every persona. While it is straightforward to

generate addresses, the real interfaces need to be abstracted on the user's terminal, so that each virtual stack is supplied with the necessary device information, such as signal strength or provider availability. This can be accomplished in software by using a Virtual Device Manager (VDM) that controls the real interfaces, and supplies the necessary primitives and information to the VIF, which contains a virtual MAC address. It is important to notice that a VNS might be extended to more than one physical device, therefore, creating several VIFs under the control of a particular identity, as show in the instantiation examples present in Sec. 4.5.3.7. From the network point of view, each VIF (with its associated virtual MAC) represents a different device, competing among each other for network access (otherwise, it would be easy to see at the link layer, especially with IEEE 802.11 protocols, that at each DIFS interval assigned to a particular device, spoofed frames were being received). While the aforementioned abstraction bloats the network stack, it is necessary to cope with the privacy issues, namely to have uncorrelated physical devices and addresses. Moreover, it provides the added benefit of enabling terminals to deal with future access technologies, such as multi-head radios and heterogeneous technologies, coping with the previously mentioned 4G scenarios.

4.5.3.3 Network Layer Pseudonyms

Using pseudonyms at the network layer is simpler than at the link layer. Network identifiers are used as locators to determine the routing path. If we consider IPv6 as our target protocol, the address is usually auto-configured. The support of a Virtual IP address can be achieved by running independent instances of the Neighbor Discovery Protocol (NDP) performing Stateless Address Auto-Configuration (SLAAC) for each virtual interface. This leads to independent addressing for each identity. The support of stateful address acquisition methods is also straightforward and requires several instances of a DHCPv6 daemon. While running several protocol instances poses a strain on the device and network, its impact is much lower than in [56], due to the fact that the instances run per persona, instead of on a per application basis.

4.5.3.4 Mobility Pseudonyms

Mobility solutions at the network layer tend to create an indirection between the locator, the IP address, and the identifier used by the transport layer, which is the actual endpoint of the transport connections. We discuss two mobility protocols that are easily extended to support the VNS concepts, MIPv6 [77] and HIP [112].

MIPv6 The Layer 3 pseudonyms mentioned in the previous section provide pseudonymity for the Care-of Address (CoA), which is in fact the aforementioned locator. For each VNS, a different HoA should be used. This means that each VNS has a different HoA and a different set of CoAs. Each HoA needs to be generated and independently registered at the Home Agent (HA). Again, added signaling is required to support privacy.

HIP To use HIP each VNS should generate its own HI, and corresponding HITs for each identity, that are passed onto the transport layer. HIP has the same problems as MIPv6, in the sense that more signaling is required to support the VNS concept.

4.5.3.5 Transport Layer Pseudonyms

When mobility is in place, using pseudonyms for mobility already makes the transport layer VNS ready, since transport protocols, such as TCP and UDP, establish their bindings with the mobility identifiers, i.e. HoA or HIT. If no mobility solution is used, one can argue that the L3 pseudonyms already provide the necessary VNS support for the transport layer, because the bindings are to the L3 identifiers, and the IP address (IPv6 in our case) is used both as locator and identifier.

4.5.3.6 Application Layer Modifications and Pseudonyms

As already discussed in the previous chapters, most Identity Management solutions work at the application level, like OpenID [122], Cardspace [109] or [20]. Nonetheless, some modifications are required to keep user privacy intact. The common use case for identity is the connection to a service presenting the necessary credentials, or identity claims, to that service (this acts as the identity selection period). The operation mechanism of these identity models considers that, upon initial connection to a service, the user is presented with a graphical user interface, known as an identity selector, from which the user selects the desired persona. While this applies to application layer privacy, it breaks lower layer privacy because the user will have a set of identifiers already in use when connecting to the desired service. In fact, the user would be bound to the present VNS, since it would be hard to change the underlying identifiers without having to start a new connection to that service. Furthermore, the change would provide an attacker with the means to link past and present identifier sets.

The user should be presented with the identity selector as part of the application startup. An example of this model is the case of a user browsing a web site and being prompted with the selection of an identity for that website. This model should be replaced by the identity selector being shown when the browser starts, leading to the paradigm that the identity is already selected upon connection to the website. The problem of identity selection has been addressed as a spin-off of the work presented in this Thesis, as shown in [126, 127], which shows pragmatic strategies to address the problem. It has also been addressed in the identity based architectures presented here, in Sec. 6.3, by linking the network identity selection to the identity selection in a SAML [20] based architecture, in the scope of the SWIFT [72] project.

The proposed IdM solutions also take care of the required pseudonyms towards different applications, includes different services, web applications, and different application layer protocols, such as SIP. The latter case is important since it can also provide mobility support on the application layer, for VoIP solutions. Using the IdM possibilities, it is possible to register different SIP identities (and identifier,s based on URIs) that use IdM based pseudonyms (e.g. SAML).

4.5.3.7 Instantiation Example

The best way to understand the VNS approach is to observe a working example of the proposed abstractions. We show a detailed example on an instantiation of a VNS across multiple layers and interfaces. Our example scenario, illustrated in Fig. 4.12, consists of two personae that run independent applications spanning over two real interfaces. In this case, two virtual stacks are instantiated to cope with the desired privacy levels.

Each network stack can use one or more application identifiers bound to the specific identity. Going through the layers, for transport, each VNS is bound to one MIPv6 HoA,

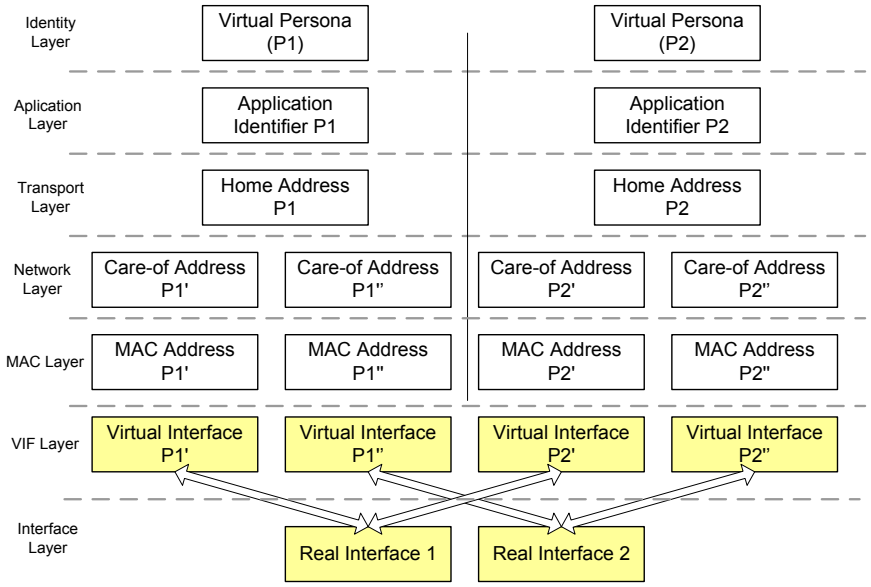


Figure 4.12: Virtual Network Stack instantiation for two personae.

providing the appropriate mobility bindings. The network and link layers depend on the number of active interfaces, that are actually connected to virtual interfaces. From this example, we observe that two real interfaces spawn into four virtual interfaces (two per identity), leading to the requirement of one MAC address and CoA for each virtual interface. We then have two MAC and IP addresses per identity: there will be no linkage between identifiers across virtual stacks, and consequently, across identities.

4.5.4 Prototype Implementation

The first step towards validating the proposed requirements demands a prototype implementation of the architecture discussed in the previous sections. To achieved this, we implemented the discussed VDM as part of the VNS proposal. This prototype [57] was initially developed as part of the IST-Daidalos project architecture [3] demonstrator, which deployed a next generation network prototype with multiple access technologies, such as WiFi, WiMax, UMTS, DVB and Ethernet, providing a playground for testing a pseudonymity approach. The prototype, along with the solution, evolved with greater IdM focus on the identity framework of project IST-Swift [90], closely coupling the solution to IdM frameworks, as shown by [93], which is one of the solutions developed in the scope of this Thesis, highlighted in Sec. 6.3.

The VDM, presented in Sec. 4.5.4.1, was implemented on a GNU/Linux operating system (Ubuntu Linux), along with solutions for the control path, and multiple access technologies. The software module follows the architecture described in Fig. 4.13 that contains the three main modules coordinated by the VDM control layer. The *Device Management* provides the creation and deletion of the virtual devices, which is in fact an abstraction that controls the *TunTap* Linux Kernel module, along with a special purpose WiFi monitor device, to listen to incoming packets not directed at the real device. The *Socket Management* module provides socket management tools to receive and transmit packets, between network, virtual device, and interprocess communication inside the system. Finally, the *Packet Forwarding*

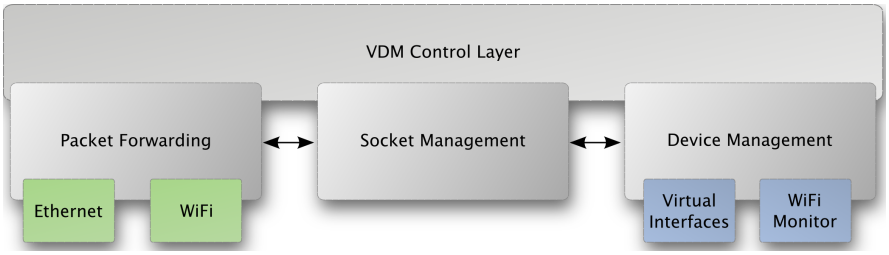


Figure 4.13: Virtual Device Manager architecture.

module provides the packet switching capabilities and primitives, along with the encapsulation functionalities, for both WiFi and Ethernet technologies. Most state is kept in the control layer that provides the "intelligence" layer to inter-relate the different modules.

The physical support of the prototype relies on generic Ethernet cards for the wired component, and IEEE 802.11b *Atheros* chipset cards for the WiFi component. The *Atheros* chipset allows packet injection through the *Madwifi* drivers for Linux, enabling the emulation of several wireless cards. Since the support for WiFi relied on packet injection, and not real driver support, the performance over the wireless interfaces is not discussed here given that it is not possible to establish relevant comparison with normal WiFi traffic. Therefore, we focus on the performance of the Ethernet card. For the computing platform, we used an AMD Athlon 64 3000+ processor to run the VDM, and a Core 2 Duo terminal to generate traffic flows - both machines pose no computational barrier. Below we present the major findings stemming from the experimental prototype, along with evaluation of pseudonym requirements, stemming from the previous pseudonym analysis (Sec. 4.4).

4.5.4.1 Implementation: Virtual Device Manager

Emulating physical devices presents a proper metaphor for enabling different pseudonym sets, because at the operating system level the network addresses are bound to a device. A virtual device is associated with a new link layer address, a pseudonym, and with different upper layer addresses for that particular device. This support separates identifier sets across different virtual devices, thus supporting multiple VIDs, while preserving operating system semantics.

From a terminal architecture perspective, the virtual device abstraction presents the necessary bindings to cope with VNS and cross-layer pseudonymity. The VDM is the core software component that enables the VNS system. Its primary function is to manage the virtual devices, present them to the applications and preserve the semantics that are available on real devices. Multiple pseudonyms are assured, as mentioned above, by maintaining the virtual devices and allowing them to interact with the system and network as real devices. Maintaining the virtual devices implies that the VDM creates and destroys network devices per identity, as required by the supervising IdM System.

Beyond these functions, the VDM provides support for identity aware applications, delivering the control packets when necessary (with identity information), and instantiating multiple protocols per virtual device. Protocols such as MIPv6 bind to virtual devices and use them as the basis for mobility. Assuming that the MAC layer technology is capable of handling such requirements, supporting this operation is straightforward. On the data path,

Virtual Network Stacks	VNS 1	VNS 2
IPv6 Global Address	3ffe::2ff:abff:fe8e:cfee/64	3ffe::2ff:3cff:fedb:78cb/64
IPv6 LL Address	fe80::2ff:abff:fe8e:cfee/64	fe80::2ff:3cff:fedb:78cb/64
MAC Address	00:FF:AB:8E:CF:EE	00:FF:3C:D8:78:CB
Virtual Interface	vif0	vif1
Real Interface	Wireless Interface / ath0	

Table 4.1: Virtual Device Manager instantiation example.

the VDM must direct packets to and from the appropriate real devices, while preserving the associated technology mechanisms, abstracted from the system. In fact, once the virtual interfaces are available, they allow multiple addresses to rely upon operating system semantics (such as routes at the network layer) - virtual devices emulate real ones. The real devices are only used as the physical access mechanism to allow actual packet transport over the link technology. For a VNS instantiation example, refer to Table 4.1 below.

4.5.4.2 Evaluation

The developed implementation enables two types of analysis: a qualitative analysis on the effective separation between different virtual identities, mostly in terms of signaling and data packets; and a quantitative analysis that sets an upper bound for performance. In Table 4.1 we can see a sample instance of the prototype including device names, link layer addresses, Link Local and Global IPv6 addresses. The VDM creates two different virtual devices on top of their physical counterpart, and consequently it creates different identifiers. The virtual devices contain the upper layer addresses, which are in fact pseudonyms for the different identities.

To analyze the performance we have chosen to consider three metrics: delay, bandwidth and interface bootstrap times. All tests were subject to at least 15 runs, of which we present the average values, later compared to the reference values for each type of test. For more volatile values we include a 99% confidence interval.

Delay is one of the most relevant metric when analyzing this type of prototype. The VDM has to process and distribute incoming packets to the correct virtual interfaces. When multiple real interfaces exist, it is also necessary to select the correct output interface: this is a one-to-one mapping, kept at each virtual interface. The process requires copying packets between different buffers [4], introducing delay beyond the processing itself that must be accounted for.

The delay presented inFig. 4.14 shows an average penalty of $36ms$. These results were obtained by directing multiple UDP traffic flows at several virtual interfaces, where each virtual interface carried one flow. The additional delay spans from 20% to 50% of the total delay, and the variation of the delay as the number of flows increases does not stray far from the reference measurements. This result presents an upper bound performance impact, resulting mostly from the implementation (using buffer copies in user space, due to Tun/Tap operation), which could be mitigated in kernel only implementation.

[4]In a production-grade implementation all copy operations would be avoided through zero-copy implementations, which is the standard for most operating systems. In this operation mode only buffer references are passed, instead of copying packets.

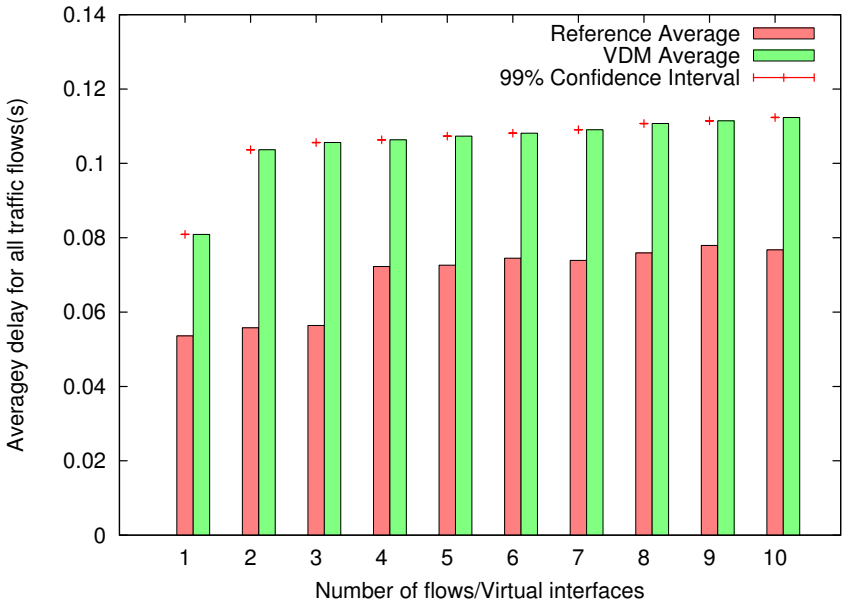


Figure 4.14: Communication delay for UDP.

The bandwidth impact analysis consists on creating multiple TCP flows and compare the average bandwidth allocation for all flows against reference measures. Each traffic flow is also assigned to an interface. Fig. 4.15 presents the measured values. We can observe that our implementation does cause a reduction in average bandwidth, as expected. The average bandwidth decreases as the number of interfaces increases, since the flow contention becomes more critical than the actual processing delay. For clarity, we also present the bandwidth difference within the graphic.

Finally, we recorded bootstrap times of the virtual interfaces to determine if the number of existing virtual interfaces affects the bootstrap of the newly created ones. These were obtained from the instant the virtual interface was created, until the instant the interface established connectivity, by acquiring addresses and routes. Fig. 4.16 presents the average bootstrap times: we clearly observe that the bootstrap time is not affected by the number of previously instantiated interfaces. This was the expected behavior for the relatively small amount of virtual interfaces created (as required by the identity model).

4.5.4.3 Technology Limitations

While we have implemented WiFi devices using a modified driver, packet injection poses drawbacks. Multiple 802.11 associations must be supported in the actual device driver or in the card firmware. Failing to do so places unnecessary strain on the system, and allows an attacker to link different pseudonyms by identifying the expected station in a Network Allocation Vector (NAV) slot. Also, with current technology all associations must connect to the same channel. An interesting alternative would be to allow the connection to different access points on different channels. However, this would require specific equipment, either with time consuming frequency hopping, or multi-radio devices that would provide added value for both privacy and mobility aspects.

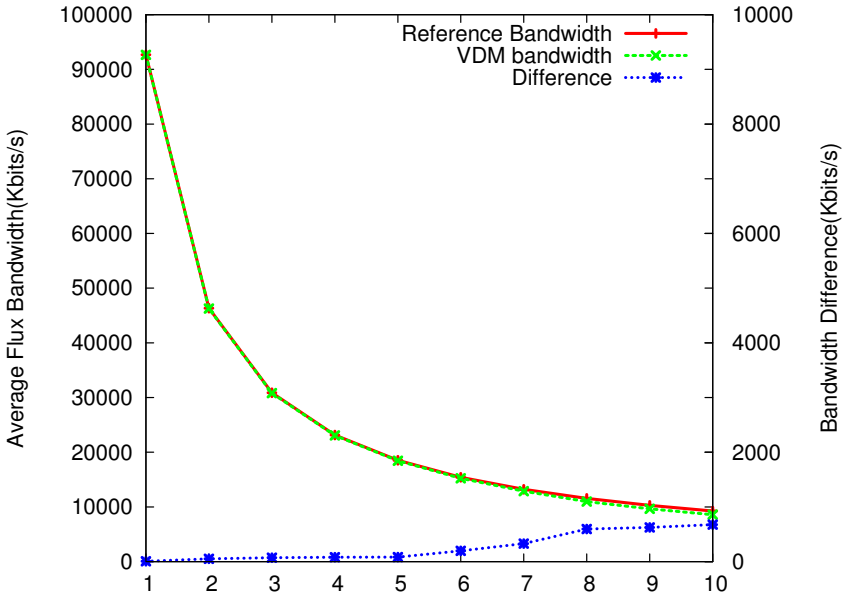


Figure 4.15: Average TCP bandwidth per flow/interface.

4.5.5 Requirement Support and Impacts

The generic analysis of network pseudonyms (Sec. 4.4) yielded a set of requirements that must be addressed when discussing pseudonym solutions.In this section we discuss the presented VNS solution and accompanying realization (through the VDM) in light of those requirements, establishing performance and deployment bounds that shed a different light on the presented requirements.

The first and foremost observation of the prototype wehn compared to the conceptual model, is that this approach avoids unwanted network identifier linkage, between different user identities. This provides an experimental validation to the proposed approach, that goes beyond mathematical proof. The link to IdM technologies satisfies both *R1* and *R1.2* simultaneously, where information is confined to each identity, and limiting the required identifiers on one per identity, rather than one per network interaction. Whenever new sets are required, the user can simply create a new view over his user identity, providing a new set of identifier towards service and network, accommodating the scalability requirement put for by *R1.2*. In fact, when properly configured, each stack runs independent instances of each protocol, using different and unlinkable identifiers, acknowledging the validity of a cross-layer pseudonymity approach. This clearly satisfies *R1.1*, given that no identifier is shared across sets. This defines the information set approach on the network layer, as a realization of the proposed concepts.

Similarly, it is in the virtual interface and identity concepts that we handle the *R2* family of requirements. The cross-layer virtual interface approach coupled with high level identifiers provides the foundation to address *R2.1* (handle all identifier layers), enabling vertical identifier sets. Since identifiers are bound to the virtual interface, we must manage them aggregately, addressing *R2.2*. The virtual interface approach also addresses *R2.3* (dynamic identifier generation functions) by reusing available operating system mechanics that gener-

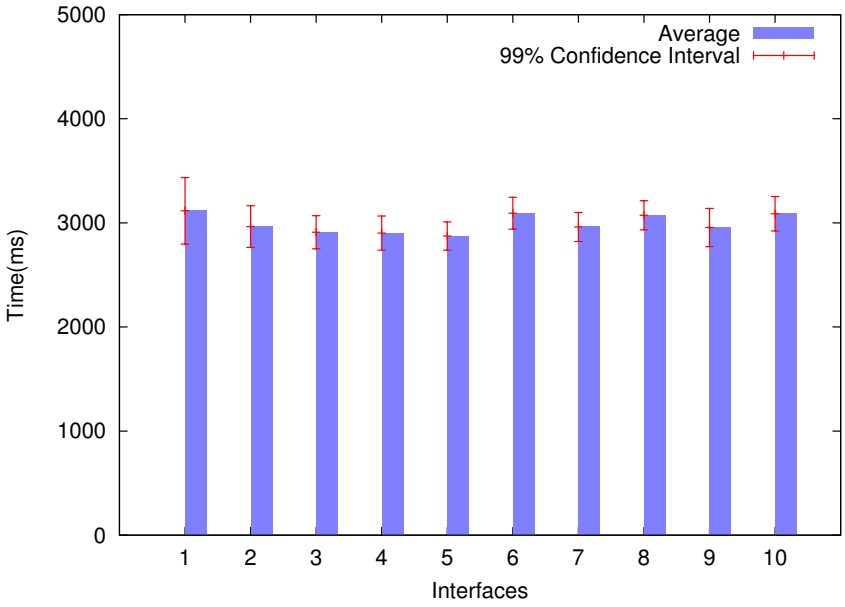


Figure 4.16: Bootstrap delay of several virtual interfaces.

ate network identifiers (MAC addresses on the TunTap kernel module, and consequently IP addresses). Also, it is worth noting that the implemented solution resorts to virtual interfaces that emulate real interfaces, retaining compatibility with current network and operating system semantics, adhering to *R3*.

However, as mentioned, there are impacts on different levels. The drawbacks must be considered, in order to prove that, beyond a privacy protecting solution, we have a feasible model fitting the requirements of privacy-preservation.

Discussing the theoretical values only makes sense in light of the identity oriented approach proposed by VNS, especially considering that it limits the amount of identities the user will have to an average of 5 [129]. This is an important aspect since it guides the expected degree of impact on the network stack and operational mechanisms, in order to conform to requirements *R1.2* (scalability) and *R3* (network semantics) and its corollary requirements.

4.5.5.1 Addressing Impacts

We have identified that one of the most critical impairments for pseudonymity systems lies in the addressing impacts, embodied by *R3.1* and *R3.2*. We must consider first and foremost that VNS provides a cross-layer approach, forcing us to consider the addressing impacts independently on each layer due to different identifier sizes. We focus on the two that cover most interactions: link layer (MAC Addresses) and network layer (IP Address). For the MAC layer address, where identifiers have 48 bits, we have a maximum of 2^{48} addresses. For IPv6 the addressing space is larger, having the dynamic part set at 2^{64} (the remaining bits represent the network).

The evaluation of the impact can be calculated as according to Eq. 4.6, using the Virtual Address Space for k sets on 48 and 64 bits, which is reflected in Fig. 4.17. This result is very similar to Fig. 4.9, but presents a nuance, since we introduce the area where the VNS

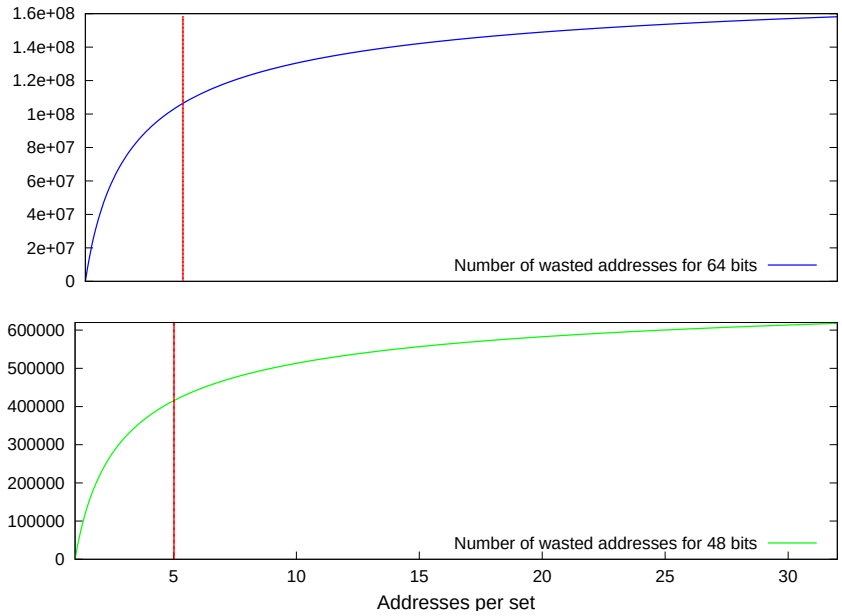


Figure 4.17: Evolution of the wasted addresses, using the Virtual Address Space, for 48 and 64 bit identifiers.

will operate, limiting the space waste to 5 addresses per set, which corresponds to wasting 4.1452×10^{05} 48 bit addresses and 1.0612×10^{08} 64 bit addresses; this is an acceptable value in a universe of respectively 2.8147×10^{14} and 1.8447×10^{19} addresses, addressing *R3.1*. It should be noted that, while both curves depicted in Fig. 4.17 follow the behavior presented in Fig. 4.9, the vertical axis represents the absolute number of addresses; as such, both curves differ by several orders of magnitude. These numbers are still within the acceptable range, especially considering that the solution is controlled by identity and this will throttle identifier consumption. Assuming IPv6 addresses are generated from MAC addresses through EUI-64 expansion, then whenever we detect a collision at the IPv6 layer, we force a complete address regeneration, thus solving *R3.2*, avoiding collision on both MAC and IPv6 layers. However, such findings lead us to conclude that a major requirement is to produce, in the future, a mechanism that considers collision detection for link layer addresses, much like Duplicate Address Detection (DAD) in IPv6, because addresses are now dynamic, with non-negligible collision probability.

4.5.5.2 Bootstrapping Impacts

An important part of *R3* relates to performance (*R3.3*). We established earlier that bootstrap times might vary depending on the number of interfaces. Ideally, the optimal setting would be one where bootstrap times are fixed and independent of the number of virtual interfaces, and do not grow with the addition of more interfaces, as opposed to growing whenever a new interface is added (e.g. linearly or exponentially). The results presented in Fig. 4.16 seem to lean towards the former case where multiple pseudonym instantiations occur in parallel, with a bootstrap time of approximately 3 seconds. This setup overhead is acceptable, even for a considerable number of stacks, given that session establishment occurs in a period where the

user is less susceptible to delay - in fact he will not be aware of it, thus partially solving *R3.3*, missing only performance impacts discussed in Sec. 4.5.5.3.

It is however necessary to point out that the bootstrap process in the prototype tests is a fairly simple one, mainly acquiring the minimal necessary network addresses. For advanced bootstrap sequences, that go beyond basic IPv6 connectivity and that might require key agreements or cryptographic negotiations, contention associated with those specific processes might significantly degrade these results. In registration or authentication processes that require cryptographic key generation from hardware devices, the key generation will most likely impose contention for a large number of pseudonyms. This has been already the focus of related work, based on VID evaluation in AAA scenarios that require proper user authentication and authorization, as shown in [13]. The main concern here is to realize which of these bootstrap steps are critical and, if possible, apply measures that circumvent its impact.

4.5.5.3 Performance Impacts

The presented implementation should be regarded as a prototype, and therefore, given its nature, it can only provide us with indicative values for performance. This allows us to frame the performance of a deployed pseudonym solution between the existing performance and the presented results. As far as prototype performance is concerned, we determined that, as expected, there is a negative impact in the communication system. However, this impact is extremely small when compared with the contention of several flows across the virtual interfaces. Delay also seems to hold up within reasonable levels. We can conclude that the pseudonymity approach has minimal network impact, properly addressing the missing gaps in *R3.3* and consequently *R3*. Given the previous findings, a commercial solution should perform very closely to the current single stack, suggesting that virtual stack implementations are feasible from an implementation perspective, as current virtualization efforts on commercial platforms begin to show.

4.5.5.4 Technology and Deployment Impact

Wireless networks are the best candidates for pseudonymity solutions, since it is difficult to determine the frame origin, i.e. physical device, (unlike wired technologies). However, wireless technologies are usually connection oriented and require explicit associations. The implementation of the WiFi interface made clear that connection oriented access technologies pose a problem when implementing a pseudonym stack approach. Packets from unassociated addresses are dropped, meaning that each virtual device enabled over such technologies needs to be registered and participate in the wireless medium contention mechanisms. Such is the case of IEEE 802.11b/g [66], requiring modified drivers to support its connection oriented nature. Association and dissociation procedures must be done for each address, therefore participating in the network allocation vector mechanisms. A solution to this issue would be a modified firmware that allows multiple associations. This would also require that the hardware is able to support a software MAC implementation, in order to control the device as a plain radio, and implement more complex features in the device protocol implementation. Generalizing the solution to all wireless technologies that will serve as the base for next generation heterogeneous networks, we reach the conclusion that the virtual device approach requires multiple association support in the device driver for connection oriented technologies. This is a crucial step toward a pseudonymity solution deployment, and should be treated as

a requirement for operating system design. The virtual devices and pseudonyms introduce a new layer of complexity and concurrency, which must be handled directly in the operating system. Concurrent access to real devices must be assured, along with new paradigms for both OS implementation and low level network access.

Application design should take into account the reality of multiple network stacks, and more importantly, of multiple user identities. Upper layer application support for identity and multiple pseudonyms is a definite requirement for evolved pseudonymity approaches, that is already being pulled by IdM, and goes beyond username and password schemes. This support translates into new paradigms in socket programming, since the current model of network sockets must be extended to enable the existence of multiple pseudonyms with different network realizations. Packets should be contained within their specific identity scope, and never cross identity boundaries, so that involuntary identifier linkage does not occur in the network, e.g. sending a packet to a virtual interface belonging to another identity. This should be guaranteed by the operating system.

A side effect that needs to be addressed in real deployment of both network drivers and operating systems is the distribution of traffic sent to the real device. Different virtual devices can have different QoS priorities introducing a new layer of traffic priority. Queuing must be applied to enable different priorities across different virtual identities, along with the appropriate management.

4.6 Conclusion

In this chapter we started dealing with the vertical aspects of privacy, because they can bring the alignment and coordination to the privacy dimensions. Based on the vertical and horizontal privacy requirements, from the PRIVED model, we attempted to build a vertical solution that aggregates user information into a coherent and manageable construction. The result was the notion of Virtual Identity, as a partial view of a digital entity. This allowed to size down the Information Set, a concept stemming from the PRIVED model, by providing the means to perceive the user in an NGN environment. This allowed not only to tackle the different dimensions of the user, but also its interactions with the network. We believe that the result is more than a model. It is a paradigm for regarding user in digital environments, opening the door to a new vision of user-centric mechanisms towards the network and services.

However, the effort cannot be simply conceptual, and must be related to the network and to the privacy model. To bridge this gap, we proposed a data model and an architecture, that effectively lay the ground work of the VID in the network. We proposed a data model closely related to the PRIVED model, building on the IS related concept, and centered around three generic constructions to model user related information: the Entity Profile, the Entity Profile Part and Entity Profile View. The EPV is the materialization of the VID, composed by a set of information blocks that are put together to form the user identity and the view that services can have over the user. The EPV has a main handle that is used as the main identity reference, the VIDID, and has filterable content, leading to the creation of different FEVP sets. The EPV, or VID, is composed of several information blocks, the EPPs, that can exist on the user local device, on the network, or in different providers (even in non-digital form). The EPV models not only the vision of a partial user identity for privacy reasons, directed identity towards services, but also an appropriate model to tackle information distribution over the different players involved in NGN. These are the tools needed to build the architecture, which

we define through a lightweight IdM-like approach. It provides the minimal components that enable the VID environments: the IDManager and the IDBroker. We showed how these functional boxes can work together to provide most of the identity and information related operations, supporting the concepts of EPP, EPV and corresponding structures and filters (through AAA).

An important step in the work evolution was the translation of these concepts into the network. The VID must have repercussions into the network. Focusing on the pseudonymity features promoted by the VID on the data layer and supported by IdM, we instantiated these proposals as network concepts. However, to establish whether pseudonyms were a feasible approach, we had to first determine the requirements and impacts using pseudonyms on every layer, along with the necessary adaptations. This led to a study on network pseudonymity, focusing on privacy, control and addressing, which are among the most pressing requirements for identifiers on the network. This study yielded several requirements that we believe capture the nature of pseudonyms on the network, and are accordingly channeled into our proposal. To accompany the VID solution we proposed Virtual Network Stacks, a mechanism to enable network based cross-layer pseudonyms, resorting to the metaphor of virtual devices. We proposed this solution aligned within an NGN network model, where control is exerted by identity (through IdM) and pseudonyms are generated per-identity (VID), according to the virtual interface approach. This allowed scaling the solution on the network side, while simultaneously respecting the vertical privacy boundaries defined by the VID. Finally, this solution was properly implemented and tested, so that it could be matched against the conceptual background outlined earlier.

The work presented throughout this chapter showed that the vertical approach is a requirement to maintain privacy, due to the need of aligning the information containers, as highlighted by the PRIVED model. In this scope, we acknowledge that the VID provides the necessary mechanics, paradigms and tools to instantiate vertical containers, and when aligned with the PRIVED model, they can be used to control and maintain user privacy. It is also important to conclude that there are means to translate the entire approach to the network by further promoting the pseudonymity solutions.

While the vertical approach is the best candidate to maintain overall privacy, it does not moot the horizontal threats to individual VID privacy. While it greatly reduces the risks, horizontal protocol threats can still be used to threaten, correlate and link identities, information, and consequently forfeiting user privacy. Therefore, we accept that there is room for improvement on each individual layer, starting right at the link layer. Taking the conceptual vertical model, implemented and tested, we move on to each individual layer, and try to complement the vertical privacy protection mechanisms with privacy enhancing technologies at all the relevant network layers. This is the focus of the following chapter.

Chapter 5

A Layered Approach to Privacy

A small leak can sink a great ship

Benjamin Franklin

A structured approach to privacy promotes evaluating privacy concepts on different layers. Reaching the limitations of vertical solutions, which cannot entrench into the unique features of individual protocols, we turn to each layer individually to guarantee privacy aware features. Individually, lower layers must not compromise the vertical privacy schemes that were outlined in previous chapters. This chapter presents a layered approach comprising link, network and application layers. To achieve this, we introduce new privacy proposals that protect the end-user and thwart the threats discussed in Chap. 3, tackling the major threats for each protocol, but always establishing the conceptual parallels to vertical relationships and the PRIVED model.

We first propose an entirely new link layer communication protocol that features a privacy-aware transport mechanism, enabling the user to communicate securely and privately. On the network layer, we promote a privacy aware routing solution. It builds on anonymity schemes to provide privacy-protecting routing that can involve service providers for widespread deployment. Finally, we provide a look into application layer protocols, and how they can leverage identity and user-centric approaches to provide privacy towards service providers, thus concluding a privacy review on each major layer.

5.1 Introduction

So far, privacy has been explored as a vertical problem, reaching across the entire network stack. However, as shown by the PRIVED model, and associated threats, it has deep horizontal ramifications. Individual layers of the OSI stack model pose different threats: some of them are serious because they provide means to track the user, learn his location and obtain personal data, which should be further explored.

The horizontal exploration of privacy enhancing solutions and improvements is justified on different dimensions. We must look at individual layer procedures, that are known to cause privacy risks, such as the ones outlined in Sec. 3.4. Given that there are serious privacy threats on individual layers, the conclusion that a flaw in one of them can lead to the corruption of the entire system is perfectly valid. This is especially true when considering the PRIVED approach (Sec. 3.3), where a single event observation, through a privacy leak, can lead to a cascade of conclusions. Also, in the VID ecosystem, having a single correlation between two separate identities can have dire consequences to the user, and at the same time, void a considerable amount of privacy support technologies (such as VNS). This shows that the vertical approach can only go so far. It is pointless to provide a complete vertical solution for privacy, if then the applied protocols can undermine the proposed solution by not respecting the PRIVED conditions.

Due to the dependency between layers, we must improve the privacy conditions to assure that there are no easy mechanisms to jeopardize the vertical information containment layers (using VIDs). By improving privacy conditions on each layer, we are contributing to an overall more robust privacy solution, where the vertical approach becomes stronger and more resilient to threats on each layer. Such privacy endangering mechanisms exist on different protocols and procedures, often in multiple places in the same layer. To overcome these threats on the different layers we must focus on specific aspects. Link layer identifiers, regardless of access technology, can be used to track and identify the user, compromising the IS. Alternatively, the network layer address not only identifies the user, but also conveys his position. Mobility mechanisms further accentuate these threats. Therefore, as discussed in Chap. 3, the mentioned layers and identifiers can be used to track, identify and locate the user, jeopardizing privacy.

In this chapter we propose solutions that tackle privacy threats on different layers of the network stack. Specifically, we propose to address tracking and location issues on the link and network layers, along with upper layer privacy integration, that do not compromise each IS as proposed in Chap. 4. For the link layer we propose a novel communication model that defines the message recipients based on cryptographic properties rather then on source and destination addresses. It focuses on addressing the communication channel instead of the involved stations, and is especially focused on wireless broadcast environments. The proposal discussed in Sec. 5.2 consists of a transport solution followed by performance simulations. It provides both end-point privacy protection, as well as location privacy protections, i.e. it is impossible to track a station by its addresses.

Similarly, for the network layer we propose a solution that relates to Onion Routing [142] focusing on hiding the real addresses of the communicating peers, and consequently their locations. By relying only on standard IPv6 mechanisms to define waypoints in the routing path, similar to hop-by-hop routing, the packets go through several privacy-enabled routers, defining special waypoint routes, with a significantly reduced costs when compared with existing solutions. Sec. 5.3.1 presents the architecture that enables the overlay solution along

with the location analysis. The solution conceals the real addresses of the users, making it harder to compromise not only the privacy threatened by IP addresses, but also breaking any possible IS correlations based on addresses.

Finally, we investigate transport and application layer privacy concepts. On the transport layer, we identify the particular privacy conditions that are caused by the very close relationship with the network layer. On the application layer, we focus on IdM solutions which provide pseudonymity. We evaluate the requirements that stem from other layers, and from the vertical space represented by the VID/VNS solution, to conclude about the best IdM protocols that fit our needs. We show how in the IdM space, SAML provides the best approach towards our goals, and is the primary candidate for vertical layer integration. This ends our chapter, showing that there are different aspects that require integration from an architectural view point, bridging the work towards the vertical aspects and architectural approaches.

5.2 Link Layer Privacy

One of the most interesting conclusions concerning link layer identifiers, highlighted in Sec. 3.4.3, is that they bear a scope greater then their design purpose: these identifiers are globally unique, yet are only employed in the logical access link. The result is a loss of privacy that stems from the fact that we are always connected, and as we change locations we can be tracked by the unique identifier used at the link layer (Sec. 3.4.4). While this is true for wired links, it becomes more relevant in wireless environments because of the encouraged nomadic behavior that the technology induces. The technology has reached such an adoption level that it is virtually present everywhere we go, but carries a hidden cost: the "always connected" environment takes a toll on privacy, which stems from the link between our physical movement and the network attachment of the devices we carry. Due to the recurring privacy arguments, users require secure and private network operation in order to trust the ubiquitous wireless access. Accordingly, to keep up with user privacy requirements, link layer technologies must evolve, and as discussed, wireless technologies require particular attention.

We focus on location privacy and identification issues at the link layer, considering 802.11 protocols as the primary use-case for the proposed solutions, in light of current technology and standards. We present an approach that sits between the physical and link layers, showing what can and cannot be protected above the standard specifications and which conveyed information constitutes a privacy threat.

5.2.1 Network and Privacy Threats

In the scope of a link layer solution, it is interesting to understand the network model, because it is important to determine how the local operations are carried out. This helps understanding the major threats, in a delimited scenario, thus reducing the complexity grasping the threats on user privacy. We start with a generic link scenario, and then instantiate into a well known technology. In our model we consider the last hop of an access network composed by one Access Point, AP and n terminals (or stations), N_i, with $i \in 1, 2, \cdots, n$. We further extrapolate the individual links between each of the terminals N_i with the AP and term them as channels, where C_i refers to the channel between node N_i and the AP. Each channel has only two endpoints, which are the addresses of N_i, MAC_i, and of the AP, MAC_{AP}. Our assumption is that, even though communication occurs only between the individual N_i and

the AP, the medium is still broadcast and all the nodes in the group can listen to all of the messages being sent over any C_i.

Focusing on this network scenario, we can highlight the major threats to link layer protocols, first shown in Sec. 3.4.4. The first threat to privacy at the link layer is that of the attacker having access to all the packets exchanged in all channels C_i. An attacker may track a device from one network to the other by moving inside the same link layer cloud and mapping the unchanging MAC address, which often leads to correlating Virtual Identities, and extending Information Sets. By applying PRIVED, we can see that these threats can extend to the network layer, by mapping different IP addresses to the same link layer identifiers. Another issue is the tracking of origin and destination of link layer messages. Currently it is easy for an outsider to determine traffic patterns and traffic direction. This information can be used to pinpoint a user, or when correlated with more information, it can provide means towards discovering the user's identity (e.g. periodically checking an IMAP server, an often repeated pattern in current software). Although the problem of anonymously linking identity to a form of certification is outside of the scope of this Thesis, we provide the mechanisms which allow the linkage of, for example, participation certificates which can be checked against a Public Key Infrastructure and AAA servers for validity and uniqueness. Such a combination would thwart attempts of multiple registrations on the behalf of the same user. In the scope of the outlined network models, and intrinsic threats, we can summarize the security objectives that a successful location and identification privacy approach for the link layer should achieve:

- Avoid using a unique link layer identifier: using the same identifier allows an attacker to track the user's location by testing the user's presence in different link layer clouds (correlating VIDs/ISs).
- Prevent linking network layer location with link layer identifiers.
- Protect communication peer identities and pseudonyms from traffic and header analysis.
- Protect users' traffic from direction inference: distinguishing traffic direction (from the AP to the terminal or vice-versa) allows an attacker to infer which service is being used and possibly the user's identity.
- Support link layer protocol operations to minimize changes to standards and implementation costs: the feasibility of our approach depends on the intrusiveness into the link layer protocol.
- Ideally, when presented with several packets in the network, the attacker should not be able to link them or even distinguish anything other than the fact that they are disjoint packets.

In fact, the last item highlighted above presents the guiding principle for the entire solution, as it provides the best protection mechanism to support the PRIVED approach. When provided with several packets in the network, an attacker should not be able to link them or even distinguish any particular information, thus protecting the user's virtual identity by keeping the information within different sets separate. This should be achieved in such a way that supports regular link layer protocol operation, minimizing changes to standards and implementation costs: Some of the benefits of our approach are its lack of intrusiveness into the link layer protocol.

5.2.1.1 Targeting 802.11 Privacy Leaks

One of the advantages of performing a layered approach, is that we can navigate from generic to concrete approaches and privacy evaluations. We have already outlined major link layer threats, that converge on identification and location. We can now analyze very detailed protocol mechanics that can yield privacy leaks, which can be divided into two categories: direct, when the data is revealed in a packet, or correlated, when the attacker needs more than one source to obtain the information. To perform an analysis on a protocol, we must take both types into consideration. In most widely used protocols, link layer addresses used to identify nodes are sent in every packet. Furthermore, a channel identifier is sometimes used and, although it cannot be used to identify the node, it aids in tracking connections. Other potentially leaked information includes sequence numbers, acknowledgment frames and round-trip times, all of which can be correlated, hence tracking the connection and the user. This leaked information is present due to protocol requirements and can not be eliminated. Some cases require that part of this information is received by all the stations. Therefore, each protocol will need careful analysis in order to determine if we can hide or otherwise obfuscate the offending fields. Since this is a generic problem, and in order to exemplify this procedure, we would like to demonstrate how information leakage can be detected in a common link layer protocol: 802.11.

As mentioned in the overall privacy threats, location privacy is threatened by an attacker having access to all packets exchanged in the channel C_i. Standard 802.11 operations require that a station (STA) listens to all on-going traffic on the wireless link. While this leaks privacy information, nevertheless some fields must be readable by everyone to ensure adequate protocol operation, such as the NAV. All 802.11 frames share a generic format, which discloses sensitive data, structured along the MAC Header or Frame Fig. 5.1, Frame Body and Frame Check Sequence. The MAC Header contains sensitive user and network data. The Frame control contains management information that carries private information, which we analyze in relation of particular threats.

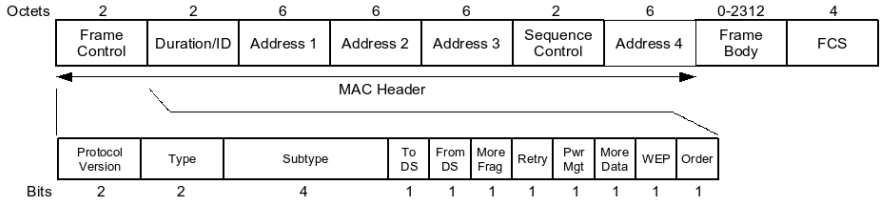


Figure 5.1: 802.11 MAC Frame and MAC Frame Control Field

Also stemming from the generic threats, comes the possibility of an attacker tracking a device from one network to the other by moving inside the same wireless link as the subject, mapping the unchanging MAC address. Hence, we must avoid using a unique link layer identifier. In 802.11, the addresses are conveyed in two or more Address Fields, which are one of the easiest ways to track stations in a wireless environment. Besides the addresses, an attacker can resort to the Sequence Control field, which is divided into a Fragment Number and a Sequence number. The sequence number is incremental and STA dependent, which allows a station to be tracked even if all other information is hidden.

Another potential threat is tracking origin and destination of link layer messages. This

information can be used to pinpoint a user. Inspecting any of the following fields, shown in Fig. 5.1, it is easy to distinguish between AP and STA: *To DS*, *From DS* (where DS represents the Distribution System, i.e. the AP), *Power Management* and *More Data*. In any of them we conclude who is sending and receiving the message. The periodically broadcasted beacons identify the presence of an AP and network parameters, but also notify sleeping STAs that they have messages queued in the AP's buffer, by using the Traffic Indication Map (TIM), a vector of Association Identifier (AID), revealing information about the present stations. Furthermore, the *Duration/Id* field has a dual purpose: it serves as the duration for the NAV, which does not reveal sensitive information, and as the AID, disclosing information about ongoing associations. Also, the attacker should not be able to tell how many active nodes exist in a network. Such information might help the attacker in performing statistic and probabilistic attacks on discovering the user's identity. He can obtain this data by observing Association and Authentication requests or by traffic analysis and data/event correlation. Analyzing the fields Type and Subtype, it is fairly easy to determine which stations are associating, re-associating or authenticating.

The discussed 802.11 fields can reveal sensitive information, unwillingly providing mechanisms to compromise the vertical identity approach. However, some fields must be read by all stations, particularly regarding the NAV mechanism. All stations must read the duration field in the frames, even if they are not the destination to perform NAV calculation, i.e. when they can send packets. This field must be sent unmodified over the air, rendering useless any solution that aims at encrypting the entire 802.11 frame.

5.2.2 Secure Transport

Our privacy proposal defines a novel transport mechanism. When active this transport protects the data and management frames against the described attacker model, assuming that keys have previously been agreed between N_i and AP. When used in parallel with classical networks, the node might obtain this key by, for example contacting his home network. As this might not always be the case, we assume the involved parties actively partake in a Diffie-Hellman authenticated key agreement[1], satisfying the objectives given in Sec. 5.2.1. The key agreement phase should only be necessary if the terminal, N_i, does not have another secure way of agreeing on a key with the AP.

5.2.2.1 Protecting the Communication Channel

In our approach, a channel C_i is identified by key K_i shared between terminal N_i and AP, as shown in Fig. 5.2. Logical channels are encrypted and only the key holders can inspect the traffic inside the logical tunnel. Thus, both MAC addresses and unique identifiers are always encrypted. But simple encryption is not sufficient: if it does not provide randomization, the same plaintext results in the same ciphered text. For this reason, we add an initialization vector iv, in order that the same plaintext results in different cipher text depending on iv. If the value of iv is synchronized on both end-points, it allows fast determination algorithms of the origin and/or destination of the packets. The most costly operation is synchronization of both end-points: a packet is lost and the sequence number s_i is no longer synchronized on both end points. When not synchronized, the nodes need to decrypt the packet with all

[1] An alternative, more privacy aware mechanism has been proposed in the form of a patent document [7], still awaiting scientific publication.

known keys and match the result against a known value. If the decryption is correct, i.e. known values match, then the connection is successfully reseeded, and iv matched to the ongoing sequence number.

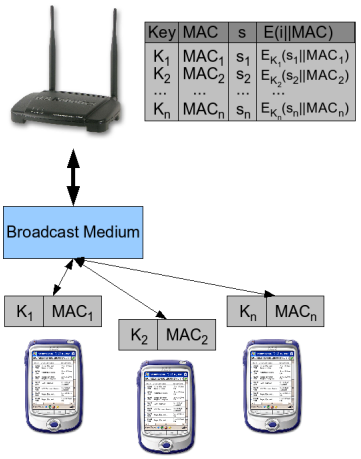


Figure 5.2: Using keys as channel identifiers in a broadcast medium.

5.2.2.2 Transport Header

When encryption occurs, it might be necessary to include padding in the packet. Since the length transported must reflect the number of bytes sent on the channel, we must include the real length of the packet, encrypted, at the end of the packet. This will allow the receiver of the packet to insert the correct length before delivering the packet to the MAC layer. This requires a transport header, shown in Fig. 5.3, which must be appended to all packets before encryption. This header contains the original length of the packet, terminates with the value of s_i and optionally contains the AID field. In packets where the AID should be sent, the AID is added to the transport header and encrypted. Before the packet is passed on to the higher layers, and after decryption, the AID will be copied on top of the duration field.

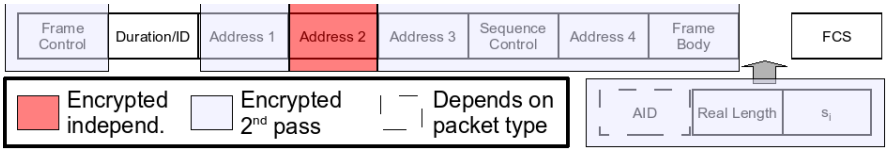


Figure 5.3: Link Layer Privacy Transport Header

Since both encryption and decryption are performed from end to beginning, we ensure that the variability in the ciphering caused by the changing s_i affects the whole packet encryption. Looking at the header we can see that all fields, with the exception of the duration field, are encrypted. Also, we observe that one of the address fields is encrypted independently. In the case of a N_i sending a packet, this field will correspond to the source address. When it is the AP that sends a packet the destination field is used in this way. The reason is that each

STA N_i must verify whether the packet is intended for it, while the AP must verify who the sender of the packet was.

5.2.2.3 Encryption and Decryption

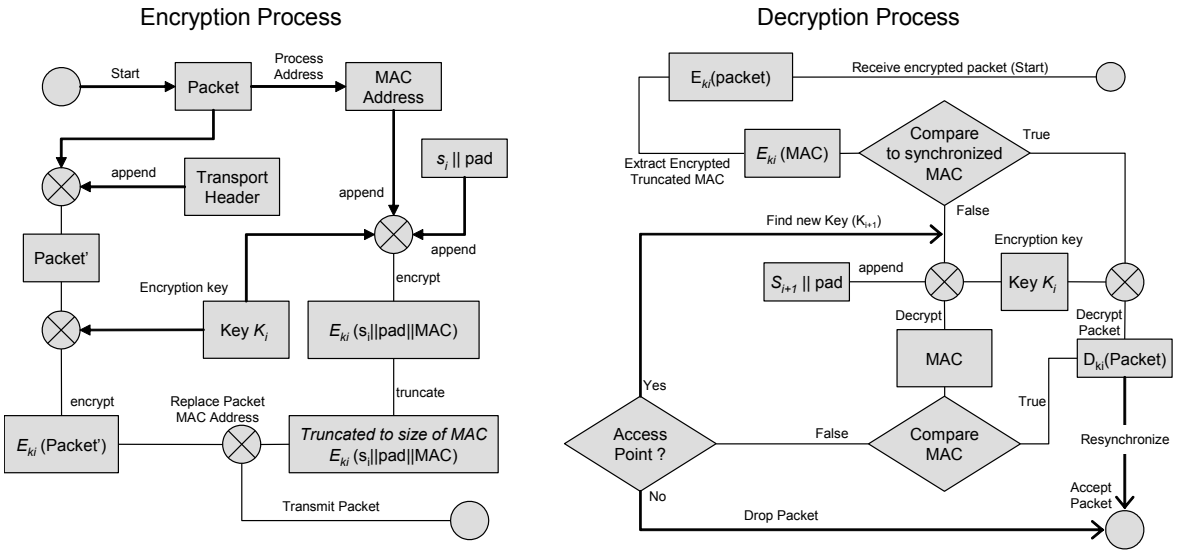


Figure 5.4: Encryption and decryption processes.

The current value of s_i is appended at the end of every sent message. Encryption is then performed from the end to the beginning using the key K_i to the corresponding channel C_i, with the exception of the fields for fast determination. These fields contain the known plaintext values of the source or destination addresses. The address (or source) field must be encrypted and decrypted independently of the rest of the packet by pre-pending the value s_i padded to the size of the block of the cipher, such that the first encrypted blocks correspond solely to s_i and can be removed before the address is re-inserted in the packet.

As depicted in Fig. 5.4, when station N_i needs to send a packet, it appends the transport header, updates the length of the packet and applies the pre-selected cipher using K_i on the identified non-plaintext fields from end to beginning. In parallel, the address of node N_i, MAC_i, must be encrypted independently because it must be pre-computed at the other side. The node encrypts MAC_i by applying $E_{K_i}(s_i||padding||MAC_i)$, where padding refers to the fact that s_i should be expanded to the block size of the cipher. After encryption, the encrypted section of s_i is truncated and only the encrypted MAC_i is added to the packet. The same encryption process can also be applied on the receiving side since s_i is synchronized.

Similarly, when the AP needs to send a packet to a node N_i, it will perform the same operations using a key table to determine which key to cipher the packet with, and pre-computes the destination address, in the same way the node pre-computed its own address to insert in the outgoing packet. When receiving a packet, we split the situation in two, depending on the receiving node type. In the case of a node N_i, "receiving a packet" means it will always try to decrypt the message and compare the address after decryption. More specifically, it will decrypt the packet, from end to beginning, and thus obtain s_i. It will then

use s_i to decrypt MAC_m by first concatenating the encrypted s_i with the padding and then performing the decryption: $D_{K_i}(E_{K_i}(s_i||padding)||E_{K_i}(s_i||MAC_i))$. This step is necessary since the MAC_m is encrypted independently using s_i as a vector for the cipher. In the end, the node can compare MAC_m with its own address, MAC_i, and check whether the packet is intended for it. Once the original values for mandatory plaintext fields and length are replaced by the ones in the transport header, the packet can be delivered to the upper layers. Finally, this will use the value s_i to update its own internal sequence number.

In the case of the AP, the initial reception sequence is similar to the case of N_i. When the AP receives a packet, it will find the right key. The optimized process is to look at the table of pre-computed encrypted MAC addresses and try to find the encrypted MAC_m present in the packet. The address which corresponds to the key K_i, will serve as the check value to determine if the decryption was successful. As this process might fail due to synchronization loss, the AP might need to test all the keys in its table. If the key is found, the AP will proceed to the decryption of the packet as described above for N_i. It will also use s_m from the packet to update its table to map next packet $(s_m + 1)$ and $E_{k_i}(s_m||padding||MAC_i)$.

An exception to the standard encryption process has to be introduced to properly handle 802.11 Beacon frames. As previously mentioned, beacons must be handled differently to prevent attacks on the *TIM*, which contains information that must be observed by all the stations in the wireless link.

To handle these particular features, we encrypt each position of the bitmap individually to every station, so that each station receives only the information required to operate, and nothing more.

Conceptually, the TIM, $\{AID_1||AID_2||\cdots||AID_n\}$, should be replaced by an encrypted version, $\{E_{K_1}(s_1||AID_1)||E_{K_2}(s_2||AID_2)||\cdots||E_{K_n}(s_n||AID_n)\}$, where the original bit value, AID_n, should be encrypted using S_n. But, each $E_{K_i}(s_i||b, Pad)$ block, , where b represents the original bit value (AID_n), has the size of an s_i, and not a single bit as required by the TIM. To convey a single bit, we simply calculate $E_K(s_n||0)$ and $E_K(s_n||1)$, compare them bitwise, and insert the bit value of the most significant bit, j, that differs between them. When N_i receives a beacon, it simply extracts its bit value b, from the expected position, calculate the $E_K(s_n|b)$, for both 0 and 1, and performs the same bitwise comparison to obtain the first different bit between them (most significant). The result indicates whether the bit belongs to the 0 value or to the 1 value, thus determining if there is any packets for that station queued at the AP.

The end result of the transport functions is that all unicast packets in the network are indistinguishable from each other. An attacker will be unable to link two different packets by using link layer information. In practice this means that packet events, from the PRIVED network instantiation, never bear correlatable information, as all information differs between them. This allows not only clearly separating the events and the link layer, but results in the desired outcome that makes it impossible to correlate any IS through link layer information, as there is no possibly shared information.

5.2.3 Performance and Scalability

There are several aspects that determine the performance issues of the proposed protocol. The transport presents a twofold problem: i) it depends on the cipher being used and ii) must obey the 802.11 timing restrictions. When choosing a cipher we must ensure that it allows the described operations and that it is also efficient, because it will be used in every packet

i.e. all nodes must decrypt at least the destination MAC address, and during time-critical events such as 802.11 Acknowledgment frames (which are in the microsecond range). The small block size and high efficient duty-cycle of RC5 makes it a perfect candidate. The block size fits the minimum encryption unit required in our scheme, which is 32 bits, which reduces the need for padding, since packets are usually 32-bit aligned. According to the NESSIE [49], RC5 encryption and decryption both take 19 clock cycles (cc) per byte in a Pentium III, which we believe is an acceptable platform for the AP. For further considerations we assume this RC5 implementation, on a Pentium III 600 MHz.

For real world deployment we consider that cryptographic primitives should be implemented in hardware, which considerably speeds up encryption and decryption times. Also, multi-core processors are becoming the common computing platform, rather than the exception, allowing us to consider a more elaborate approach, where the proposed scheme is implemented with a two-queue solution for packet processing. The first queue handles synchronized packet streams, while the second queue is used when re-synchronization is necessary, which takes longer to process. This enables us to minimize the delay imposed by out-of-sync packets in the cryptographic scheme. Another problem we consider is the placement of the transport header, at the end of the packet, enabling faster processing. When packets are received, there is no need to iterate over the mandatory 802.11 MAC header. Including a fixed size header at the end of the packet enables the packet receiver to quickly locate the necessary bits relative to the end of the packet, which is clearly marked with a Frame Check Sequence, boosting the speed at which nodes can discard or accept packets, without any further processing of 802.11 fields.

The overall system scalability also depends on how many nodes can coexist in the same AP, since it must go through $N/2$ keys on average to decrypt an unsynchronized packet. This effect can be mitigated through using first, a hash table based on the expected cryptogram that is separately encrypted, so that the proper key can be quickly retrieved base on a $O(\log n)$ operation. The second factor to consider is to order the keys smartly to process out-of-sync packets coupled with a proper caching policy. The most recently used keys will have a much higher probability of being reused in the near future, since a station has a particular time window to transmit and given that the traffic usually follows burst patterns, it will require more than one time slot to transmit all the data. Meanwhile, nodes which tend to enter sleep mode, and therefore not transmit, will drop further down in the key cache. The Most Recently Used policy will also enable the keys belonging to users who transmit the most packets, to be closer to the top of the key table, enabling a quicker lookup.

Performing cryptographic operations on (parts of) every packet bears a hidden power cost. The power consumption of the 802.11 cards will increase even though this can be mitigated with optimized low-power hardware.

5.2.3.1 Traffic Impact Analysis

For our test scenario we consider one AP, one Correspondent Node, which is the destination for all communications in the wireless channel, and an increasing number of nodes (N_i). CN is attached to the AP via a cable as not to interfere with the radio part of our simulations. Each added node N_i also increases the load on the network and reduces the opportunities of a node to find the medium free (which will induce collisions and, due to the nature of our protocol, cause loss of synchronization between the nodes).

We have performed all our simulations in NS-2 [115] 2.29, using the following parameters:

The nodes (NN) vary from 1 to 20; the UDP data rate is 67.8 Kb/s, using 178 byte packets; the TCP packet size is of 512bytes; the simulation lasts 60 seconds (after warmup), over 10 runs.

To perform our simulations we inserted a computation delay at the AP which depends on whether or not the AP is synchronized with this node. The mechanism used to check whether the node is not synchronized with the AP is based on whether the MAC layer has re-transmitted a packet due to collision. In cases where a re-transmission has occurred, the AP will take the average time of finding an entry in a table which is of size $NN/2$ (where NN corresponds to the number of simulated nodes)[2]. Once the key is found, we assume the node to be synchronized once again with the AP.

5.2.3.2 Impact on Real-Time Traffic

In this scenario, we are interested in the behavior of real-time applications, such as audio and video, and make use of UDP with constant bit-rate traffic. We are interested in how our scheme affects both end-to-end delay and jitter. Source traffic at each node transmits at *67.8 Kb/s*, with *178 byte* packets. This simulates a *64Kb/s* voice call and the RTP overhead.

The performed simulations cover three different cases. The first, for comparison purposes, is a plain 802.11b simulation. The two other scenarios implement higher processing delays at the nodes, with one and two queue variants. a ref to where the queues are discussed.

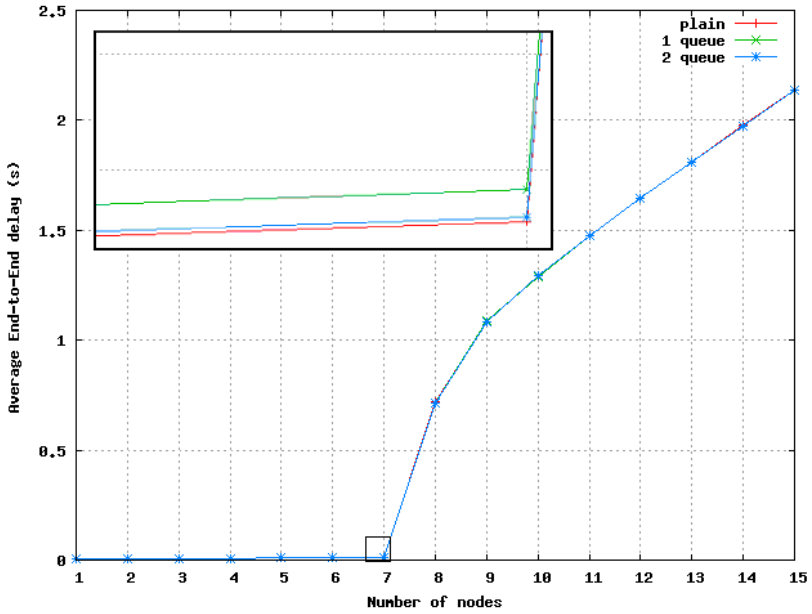


Figure 5.5: End-to-end delay, constant bitrate 67.8 Kb/s per node from 1 to 15 nodes, zoomed is the saturation point

Fig. 5.5 and Fig. 5.6 show the end-to-end delay and jitter for the real-time traffic. We can observe that the saturation point of the 802.11b network is located at 7 nodes per base station. With 7 nodes and more, both the delay and jitter escalate rapidly. Fig. 5.5 further illustrates the jump point for the delay, in the transition from 6 to 7 nodes. This observation

[2]Please note that we do not assume any optimization or ordering of this table.

is valid for all the scenarios, leading to the conclusion that the processing times do not impact the saturation point of the network, even though a slightly higher delay is noticeable. The observed delay behavior is consistent with the introduced encryption and decryption times, since the double queue performs slightly better than the single queue, even though both are higher than the plain scenario delay. Under the saturation point, the double queue delay is very similar to the plain scenario, due to the fact that the network is not performing a large amount of retransmissions. The small amount of retransmitted packets, that have a higher processing time, do not affect the synchronized packets, which have a small processing time. Above the saturation point the second queue shows even greater value by providing a significantly smaller delay than the single queue, because the collision/retransmission frequency increases due to network congestion. However, above the saturation point, the bottleneck is the access to the network medium and therefore the delay can present oscillations regardless of the simulated scenario.

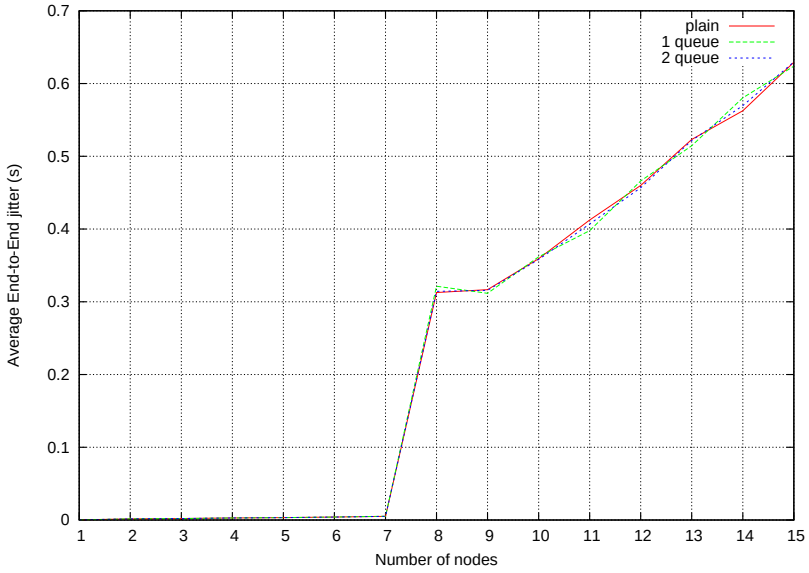


Figure 5.6: End-to-end jitter, constant bitrate 67.8 Kb/s per node from 1 to 15 nodes

Fig. 5.6 presents the jitter values for all the tested scenarios. We can again clearly notice that the most influential parameter for the jitter increase is the node number, and consequently the saturation point. As seen in the figure, for up to 6 nodes, the jitter increases slowly and steadily, whereas after the 6th node threshold we notice a rapid increase, due to network congestion. The jitter values conform with the delay, showing a smaller jitter for the double queue when compared to the single queue scenario, but both are still higher than the delay for a plain 802.11b network. Above the saturation point, the jitter variation is more erratic, for the same reasons that the delay behavior above the saturation point: network congestion.

5.2.3.3 Impact on TCP Throughput

In this scenario, we are interested in the degradation of the TCP connection and how it affects the throughput. For this effect we consider a TCP connection which will try to make maximum use of the available bandwidth and introduce background noise in the form

of stations transmitting constant bitrate traffic. Figure 5.7 shows the different available bandwidth for the three proposed scenarios. We can see that increasing the number of nodes producing background noise degrades the TCP throughput, as would be expected. The double queue simulation shows the worst performance due to the fact that it introduces unordered packets into the network, causing the TCP congestion control mechanisms to reduce the TCP window and consequently the throughput. For the single queue scenario, the introduced delay actually helps the TCP congestion control, smoothing the TCP window increase and therefore introducing less back-off operations, which has a positive effect on the bandwidth that on average is higher than in the plain 802.11b simulations. As a conclusion, TCP bandwidth is not greatly affected by the introduced processing delay, and the discrepancies between the plain and one queue scenarios could be solved by using a more adequate Congestion Control Algorithm, giving a fairly similar overall performance.

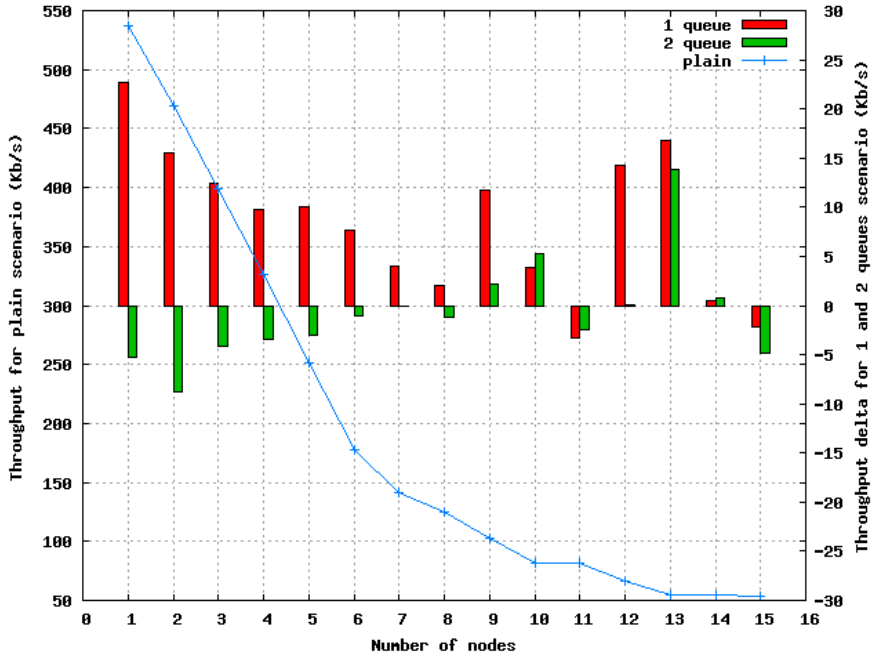


Figure 5.7: TCP throughput, constant bitrate 67.8 Kb/s per node from 1 to 14 nodes of background noise

5.2.4 Link Layer Vertical Interactions

We provided a mechanism that enables different features at the link layer. We highlight anonymity from passive eavesdroppers, hiding any ongoing communication towards the access point. In fact, the ability of hiding every identifier that might yield tracking information is a powerful feature that does not compromise the vertical approach, nor the privacy of upper layers. By focusing on wireless technologies, we showed the proposal's feasibility by using a well known protocol (IEEE 802.11) as an instantiation. The 802.11 stations have assured location privacy, along with data integrity protection. The proposed scheme can still be used with other 802.11 security mechanisms, such as 802.11 [67], but can obsolete them. Furthermore, given that WiFi cards can have embedded hardware cyphering components,

the highlighted advantages can probably offset the costs of replacing the existing commercial infrastructures. Our approach should be used in conjunction with a pseudonym mechanism to prevent tracking by active communicating peers.

Through the outlined conclusions we proved that the link layer does not need to be a correlation mechanism for upper layers. This effectively defers potential correlation problems entirely to upper layers, given that it becomes impossible to use upper layer identifier to correlate lower layer identifier (e.g reusing an IP address over multiple link layer identifiers), as no identifier is actually reused. Going back to the PRIVED model, we can say that this solution provides the most protection, given that no link layer addresses are seen by eavesdroppers, and hence no relationship can be made to other information in the IS.

Note that, although our proposal addresses link layer threats, there are highly specialized physical attacks which are not covered by our approach. In radio based technologies attacks may rely on the physical characteristics of the radio channel. Such attacks include finding the nearest station and triangulation or trilateration, by analyzing the signal strength, signal-to-noise ratio and radio-frequency fingerprinting. Some vendors even support certain protocols for location services which can be used against the users. Corbett et al [26] recently proposed a passive method to determine the vendor of a certain card by analyzing the way in which the station adapts its rate (in this case for WLAN). These attacks may erode the protection offered by the proposed solution leading to the idea that this proposal should be taken in conjunction with techniques which also protect the physical layer when necessary (considering that these attacks usually require expensive equipment and are hard to perform).

Beyond the lower level interaction, there are important interactions which require upper layer attention. One example that needs to take into account the vertical approach is when the access point is able to identify the link layer address of the user. It is interesting to see that we cannot use this mechanism to protect the user from the access point, because it needs to identify the STA at link layer, leading to the idea that the network should cooperate in the privacy protection mechanisms, rather than the opposite.

Regardless of the network's role, the vertical pseudonymity approach, represented by the combined VNS and VID concepts, can complement our link solution to the respect that for different operations, new link layer identifiers can be used, following the reasoning presented in Chap. 4. Therefore, when protecting from the network and other services, it is possible to use different pseudonyms at link layer, following the VNS approach, coordinated by VIDs. This guarantees that the vertical solution not only ensures privacy protection, but is properly complemented and reinforced by the link layer approach.

But this vertical relationship leads to the idea that, even with powerful layered privacy measures, the vertical solution still needs to be properly accounted for on above layers. Consequently, the network layer still presents threats concerning user identification and location, forcing further consideration on the network layer despite the threats already tackled at link layer.

5.3 Network Layer Privacy

On the network layer it is important to protect the communication contents from a vertical perspective, since it transports all upper layer information. It is also important to protect the endpoint and its location, because the IP address uniquely identifies both the peer behind the communication and its topological location (Sec. 3.4.4). Location privacy becomes important

because the topological position expressed in addresses can be converted into geographical location. But more importantly, the unique addresses, through correlation properties can become a major tool to link different VIDs on the network. As discussed in Chap. 3, network addresses can be a simple, yet effective tool for correlation and IS construction.

Most solutions that address network privacy threats usually rely on the concepts of Chaum Mixes [22] to provide anonymity, mitigating the side effects of using IP addresses: if packets are anonymous, the observed address cannot be traced back to the original sender, voiding location information an correlation techniques, since it does not concern the message sender. However, such protocols usually require trade offs between privacy and performance, or sacrifice compliance with standard routing protocols. The most prominent example of such solutions is TOR [37], which delivers privacy at a cost. TOR employs multiple encryption layers on every packets to conceal the packet origins and routing hops, incurring in a hefty performance penalty. It is also used on top of IP, due to its non-standard requirements, such as circuit establishment and fixed data cell size. The performance overhead and the lack of integration with routing schemes (IPv4 and IPv6) undermine the adoption of such privacy solutions on the network. Therefore, privacy has a price that only some end-users pay, becoming a cooperative peer-to-peer effort. This is what we call pushing privacy support into the edges of the network, resulting not only in crippling performance, but also on compromised privacy: packets travel to the network edges where forwarding is performed by untrusted endpoints in a peer-to-peer system, that can inspect the ongoing traffic [131]. Distributed environments also pose severe difficulties concerning lawful interception, which is a requirement for any commercial network.

To create a truly adoptable solution, we must reduce performance costs, allowing a solution that can be easily deployed inside the network, compatible with current routing schemes. In this scope, we propose Waypoint (WP) Routing, a lightweight framework that through a novel cryptographic routing scheme, enables treating privacy as a value added service (VAS) that can be provided by the network. We provide end-user privacy by hiding the original sender, through several encryption points, mandatory waypoints inside the network, making it impossible to identify the end user or his location. This directly results in voiding any correlation means, based on network addresses, as already provided for the link layer. Since each address can be hidden, as well as its relationship to the user, it does not compromise the user's VID, nor extends existing information sets.

This is achieved by using encrypted IPv6 Routing and Extension headers, based on the concepts of Onion Routing [142] and TOR [37] mechanisms. The encrypted extension headers define lightweight overlay privacy routes, where each router is aware only of the next hop in the route, at a reduced cost, thus minimizing performance impact because it requires less encryption (only extension headers are encrypted). Also, by keeping the packets inside the network, and routed through trusted entities, we avoid traffic inspection by untrusted peers [131], along with smaller delay.

Introducing entities along the communication path that anonymize the source of the traffic is a proven privacy approach. In Waypoint Routing[3] we introduce mandatory waypoints in the communication path, that are only aware of the next hop in a route and replace source and destination addresses, diluting any information beyond next and previous hop.

As each router only knows the next hop of the defined path, privacy is assured by dis-

[3]The waypoint naming stems from the use of waypoints as geographical references for navigation purposes, much like we use in our cars everyday with GPS systems.

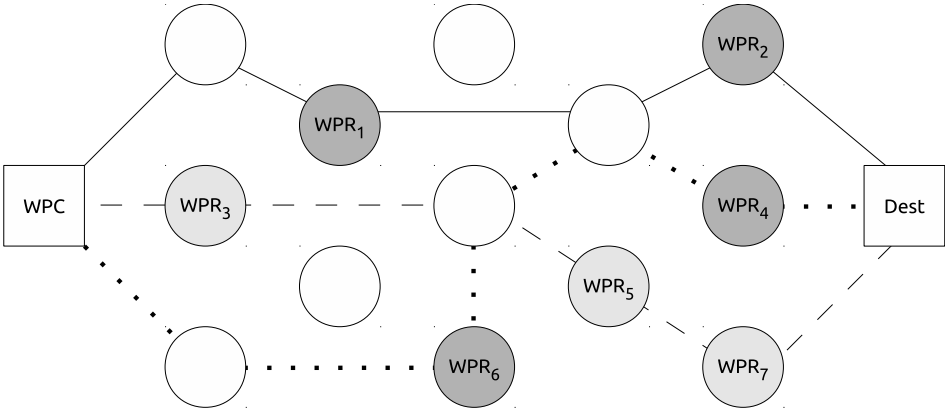


Figure 5.8: Example Routes with defined Waypoints

tributing routing information. In this case, no entity has knowledge about anything beyond the next hop. The only entity that might have a broader picture of the privacy associated with the routes is the network provider, by controlling multiple network hops. But it this case, we use this knowledge to our advantage, making the network a part of the privacy enhancing process, transforming privacy into a network provided service.

The waypoint mechanism effectively conceal the packet origin and consequently the sender's identification and location, which is tied to the original address. In most cases, no assumptions can be made about the observed addresses regarding the original source and destination, in line with the Chaum Mix [22] approach, and complying with the presented PRIVED requirements. Replacing addresses is only possible through the use of encrypted routing hints included in IPv6 extension header options, as shown in Fig. 5.9, and can play a major role in preserving the vertical boundaries defined by information sets.

5.3.1 Waypoint Routing Overview

As seen in Fig. 5.8, an overlay route is defined by several Waypoint Routers (WPR) that anonymize the traffic flowing between the end-user, the Waypoint Client (WPC), and a selected destination. To establish the route, the WPC contacts each selected WPR, previously retrieved from a directory or discovery service as proposed in Sec. 5.3.2.1, establishing authentication and shared cryptographic material. The selected WPRs compose a virtual circuit defining how packets flow (i.e. which waypoints are visited).

To use the circuit, a client encrypts the next hop as well as the destination address, placing them in an IPv6 extension header and forwarding the packet to the first WPR in a route. Upon receiving a packet, each WPR decrypts the addresses conveyed in the extension header, the Routing Hint, thus determining the next hop which becomes the new destination. To finish the forwarding process, the WPR replaces the routing hint (in the extension header) with the one corresponding to the next hop, as well as the packet's source and destination. The WPR also decrypts the Circuit Identifier, as discussed in Sec. 5.3.1.1, to ensure that the packet reaches the proper destination. Both these fields have the same size of an IPv6 address, fitting in the IPv6 extension header, shown in Fig. 5.9, which is transparent to standard routers.

The WP Routing process can be summarized as a cryptographic source-based routing

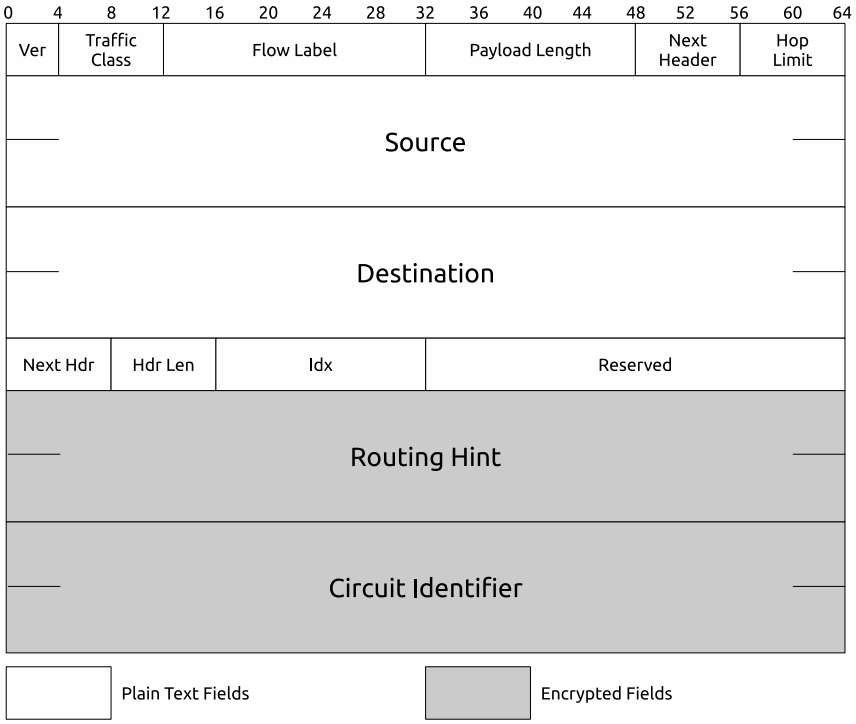


Figure 5.9: IPv6 Packet header with cryptographic extension header.

mechanism for IPv6, where the packet's source and destination change at each WPR. The routes, defined in Sec. 5.3.1.2, can be established during the authentication process using the shared keys between WPC and WPR (obtained through a Diffie-Hellman exchange) and are composed by the encrypted hints and reusable IP-level circuits, as shown next.

5.3.1.1 Hints and Circuits

The two main Waypoint Routing concepts are hints and circuits, which must be integrated into the routing infrastructure. These mechanisms, compatible with standard IPv6 routing rules, are necessary to properly route packets, as presented below.

$$H = E_k(IP_T). \tag{5.1}$$

The Routing Hint, defined in Eq. 5.1 and present in each packet, encrypts the address of the next hop in the route (T). According to IPv6 destination option rules, it is processed by the packet's destination, which decrypts the hint to obtain the next WPR in route. Therefore, in general terms, a node WPR_n will receive a hint H_n, corresponding to the encrypted address of the following WPR, $IP_{WPR_{n+1}}$, with key k_n. The hint H_n will be stored at the previous router, WPR_{n-1}, and sent to WPR_n, yielding the address of WPR_{n+1}. This process transforms WPR_n into a shield that anonymizes communication between adjacent routers. The routing process is further clarified in the next section.

If only the next hop was included in the packet, a different circuit would be necessary for each target, as the only available mechanisms to differentiate targets would be the associated

keys (which define a circuit). To reuse circuits along the path, we introduce the Circuit Identifier (CID), inspired by Onion Routing [142]: the CID encrypts the final destination, eliminating the need to keep state on every WPR for a specific target and allowing the reuse of circuits between routes (which only requiring hop-by-hop hints). It uses several layers of encryption, as described by Eq. 5.2, one for each WPR, ensuring that packets traverse the defined route, and that the CID changes (through different encryption keys) between hops (preventing tracking based on the CID), similar to TOR. The example CID_n, in Eq. 5.2, with n layers of encryption, generically represents a three waypoint route, used in the example presented in the next section, where the IP address of the target is encrypted with keys k_n, k_{n-1} and k_{n-2}, to form the CID.

$$CID_n = E_{k_{n-2}}(E_{k_{n-1}}(E_{k_n}(IP_{Target}))) \tag{5.2}$$

The CID is progressively decrypted, as it passes through each WPR, creating a hint index [4], used where circuit multiplexing is required. Wherever multiplexing is not required, the WPR decrypts the CID and forwards the messaged to the hint stored for the (idx_n, k_n), as described before, resulting in a default hint behavior (each required multiplexing operation should be explicitly registered at each WPR). In fact, with the introduced CID, extending or splitting a route (similar to route "telescoping" [37]) only implies registering a new hint and a corresponding CID in the required WPR, eliminating the need of an entirely new route. This process also guarantees that the global target is only obtainable trough the collaboration of every selected WPR. This raises the number of encrypted addresses in the packet to 2, where one is a hop-by-hop encrypted address, the routing hint, and the other is the encrypted final destination, the CID.

5.3.1.2 Routes

In order to understand how routes are formed, and the exchanged information (Hints and CIDs), we present an example of a three waypoint route setup and the associated forwarding process. The example is shown in Fig. 5.10, representing the information present at each involved element. Reaching this state is an iterative process: first, it is necessary to establish a WPR route including WPR_1, to reach T_1, then it is necessary to include WPR_2, forming a two WPR route, and later WPR_3, forming the final route, with three elements.

To start the first phase, WPC contacts WPR_1, establishing a symmetric key (k_1) for hint encryption, through an authenticated Diffie-Hellman exchange. This process is repeated for every WPR. To facilitate the key retrieval mechanisms at the WPR, a key index, idx_1 for WPR_1, is also agreed upon for future inclusion in the messages.

To reach T_1 via WPR_1, the WPC encrypts the address of T_1 with key k_1, forming the routing according to Eq. 5.1, where routing hint H_1 is the result of encrypting the IP address of T_1 with key k_1. The hint is then inserted into an IPv6 extension header, included in the packet. At this point, since the Hint is similar to the CID, the CID can be omitted, indicating that WPR_1 is in fact the exit router. The packets are sent to WPR_1, with the required index, idx_1, consistent with Fig. 5.9.

[4] Using an identifier for each encryption circuit can lead to label switching mechanisms. This would allow a more adequate switching mechanism, requiring hop-by-hop or onion-encrypted labels. However, we defer this option for future work, in order to simplify the current proposal.

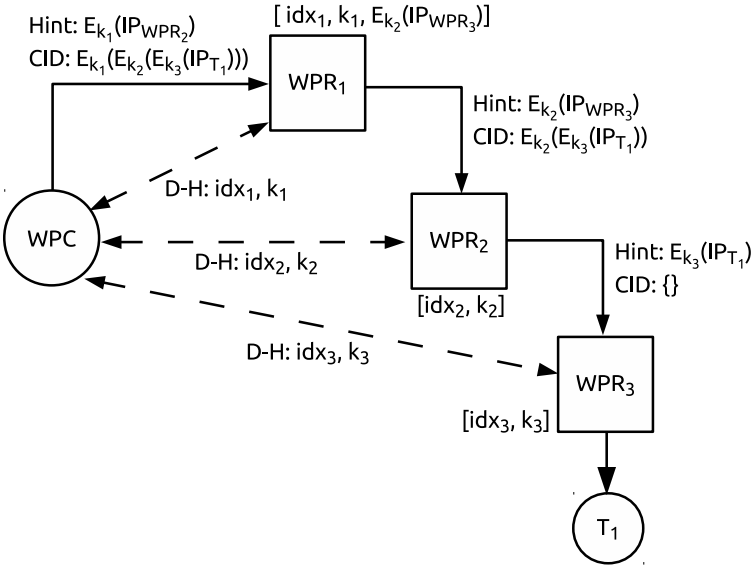


Figure 5.10: Instantiation of a 3-hop route with associated information.

To increase privacy, the client should include more WPR elements in the route. Therefore, the route can be extended to WPR_2, where the client registers, establishing (idx_2, k_2). Consequently, a new hint must be added at WPR_1, in the form of $E_{k_2}(IP_{T_1})$, to be included in the extension header when WPR_1 forwards packets to WPR_2. Also, the new CID is the target address, encrypted by k_2 and k_1, respectively. When WPR_1 receives a packet, it decrypts the hint present in the extension header, $E_{k_1}(IP_{T_1})$, with the corresponding key, indexed by idx_1, thus determining the next hop. Before forwarding the packet, WPR_1 decrypts the CID, concluding that it is not the last WPR in the route. As a result, before forwarding the packet, the decrypted CID, $E_{k_2}(IP_{T_1})$, is copied into the hint field.

To introduce WPR_3 in the route, the process is similar: the client registers at WPR_3, obtaining (idx_3, k_3) and updates the hint present at WPR_1, which becomes $E_{k_2}(IP_{WPR_3})$. Also, the initial CID becomes subjected to three layers of encryption, shown in Fig. 5.10 as $E_{k_1}(E_{k_2}(E_{k_3}(T_1)))$. At this point, when WPR_2 receives a packet, it decrypts the contained hint, $E_{k_2}(IP_{WPR_3})$, determining WPR_3 as the next hop. As it has no further hints for this CID, it decrypts the CID and inserts it into the Hint field, corresponding to $E_{k_3}(IP_{T_1})$. The behavior at WPR_3 requires decrypting the received hint, resulting in T_1. At this point, since no CID was provided, WPR_3 becomes the exit node, performing Network Address Translation (NAT) for the WPC, the last shield in the process.

The final state of this process is represented by Fig. 5.10, which details the defined route using three waypoints. It shows the shared keys, the information present at each point, and the content of the headers on each hop of the route.

5.3.2 Privacy as a Service

The WP Routing mechanism enables a lightweight approach to privacy at the network level. By requiring less encryption and being seamlessly integrated into IPv6, it becomes simpler to deploy. The reduced performance cost, analyzed in Sec. 5.3.3, allows privacy to be deployed

inside the network, on core routers, as opposed to the edges, by end-users. This can contribute overall adoption and also enables a new paradigm of perceiving privacy as a Value Added Service (VAS). Because it can be delivered by existing network providers, WP Routing permits the creation of a Privacy Service (PS). The PS requires the introduction of a logical entity, the Privacy Controller, that together with a set of WPR, forms the PS architecture illustrated in Fig. 5.11, thus completing the framework.

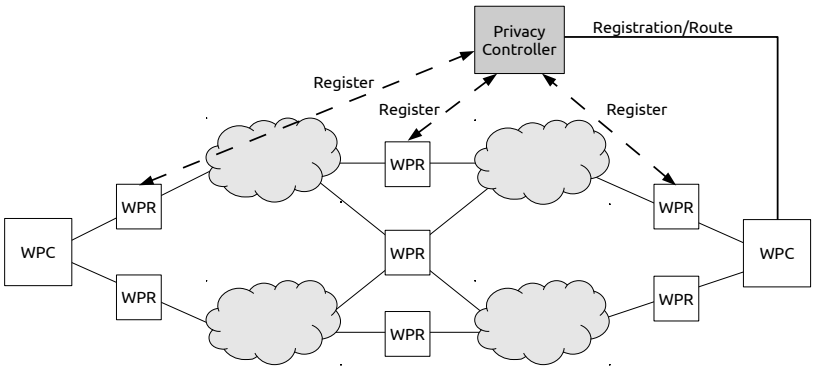


Figure 5.11: Sample architecture operation coordinated by the Privacy Controller.

5.3.2.1 Privacy Controller

As part of the PS, the Privacy Controller (PC) is the entity responsible for managing the privacy infrastructure composed of WPRs, and for providing end-user authentication and access control. As the primary privacy control entity, the PC handles WPR discovery, allowing route establishment. Moreover, it can contribute to hint and address management in different scenarios. In fact, the usage of a trusted entity for privacy control functions has been presented before in Chap. 4. In the VID model, the IdP, Identity Provider, assumes these functions for user related information. Therefore, it is only natural, that in a real deployment, the alignment between Waypoint Routing and the VID framework should rely on the IdP assuming the functions of PC. This close coupling of the IdP to the network is also proposed in Sec. 6.5.4, where the IdP manages HIP identities towards the network.

As a service, the PS holds a tight relationship with its users, enabling authentication and access control to the waypoint routing service. The trust association created between the provider and user leads to several outcomes: as a service, the PS must meet often ignored legal requirements, such as lawful interception and data records preservation; also, the user enjoys a new privacy environment, where the PS's design goal is to provide privacy, where in other scenarios this is a user provided best-effort feature. The second core PC competency is to enable route setup by providing the user with WPR information. By managing a set of WPRs, the PC can easily interact with the user to provide an adequate set of WPRs for the desired destination. This takes a centralized view of WPR discovery, putting PS in the driver seat of route establishment. It is important to note that, when requesting routes, the WPC can insert vectors (IP addresses in the vicinity of the target) to allow a customer answer from the PC, better suited for the target address. The PS, through the PC, can also assume the support role of resolution or rendezvous service, providing hints to entities seeking to contact

a WPC. The PC can store public (or global) hints that enable a resolving peer to reach the required WPC.

To complement privacy, multiple privacy services can be used, even belonging to different (privacy) operators. This helps increasing privacy by distributing user information, but also prevents the PS from becomings a single point of failure. If a PS fails, established routes are maintained, only failing for new resolution requests. Nevertheless, when a PS fails, the WPC can simply establish a new binding for the particular route, or simply divert the route through the path mechanisms described in the following section.

5.3.2.2 Route Selection

Route (and WPR) selection can be one of the biggest contributions made by the PS, through the PC. We separate routes into explicit, User Generated Routes, and implicit, Service Generated Routes. In User Generate Routes, the endpoint undertakes the bulk of the effort for WPR selection and route setup, using as many PS as desired. Here, the PC functions as a WPR discovery service. In Service Generated Routes, the PS can take advantage of the Hint routing mechanism to provide on-demand and implicit user-independent routes.

Using explicit routes, the WPC has the responsibility of establishing the route. This implies authenticating at the PC and obtaining a list of WPR, highlighted in Fig. 5.11. The node can include an IP address in the Route Request as an indication of destination, obtaining routers closer to the optimal solution provided by IP mechanisms (see Sec. 5.3.3.2). Afterwards, the WPC contacts each selected WPR, authenticating and establishing the required keys. In this scenario, the PS facilitates WPR discovery. However, the node can use several PS for a single connection, increasing his privacy.

Service generated routes are implicit routes that can be used as a mean of increasing user privacy, without requiring the user to actively partake in the process. This comes as additional privacy provided by the PS, without involving the user. This is only possible due to the characteristics of the proposed cryptographic solution presented in Sec. 5.3.1.1. Given that hints are defined on a hop by hop basis, at any point the WPR or PC can extend the existing route by pushing forward its hint, and generating a new hint towards the WPR that will be receiving the original hint. This process is deemed *subpathing*. It should be noted that, because the original sender is not aware about added WPRs, it will not share a key with them and will not be part of the CID layered encryption. To solve this issue, the WPR in the extended subpath must not attempt to decrypt the CID and only forward it.

5.3.3 Routing Impact Evaluation

To understand and discuss the benefits of the Waypoint Routing, we must evaluate its performance. We provide a two-fold analysis based on overhead and on path optimality, which determines how "close" the privacy aware routes are to the optimal ones obtained with current routing mechanisms. We later show how this proves to be one of the WP Routing main advantages.

5.3.3.1 Performance

To assess the overhead, we perform an analytical evaluation of the control overhead in data packets, and a ratio of encrypted bytes over transmitted information, allowing a better understanding of the privacy costs. Assuming a packet with the Maximum Transfer Unit (MTU)

size of 1500 bytes, we calculate the percentage of control overhead. In our scheme, a packet is composed by the IPv6 header, 40 bytes long, and two IPv6 options, of 24 bytes each: a modified routing header to include the hint, and a destination option to include the Circuit Identifier (conveyed in the reserved space). This represents a 48 byte increase on maximum size data packets, which is a 3.2% percent overhead increase. In other solutions this overhead is larger: TOR uses a 500 byte fixed cell size for data, which is then appended with headers, compounding a 960 byte overhead (not factoring in IPv4/IPv6 headers). This represents $65,8\%$ overhead increase. As far as encryption is concerned, the proposed solution mandates that the two hints are encrypted at every point. This requires 32 bytes at most, while other solutions require much more. As an example, TOR encrypts/decrypt the 500 byte cell at every hop (not counting management data structures).

5.3.3.2 Path Optimality

We propose a path optimality measure, which is the difference in path cost between the optimal route defined by the routing mechanisms (e.g RIP, OSPF, BGP), and the "privacy-aware" route. The objective is to provide a broad idea of how privacy impacts the route, and then to draw conclusions on the overhead imposed on the users to achieve privacy. We compare, in a simulated environment, three different approaches against the optimal route: 1) an *edge Path*, which is a random route using the networks edges (end-to-end routes, as they exist in the current TOR deployment); 2) a *core path*, using random routes based on the network core (to support privacy in the network); 3) and the *composed path*, which is a simple route selection mechanism that uses network knowledge to build privacy-aware routes minimizing path cost, resulting in less overhead.

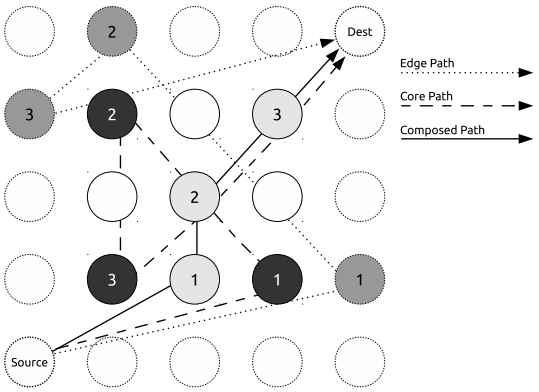


Figure 5.12: Simulation Scenarios for path optimality considerations.

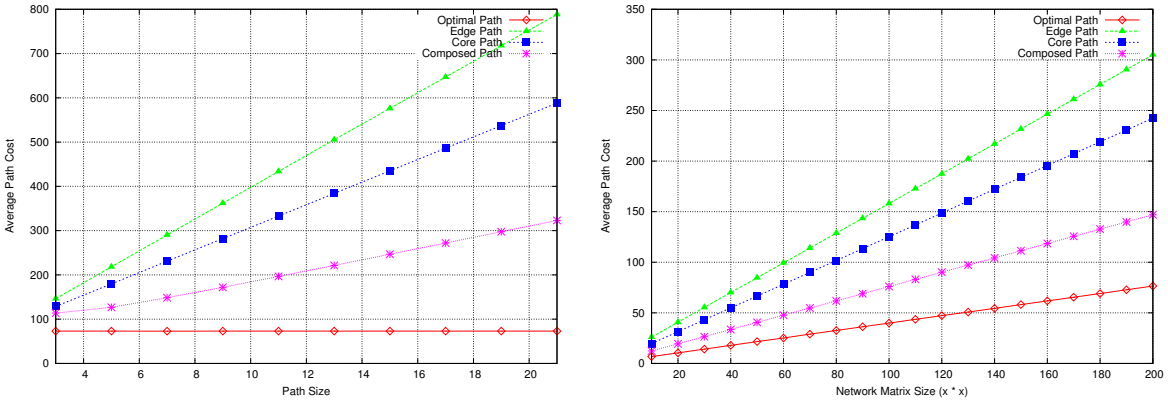
In the simulation, we consider the network to be a square node matrix, shown in Fig. 5.12, where edges (dashed) represent endpoints and the inner matrix nodes represent routers, with fixed coordinates. We define the path cost as the cartesian distance, calculated hop-by-hop. In a three-node route, the path cost would be the sum of the cartesian distances between each of the three nodes. The *optimal path* cost is the cartesian distance between source and destination, a simplification of the shortest path. The *edge path* uses random edge-nodes to act as WPRs (only dashed nodes in Fig. 5.12, whereas the *core path* resorts to random core nodes (inner matrix nodes). The *composed path*, in the simulation, uses a divide and conquer

strategy: it selects a mix of nodes in the source's quadrant, a center node, and nodes in the destination's quadrant, forming a composed route. Every result presented is the average of 10^4 executions in a custom simulation environment.

Fig. 5.13b presents the findings concerning path cost, with a 3 hop route and varying network size. This shows that the *core path* presents an overhead decrease of 28.4% on average, when compared to the *edge path*. Even more, the *composed path* reduces the average overhead by 69.6% when compared to the *edge path* approach, and 57% when compared to the *core path*. The *composed path* approach is the closest to the optimal path, with roughly double the path cost.

We apply the same scenario on a fixed network size of 10^4 (100x100 matrix) nodes and varying route size, as shown in Fig. 5.13a. This graph shows the same type of gains: the *core path* strategy decreases overhead in more than 27.4%, while the *composed path* brings the overhead down by more than 63%, when compared to the *edge path*.

With a generalized performance increase, the exciting conclusion is that using the network core (e.g. both *core path* and *composed path* approaches) can lead to a significant performance boost and overhead reduction, which in turn can result in better adoption.



(a) Variable path size impact.

(b) Variable network size impact.

Figure 5.13: Impact on abstract path cost.

5.3.3.3 Building deployable solutions

In the previous sections we presented the Waypoint routing concepts and architecture to define a new routing infrastructure that enables privacy support, along with an evaluation of the proposed solutions. Derived from Chaum Mix concepts, our proposal resembles other mix network approaches such as TOR, by anonymity standards. However, by preventing packets from reaching the network edges (end-hosts), we provide a more effective privacy solution that is not exposed to the perils of malicious exit nodes or traffic inspection [131]. Instead, we rely on the PS, which is designed to secure user data. This brings the network operator back into the privacy game, enabling it as a privacy provider, capable of delivering optimized traffic and privacy data paths stemming from his deep knowledge of the network. This not only provides better performance, but removes the peers from providing security. While in theory anonymity provides a better solution, in practice malicious exit nodes can subvert the

system, whereas here the exit nodes (and all peers in the path) are trusted. Thus, the user is trading part of his anonymity towards the different privacy services for trusted privacy. This also proves to be an advantage over peer-to-peer systems like Tarzan [51], because here the WPR entities can be certified and trusted, whereas in systems like Tarzan, and even TOR, attackers can try to operate nodes to gain knowledge about the ongoing communication, and even inspect traffic as the last hop.

Beyond the privacy gains, the other key advantage is performance. When compared to approaches that favor user involvement, the evaluation showed that just by moving from the edges to network routers can yield a 30% overhead decrease. And when applying a simple strategy of quadrant allocation, the overhead can be reduced by as much as 70%. This evaluation showed that the approach has a significant relative advantage by allowing WPR selection, further highlighting the benefits of converting privacy into a service, provided either by the network operator or by a VASP, which only benefits from network knowledge and smart route construction. Also, by diversifying WPR and PS, user privacy can increase due to the added confusion.

When compared to TOR, layered encryption in WPR is restricted to the global hint, while hop hints use normal encryption. This shows that the encryption in our scheme is only a small fraction of what is required in TOR, which encrypts the entire cell. Also, by working with IP-level mechanisms, as opposed to TOR that works on the transport layer, using only IPv6 headers results in massive overhead reduction, from the estimated 65% of TOR to an acceptable 3%. We conclude that the proposal incurs in a privacy overhead penalty, but with significantly better results than other proposals due to the low amount of extra information required.

One of the most fundamental differences is that TOR does not make assumptions on how the Onion Routers are selected, whereas WP Routing makes strong assumptions on the architecture definition, to keep traffic out of the network's rims, and to maintain communication on an optimal and secure level (avoiding traffic from being routed by endpoints). If we approximate TOR to the random model analyzed in Sec. 5.3.3, and Waypoint Routing to the other analyzed variants, *core path* and *composed path*, we can see that a core-oriented scheme can boost the performance of the overall privacy scheme, greatly reducing the performance impacts on expected delay, in some cases as much as 70%.

It is possible to summarize the proposed solution as source based routing mechanism for IPv6 using encrypted addresses within extension headers, concealing unique addresses and location through the use of Waypoint Routers. Beyond the end-user benefits, and due to its lightweight properties, the framework fosters privacy as a value added service. This promotes using the network infrastructure as the driver for privacy, getting the network operator off the sidelines in the process and leading to considerable privacy and performance gains.

The proposed Waypoint Routing mechanism allows scalable performance by bringing privacy mechanisms into the network layer. In doing so, we allow the construction of "privacy-aware" routes, based on waypoints. WPR uses less encryption, creating less overhead, and routes closer to the optimal network mechanisms. Even though the simulations use simple routing scenarios (starting with an iterated triangle inequality), they already indicate that transferring privacy towards the inner network nodes, along with smart route selection, can lead to a smaller privacy performance impact.

5.3.4 Waypoint Vertical Integration

The Waypoint Routing solution solution addresses specific network layers threats. But the dichotomy of vertical vs horizontal privacy approaches indicates that it is still necessary to analyze the solution from a vertical perspective. This means that it is necessary to consider the WP Routing implications in the PRIVED construction, and also in the integration with the VID model and the complementary VNS pseudonymity solution.

From a PRIVED perspective, it is necessary to determine the impacts on the overall privacy model. The first interesting concept is that we are breaking the uniqueness of IP addresses, which carries both user identification and location. By removing identification, and only providing location limited with the understandable hop, we solve two issues: i) we provide less means, and one less global identifier on which to perform IS information correlation; ii) we solve an orthogonal privacy solution that is created by the semantics of the identifier, rather from an identification problem that directly relates to the IS in the PRIVED model. In practice, this reduces the meaningful information that can be extracted from a packet event, which now carries less obvious relationships between upper and lower layer identifiers. This advantage also aids in preserving the different IS boundaries, where we have one less threat pending on the correlation of multiple information sets.

Similarly to the proposed link layer solution, the Waypoint approach protects against several threats on the specific network layer. But, it also assumes that there are certain entities in the network that can be aware of the identification of the node. In the previous solution this was the Access Point. In the Waypoint model it is the Privacy Controller, which does not protect against the PS. Also, as the communication is intended for a peer, or service, the waypoint solution does not protect the exchanged user information. The solution for this problem, as for the link layer, must come from the VID/VNS combination to enhance user privacy.

The VID framework should be used as a complement, from an information protection point of view. The different virtual identities can help keep separate views of the user, which in turn is supported by the Waypoint solution on the network layer, creating a symbiotic relationship. This is further supported by the fact that, in combination with the VNS approach, it can provide several tools to disguise the user, providing a layer of protection towards correlation from the network operator, and in this case, from the privacy provider. By using several network layer addresses, pseudonyms, the probabilities of having a correlation between already used addresses, at the access network, at the PC, or even at a WPR, can decrease due to the properties of the pseudonymity solution.

Furthermore, it is important to acknowledge that a network cooperating entity is required. As discussed, the PS shares many trust goals with the IdP, and an aggregation of both provides an aligned view on privacy, as well as the bridge between horizontal and vertical solutions.

The main conclusion from the complementarity of both solutions is that the network information can be used as contextual information to threaten user privacy, which is now safer due to the proposed solution. The Waypoint solution assists these problems at the network layer. However, it does not eliminate the need for vertical coordination, through virtual identities and virtual network stacks. Privacy threats are still very relevant from a vertical perspective, and even though we present solution to handle the layer below, there are still layers above that can jeopardize privacy: the transport layer and, more importantly, the application layer, as discussed in the next section.

5.4 Transport and Application Layer Privacy

The previously discussed layers, link and network, required an explicit solution to cope with privacy requirements stemming from the PRIVED approach. But, as we move up in the network stack, we face more implicit solutions that must be properly analyzed.

The implicit properties result first, from the tight bounds between the network and transport layers, and second, from the close relationships between the application layer and the VID concepts supported by IdM.

In this section we provide an analysis of the complexities of the transport and application layers, to align them not only in a horizontal perspective, but more importantly, with the vertical privacy landscape defined so far.

From the transport perspective, it is worth discussing the tight coupling with the network layer, and the consequences for privacy. On top of this, it is still important to further extend the threats defined in Sec. 3.4.4, as well as the consequences of decoupling the dual functionality of the role of locator and identifier, introduced by the IP address, that reaches the transport layer.

The application layer presents challenges that are tightly related to the vertical solution. As this layer deals with different protocols and services, it is hard to provide a single targeted solution. To address this, we extend the idea of using IdM as an application layer solution, closely related to the VID model. As such, we identify and select an IdM solution that is capable of satisfying the VID approach, and can serve as the basis for the architectural instantiation in the following chapters.

To achieve these goals, we first look at the transport layer, and then the application layer, and analyze them considering the protocols employed and the relationships with the vertical approaches, especially the VID concepts in the light of the PRIVED model.

5.4.1 Transport Layer

One of the most important features of the transport layer is that is shares identifiers with the network layer. This creates a dependency relationship that has privacy implication. The IP, which is used for the topological location in the network, serves to identify a host on the transport layer. In fact, as previously discussed in Sec. 3.4.4, the transport layer becomes a network address (IP) associated with a well know port, to identify to which application should the packets be delivered. This is the dual meaning that forces a tight relationship between network and transport, that must be further analyzed.

The ports used on the transport layer also raise privacy concerns, in the measure that they identify an application. Even though is is only a tacit definition, an agreement supported by the Internet Assigned Numbers Authority (IANA), it provides a relationship between port and application that in most cases allows the quick fingerprinting of the traffic type, and of the user's desired action. But this is only a tacit definition. In fact, several firewalls work by filtering out well known ports.

Despite this usage, it would be possible, from a privacy perspective, to decouple port numbering from the actual application selection. As an example, it is possible to run an HTTP server on port 8000, a port which is not usually associated with HTTP access. However, this requires that clients know beforehand to which port to connect, implying a discovery process. This discovery process would require a shift in how we use network ports, towards a semantic discovery of applications. In this case a node would first discover the "HTTP application", in

order to learn the correct port to connect. This would break the privacy concerns of ports.

The feasibility of this approach is already supported by the fact that several applications only use this well-known port as a starting mechanism. Once the connection is accepted, the communication is switched to other ports, negotiated upon connection, e.g. FTP in passive (PASV) mode. With a proper discovery protocol, it would be trivial to adapt services towards this paradigm. Furthermore, an advantage of such an approach would be that, with the availability of a discovery process for port numbers, we are not limited to port numbers, and can use any identifier (which could comply with the VNS approach, of using pseudonyms on the port level, to differentiate the connections without compromising the IS).

But as discussed, this is not a limitation of the way the transport layer is built, but rather a usage problem, on how we associate well known port number to specific application. As such, we focus on the duality presented by the reuse of the network identifier, in this case the IP address. We explore two different situations that concern the reuse of the identifier from the network layer, and when there is a solution that enables decoupling between the two layers, introducing a new namespace of addressing.

5.4.1.1 Network Layer Dependencies

The most common situation for the transport layer relies on reusing the network identifiers - the IP address. As discussed, this standard use of the network stack results in a dependency that also defines the privacy properties at the transport layer. In fact, the overlapping role between locator and identifier, and its reuse leads to the same location and identification problems tackled in Sec. 5.3, for the network layer.

But this tight coupling allows a simple conclusion: in the presence of a solution for the network problems, the transport layer issues are matched by the network properties. This creates the implicit solution that guides our discussion.

From the PRIVED model perspective, if the correlation is always true, then the privacy problems are not exacerbated as long as there is a solution that addresses the network layer issues. This means that there is no extra information aggregated into the IS, besides a port number, that further compromises the IS. Since port numbers are the same for all hosts, extracting further information from this event observation is only interesting from a data mining perspective, as it requires deducing knowledge not only from the observed identifier but also about the meaning of the action. While this sort of analysis is possible, it is out of the scope of our proposed solutions, since it requires probabilist and semantic inference of knowledge.

5.4.1.2 Locator and Identifier Decoupling

A different approach is to consider that the network and transport identifiers can be separated. This can happen in certain circumstances, involving a locator-identifier split. As already presented, this can be the case when using mobility solutions such as MIPv6 or HIP. In the first case, a new addressing layer based on IP addresses (Home addresses) is used as transport identifiers. In the latter case, hashes of public keys (Host Identity Tag) serve the purpose of transport identifier.

These are just examples that allow separating the network from the transport layers. This decoupling can become a two-wedge sword, as on one hand it can create the opportunity to truly solve the network problem as a pure locator solution, but on the other hand it requires

that we solve the transport layer identifier privacy issues. Assuming that there is a solution which splits identifiers and locators, a new opportunity is created within the addressing structures that requires an alignment with the VID and pseudonymity solution.

When properly separated, the transport layer identifiers assume a pure identification role which, in most cases (e.g. HIP and MIPv6), introduces a globally unique identifier. This was discussed as part of the mobility pseudonyms in the VNS approach, in Sec. 4.5.3.4. From this section we can extract the generic solution for a split solution, that relies on using pseudonyms, guided by the VID model, which should then be coupled with both the upper (application) and lower layer identifiers. From the PRIVED perspective, this is similar to the previously presented solutions, since it requires pseudonyms to break the correlation mechanisms introduced at the transport layer.

5.4.1.3 Vertical interactions of Transport identifiers

From a vertical privacy perspective, regardless of the two discussed scenarios, coupled or decoupled layer, there is still the need for a vertical solution. In the first case, where both layers are tightly coupled, this coordination is assured by the network layer solution. In a possible instantiation scenario, this would be entirely assured by Waypoint Routing.

If a layer decoupling is in effect, like the one proposed in the next chapter instantiating mobility and privacy solutions (Sec. 6.5), solving the network layer becomes simpler, but there is a need to address the transport layer independently. The mechanisms to achieve this would require to properly align the transport addressing solution with the pseudonymity approach from the VID model, in order to respect the PRIVED correlation properties.

Regardless of the situation, this layer still transports the application layer information. As the other vertical dependencies, solving the transport layer does not entirely solve correlation through upper layer identifiers, nor does it handle the information exchanges with peers or services. Again the solution is two-fold: first, it is necessary to address the application layer identification problems, and second, these solutions must be properly scoped within the VID approach to enable a vertical privacy interaction.

5.4.2 Application Layer

As argued for the vertical approach in Chap. 4, the privacy solution for the application layer lies with Identity Management. How to introduce IdM as a privacy solution has been preliminary discussed in Sec. 2.4. The IdM paradigm brings the notion of user or digital identities. We have relied in this construction to build the vertical VID solution, as well as the support for network pseudonyms. Therefore, it makes sense that this is the approach followed at the application layer, thus reusing the solutions provided by IdM. However, despite relying on IdM for several aspects, most of the application layer specific solutions have been assumed throughout the Thesis. It is only natural that we align the application layer solutions to fit the key concepts that we have been exploring up to now, namely pseudonymity.

While the focus of this Thesis does not particularly fall within the application layer, the impacts and threats present on this layer must be understood, so that we can rely IdM technologies and an architectural corner stone for the VID model, especially as we seek to instantiate the IdM functionality, rather than stay at a purely conceptual level. In fact, this is what we perceive to be privacy at the application layer, making it important to explore.

Many technologies that could drive our IdM privacy aware interaction exist at the application layer. We have explored several solutions [122, 70, 116, 20, 109] in Sec. 2.4. However, out of the range of the cited approaches, we must select the one which best fits our proposed concepts and requirements. As outlined in other sections, the most important factor that drives the proposed vertical solutions is the pseudonymity support, which is the main privacy requirement for the application layer. Pseudonymity assures, according to the PRIVED model, that information is not correlated between different providers through unique identifiers. But, at this point we are not concerned with the vertical implications of this statement, but rather with the horizontal support at the application layer, enabling us to explore different solutions.

At the application layer, we are mostly concerned with service interaction and the support of multiple pseudonyms (ideally per SP). In fact, we can generalize this requirement to a claim concerning all interactions: the application layer solution must enable the use of generalized pseudonyms towards service interactions. With this main goal, we take SAML 2.0 [20] as the primary application layer technique towards privacy preservation, and generalize it, so that its benefits can be extended onto other IdM and privacy solutions, opening the possibilities on the application layer. It is worth noting that the overview presented here is entirely biased towards the network point of view. The objective is not to discuss the information exchange, but rather the information leading up to it, involving the network.

5.4.2.1 The SAML Use Case

Over different iterations SAML became synonymous for a set of protocols, which together define the core SAML 2.0 specification [20]. These include protocols to handle Assertions, Authentication, Artifact and Name resolution and mapping, as well as Single Logout (Single Sign-On is implicit, given that it is one of the most basic requirements). The two most important features provided by SAML are the already discussed, and important, SSO and Federation mechanisms, which make it useful to support interoperability among different providers. Beyond the User entity, the SAML architecture mandates the Identity Provider (IdP) and the Service Provider (SP). The proposed naming also includes SAML Authorities or Asserting Parties as alias for the IdP, and Relying Parties as alias for the SP.

One of the fundamental aspects about SAML is that it provides pseudonyms, which are "opaque pseudo-random identifier with no discernible correspondence with meaningful identifiers", citing the SAML specification. By employing these pseudonyms, SAML promotes privacy by preventing any type of collusion or linkage by multiple providers or eavesdroppers on the network, that would be possible through global identifiers (Sec. 3.4). The key aspect is that SAML uses these identifiers when establishing a relationship between the IdP and the SP. Therefore, every bond between SP/IdP uses a new, uncorrelatable pseudonym, thus safeguarding user privacy. And because this is managed through the IdP, the user is not burdened with identifier management. Beyond privacy protection, SAML also provides identifier protection as defined in Sec. 3.5, through encryption of the different exchanges which contain attributes, identifiers or assertions, thus ensuring confidentiality. Also, SAML complements this with SSL/TLS tunnels, providing the sought after identifier protection. Another key aspect that emerges from SAML is the use of profiles: SAML provides a generic framework that is instantiated as required by specifying different profiles (e.g. Web access). This is generic enough to cover most use cases provided by any other mechanisms, at the cost of simplicity. This aspect makes it more flexible other web based solutions, like OpenID.

Beyond these network oriented mechanisms, SAML includes strong privacy features which

allow providers to exchange privacy policies and settings. This type of privacy is what we refer to as the contextual opportunities, because it deals with the contents of the exchanged information. SAML enables the definition of access and exchange policies that limit the nature, type and form of the exchanged information. This is a powerful mechanism, that even though escapes the objectives of our work, provides an important privacy feature which is currently under research.

The most relevant part of promoting a SAML approach, from our point of view, is that, when properly aligned with the network layer, SAML can provide application layer support for most of the privacy principles proposed by PRIVED. It is pseudonym oriented approach, enables the information reduction principle even at the information exchange level, and provides great tools for vertical control aspects. This makes SAML a powerful composition of feature rolled into a single protocol. But, strictly speaking at the application layer, the baseline lies in the fact that SAML provides the required pseudonym granularity to interact with different services, providing a very adequate privacy solution in light of PRIVED. It would even be possible, with a collaborating IdP, to support anonymity in the interaction with services.

5.4.2.2 IdM Support on the Application Layer

After the observation of the SAML use case, we can extract several concepts that must be supported at the application layer. These can be used, first to rule out the remaining solutions as privacy drivers for our proposals, and second to enhance those same application protocols with the missing features. A generic IdM solution would have to provide, beyond the fundamental IdM issues, a strong component for pseudonym generation and management. This includes identifier management for the user, and controlled interactions with SPs. It is in these two aspects that most solutions do not provide any guidance or support. We extract from SAML the necessity of establishing non-traceable pseudonyms that should not be subjected to linkage or correlation to other identifiers. This is fundamental, in order to support the information reduction argued within PRIVED, as well as to prevent the correlation of different IS. The second aspect that is worth mentioning is the granularity at which IdM solutions must generate pseudonyms. In line with the PRIVED concepts, SAML provides a unique pseudonym between SP/IdP. This a very good approach, and IdM solutions that are privacy focused must provide such granularity, so that it is possible to avoid any type of privacy threats. Furthermore, it should be even possible to generate more than one SP/IdP pseudonym per user, according to the needs of the user. Only in this flexible environment can we guarantee that the information sets are not compromised. This is also vital for any vertical integration with solutions such as VNS.

However, currently only a few IdM solutions feature such approaches. This is particular visible in web environments, where solutions are mostly aimed at SSO, and data management i.e. to simplify the user's life, which is a valid argument in itself, but does not yield the most adequate privacy environment to support cross-layer network privacy.

5.4.2.3 Vertical Integration of IdM

There is still one topic that escapes the horizontal approach. We have argued in Chap. 4 that we turn to the IdM layer for control, both for identity information management, and also for cross-layer pseudonym management. But so far, there has been no indication on how to

actually achieve this using a specific IdM technology. While we propose such a solution in the next chapter, it is important to establish two basic requirements that are entirely imposed on the application layer protocol: i) there must be a strong relationship between the IdM layer and the remaining identifier and protocols in the network stack; ii) there must be a strong relationship towards the user.

The first requirement establishes the need for control mechanisms regarding the vertical approach. If the IdM solution simply aims at SSO and nothing else, it misses the objective of providing a vertical layer that enables several of the benefits discussed so far. And this is directly tied to the user relationship that is granted by the IdM solution. Solutions like SAML promote the usage of policies and strong access control to limit access to user information, and to promote different privacy mechanisms. This must be a part of the solution, which will enable not only vertical policies, but also pseudonym related policies.

But, the cross layer integration requires that we explore a vertical model for privacy management, which influences how we treat the IdM solution: using an IdM solution as an application layer protocol requires that all communications are established through lower layer protocols. However, establishing lower layer protocol connection requires that we have already instantiated a VID onto the network, which causes an inter-dependency between layers and identity selection. This problem cannot be solved horizontally at the application layer, requiring once again a vertical solution, explained in the next chapter, where we integrate SAML with VNS.

5.5 Conclusion

At the beginning of this chapter we set out to perform individual enhancements to each layer. The reasoning behind this course of action was clear: having proposed a vertical solution, the need to tackle individual threats at each layer becomes more pressing, so that they do not compromise the overall privacy approach. The motivation behind this approach led to seeking out individual threats on different layers, and to the analysis of which protocol mechanisms lead to privacy loss. When contrasting the work highlighted in Sec. 2.5, which dealt with the network aware privacy aspects presented in the related work, the focus was evident on both the link and network layers. On the link layer, the major threats deal with protecting the network access, from identification and tracking. On the network layer, the main focus is about protecting user identification as well as location.

However, these are the overall threats. We started to dig deeper on the link layer, and as we analyzed the privacy threats, we discovered that each individual protocol can yield a large amount of information, if not properly handled. We explored 802.11 to demonstrate our work on the link layer, thus providing an in-depth analysis of what can be extracted just by protocol observation. This work resulted in a clear definition of all the information that can be seen on the network, and that can threaten user privacy. After carrying out this analysis we can only conclude that to take privacy seriously, this should be performed for every protocol that is the target of privacy threats, and that can identify the user.

Then we proposed a novel solution that involves the secure transport of link layer packets through a novel encryption scheme. In our proposal we identify the communication channel between two peers by the employed key, rather than by any end-point identifiers (such as MAC addresses). This proved to be a very effective scheme given that only the channel key holders are capable of looking at the contents of the stream, thus releasing no private

information. But, this required a few changes on the link layer protocol operations, and so we were faced with the necessity of evaluating the feasibility and performance of the proposed solution. This was done through simulations, considering several aspects. The bottom line is that, using the proposed scheme effectively provides link layer privacy, but carries a small penalty that can be minimized with the right tools and proper implementation. This led us to believe that our solution, not only meets the proposed privacy requirements, but succeeds in not compromising the vertical aspects discussed in Chap. 4.

Afterwards, we naturally shifted the focus towards the network layer, to deal with identification and location issues. We found that the best way to provide privacy relies on voiding the IP addresses of topological meaning, or at least, restricting that meaning to a very narrow location range. With this in mind, we turned to anonymity based solutions to find the answer to location and identification problems. However, current anonymity schemes provide significant penalties for adoption, relying on network edges and users for routing, and providing large performance overhead. With this in mind, we proposed a lightweight routing mechanism that can be described as a source based routing mechanism for IPv6 using encrypted addresses within extension headers, concealing unique addresses and location through the use of Waypoint Routers. The result was in fact a lightweight routing mechanism that only encrypts extension headers, pushing the packet along the route by using routing Hints. This enables the paradigm shift that we deem important: by creating lightweight schemes, we can push privacy into the network core, and network operators, thus increasing the adoption of privacy technologies and reducing the costs of privacy as a whole. This opened towards positioning the network provider (where the operator is a primary candidate) as a privacy provider, using its network knowledge to aid in route selection and privacy provision, resulting in much better and efficient routes. To provide accurate values for this, we proposed a simple evaluation mechanism based on path optimality, that tries to compare different routing solutions to an optimal route, thus estimating the impact on network routing schemes, caused by different privacy solutions. The results were indeed encouraging, resulting in performance gains over existing solutions that indulge our best expectations. The conclusion was that, not only it is worth considering the proposed scheme, but that the path optimality measurement scheme can provide interesting metrics for privacy comparison.

Lastly, we performed a two-step analysis of the transport and application layer. On the transport layer, we uncovered the identifier dependency between network and transport layer, along with the threats introduced by transport mechanisms (ports), as wells as those that can result from an identifier decoupling. The conclusion that the transport layer can be indistinguishable, from a privacy perspective, from the network layer due to overlapping identifiers, allowed us to step into the application layer, simply to understand the assumptions and features required to support the outlined privacy mechanisms. In the light of pseudonymity, we found that among the analyzed protocols, SAML provides the best match for any application layer solution. However, we established some generic concepts that guide the selection, adoption and even modification of application layer mechanisms: i) the application layer IdM protocol must be strongly pseudonym oriented, establishing non traceable pseudonyms that bare no resemblance to any global identifiers the user might have, thus undermining any possible correlation; ii) the IdM solution must support a granularity, that at best, supports creating per SP pseudonyms, thus generating no basis for collusion on the SP side, as well as for eavesdroppers on the application layer. We also found that both these requirements are well established within SAML, and that it is reasonable to look once more at the vertical aspects of integrating a SAML based on a joint VID and VNS approach, thus tackling the

remaining network integration issues.

These issues stem from identity selection and network utilization problems that must be met with a vertical (architectural) approach. These concepts are considered in the next chapter, showing how different architectures can abide by these principles as we build architectural instantiations of these paradigms.

Chapter 6

Architectural Instantiations

> Once we accept our limits, we go beyond them.
>
> **Albert Einstein**

From the previous chapters, identity emerged as a path towards increasing privacy to the end-user. As a privacy tool, identity simplifies information control, while simultaneously allowing the development of user-centric privacy solutions. This is visible in both the horizontal and vertical approaches presented before, creating a trend that can be synthesized through architectural privacy drivers.

In this chapter we present four architectural drivers that define conceptual abstractions for network architecture design, resorting to the underlying principle of using identity as a privacy enabler. For each of the proposed drivers, we introduce concrete instantiation examples that can be seen as the results of the concepts explored so far.

As particular instantiation examples we first propose a solution that integrates a vertical privacy layer using a specific IdM protocol. This is followed by a solution that decouples control and execution in the network, allowing the use of identity as a privacy-aware control layer. We then explore a privacy architecture focused on operational aspects, supported by an identity-enabled layer separation mechanism that can increase network privacy. Finally, we propose a privacy-oriented architecture that uses identity to simplify the cross-layer identification and control through identity references, closely related to the VID and VNS approaches.

6.1 Introduction

The last two chapters explored the idea of a vertical dimension to privacy, complemented by horizontal considerations where necessary. The vertical dimension deals with user-related privacy concepts, relying on a cross-layer approach to support the necessary privacy requirements. This vertical function was designed using the concept of user identities, using IdM as a realization of the proposed approach. The horizontal dimension targeted specific layer and protocol threats, focusing on eliminating unwanted relationships and mechanisms that compromise user privacy, on different network layers. These solutions usually required using properties from the vertical considerations as means towards supporting end-user privacy.

In this context, identity appeared as a privacy tool that enabled controlling user-related information, aggregating it into well-defined sets, as well as defining a mechanism that supports linking different protocols to the user identity (e.g. through pseudonyms). By acknowledging that identity can function as a privacy support technology, it is possible to combine the concepts implicit in other chapters, into a specific set of privacy architectural drivers. These drivers can become the principles that introduce the possibility of designing privacy-aware network architectures.

In this chapter we propose to design privacy as an integral part of the network architecture, supported through identity concepts. We first identify these concepts as the mentioned architectural drivers, and how they can be converted into concrete instantiations. These different aspects are proposed in Sec. 6.2, where identity is used as the main support tool, resulting in four different concepts that are instantiated into different privacy-aware architectures.

The first step consists in defining an architecture that instantiates identity as a vertical privacy enabler. As presented in Sec. 6.3, this proposal integrates a VID/VNS combined approach with specific application layer IdM technology, in our case, SAML 2.0 as indicated by the previous chapter.

The second approach proposes a clear separation between the control and execution of network related mechanisms. It builds on the concept of using identity as a privacy-aware control layer, separating user-related information from the actual protocol execution mechanisms. This separation attempts to avoid the correlation of different events by isolating the actions on the network, properly controlled by a privacy-enabled layer. In Sec. 6.4 this concept is instantiated into a mobility related architecture that splits control and execution, outsourcing control to the IdM layer and leaving the execution up to the different mobility protocols. With the control formalized on a vertical layer, we explore the operational aspects of different protocols, exploring the role assumed by identity in the context of mobility aspects (a common concern in NGN).

Closely following the concept of separating control and execution, we also propose a separation between layers. By breaking the dependencies between layers, properly tied to user information, and in this case, identity, it is possible to develop privacy-aware protocol instantiations. Accordingly, identity can become an operational driver when properly tied to layer separation mechanisms. This is presented as part of a locator-identifier split, in Sec. 6.5, where we explore the role of identity as a privacy-aware mechanism that can be incorporated into different protocols, in this case the Host Identity Protocol. HIP is a protocol that already presents identity-related concepts, along with the proposition of clearly separating the network from the transport layer, a property that we use to the benefit of privacy-aware mechanisms.

From a privacy perspective, based on the previous concepts, identity can be classified as a

user-related control layer that can be present in different network protocols. Combining these properties can lead to a simpler and more effective privacy control mechanism. To achieve these goals, we provide a last architectural driver that defines identity as a privacy-aware interaction driver in NGN. In practice, this means that to improve the privacy control and center it on the user, we convert the current network stack (and protocols) into a user-centric operation that is enriched by user information. By using mechanisms stemming from the VID model, the VIDID, it is possible to endow the network with privacy-aware identifiers, controlled through identity. The result of this application is an architecture that embeds identity into existing protocols, as means of providing simpler control over important privacy-related information. In our proposed approach, presented in Sec. 6.6, network protocols become in some way related to the user identity, introducing a privacy control mechanism over the interactions that occur on multiple protocols.

To better understand the different instantiations, it is important to concretely define and summarize the four different architectural concepts that drive the different proposals. This is presented in the following section.

6.2 Architectural Concepts

The solutions proposed in the previous chapters have underlying recurring principles. These principles can be summarized as the concepts that drive architectural design, which we emphasize as architectural drivers. These approaches must be systematized, so that the resulting concepts can be applied when proposing privacy-enabled architecture instantiations.

The identification and application of the conceptual drivers can be seen as an outcome of the privacy proposals present in the previous chapters, which already indicated the possibility of leveraging broader concepts towards new privacy-oriented architectures.

One of the important recurring concepts is using identity as a privacy tool. Because identity simplifies the control of user-related information, it provides adequate tools upon which to design privacy-aware systems that take into account user-related data. It can also create a vertical space that can relate both network and user related identifiers.

As a cross-layer privacy tool, identity can cover several technologies, ranging from low level authentication to application layer services, which makes it applicable to users, devices and services. Accordingly, identity becomes more that just a hub for user information. Instead, it becomes an important tool for intrinsic privacy support, from where we can draw several privacy related architectural concepts.

Based on the cross-layer capabilities, it is possible to envision identity as a vector that drives privacy design. This results in the first concept, which highlights identity as a vertical privacy enabler. Identity can provide a privacy-oriented vertical layer, using the IdM protocol to control pseudonyms on different layers. This requires an integration between the cross-layer pseudonymity solutions and the IdM solution at the application layer, thus providing a vertical architecture. It is important to notice that, while we sometimes purposely confuse identity with IdM, it is not restricted to IdM. Identity can embody a design philosophy, whereas IdM can be a core component (among others) used to instantiate the outlined philosophy.

The integration of the vertical privacy enabler concept opens the door to a new design feature that outsources control decisions on the network to a privacy enabled layer. Aligned with the previous concept, we can resort to the vertical identity layer as the control mechanism for different network procedures, thus making them privacy aware. This is proposed through

the second architectural driver, which favors identity as the main control mechanism within the network, thus creating a privacy (aware) control layer. The reasoning behind this derives from the fact that the identity layer is capable of handling user information, making it an information cluster (relative to the user) in the network. By aggregating user information and collecting network data, it can provide the conditions for strong decisions regarding any protocol in the network. Furthermore, the decision capability is already partially reflected in the IdM layer, since it comprises several policy management considerations. The practical outcome is an architectural split that differentiates between control and execution, with the control part relying on identity.

Separating different functional aspects requires a more tight control on the network in order to maintain the consistency across different operations. While this is achieved through identity for the control and execution, it can also be done for separating different layers in the network stack. This introduces value not only for the control part, but also for the operational aspects of the network, which can be related to a privacy-aware control structure. Following this reasoning, we take advantage of identity to define a privacy-oriented layer separation that breaks the inter-layer dependencies, and thus minimizes the correlation opportunities.

Using identity-related aspects in the operational part of the network results in the third privacy driver that defines identity as privacy-enabling operational driver. The value of this concept lies in the fact that, by separating different layers with user-centric references, it is possible to include privacy considerations in the different network protocols.

One possible instantiation of this driver is using HIP, which already supports identity concepts that enable easily switching network identifier (IP addresses), without compromising ongoing connections, which are tied to the Host Identity (that can be linked to IdM).

Finally, we consider that identity can become the primary privacy-aware interaction driver in NGN. This driver can be seen as traversal idea, present in the previous concepts. Its main purpose is to promote identity-related information in every network interaction, by inserting references in existing protocols. This enables a two-fold approach where, on one hand, every network interaction transaction can be simplified by resorting to common and cross-layer user information and, on the other hand, every transaction can be governed by privacy-aware mechanisms. This also promotes the use of identity as an integral part of network stack management, enriching several protocols and making them privacy aware by providing an easy relationship to pseudonym based solutions.

Proposing several architectural privacy drivers means that all the current models for network interactions can be redesigned to stand any of the proposed principles, which are closely related to the user. We can summarize the different drivers as follows:

- Privacy through an identity vertical enabler (Sec. 6.3)
- Privacy through an identity control layer (Sec. 6.4)
- Privacy though identity as operational driver (Sec. 6.5)
- Privacy through identity driven interactions in NGN (Sec. 6.6)

Using these conceptual guidelines, it is possible to design different privacy-aware architectures promoting the identity aspects outlined above. The result is a set of proposed architectures that show the advantages of the different drivers, instantiated onto network paradigms.

6.3 Cross Layer Privacy support For IdM

As part of an architectural effort to increase privacy through a vertical layer, it is necessary to integrate pseudonyms as an extension of user identities. This concept is an underlying approach of the VID framework, aligned with the virtual network stacks to provide pseudonymity. Therefore, it must be instantiated through a concrete IdM proposal that integrates with the network, leveraging the advantages that IdM provides, such as an authentication and authorization framework aimed at retaining user privacy. This complements the use of a privacy-oriented vertical solution to control pseudonyms on different layers.

Identity management systems focus on privacy aspects that deal with the interaction between users and services. In this context we use a SAML [20] based IdM framework [76] that employs pseudonyms between user and service provider (SP), providing privacy to the user by avoiding correlation across different SPs. Through minimum disclosure policies and technologies for user attributes, this approach protects every user interaction by hiding or withholding sensitive information. In this section we present a SAML-based IdM solution that supports the VID concept, designed to integrate with the VNS capabilities, especially focusing on how to support the major IdM interactions. The main objective is to prevent IdM related pseudonyms (SAML pseudonyms) from being linked through network stack identifiers.

Since the network vertical threats that relate to the PRIVED model fall in the category of the vertical solution proposed earlier, addressed by the VNS approach, we focus on a solution that handles the threats of correlating IdM pseudonyms through specific network interactions: by inspecting network events, such as the initial contact towards a service provider or selecting an identity to use on the network, it is possible to compromise IdM pseudonyms. We take advantage of cross layer pseudonyms to mitigate the threats outlined in Sec. 6.3.2, thus enabling a privacy solution in IdM scenarios that integrates with the proposed network pseudonym solution. However, it is important to first characterize the IdM architecture, which was also defined in the scope of the SWIFT [72] project.

6.3.1 Cross-layer IdM Architecture

The IdM architecture upon which we base our network pseudonymity integration support, already attempts to cover a wide range of privacy issues by employing a cross layer identity solution. With a SAML framework at its heart, it uses IdM concepts to tackle network based interactions, well beyond SP interactions. We focus on the relevant aspects that can be used to drive cross layer privacy, addressing issues that range from the network layer to the application layer, providing an approach to ensure the user privacy. Some of the most important features provided by the SAML framework are the pseudonyms established between the SP and the user, shown as the End User[1] (EU) in Fig. 6.1, which protects the user identity, making the user anonymous at the SP. Single Sign-On also provides an important value for the end user by enabling a more secure and controlled environment, entirely supported by SAML , ensuring user authentication and identity control. Also important is the use of the VID concepts, which means that the user only discloses the relevant parts of his identity, appearing as different virtual identities exist, instead of a single user. This increases the

[1]The term End User stems from SAML naming, and commonly refers to the end user as a manifestation of the user on a particular device. While we adopt this naming here for proper alignment with SAML, this commonly refers to the user identity in our proposal, or simply the user, and we differentiate from the user terminal when necessary.

overall privacy of the user.

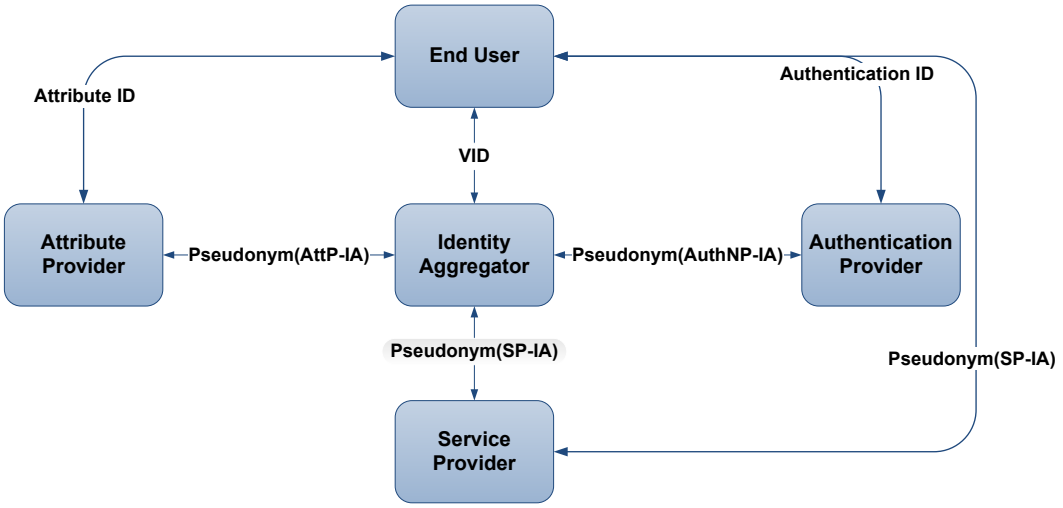


Figure 6.1: A Cross Layer Identity Management Architecture.

The architecture extends the use of the SAML framework to the different layers of the network. This means that the IdM framework is not limited to service interaction and establishes a consistent cross-layer approach for security and privacy. It is composed of five main functional elements, as seen in Fig. 6.1.

The user plays a central role in the IdM architecture. As the owner of the identity information, the EU selects appropriate policies and mechanisms that handle how information is disclosed within the system. By using several identities, the EU is able to consume services made available by an SP, which in this scope is anything that provides additional value to an EU, including network services. The SP consumes EU's identity information in the form of *Authentication Statements* to ensure that the EU is really who he claims to be, and *Attribute Statements* that ensure the EU meets all the necessary authorization requirements to consume the services. The Authentication Provider (AuthNP) is the element responsible for the actual EU authentication. It creates the necessary authentications statements that also act as SSO Tokens, and are used by the EU in order to prove that he is the rightful owner of a given VID, thus proving its authenticity to SP and network services. This description fits perfectly with the envisioned features for the Identity Manager presented in the VID framework[2] (Chap. 4).

The described processes occur through the IdAgg that acts as the coordinating element of all the EU identity related information. In the VID context the IdAgg should be mapped directly into the role of the IdBroker, proposed in the VID framework. In fact, the IdAgg is should not be the information holder of user related information (e.g. attributes, credentials), but instead should retrieve it from an AttP. In this case, each attribute should be consider an Entity Profile Part, described in Sec. 4.3.1.2. Accordingly, the AttP corresponds to the Entity Profile Part holder.

By interacting on the behalf of the user, using different pseudonyms for each peer entity

[2]Even though the functionality is similar, the naming is modified to be properly aligned with the SAML 2.0 specification, as well as the expected functions stemming from the cross-layer IdM framework proposed within the SWIFT project. This provides a seamless integration and protocol re-use.

within the framework, the Identity Aggregator (IdAgg) enhances EU privacy. One IdAgg keeps track of the different AttP entities, responsible for storing user attributes, fetching the necessary attributes needed by the EU to consume a service. The framework employs a minimum disclosure policy when it comes to disclosing EU's attributes to the SP. This is particularly interesting in the light of the PRIVED model, where each Information Set should be as small as possible (which can be achieve through minimal disclosure) to prevent unwanted correlations between different sets.

The cross-layer IdM approach provides a user centric framework where the EU has control over its identity. It protects the user information against unauthorized access and exercises control on information access, by disclosing only what is absolutely necessary for the EU to consume a service. It also uses pseudonyms in the communication channel between all the entities, being the IdAgg the only element that can link different VIDs. However, the network can jeopardize the aforementioned privacy efforts by presenting several threats as discussed in the following sections. Also noteworthy is that part of the lure of IdM systems is the creation of distributed policy environments [87], where policies play an important role in the framework, for access control, information, context or networks, among other resources.

6.3.2 Linking SAML pseudonyms

As an IdM property, SAML supports the use of pseudonyms to ensure that a user can perform multiple uncorrelated interactions with the same service provider. This follows the philosophy of the PRIVED privacy model, as well as indicating that it can be aligned with the pseudonymity approaches outlined in Chap. 4. While guaranteeing that correlation at the SP is impossible using the mentioned pseudonyms alone, it also clearly states that "correlation may be possible through non-SAML handles" [20]. Maintaining different pseudonyms for a certain layer will not ensure privacy, if through vertical linkage we can create a relationship between a lower layer identifier with a higher layer identifier, thus inferring that those pseudonyms belong to the same user, as shown through the PRIVED model. In practice, a user may present many SAML pseudonyms to the same SP, but if all interactions are made using the same IP address the SP could infer that all actions were performed by the same user, defeating the purpose of the pseudonyms. This issue does not stem from SAML, but rather from the fact that SAML interactions are carried atop identifiers over which SAML has no control, adhering to the vertical correlation properties in PRIVED,

The aforementioned correlation events highlighted in the PRIVED approach will always occur in two situations on the SAML framework: upon the first authentication of the user against the IdAgg, and when contacting an SP. This procedure maps well to an IdAgg-initiated authentication scenario: when the EU is already authenticated with the IdAgg, and wishes to access an SP, the EU asks the IdAgg to issue an Authentication Statement for the SP. This Authentication Statement is now bound to a SAML pseudonym (created during the initial enrollment/subscription) used by the SP to identify the EU. Assuming that the EU has more than one subscription with the same SP it would be easy for the SP to, through vertical linkage, infer that those pseudonyms belong to the same user. We consider the set of identifying information that SAML accredits to one identity (one pseudonym) and expand it to include the identifiers from the lower layers. With this expansion we can align the pseudonymity features from SAML with equivalents in the network stack.

These interactions that can lead to a privacy breach, are directly related to identity selection (as discussed in Sec. 4.5.3.6). In SAML based interactions, the user chooses his identity,

represented by the employed pseudonyms, when accessing a SP. Before the authentication process occurs, SAML has no considerations regarding pseudonymity, considering that all non-authenticated users are equal. However, the same is not true for network interactions: when a terminal sends a message into the network, it is immediately disclosing identifiers and therefore asserting an identity. For IdM it means that upon the moment a terminal starts sending packets it selecting an identity, in the form of network identifiers (IP and MAC addresses). This implies that, prior to contacting a service, the user's identity must already be instantiated in terms of local identifiers (at the terminal). If not, when the EU contacts an SP, it is already presenting a set of network identifiers without an associated SAML pseudonym. After the initial SP contact, the SP will initiate an EU authentication to determine the user's identity (pseudonym). This is another important use-case defined as SP-initiated authentication [124]. From now on, any pseudonym that the EU presents to the SP can be correlated, through network stack identifiers that the EU presented to the SP on its initial contact.

The contradiction between the typical SAML authentication and the exposed interaction model dictates two main cases that must be handled: 1) the user has already chosen the identity to be employed (and consequently all associated pseudonyms); 2) the user has contacted a SP but has not yet selected an identity, and special considerations must be made to circumvent this use-case.

6.3.3 Supporting Cross Layer Pseudonymity

Pseudonymity is a core SAML feature, and therefore a prime candidate to be integrated into the cross-layer IdM framework by following the VNS approach. We re-use the concept of virtual interfaces instantiated per identity, as presented initially in Sec. 4.5.3.1, creating a set of network identifiers based on identity. As discussed, the IdM paradigm is important to instantiate the control proposed by the VID model over the different VNSs, as emphasized throughout in Chap. 4. The proposed control plane interacts with applications, providing information for network stack management (Sec. 4.5.2). However, the challenge for the integration of VNS and SAML, is to determine when and how to apply a new VNS, either when contacting the IdAgg or an SP.

6.3.3.1 Network Pseudonyms and the IdM Framework

To integrate network pseudonymity support in the SAML-based architecture, it is necessary to generate different identifiers for each SP (or group of SPs). Managing the complexity of using per-SP pseudonyms should be supported by IdM control layer, to enable the required granularity for pseudonyms. In this context, it is also important to balance privacy and performance, given that creating multiple network stacks can have an impact on the network (Sec. 4.4): creating multiple pseudonyms will impact the access technology because the device will now have to take care of multiple ongoing communications (e.g. multiple 802.11 [66] associations). Addressing impacts derive from the fact that the user will employ a number of identifiers proportional to the number of VIDs in use, reducing the address space available in the network. Because of this, it is important save resources by re-using network stacks for different operations, while still preserving the user privacy.

For most operations, such as contacting an SP, different SAML pseudonyms will be used, assuming that the identity was already selected and properly authenticated, leaving us in the IdAgg-initiated authentication scenario. The decision to generate or reuse a VNS should

come from the IdM layer, which is aware of authentication, VID, pseudonym and SP. The most straightforward strategy would be to instantiate a stack for each SAML pseudonym. Alternatively, it is possible to instantiate one stack per VID [129] (Sec. 4.2).

Regardless of the adopted granularity, there is always information divulged at the application layer (e.g. presenting the user's real name at different SP links different pseudonyms), a situation that is foreseen in the IdM application scenarios.. By reusing the same privacy policies that guide this process, it is possible to select one VNS for different providers. Such policies should be enforced by the IdM control layer, which handles the mapping between the SAML identifiers and the network stack identifiers.

The IdAgg can be considered a special case, given its implicit trust properties on behalf of the user. The initial authentication with the IdAgg requires a VID, along with a VNS, and yielding an SSO Token [124] that may be used to contact several SPs. Subsequent VID authentications at the same IdAgg can use the same VNS, given that the IdAgg is trusted entity and is already capable of correlating different pseudonyms to the same VID, thus saving resources that yield no privacy increase. This is a recurring trend throughout the proposed solutions, where the network provider is required to be a trusted entity: the Identity Manager in VID framework; the Access Point in the link layer privacy solution; and the Privacy Service in the Waypoint Routing proposal. In all of these cases there is a common trusted entity that is able to map the different identities or privacy features, something that must be included in the architecture design.

Nevertheless, in this case, supplemental privacy countermeasures can be taken to protect against eavesdroppers, such as a user creating a set of predetermined VNS instances, used in a round robin fashion, defining a VNS pool for IdAgg authentication. While this has no privacy effect towards the IdAgg, it can be sufficient to defeat eavesdroppers on the network, while still saving a fair amount of resources.

6.3.3.2 Contacting the Service Provider

The proposed generic policies that cover creating or reusing a VNS to contact an SP assume the user is authenticated and has already selected a VID, representing the IdAgg-initiated authentication scenario. However, this only covers the case where the identity to use is already clear. When the first contact is towards the SP, defining the SP-initiated scenario which is the one first contacting the IdAgg, the risk of correlation increases, given that a network stack is required to first contact the SP. Protecting the user in this scenario requires two approaches to be considered: 1) use a new VNS for such events, preventing correlation but triggering a performance bottleneck or 2) use a VNS pool dedicated to this case, from which a VNS is selected on a round robin basis on each occurrence. As soon as the user selects his identity, the appropriate VNS can be used instead.

The later approach needs to be carefully evaluated as there will be a chance of correlation, even if small. The positive uptake is that this reduces the number of required stacks to maintain the unlinkability across different identities. By using a round robin approach, there will be a large period of time until stack reuse occurs and linkage is possible. Nevertheless, since this is policy driven, it can always state that a new VNS will be used in such cases, incurring in the performance penalties as mentioned in Sec. 4.5.5, but safeguarding user privacy, especially if we considered this to be the non-standard case as opposed to first authenticating at the IdAgg.

6.3.4 A SAML based Architecture

By itself, the VNS is only a tool that can prevent linking attacks on the network. The strategies that allow control of the pseudonymity solution are just as important as the tool itself. For user privacy protection, there must be a clear definition of how the VNS is used by the IdM framework, and more importantly, we must define the entities that provide such control and interaction, and the functions they must support. The goal of the privacy proposal is to establish a relationship between SAML pseudonyms in the IdM layer, and corresponding virtual network stacks. To do so, we connect the SAML operations that require network resources to the creation or re-usage of a VNS. An example of this process can be a user contacting an SP which will cause a VID to be chosen, consequently triggering instantiation of SAML pseudonym. This pseudonym instantiation must be properly conveyed to the cross-layer IdM framework, where a decision must be taken to either create a VNS or use an existing VNS. Below we present the functional elements from which the architecture to support cross layer privacy was built on, along with an example of how this solution maps into an IdM practical use-case.

6.3.4.1 Functional elements

The functional elements that compose the privacy architecture are built around the principle that there must be a tight integration between the VNS solution and the IdM support functions.

This integration is done through the SAML Core which intercepts SAML operations that require network resources for the interaction between a user and a SP. Fig. 6.2 details the cross layer privacy architecture composed by three main components:

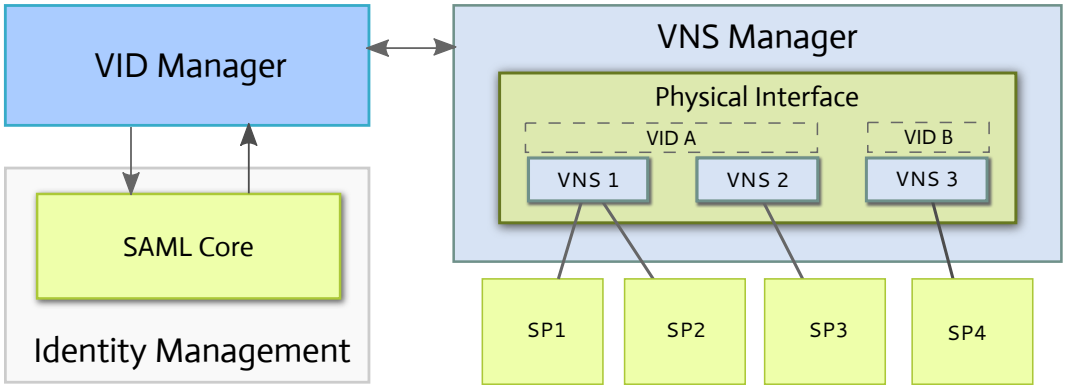


Figure 6.2: Cross layer pseudonym support architecture.

VID Manager (VID-M) The VID-M is the central element of this architecture and interacts with all the other elements. It is the interface point between the IdM layer and the VNS functionality and intercepts SAML pseudonym instantiations through the SAML Core. If network resources are needed, the VID-M triggers the policy mechanisms that will decide if, for a given pseudonym, it should use a new VNS or reuse an existing one.

VNS Manager (VNS-M) The VNS-M element is the primary point of interaction with the network stack. It is responsible for the proper usage of the VNS. The VNS-M enables

the creation of the necessary network pseudonyms along with the configuration of the underlying network stack. It also provides the virtual interfaces for each VNS, which simulates the existence of multiple devices in the user's terminal.

SAML Core (SAML-C) The SAML-C element is responsible for bridging the SAML operations performed on the IdM layer that involve to the creation or usage of VIDs, pseudonyms or assertions, to the VID-M. The SAML-C also creates the link to any identity management policies that can come from the IdM layer.

6.3.4.2 Functional Overview

Using the defined functional elements of the cross layer privacy architecture, we describe how they interact to provide the desired functionality from a privacy perspective. This operational overview describes using different VNS stacks in two scenarios that relate to user authentication, either IdAgg-initiated or SP-initiated [124]. We only present the case of web based services, thus assuming that the user already has a network connection but it is not consuming a service. Whenever a user attempts to access a service, the SP will initiate an authentication towards the IdAgg. The identity selection and the VNS interaction must be considered before contacting the SP. This interaction should be policy controlled and can be described in the following three step process:

1. When an EU requires access to a service, he must first select a VID to use, triggering the creation of a VNS.

2. In the case that the EU needs simultaneous access to a different SP with the same VID, it should use the IdAgg initiated authentication, given that he is already authenticated and possesses a SSO Token. In this step, the EU has two different options: 1) use the same VNS (one VNS per VID scenario), or 2) create a new VNS in order to prevent correlation of the same VID with the two SPs (one VNS per SP scenario).

3. Finally, if the EU requires access to a new SP with a different VID, step 1 is repeated and a new VNS is created.

The proposed three step process can be mapped to the interaction shown in Fig. 6.3, which exemplifies the authentication and subsequent consumption of a service. In this interaction the EU starts by selecting a VID, triggering the creation of a new VNS. Afterwards, the EU performs a Service Access operation to consume the service provided by SP 1. While consuming this service, the EU wishes to access SP 2. Due to the fact that he is already authenticated with the IdAgg, only a new VNS is created and used alongside the already existent SSO Token to obtain an Authentication Statement for SP 2. The end result is the instantiation of two VNSs by the user, one per service, making it impossible to link the SAML pseudonyms to the same EU terminal. This ensures user privacy even while consuming services with different VIDs on a single terminal.

6.4 Identity Driven Mobility Architecture

Using identity can increase the privacy of a network architecture, by providing the glue between application layer and the lower layers, providing a vertical approach to privacy. The

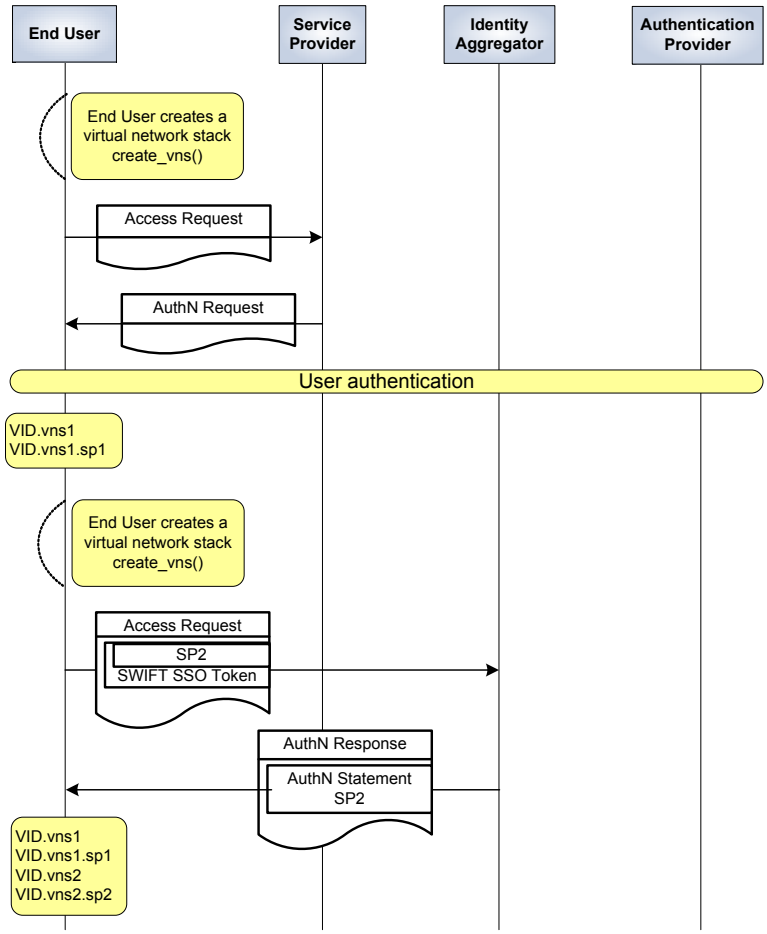


Figure 6.3: Accessing multiple services using Single-Sign-On.

vertical privacy integration reinforced the use of identity as a privacy tool. As we further outsource the control for privacy-aware mechanisms, we identify the feature of increasing privacy by using a common control layer for all network and user related aspects. This role can be assumed by the identity layer, becoming even more important as a privacy-providing mechanism.

By focusing on the feature of using a common privacy-aware control layer, we highlight the second architectural driver of outsourcing the control functions to identity, through a separation between control and execution. Here, we instantiate this design concept by separating the decision and execution of mobility mechanisms in a user-centric scenario.

Mobility can benefit directly from identity related information, as it requires many user-related inputs (e.g. user profile, preferred networks, experience preferences, etc.), either on end-user devices or in the network, to perform adequate and informed handover decisions. This relationship defines a context where it becomes interesting to explore the separation of control and execution.

With identity at the decision helm, a separation between decision and enforcement can lead to an environment that supports information demanding mechanisms, like mobility. The common privacy denominator drops to identity, where most information is now directly or

indirectly tied to the user identity, isolating IdM as an important tool for cross-layer privacy support in the network, and consequently, for mechanisms that require both privacy user-related information, like mobility. Having identity at the core of the mobility architecture enables a user centric approach to information and policy distribution capabilities: the identity framework naturally fits a control view of mobility, since it stores many user and network related policies, along with attributes.

When considering mobility management, the natural bias is towards the operational aspects. Attention is usually devoted to classical mobility solutions, such as MIPv6 [77] and derivate protocols [41, 135, 121], which have become the standard approach. These solutions focus almost entirely on the operational aspects of device mobility, and can be considered a tool that solves the data layer aspects of mobility. But they do not provide a strong framework for handling the informational and control aspects of mobility. Similarly, other mobility mechanisms, like SIP [125], used to implement terminal, service, session and even personal mobility, mostly target operational aspects as well. These examples indicate that there is no common approach that aggregates different mobility protocols, thus creating a gap on how to integrate them together both in control and operational views. Consequently, such different aspects, all operating on the network, have different control mechanisms and can unavoidably lead to more identifier correlation opportunities, further increasing the network privacy issues.

On an opposite approach, as the handover and mobility mechanisms increase in complexity, they have increasing information requirements, to enable more efficient technical solutions for the mobility operational aspects. This leads to an ecosystem where protocols require more information, but can simultaneously further jeopardize privacy. An example of such a protocol is IEEE 802.21 [68], which defines Media Independent Handover functions that enable a low level information distribution mechanism to assist the mobility process. By using the Media Independent Information Service, it provides a network oriented approach for information distribution, which enables mechanisms for information distribution for the handover process, but fails at providing a cross layer mechanism given its narrow applicability. This is another useful tool in mobility management, but not a vertical control layer. By manipulating 802.21 as a tool, the mobility extensions presented in [75] try to gather as much input as possible both on the network and on the terminal, covering QoS and network related user preferences to perform "smarter" mobility decisions. While interesting, and a step in the right direction, this solution falls short of the cross layer approach that is required for the Future Internet. Such solutions leverage the operational aspects of the different protocols, but do not make a strong argument for a common and vertical mobility and informational management layer.

Most of the aforementioned protocols fail to acknowledge that the important aspect is not moving devices, the strictly operational view, but instead it is the user needs, where privacy plays a major role. In this context, mobility is just an action, like any other, that requires a strong control layer, which does not focus on signal quality or similar metrics, but instead enables a cross layer approach that empowers the user and preserves privacy.

6.4.1 Identity Centric Mobility Management

As part of the architectural drivers, identity provides several advantages to user and service interaction. The concepts that stem from IdM framework, discussed in the previous section, revolve around enhanced security and privacy as part of the core system value, introducing a cross-layer architecture with a vertical notion of privacy, which goes beyond current IdM systems. In the presented instantiation example (Sec. 6.3), this was achieved by integrating

SAML, in a composed VID and VNS approach. Consequently, every network interaction is made privacy-aware and influenced by identity information, a key concept of the VID framework, where identities become the communication endpoints. In this ecosystem, the IdM system gains a new dimension by defining itself as a core technology, making it a suitable candidate for a stronger control layer.

When discussing IdM systems, it is worth nothing that we are considering the specific subset of entities that provide the basic IdM functions: strong authentication and between all the involved entities; secure attribute exchange and information storage; policy oriented mechanisms as privacy and decision enablers for the aforementioned functions. These features are present in the cross-layer IdM architecture presented in Sec. 6.3.1. In fact, we build the presented mobility enhancements around the same architecture which was presented for integration of IdM SAML pseudonyms and the VNS approach.

6.4.1.1 Identity Driven Mobility

Mobility is becoming less about maintaining sessions and more about enabling an improved user experience. This user centric characteristic turns mobility into an identity driven process. The paradigm is becoming about user centric information, and applying network functions towards user needs, rather than centering the mechanisms on the network itself. As such, a plethora of vectors contribute to the decision of where a terminal attaches and whether it is necessary to change point-of-attachment. To accompany the shift, it becomes clear that the mobility functions must be defined according to the notion of identity. It should be possible to formulate identity dependent mobility decisions. We rely on a rich information set and the application of mobility protocols as tools transformed into a cohesive architecture by the vertical IdM layer that controls user privacy.

To provide a rich information environment we use the attribute server as storage for mobility and user related information, rather than creating protocol dependent entities that store only a subset of information (e.g. Media Independent Information Service in 802.21 [68]). The dynamic nature of the information that contributes to the handover and mobility decisions implies the use of a dynamic structure (not the semi-static information types defined by current protocols), that adapts to shifting information requirements, important for mobility in the Future Internet. Also, the information can be shared across different protocols, rather than a protocol-specific silo.

Focusing on specific tools, no single protocol has proven superior to others, especially considering different layers. Therefore, it becomes apparent that the Future Internet will not be made of a single protocol, but of many performing individual functions. This creates the need for a coherent control layer that can naturally fall on the IdM plane. An advantage of using IdM as a primary element in the mobility management architecture is that, beyond a controlled authorization and access control, it defines a privacy-preserving environment, which is particularly interesting when defining complex mobility scenarios that involve multiple network providers with different identity-dependent attributes.

6.4.2 Splitting Mobility: Control and Action

We propose a definition of mobility as a two-step process, consisting first, on the decision, and second, on the actual process of triggering and executing mobility. While the decision process that will guide user or session movement is ultimately protocol independent, mobility

management is clearly linked with the protocols used at different levels in the network.

The entire mobility decision process forms the control layer, while the process of triggering mobility, i.e. determining the necessary actions and performing them, is the execution layer. This leads to the separation shown in Fig. 6.4, where the mobility process is divided in two different layers [123]: the Mobility Control Layer and the Mobility Execution Layer.

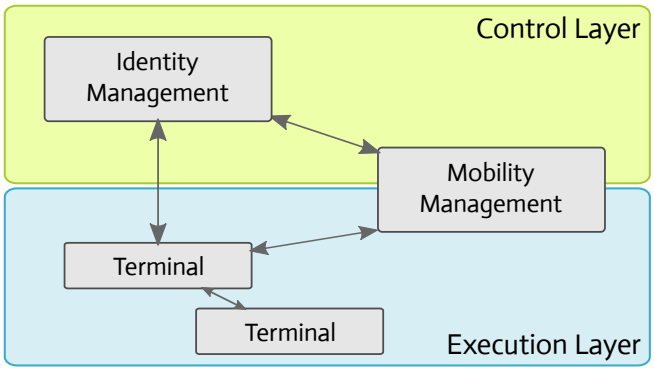


Figure 6.4: Control and Execution duality.

6.4.2.1 Mobility Control Layer

The first part of the proposed mobility management scheme is protocol independent, serving a dual purpose: it acts as the information repository for storage; and as the decision plane, based on cross-layer information focusing on mobility as a policy based mechanism. Acting as the information repository, the control layer is concerned with both static and dynamic information. Static information can be characterized as capabilities or features of user (e.g. preferences), network (e.g. contracted bandwidth) and devices (e.g. display size). On the other hand, dynamic information deals with the surrounding environment or conditions (e.g. network load or user location), as well as with policies that define the guidelines over existing capabilities and environment. Consequently, the identity control layer should act as the information hub for all data relating to mobility protocols and decisions.

The information collected from both network and user should then be coalesced into a control pattern that establishes what actions should be taken (policy trigger and driven). The resulting actions, which get pushed onto the execution layer, enable the mobility processes and are a direct consequence of the results of policy executions. In practice this process can outlined by a "control decision" taken in the IdM component, which is then is translated into an "action" of moving the session from one identity to another, therefore establishing a barrier between the control functions and the action of moving the session between identities.

6.4.2.2 Mobility Execution Layer

The high level mobility process decisions are outsourced to identity-aware components, through IdM, that are capable of outlining privacy-aware decisions, using user-centric information. Once a decision is sent from the control layer, it will need to be converted to protocol actions. Consistent with the two step process, the execution layer is able to determine what actually needs to be done in results of a decision, and how to realize those actions. This is achieved

by introducing two abstractions, that come together as shown in Fig. 6.5, which are protocol independent adaptation layer, and a protocol dependent action enabler:

Generic Execution Layer (GEL) : The GEL provides high level abstractions that can be used in mobility centric decisions, and breakdown generic identity driven decisions into protocol and layer oriented decisions.

Protocol Executors : The executions take the parsed decisions conveyed by the GEL and propagate them as protocol specific operations.

By taking advantage of semantics that generically describe the mobility process, the identity management components are able to apply policies and convey decisions to the mobility management components. But, the semantic should not be focused on specific mobility action that must be taken. Therefore, the GEL converts the abstract decisions into actions that fit the granularity of the mobility protocols, by clearly identifying the available identities, sessions and devices. The conversion process should also be both privacy and identity aware, in the sense that it is responsible for identifying the potential impacts and conflicts on the different virtual network stacks that might be affected by the decision. Similarly, it is also in a position to determine whether or not a specific mobility action can compromise the correlation of different identifiers. The GEL is a particular example of how a privacy-aware control layer can function both vertically, by integrating with the identity approach, as well as providing a privacy-oriented course of action for the different protocols that are executed on the same layer (horizontally).

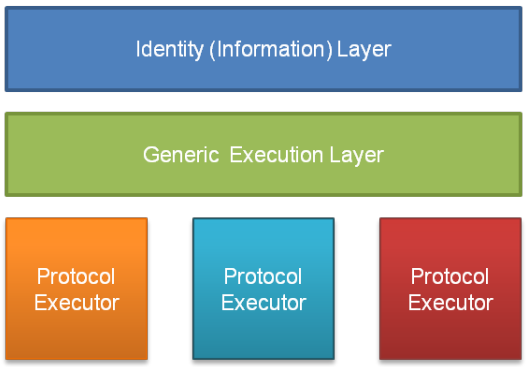


Figure 6.5: Generic mobility and identity abstractions.

One of the important features of this approach is that it enables mobility to be expressed independently of the protocol, as well as permitting a concurrent execution process on the by the different protocol executors, by directly triggering the required protocol mechanisms. From an abstract point of view, we must provide a common semantic approach to mobility, enabling the dissemination decisions, understood at the GEL, that translate into particular protocol actions.

6.4.3 Generic Mobility Architecture

Control and execution define the two concepts that can become the cornerstone of the proposed identity driven mobility architecture. But to constitute a feasible approach, they must

be translated into functionality supported by network entities that can be instantiated in the IdM architecture. Moreover, these abstract roles will allow modeling the mobility process encompassing multiple protocols and still remaining consistent through a common control.

To properly instantiate the generic entities, we assume that control is done through information and decisions, while the execution process is the enforcement of the decisions originating in the control layer. The specified requirements can also be observed in generic access control frameworks, from which we reuse the decision and enforcement concepts resulting in the three entities described below:

Mobility Information Point The entity that stores information that is pertinent to the mobility process. It stores domain related information depending on the level it operates. These entities can be distributed across the network, targeting several IdM specific functional boxes, as well as user devices, for user generated information. This can be the attribute server, the identity aggregator or a new (distributed) component.

Mobility Decision Point This is the entity that gathers both the static and dynamic information, executing the decision process controlling mobility. This entity can be distributed over the network where mobility information has relevance. A few examples are the access network, the local mobility domain, the global mobility domain, as well as the device for user centric information and mobility events.

Mobility Enforcement Point The mobility enforcement point should interact with the decision point, by sharing the abstract interface layer, so that it effectively bridges the decision into protocol operation. This should be mostly network entities, protocol specific, and the user devices, which will be part of the focus of the actual mobility process.

This approach enables us to model the mobility process, while reusing the entities and protocol which are already in place. Mobility is triggered by the Mobility Enforcement Point (MEP) using the GEL to collect and transform network and user event into the correct mobility semantics. A mobility decision request is then sent to Mobility Decision Point (MDP) that decides if a mobility action should be performed, based on information collected from the Mobility Information Point (MInP). If the result is to perform a mobility action, the inverse flow occurs: the MDP conveys a decision to the MEP in generic mobility terms, which is then properly translated by the GEL into semantics specific to the mobility protocols. When the translated decision reaches the MEP,it executes the various required steps to carry out the necessary changes in the network to complete the mobility task.

6.4.3.1 Identity Control Plane

Given the generic nature of the defined entities, they can be treated as roles assumed by already deployed entities. But, before instantiating these roles (which is done next in Sec. 6.4.4) into concrete entities, either from VID model or from the SAML-based cross-layer framework, it is necessary to further clarify the different required roles and purposes. Based on the two critical roles (decision and enforcement), we divide the control plane into an information management component and a policy execution component, assumed by each relevant entity on the network. Most of the mobility decision process can be offloaded to network entities, residing in the control plane with cross-layer information access. But, this is not the only

location where mobility decisions occur: when the local network or visited domain also controls the mobility within its networks, i.e. network based mobility management, they are also mobility decision points. This process is based on network conditions, which will not directly involve information related to the end-user. The result is that local mobility management entities become specialized MDPs that focus on specific aspects of the network. Furthermore, the terminal can also become an MDP, by apply user policies and inter-provider policies that cannot be handled by a single network or identity provider. The mobility control plane then results in set of decision and enforcement points scattered through different places: in the IdM system, in the mobility system and on the terminal (which is a part of both). This is exactly what is shown as in Fig. 6.6, where the defined elements interact through control primitives.

It should be noted that policy evaluation can be a daunting task, taking much longer than a few seconds. For particular cases, where movement decisions are time constrained, there should be a fallback mechanism or deadline definition for distributed decisions or policy executions (similar to real time operating systems) assuring a valid response in useful time.

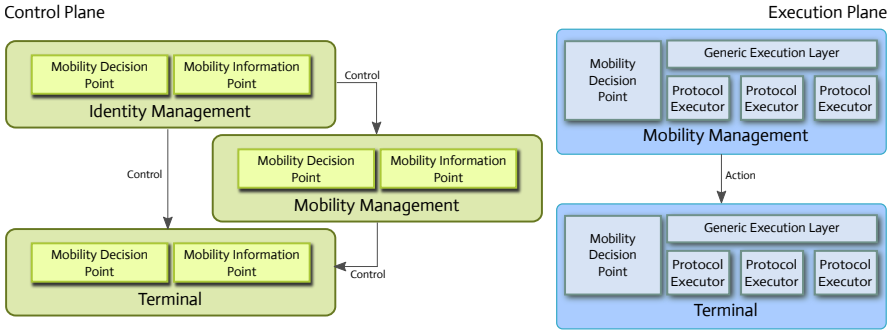


Figure 6.6: Control (left) and Execution (right) plane views.

6.4.3.2 Mobility Execution Plane

The execution plane involves different entities and implies the extended use of the GEL to transform control plane decisions into concrete protocol operations at the MEP. Fig. 6.6 presents the mobility management backend and the terminal interacting through the defined abstractions. There are enforcement points in all entities involved in the mobility process (i.e. signaling and operations), that is the mobility management system in usage, and the terminals on the user end. The GEL will then transform these commands into the appropriate protocol executors that pass them to the involved terminals.

The protocol executors, closely coupled to the MEP, provide the actual modular view that enables our proposed architecture to act as a control blanket over current and future mobility solutions. Therefore, each protocol executor should correspond to a different technology (e.g. MIPv6, HIP or SIP), with the benefits of reusing currently established protocols with similar conditions and performance. In fact, performance issues become orthogonal to the management system, given that the modular system cannot improve the performance of each individual building block.

6.4.4 Instantiation

The proposed architecture is capable of modeling several mobility scenarios, where different entities can play important roles. For this specific instantiation, concrete decisions on the mobility process must be made. Following the reasoning in Sec. 6.4.3, we assign the designated roles to the different entities of the cross-layer IdM architecture presented in Sec. 6.3.1, using the IdAgg, AuthNP and AttP, further clarifying the abstract definitions. As also mentioned in Sec. 6.3.1, the IdAgg is similar to the IdBroker in the VID framework, whereas the AuthNP maps to the IdManager, and finally the AttP, corresponds to the EPP holder. This mapping creates a common alignment over the different provided solutions, starting from the VID framework, and leading up to the different instantiations.

The central element in the scenario is the IdAgg, which acts as the MDP. It controls the mobility process by resorting to user identity (e.g. for retrieving preferences and contracts information), associated policies, and network status. This information is mostly retrieved from the Attribute Server (AttS), which assumes the role of MInP. The policy engine, which should be part of any IdM architecture, is divided between the MDP, which acts as the Policy Decision Point, and the MEP, which acts as the Policy Enforcement Point, thus disseminating and enforcing mobility related decisions and policies. For the mobility protocol, we propose including PMIPv6 [41] as the primary mapping protocol due to its network based approach for handling mobility, which allows outsourcing mobility decisions to the network. Therefore, both PMIPv6 entities, the Local Mobility Anchor (LMA) and the Mobile Access Gateway (MAG), compose a single protocol executor connected to the MEP.

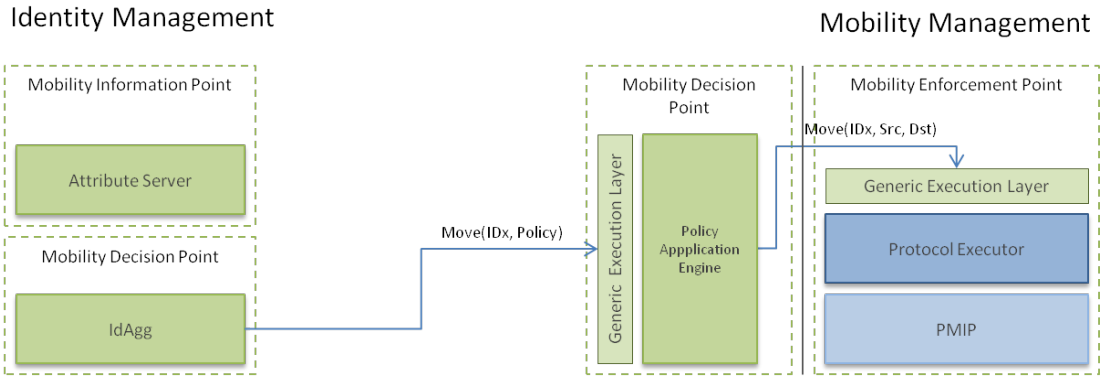


Figure 6.7: Example architecture instantiation.

When mobility is triggered by PMIPv6, the LMA uses the GEL to convert the event into the correct semantics and to request a mobility decision to the IdAgg (acting as MDP). The IdAgg collects the necessary information (e.g. user attributes, network information or policies) from the AttS (acting as MinP). If the result of the information decision is to perform a mobility action, e.g. change point of attachment, provider, or session condition, then the IdAgg will format the resulting decision and distribute the appropriate decisions to achieve the desired state in the network, which will be processed at the MEP by the policy engine. The resulting actions will be sent to the correct GEL, where they will be transformed from the mobility semantic decision into the correct protocol action and towards the LMA/MAG. This process summarizes the nature of the mobility process in the architecture.

6.4.5 Evolving Identity Paradigms

From the presented architecture, the most fundamental achievement is that a clear separation between control and execution seems feasible. This can become a tool for better privacy and identity integration, as well as a door for further protocol optimizations that do not disrupt current network functionality. With this separation it becomes possible to introduce privacy as a design feature of different network architectures by using an identity-oriented control layer.

One of the key benefits of this separation is the common mobility semantic across different elements, taking full advantage of the strong policy oriented mechanisms used through the network. And as such, policy based mechanisms can be made user or identity-centric, moving control to a rich information layer. In fact, the movement primitives realized by the execution layer are a consequence of distributed policy application that can follow a hierarchical policy evaluation approach, adding flexibility to mobility management. However, the presented architecture should take into account the possible overhead of distributed functions when considering mobility (time) requirements: even though individual protocol performance should remain unchanged (because protocols are not modified), the overall performance also depends on the different modules present on the execution layer.

Regardless of the potential benefits, the separation in the mobility framework shows that there is margin for progression in the mobility aspects. Current semantics for mobility become part of a new framework that can support existing communication paradigms, and simultaneously allow future iterations of new or existing protocols, given the modular approach.

The proposed separation introduces new networking paradigms, due to the fact that not only can the user-privacy and information controls benefit from this approach, but also that a common control layer (in this case identity-driven) can support cross layer (mobility) concepts. It also indicates that, by addressing inter-dependencies between layers, as done here for mobility purposes, a two-fold scenario appears: i) privacy can be independently improved on each layer, trying to mitigate individual threats and perils; ii) replacements for recurring concepts, such as current mobility schemes, can be used transparently due to the abstractions put forth by separation layer and constructing of common paradigms.

6.5 Instantiating Privacy and Mobility through Identity

One of the proposed architectural drivers focused on achieving privacy through identity as an operational driver. The value of such a proposition stemmed from the fact that using identity as a privacy tool can be useful, not only on the more control related aspects of the network, but also on the operational side of protocol instantiations. It is possible to explore an identity approach that promotes a separation between functional layers, to promote a privacy-oriented environment less prone to identifier correlation, where it is possible to defined new solutions for privacy threats on the different network layers.

In this context, identity makes sense as a tool to achieve privacy, since it can provide a separation between different layers, as well as a relationship towards the user, establishing the opportunity to reuse the vertical privacy framework. An example that introduces a separation between layers is the Host Identity Protocol (HIP). HIP introduces an explicit differentiation between locator and identifier, by splitting the network and transport layer through the introduction of a new namespace that assigns identities to hosts. The Host Identity (HI) is the public part of an asymmetric key pair. The main idea behind HIP is to decouple the dual

functionality provided by IP addresses, which results in seamless mobility support through IP address renumbering supported by strong security features.

This is a direct application of identity as an operational procedure that can be explored with a privacy intent. The identity metaphor serves to create a layer separation and a new endpoint for the network communication. The key concept to extract is that, when a separation is created between layers, it is possible to reduce the correlation options, since it becomes possible to employ the proposed solutions, such as pseudonymity, to create a privacy-preserving environment. From the PRIVED perspective, the upper layer identifiers, which are above the separation layer, do not need to be bound to a single lower layer identifier. Consequently, it is possible to create Information sets that are not correlatable through network identifiers, since it is possible to use different ones for corresponding upper layer identifiers. In this separation, identity becomes the tool that enables the paradigm to which the communication becomes linked, making it possible to apply a VID approach supported through pseudonyms. Furthermore, from this separation, it is possible to address individual network problems on each layer, since they can be decoupled from other protocols.

In this section we use the split proposed by HIP, along with identity as the operational driver, to introduce new ways to tackle mobility and privacy issues. First, we solve specific network issues such as location tracking of users, by concealing the use of IP addresses. This can be done by using the HIP identity mechanisms that stem from the locator/identifier split, to outsource the locator to identifier conversion into the network, making sure that the user's location is never revealed within topological boundary. This provides location privacy to end-nodes. Second, we use the identity properties of HIP to establish a relationship to the IdM environment. This creates a scenario that can take advantage of Virtual Identities and Virtual Network Stacks to provide a vertical privacy solution, with horizontal ramifications.

6.5.1 Revisiting Network Privacy Issues: the HIP use-case

As we have seen, there are two actions associated with the IP address: one is identifying the node and the second is providing its topological location. If this functionality is divided in two different roles, with a protocol like HIP, we can provide orthogonal solutions that can be complemented to solve these threats without requiring complicated solutions. We define a network framework that uses HIP as the base protocol and is able to conceal the IP addresses of communicating HIP nodes. From a privacy perspective, we can take advantage of the discussed separation to reduce the correlation potential of identifiers, but also, it is possible to address specific location privacy at the network layer. Using an architectural approach based on protected geographical areas (IP network domains), beyond which the locators have no meaning, it is possible to design a mechanism that provides location privacy to end-users. This is done by taking advantage of the locator/identifier separation provided by HIP, that currently has few privacy considerations. Our proposal relies on an architectural solution for location privacy without requiring modifications on the core network, while still supporting mobility.

The current HIP architecture does not take into account location privacy issues, since it requires a node to send its locator to every peer, similar to the Binding Update messages exchanged in MIPv6 [77]. In fact, HIP ultimately suffers from the same location privacy issues as MIPv6 described in [44, 60], discussed in Sec. 2.5, by exchanging locator parameters in the HIP Base Exchange (BE), the initial handshake. This also occurs on mobility events, where update messages with the locator parameter must be sent. HIP promotes an end-to-

end paradigm, where both Initiator and Responder learn each other's current IP address once the BE is completed. But, in an architecture which supports location privacy, hosts should never be able to map the identifier to the real locator of the node.

There have been previous proposals to introduce network elements that shield a HIP node's location [85]. We aim to extend this, and provide new functional units that enhance the protocol operation and provide location privacy to the nodes in the network [100].

6.5.2 HIP Location Privacy Architecture

We can take the separation aspects of HIP to provide an environment that has a strong relationship with identity (and the VID layer), but that also has an impact on network related privacy. As suggested in [85], location privacy is provided by delegating the HIT to IP resolution to a network entity called the Rendezvous Agent (RVA). Moving the resolution upwards in the network topology, from the HIP Mobile Node (HMN) to the RVA has the added benefit that locators do not need to be disclosed in the Access Network. The core feature of our solution is the concept of RVA protected areas, which are access networks where global locators are either concealed or not used at all. Instead, HITs or local addresses are used to identify the traffic path. The RVA is also responsible for local mobility, i.e. under its protected area. Rather than defining a specific transport layer for our approach, we define a set of basic requirements which must be met for the protocol to function properly. The only assumption made is that the core network is IP based. We do not specify any particular technology under RVA protected areas. In fact, we consider possible instantiations based on direct IPv6 address translations, tunnels or semantical adaptations (replacing IPv6 addresses with HITs). In Sec. 6.5.3 an IPv6 based solution is described.

A simple example of the proposed topology is shown in Fig. 6.8, consisting of two RVA protected areas connected to the Internet. An RVA protected area is composed by multiple ARs which are directly connected to an RVA. There are no assumptions made about the size and number of necessary RVA protected areas, although it is reasonable to think that an RVA can cover a large number of access routers. A wider coverage area, geographical or topological, limits the amount of location information revealed to an external eavesdropper. The RVS and DNS servers are located in the core. The AR and the RVA are functional entities, thus they can also be collocated in the same machine, but at the expense of some limitation on location privacy.

Depending on the chosen solution for routing in the access networks, already existing HIP elements may require modifications, since hosts and routers depend on the protocol used in the access network. If some form of identity based routing is used, then the amount of information to be kept at each node (e.g. AR keeps HIT based neighbor tables) is larger. If the access network remains IPv6 based, then no modification is required other than enhancing the neighbor advertisement protocol. Another element of the HIP architecture which requires minor modifications is the RVS. The RVS should be capable of performing a double resolution: translating a received HIT of a host into the HIT of its designated RVA, and resolving the corresponding RVA address. However, to understand the architecture we must further understand the RVA, along with the extensions necessary to the BE performed between nodes.

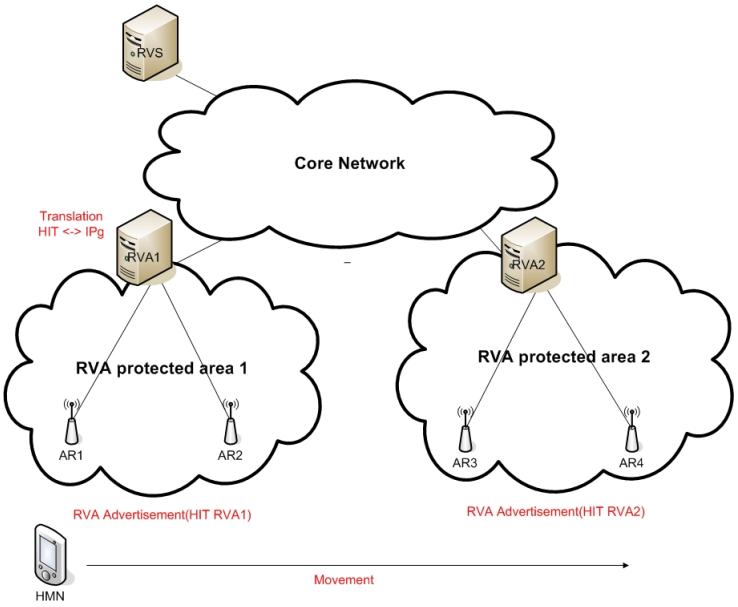


Figure 6.8: Basic architecture topology example

6.5.2.1 Rendezvous Agent

The RVA is an enhanced Rendezvous Server (RVS) which performs the IP-HIT or IP-IP address translations. This functionality split provides location privacy to the HMNs behind it. The mechanism consists of re-addressing packets flowing from and to the core network. To forward packets to a host outside the RVA protected area, the RVA addresses a globally routable IPv6 address previously assigned by another RVA to the destination host. When an RVA receives packets from the outside network to a host belonging to its RVA protected area, it readdresses them to HITs, or local addresses, and forwards the packet to the destination. Note that the RVA is the entity which assigns globally routable IP addresses to the hosts under it, and the only one capable of mapping HIT, or local address, to global addresses. The RVA is capable of forwarding packets based on HITs through maintaining a mapping for every HMN in the protected area to its point of attachment, which is the AR. The RVA is responsible for handling mobility for the nodes in the protected area, raising the possibility that the RVA might have to signal other RVAs or HIP nodes, on behalf of the HMNs, for location updates.

6.5.2.2 Base Exchange Extensions

When an HMN first arrives to a protected area, it registers with the responsible RVA. The HIT_{rva} is retrieved from the Advertisement messages sent by the AR. The registration takes the shape of a BE with the RVA (Fig. 6.9) using registration extension [83]. Once this phase is over, the RVA assigns a global IPv6 address (IPg) that is used for the registering node in the core network. The IPg should be generated from a pool of addresses assigned to the RVA. In case we are using identity based routing, during this phase the RVA learns the HIT-AR mapping necessary for packet forwarding. Once the BE with the RVA described above is completed, the HMN has to register with its RVS or update it. Afterwards, if registration at the RVS is needed, a BE is performed, which informs the RVS of which RVA is being used,

by the inclusion of the area RVA identifier in an RVA parameter in the I2 packet.

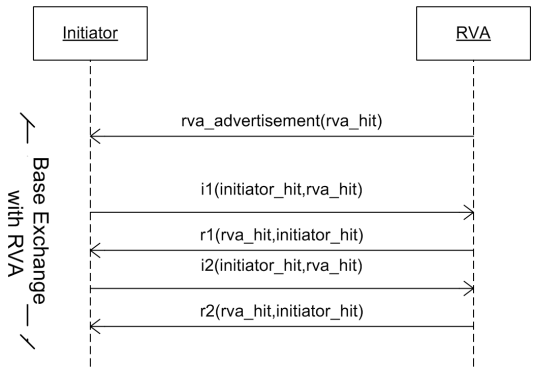


Figure 6.9: Base exchange with Rendezvous Agent

The HIP base exchange between an initiator and a responder remains mostly unchanged, with minor differences at the network layer: because the BE packets now travel through each node's RVA, all the traffic is now proxied.

6.5.2.3 Mobility Support

In our architecture we can define two types of handover: Intra and Inter RVA. Both these procedures are triggered by the advertisement system when a new access router or a new RVA identifier is detected. Moving within a protected area constitutes an intra RVA handover, and requires only updating the RVA binding. The handover is transparent to all communicating peers. If the HIP node changes protected area, then an inter RVA handover occurs. In this scenario the host has to register with the new RVA (performing a Base Exchange), and updating its RVS entry with the new responsible identity. The new RVA should also inform the old RVA of the handover, so that packets are forwarded correctly and the connection is not severed. When the new RVA receives a forwarded packet from another RVA, it updates the location to the forwarding RVA. Forwarded packets need to be differentiated from the normal traffic, allowing a RVA to decide whether mobility updates are needed or not.

As stated before, our framework also requires RVA-to-RVA communication for location updates. The RVA-to-RVA communication should be preceded by a HIP base exchange, allowing secure communication and, more importantly, authentication. But depending on the scenario, the trust relation between the RVAs may be different. For instance, in a network operator scenario, all RVAs may be certified by a common Certification Authority, allowing only trusted RVAs to signal each other. A more flexible solution resides on the HMN providing a certificate to the RVA during the registration process, thus enabling them to prove to each other that they are acting on behalf of the HMN.

6.5.3 Instantiation and Analysis

The framework definition, as described in [100, 102], does not make assumptions on packet transport mechanisms within the RVA protected area, although IPv6 is assumed in the core network. The most feasible approach is to also use IPv6 within the areas where locators have meaning. These areas, named protected areas, define the geographical space beyond which locators are replaced with global identifiers, as discussed before, and also require topologically

bound identifiers. This means that communication inside the protected area is done with IPv6 addresses with a global format, but with local scope, translated by the RVA. The main advantage of this solution is that it requires no changes to routing mechanisms within the access network. With the IPv6 access network, the deployment of the RVA advertisement system consists on enhancing the Router Advertisement [144] messages to carry HIP parameters as options. Just like a HIP parameter, a neighbor discovery option has a type-length-value format. The new HIP parameter (RVA_INFO) is an option that advertises the HIT_{rva}, along with the advertisement lifetime. Through the advertisement mechanism, the HMN can detect different protected areas. Address configuration and mobility detection should be done according to [144], with the extension of RVA detection by HIT_{rva} announcements. The registration with the RVA is performed according to the previous section, followed by the assignment of a global IPv6 address to the HMN. Later, the HMN registers the acquired HIT_{rva} with the RVS, for correct I1 packet forwarding. When movement is detected, the HMN updates the binding with the RVA, in case of intra RVA mobility, or registers with the new RVA, in case of inter RVA mobility, updating the RVS afterwards.

6.5.3.1 Prototype Implementation

To validate our approach through practical testing, we implemented the global IPv6 assignment and packet translation mechanisms, discussed above. Registration procedures are similar to those performed when registering with the RVS, and therefore are of secondary importance to a prototype. The implementation was developed under Linux, kernel 2.6.15, using as basis the HIPL implementation provided by the InfraHIP project [48]. It consists of a manually triggered registration process, with an IPv6 global address being assigned upon completion for the registering local address. After this, the RVA performs the necessary translations. The work is performed by a module which stores the registered addresses, under a hash table, resorting to Linux IPTables for packet capture. In Fig. 6.10 we depict the scenario used for the performance evaluation. This scenario consists of two access networks, both served by the same RVA. This RVA is capable of handling multiple protected areas. After each node registers, the RVA generates addresses from a pool of available prefixes. For simplicity, we use only one prefix for both areas. The testing procedure is composed of two phases: first we show the results of a base exchange between the two endpoints and the assigned addresses in the several areas and nodes. We then perform measurements on the responsiveness of the system, using Round Trip Time (RTT) values of ICMPv6 echo request and response packets, and on how our protocol impacts the overall system performance, by measuring the throughput of TCP when our scheme is in place. This prototype aims at showing the flow of addresses through networks between communicating peers, and also how the translation overhead impacts the implementation, allowing us to generalize to protocol operation delays.

6.5.3.2 Location Leakage Analysis

Since both endpoints are assigned and use global IPv6 addresses for communicating with each other, they cannot determine the actual address, and consequently location, within the RVA protected area. In Table 6.1 we represent the information gathered at each point in our reference scenario. As an example, we can see that Node A is sending to the global address of B, 2001::6ada:1e65:93f3:ff00, but is unaware of his local address, 3ffa::1, where the packets actually get delivered.

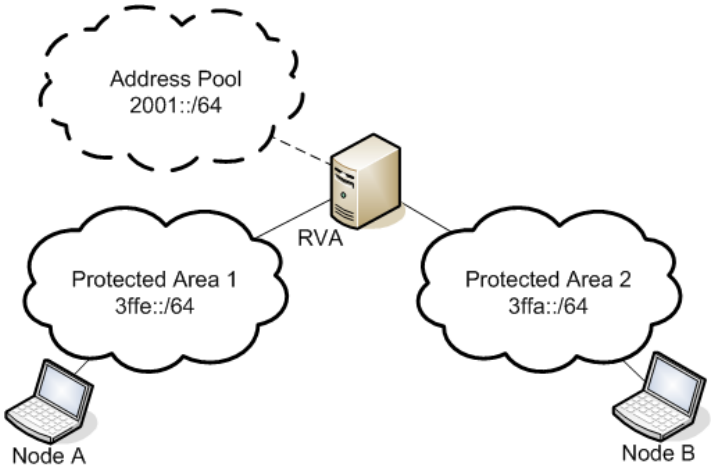


Figure 6.10: Implementation testing scenario.

Networks	Node A	Node B
Area 1	3ffe::1	2001::6ada:1e65:93f3:ff00
Core	2001::ded8:ce89:6390:eb00	2001::6ada:1e65:93f3:ff00
Area 2	2001::ded8:ce89:6390:eb00	3ffa::1

Table 6.1: Summary of the addresses observed on each network segment

Protecting the identity of the nodes is not the primary concerned of the proposed scheme, which targets specific layered issues. An attacker can still identify the HMN and its peers, but not their network topological position. To actually protect the user identity associated with the HIP identity, increasing the privacy of the end user, requires establishing a relationship towards the VID and VNS approaches, discussed in the following sections (Sec. 6.5.4).

The RVA is a point of information gathering for the network and, if compromised, reveal the identity and location of registered nodes. Another mechanism that provides information on the location of the node is hop count. Even if the node is behind an RVA, the hop count, together with the topology of the underlying network, can reveal information on the whereabouts of the node. One mechanism to thwart such attempts is to keep the hop count value between the RVA and the node to 1. This can be achieved by tunneling.

6.5.3.3 Performance Evaluation

From the prototype we analyze two different performance metrics: the first is the RTT of ICMPv6 Echo Request and Response packets, to evaluate the real impact of packet translation. In fact, we measure the implementation delay to infer the real protocol impact, although aware that this is not an optimized implementation. We use the RTT to measure the delay introduced by all the necessary translations. The other proposed measurement is the TCP throughput that shows the impact on bandwidth caused by the translation delay. In both cases, we compare the performance with RVA intervention to normal HIP operation.

Regarding the RTT, we perform 100 measurements, and present the averages for each run in Fig. 6.11, matched against a plain HIP scenario, under the same conditions. As we can see the RVA introduces a slight delay in packet delivery. Using our protocol, the average

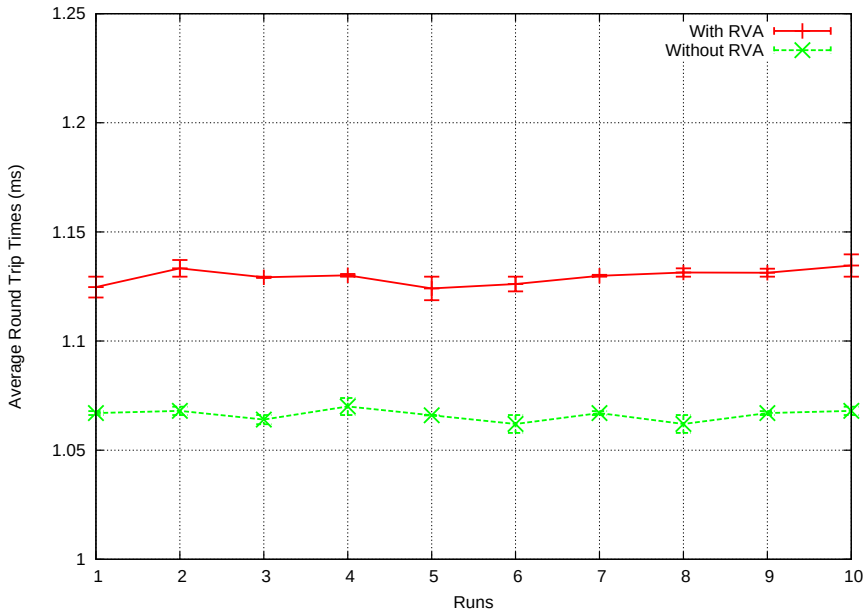


Figure 6.11: Round Trip Time (RTT) Impact

value for the RTT is of $1.130 \pm 0.003ms$, whereas, without RVA, it is of $1.067 \pm 0.002ms$. However, the difference is sufficiently small that we can discard the impact on network traffic. We measure the maximum available bandwidth for a TCP connection in two scenarios with and without RVA processing. Each run consists of starting a TCP traffic generator for 30 seconds, and obtaining the bandwidth of that flow. The results of an average of 10 runs, show that the total bandwidth used with normal HIP is of $6.43 \pm 0.03Mbit$, and for HIP with RVA translations is of $6.44 \pm 0.07Mbit$. The similarity in these values allow us to neglect the impact on network throughput performance, meaning that the added delay has very little impact on the TCP throughput.

6.5.4 Integrating HIP with IdM

At a first glance, user and host identities, each using distinct identifiers, are unrelated. But, user and host identities cannot be considered independent as they can become part of the same Information set. In fact, if a user is represented by a VID, then it is only natural that the host identifier becomes part of a VNS associated with that VID, instantiated to consume a service. This happens because the HI identifies the host, which belongs to the information virtualized by the VNs approach. This indicates that an integrated view on identities across the user and host level is required, promoting the vertical privacy integration, taking advantage of the proposed privacy concepts that focus on identity as an operational tool.

Host identities are coupled to user identities, making the integration of both a necessity achieved through an architecture that considers HIP as a network level protocol capable of delivering mobility and multihoming heavily based on identity concepts. HIP and user IdM are very different when considering the layer on which they operate, the identifiers employed and the problems they solve. This creates several challenges, where an integrated solution can leverage properties of the different solutions, tackling key issues such as security of IdM transaction through HIP, a distributed trust mechanism for HIP through IdM, and the integration

of both technologies for enhanced user mobility and cross-layer Attribute Exchange. However, the integration of HIP and IdM requires addressing two problems: first, the integration requires appropriate security mechanisms that allow establishing trust information, based on the presented Host Identities and the IdM systems; second, it is necessary to integrate the different structures and identifiers stemming from the HIP and IdM namespaces.

6.5.4.1 HIP IdM enabled Architecture

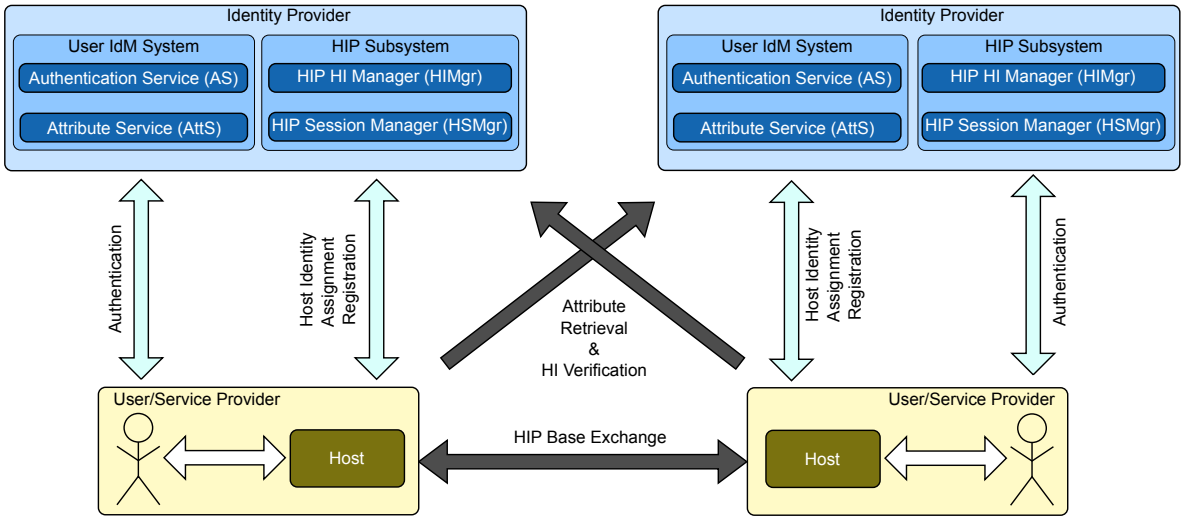


Figure 6.12: Architecture and Interaction among Components

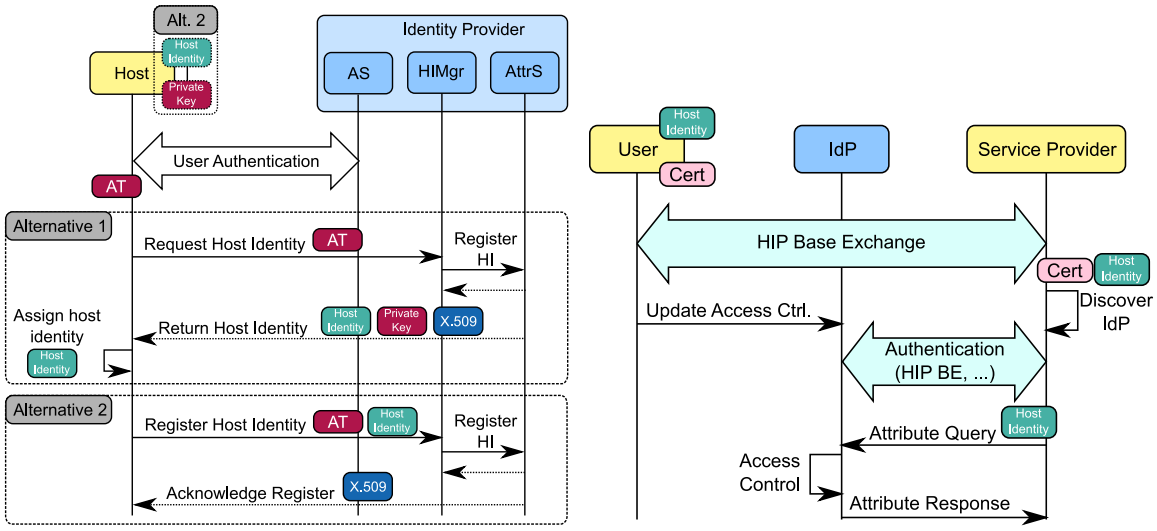
The proposed integration presents architectural challenges that must be addressed within the architecture. The key operations that must be outlined fall under the category of Host Identity Management, carried out by the IdP. Beyond this, the IdP must also provide attribute retrieval mechanism using HIT, and verifying Host Identities, belonging to different hosts. This makes the IdP the main entity that allows the interconnection of both host and user identities. Fig. 6.12 illustrates the supported IdP services and interactions between users and IdPs. The user IdM services, based on an extended SAML architecture, allow the integration of the host identity namespace. The main user-centric IdM services are similar to those highlighted in Sec. 6.3, covering authentication and attributes, which are described here in terms of services. The Authentication Service (AS) provides user authentication by generating a user-specific Authentication Token (AT) that can be used to consume other services i.e. SSO. The Attribute Service (AttS) manages user and host related attributes, indexed by either HI/HIT or user identifiers.

The most important part of Host Identity integration is a two part addition to the IdP, that enables creating or assigning Host Identities. Additionally, it provides session management to allow information retrieval about ongoing HIP sessions. The HIP subsystem at the IdP is composed by the HI Manager (HIMgr), which provides the HI assignment (creation and host assignment) and HI registration function (for self-assigned HI). The second part is the HIP Session Manager (HSMgr), which keeps track of ongoing HIP sessions relating them to access control.

6.5.4.2 Host Identity Management

In the proposed architecture, the IdP is responsible for HI management, which refers to HI generation at the IdP. Alternatively, this role can be replaced by user-generated HIs that are later registered at the IdP. In both cases, the user first authenticates against the AS to obtain an AT for further interaction with the HIP subsystem as illustrated in Fig. 6.13a. In case of an existing HI, the authentication process can use HIP to increase the registration security on top of a secure channel.

When the user requests an HI, the IdP assigns one based on the provisioning of the obtained AT, which can be newly created or already existing, depending on the AT. The creation process includes generating a public/private key pair, whereas the assignment is just transmitting the keys towards the user. This transmission triggers the establishment of a relationship between Host and User Identity by registering the HI and HIT as attributes at the AttS. HI verification can be done through X.509 certificates created by the HIMgr, and provided to the hosts. When the HI is assigned by the IdP, it can always assign the same HI independent of the actually used device, creating a long term relationship between the HI and the user identity. Moreover, the IdP has to store the private key and thus act as a key escrow for HIP, in scenarios where this feature is required (e.g. lawful interception through a trusted entity).



(a) Host Identity IdP Registration

(b) Extended Base Exchange and attribute retrieval

Figure 6.13: Host Identity Management Process

When key escrowing at the IdP is not desired, the user can generate its own HI and register the public part at the IdP, resulting in increased network privacy. The authenticated user contacts the IdP, providing an HI to the HIMgr that properly calculates the HIT, registering both at the AttS. This enables the mapping between identities, and provides the knowledge for X.509 certificate generation.

The HI verification creates the new role of Public Key Infrastructure for the IdP, delegating trust to the IdM plane. Whenever these verifications are needed, the certificates can be verified by contacting IdP, either directly or through federation mechanisms. A host can

provide its certificate to a peer during the base exchange, as explained in [113], which defines a type-length-value field "CERT" for X.509 certificate transport [64]. The inclusion of an IdP reference in the certificate also allows it to function as a pointer for lookup services, allowing attribute retrieval. Based on this reference and on the received HI/HIT, it is possible to use SAML for attribute retrieval, as exemplified in Fig. 6.13b.

This instantiation presents an approach of how to explore the horizontal privacy concepts. Not only does it provide the mechanisms to increase localized privacy threats in the network (e.g. location privacy) but, through the use of identity, it provides a seamless integration with the VID model. In this particular case, the IdP is part of the IdM system. The IdP is the aggregate set of the IdManager and the IdBroker, which in terms of the cross-layer IdM architecture, it represents the IdAgg and the AuthNP. Therefore, there is a common instantiation trend over the different proposed functions, as well as a clear space under which to introduce user Virtual Identities, as well as the Virtual Network Stacks, supporting the pseudonymity approach.

6.6 Intrinsic Privacy-aware Identification

One way of instantiating a privacy-aware network architecture is to create a close relationship between the network and the user, which has been done so far by introducing identity as a privacy control mechanism. Based on the previous instantiation examples, resorting to identity as a path towards privacy also works on an individual layer approach.

By including identity information in the different network operations, it is possible to create a privacy-aware environment, where the user identity, and consequently the VID layer, becomes the driver for the different network interactions in NGN. This has been presented in Sec. 6.2 as the final architectural driver.

To truly instantiate identity, as means to provide cross-layer privacy support, it requires that the different interactions in the network become user-centric, resorting to the common control layer. This requirement includes not only the protocols that explicitly convey an identity concept, such as HIP (explored in the previous section), but also legacy protocols working on the network today. This allows us not only to instantiate pseudonyms as part of the VNS approach, but also to provide a close relationship to the vertical privacy control layer. On one hand, this approach can simplify the use of cross-layer (user) information, resulting in more available information for protocol usage and simultaneously providing simpler and more effective control mechanisms for the information set or VID, enhancing the privacy aspects. On the other hand, by relating (most of) the network operations to user identity, it is possible to create a tighter privacy control on the information that is exchanged with every transaction performed on the network. This approach reinforces the use of identity as a tool for privacy-aware mechanisms at the various layers in the network stack.

To provide this relationship, it is necessary to include the different protocols in the network, even if they did not previously have an identity connection. Even though the lower network layers are important, they have been mostly neglected in terms of identity concepts beyond using the vertical layer as a control source. User related information can be explored by lower layers, not only for control purposes, but also for user relationship. It is possible to devise a solution where identity information becomes easily accessible by all layers in the OSI model, enhancing network protocols and providing better cross-layer integration, starting with AAA mechanisms. This dependency illustrates identity as the interaction driver

in NGN, where user information and identity information become fundamental aspects of network operation, defining a paradigm that supports privacy-aware networks.

We propose to integrate privacy considerations in the different network protocols and layers by creating an identification system that supports intrinsic identity references. This is feasible through the VIDID structure, presented in Sec. 4.3.1, enabling identity information to be referenced from any protocol when properly enhanced with resolution mechanisms, promoting the use of identity in lower layers, such as transport and network. The protocol operations that can benefit the most from this interconnection are those which require more user related information to operate, like mobility protocols in NGN scenarios. Through an identity-dependent design, a new privacy paradigm can be created by embedding identity related identifiers into existing protocols. However, this integration can become complex when handling a large array of protocols, of which several are at the core of 4G networks and have been mostly indifferent to the privacy and identity aspects. NGN encompasses services ranging from VoIP to IPTV, and span across many different access technologies, such as WiFi, WiMax or even UMTS. Such scenarios are very volatile in terms of user mobility, making session continuity an important issue, and increasingly require more user and network information. These requirements include QoS and AAA information, which brings added complexity to the environment that must be made identity aware.

As part of the analysis performed in Sec. 2.4.2 to determine the existence of identity related concepts in the network, we concluded that protocols have used identity in different formats, but none of the presented protocols can express a true relationship towards identity or cannot easily integrate with legacy (identity unaware) protocols. A considerable improvement can be made through the VID approach, along with VNS support, bringing the control of identity information into the network. But even with the proposed solutions to support a different namespace or identity integration mechanism, none is able to provide a unified namespace with a cross-layer design centered on identity and supporting mobile environments. We propose an identity relationship by adding an Identity Pointer[3] to the generated pseudonyms, enabling a strong connection between the VID solution and the different network protocols through their identifiers. The proposed identifier needs to appear on different layers, enabling a clean identity integration scheme, without requiring modifications to the entire network stack and protocols. This will enable the support of a distributed database model, indexed by common identifiers enabling a tighter and simpler privacy control on the different aspects of the network, such as resource authorization, QoS information, as well as new mobility paradigms.

The support for such identifiers requires an enhanced architecture, for both terminal and the network, to support the identity approach: regardless the addresses used, the same identity material is always provided, greatly simplifying the network processes such as accounting, authorization, QoS reservation. Moreover, using the same identifiers across different attachment points provides a consistent mobility approach across the network, enabling a new identity-dependent mobility paradigm.

[3]The ID-Pointer is a synonym for the VIDID proposed in the VID framework. In this section we propose using a more generic term, ID-Pointer, to enable a broader application scope that is not necessarily constrained to the VID framework, but can work as a generic identification structure for any solution that requires establishing a relationship to an Identity oriented namespace.

6.6.1 Supporting Privacy through an Identity Oriented Architecture

Integrating identity in the different levels of the network requires a well defined architecture. This introduces a necessity to rethink the current the terminal architecture, along with the need for abstractions that transform NGN operations (e.g. mobility) into privacy-aware paradigms controlled to the vertical layer resorting to user information.

The architecture upon which we propose these new abstractions is shown in Fig. 6.14, and was first introduced in Sec. 1.2.1 where we have addressed typical 4G heterogeneous network scenarios [1, 3, 4, 73]. Recovering this characterization is important because the architecture summarizes several important features to the network mechanisms that we review, especially considering controlled resources and authorization: Bandwidth Brokers control network resources and their utilization; Authorization, Authentication, Accounting, Auditing and Charging (A4C) servers take care of terminal authentication and authorization; and SIP proxies handle SIP based applications, especially VoIP. Also, a service pool is available either local, or remote e.g. Internet. In this 4G network representation we also include and Identity Manager to allow for the support of identity based access to the services and mobility.

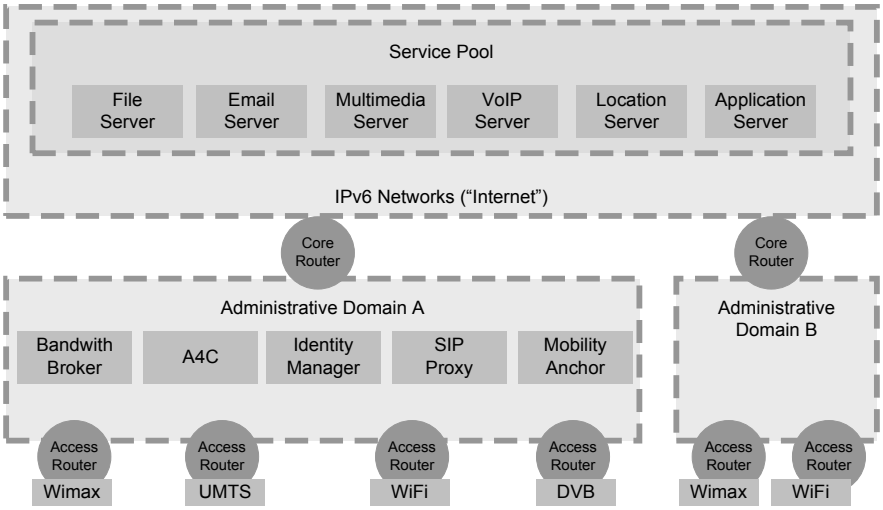


Figure 6.14: Network Architecture Model.

The presented functional units have distinct planes of action that culminate in different namespaces for each independent area. But most of them, including services and applications, are to some extent user oriented, either for control, management or measurement. This means that there is relevant user information, potentially containing private information, scattered in the network. Through this property, it becomes obvious that the user must regain control of the information that relates to him, and is crucial to retaining his privacy. This can only be achieved by linking this information back to the user, through identity concepts, and adequately handling it in a privacy-aware location, such as a vertical privacy layer that adheres to the VID concepts and respects the PRIVED model.

The information distribution leads to different databases over multiple entities, with information pertaining to the user. This paradigm can create information conflicts, than only increase when outlined in the context of mobile environments, such as in the targeted 4G scenarios, where terminal roams across multi-operator heterogeneous networks with controlled

access and resources.

In this scenario, the user has several hurdles to overcome in cooperation with the network, and the first is at the access point, where authentication is required at an A4C server, eventually provided by the Protocol for Carrying Authentication for Network Access (PANA) [74]. With the generated credentials, the A4C creates state at a MIPv6 HA, introducing a binding between authentication material and address information. By now several namespaces and identifiers are in play: link layer access between the user terminal and AP; the PANA protocol between client, AP and A4C; Diameter (or similar) between the A4C and the HA, involving Network Layer identifiers. Afterwards, the terminal will register its Network layer addresses with the A4C and the HA, completing the interactions on this plane. But, there are still the interactions with the Bandwidth Broker for QoS purposes. This is established by the terminal, AP or Access Router with the Bandwidth Broker, coordinated with the network entities, using MAC addresses, local IPv6 addresses, CoA or HoA.

Going into NGN environments that support multihoming and heterogeneity, the previous mentioned bindings are multiplied by each active interface, yielding a multidimensional control and data plane where several and different identifiers are used, further accentuated by the VNS approach which creates virtual devices to support different network stacks. Using identity as a driving concept can help to simplify the control required for each of these mappings, that carries their own identifiers and protocols. Failing to provide a common control point, e.g. VID, can lead to unnecessary and costly mappings, followed by unnecessary database replication, where several entries exist across different planes that in fact deal with the same entity. In fact, this is the reasoning behind the privacy-oriented vertical layer presented in Chap. 4.

In this complex environment, we can rely on identity to provide proper namespace integration that creates a viable approach for user privacy: the namespace provides a rich set of information that directly relates to the user, enabling abstractions that rely on the user-centric paradigms instead of network devices and stack elements

6.6.1.1 Identity Referral

To instantiate the aforementioned views of identity onto the network, enabling the same conceptual views to be used across administrative domains for user-centric functions (e.g. QoS or A4C), requires two concepts: an IDManager and an ID-Pointer. The IDManager, similar to that employed by the VID framework, stores identity information along with user policies and provides a common view over user information to other network entities, such as domain functions or service providers. It acts upon an identifier that refers to stored identity information. The identifier, ID-Pointer, provides the integration between the Identity Namespace, and consequently the IDManager that represents that namespace, and network protocols. It is used as a handle, derived from identity information, and understandable at the IDManager.

The ID-Pointer is nothing more than the realization of the VIDID (Sec. 4.3.1.3) from the VID model, presented as the link towards identity, that can be included in legacy protocols. The structure of the ID-Pointer is shown in Fig. 6.15. It is composed of 2 fields: identity realm (16 bits long) capable of encoding 2^{16} (65536) different identity realms (which can be viewed as an administrative trust domain in Figure 1); and an Index of 48 bits capable of indexing 2^{48} (or $2.8 * 10^{14}$) different identity registers. Further study should be devoted towards achieving a better balance on field sizes.

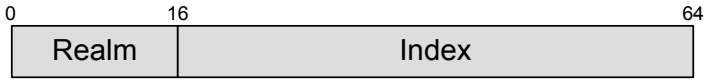


Figure 6.15: Identity Pointer (ID-Pointer).

The proposed configuration for the ID-Pointer allows any entity on the network to quickly locate the realm to which an identity belongs. Converting a Realm into an IDManager's address requires a resolution mechanism. While different identifier resolution strategies are not a primary concern for the proposed solutions, the realm could be obtained through a DNS-like mechanism using reverse lookups, or through Distributed Hash Table (DHT) mechanisms. Through the resolved realm an entity should obtain the necessary information to access the correct IDManager and reach the desired identity information. The ID-Pointer is only truly useful if integrated across the network stack. This requires an extension or change on the current protocols and layers, either by using explicit negotiation or by modifying the different layer identifiers to include the ID-Pointer. Besides an identity scheme, the cross-layer integration of the ID-Pointer does not require major modifications to the entire network stack and protocols, as described in the next section.

6.6.1.2 Identity Bindings

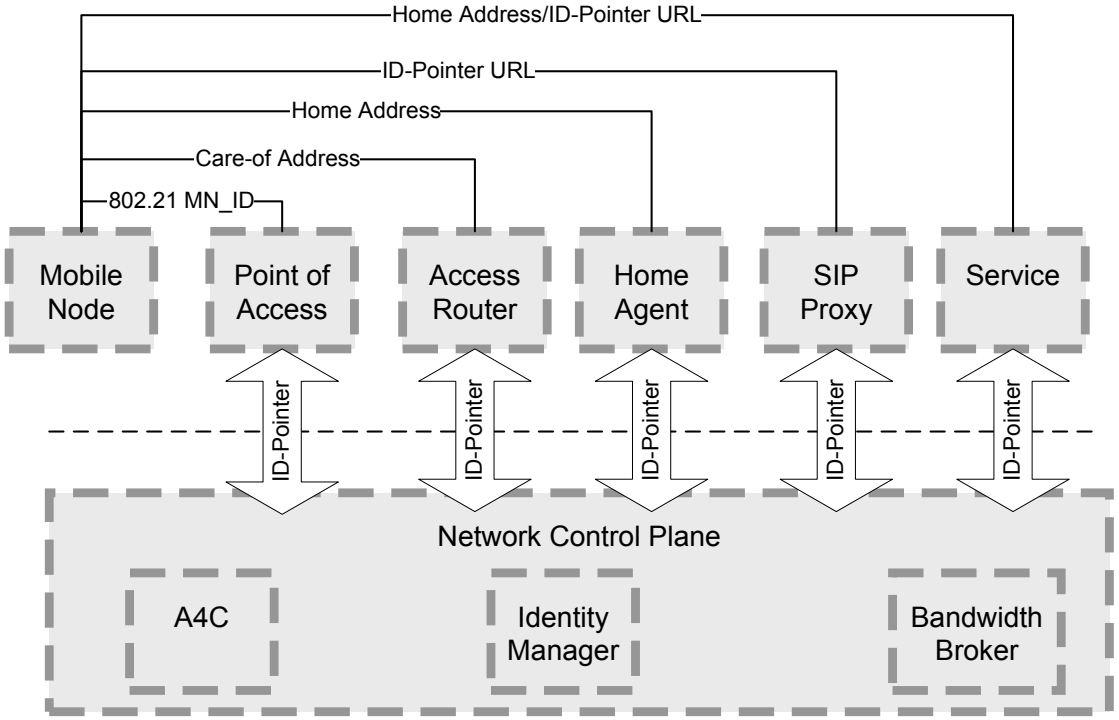


Figure 6.16: Identity integration at different layers.

Upon establishing methods to quickly and easily refer to identities, we need to provide

the correct mappings and bindings to be used in a cross-layer design. Using implicit pointers embedded in the protocol, network entities are able to retrieve the identity handle and resolve it without any functional modification to each protocol. Alternatively, identifiers can be exchanged out of band, e.g. using negotiation protocols, requiring network entities to exchange this information deliberately. The solution depends on where and what level we are integrating the ID-Pointer. We present a bottom up approach, covering from the link layer up to the application layer, including mobility, for the integration of ID-Pointer structures, either in negotiation phase or imbued in the native protocol identifiers. Fig. 6.16 presents an overview of the ID-Pointer integration at different layers, for which we offer a layer by layer description:

Link Layer Since L2 addresses are 48 bit long, there is no space to convey the complete identifier in the addressing structure itself. But, next generation heterogeneous scenarios are using IEEE 802.21 [68] to provide link layer Independent Media Services. We can include in the 802.21 negotiation procedures, the ID-Pointer as the Media Independent Handover Identifier, or in a PANA [74] negotiation phase. Assuming that the reference architecture is 802.21 capable, we replace the MN_ID, which is a type-length-value field, with the ID-Pointer: this enables the linkage between the MAC address and an MN_ID, therefore providing the reference to the ID Layer.

Network Layer At layer 3, the IPv6 address provides a proper space to include the ID-Pointer, carried inside the actual locator. The last 64 bits are used to identify the owner of the address, and could be replaced by the ID-Pointer. Fig. 6.17 shows the IPv6 address configuration, built after stateless address auto-configuration, and providing the ID-Pointer to the AR. This generated address is in fact the MIPv6 CoA that will be later registered with the HA. Through the ID-Pointer, the AR has sufficient information to access the network control plane in order to retrieve the required mobile node information, such as QoS, authorization and user preferences.

Transport The HoA also follows the same structure, enabling the HA to also have easy access to the mobile node information. Since transport protocols will establish their bindings using the HoA, which acts as the endpoint identifier, we are in fact integrating the ID-Pointer in the transport layer, implicitly conveying identity information also to the services (through HoA).

Application Layer The application layer has a rich variety of protocols, most of which are URL based (Sec. 5.4). Some protocols that work at the application layer, e.g. SIP [125], allow changing the communication endpoint while still maintaining ongoing sessions. Using SIP as an example use-case for identity integration, it is possible to outline an approach of how to integrate the identity based identifiers. Integrating identity into SIP requires breaking the resolution of SIP identifiers into several stages. The terminal registers its HoA with the SIP Proxy, which provides the ID-Pointer. Afterwards, to communicate with a user, one must know the URL, in the form of *johndoe@domain.tld*. The first step is to resolve the *domain.tld* to identify the IDManager, using a DNS record similar to MX, as done for email. The username could then be resolved on the IDManager, obtaining an ID-Pointer, for the target user. This allows redirecting the initiator at the SIP level to the correct resource. This process implicitly links URLs and identities, allowing the initiator to retrieve information from the destination, if allowed, and providing a verifiable identity to the responder - the initiator's ID-Pointer.

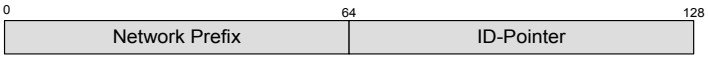


Figure 6.17: IPv6 ID-Pointer.

As the descriptions for the different layers show, the ID-Pointer is a flexible structure that can be adapted to different protocols. The adaptation only requires minor modifications to fit most protocols, and provides an invaluable identity reference that can be used at any point in the network.

6.6.1.3 Terminal and Network Support

To have a privacy oriented design supported through identity, both terminal and network need to undergo modifications affecting the network control plane. The terminal requires a control layer that instantiates identity functionality, interacting with applications in order to provide inputs for network stack management. As seen in Fig. 6.18, applications might be identity aware and provide specific inputs to the management plane, or legacy applications, where the management decisions will be inferred by a legacy interface component, which is in-line with the proposal made for VNS and for SAML IdM integration, discussed in the previous section (Sec. 6.3).

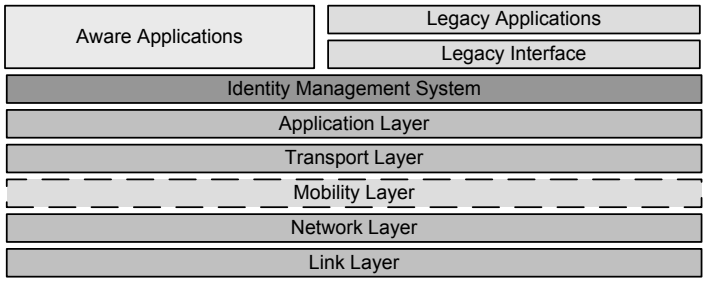


Figure 6.18: Terminal Control Plane.

Mobility protocols should refrain providing triggers for mobility, but just reacting to control plane commands, following the identity oriented operations to maintain connectivity. While the control plane has a direct path through the identity management layer, the data plane is orthogonal. Considering that identity management is mainly a control plane task, its repercussion on the data path is to keep each layer consistent with the current identity and mobility policies. This dichotomy is further explored in Sec. 6.4, that targets specifically the separation of control and action using identity driven protocols, to handle the increased information hardships that result from heterogenous mobility management solutions in 4G scenarios.

On the network side the modifications are more operational than functional. The normal network operation is based on distributed information, which can be modeled as relational databases. These databases are unrelated among them, since they use different identifiers for each piece of information, relating to the same user. As shown in Fig. 6.19, we propose to change the way these databases are organized, making them identity oriented, by using the

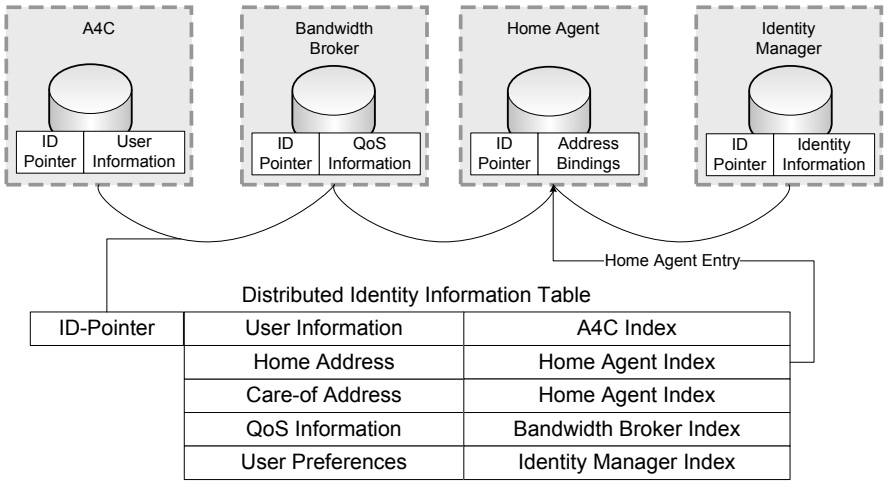


Figure 6.19: Identity Oriented Network Database Model.

same indexing material, the ID-Pointer, across all the databases: the same ID-Pointer grants access to the relevant database. As an example, using the ID-Pointer at the Home Agent enables the requester to access the HoA and the set of CoAs for a particular database; using it at the Bandwidth Broker grants access to a user's profile.

6.6.2 Instantiation Example: Identity and Mobility Management

After introducing the architectural cornerstones to accommodate architecture with intrinsic privacy support through identity mechanisms, we consider a mobility scenario as an instantiation example, since it plays a major role in NGN networks. Even the proposed identity-centric approaches could be mapped to a wide range of protocols, we selected a few use case protocols to provide clear examples: PANA [74] for authentication, MIPv6 [77] for network layer mobility and NSIS [52] for QoS. Using these specific protocols, we present generic procedures for the handover phase, allowing a clear view of how identity leverages and simplifies protocol mechanisms.

Optimizing user and flow distribution, that governs the mobility process, requires information at several levels. Collecting information about an identity combines the layered view over a user, which is indexed based on ID-Pointers. At the link layer, an 802.21 based framework collects information, such as network availability and provider information, e.g. L2 QoS capabilities. Higher layer information is also easy to retrieve: the ID-Pointer can be used to access QoS information at the Bandwidth Broker, accounting and authorization information at the A4C. User policies can also be involved in the decision process, stored at the IDManager, along with top-level user information. It becomes very easy for the network and user to gather all the necessary information to start the mobility process, represented by an handover event, and run algorithms that effectively distribute the load and optimize resources.

These processes can be exemplified by a seamless handover process, leveraged through identity mechanisms, and therefore containing the much needed privacy considerations. In a network handover the trigger to move an identity can have two origins: network or terminal

initiated [4]. Identity based mobility has advantages in both cases. When performing terminal (represented by an identity in our case) initiated mobility, the terminal decides to seek out a new point of attachment. In this case, the signaling load can be reduced: the network easily collects the flow information about an identity, leaving up to the identity or device just the need to signal that a particular identity will be relocated, followed by the actual movement. In network initiated mobility, the network decides that a particular identity should change the point of attachment. The benefit resides in the ease of information retrieval and signaling, since two access routers can easily share information about an identity, by sharing ID-Pointers and related flow information.

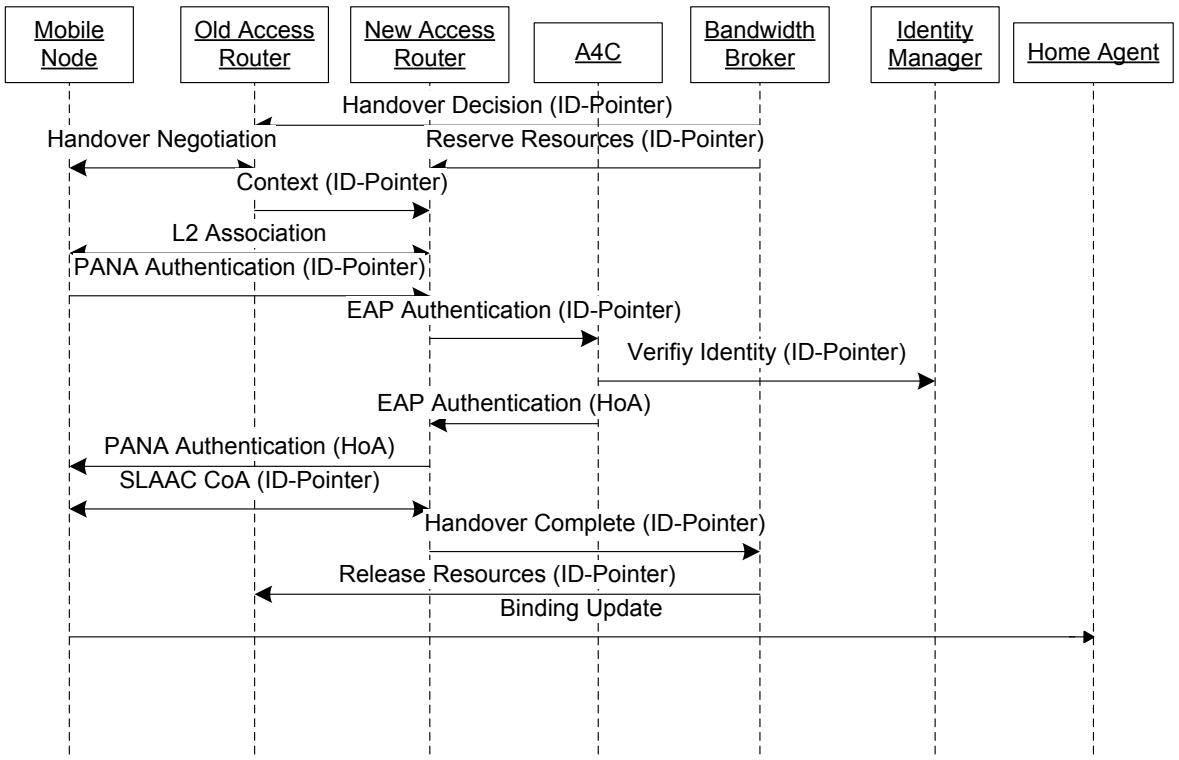


Figure 6.20: Generalized Network Initiated Handover Procedure.

Fig. 6.20 depicts the information flow process of a simplified network initiated handover, but with 802.21 signaling suppressed for simplification purposes. A network control point, possibly the Bandwidth Broker, decides to perform the handover of a flow and informs the old and new access routers that the Identity identified by ID-Pointer will change its point of attachment. This is performed by preparing the reservations on the new link, and by the old access router transferring all context relative to ID-Pointer, to the new one. After this process, the node moves and performs a PANA authentication once again (this process could be optimized by pre-authentication schemes before handover). Then, both an address configuration and mobility update are required. In the mobility process, no other elements need to be informed, since the ID-Pointer is still valid and network elements are Identity

[4]Both types of mobility triggers, network and terminal initiated, along with strategies to support both, are explained with further detail in [75]

oriented.

6.6.3 Network Impacts

The discussed identity bindings provide an integrated view over the network and user environments that takes into account privacy, while still retaining existing protocol properties and assuring backward compatibility. Therefore, we reuse each protocol without modifications, through the same identifiers across the network, in the form of the ID-Pointer. Accordingly, it becomes important to discuss the benefits and costs of such approach and the effect of supporting multiple identities.

This identity approach allows the support of a distributed database model, indexed at each network element by the same identifier, providing the necessary cross-layer and cross-protocol integration. This distributed meta-system is not bound by a particular protocol identifier, as opposed to today's systems, which require different identifiers at different network points, increasing the problems of the scattered information model. Using these different identifiers consistently is possible through the identity (VID), which assures an aggregation point. As such, it becomes feasible to use different pseudonym identifiers on the different layers, providing a VNS approach, without breaking existing protocols, and taking advantage of all the privacy features proposed in Chap. 4.

Moreover, regardless the addresses used, the same identity material is provided, greatly simplifying network processes such as accounting and authorization. Each layer contains information indexed by the same ID-Pointer on both remote and local entities. This means that the index used at the Access Point is the same used at a Bandwidth Broker, and requests and responses are performed based on the same identifier. In complex information environments, this enables a uniform view over a user and its sessions, regardless of where they are occurring. In a normal scenario a Bandwidth Broker that keeps track of L2 and L3 assignment would, for the same user, require a list of the MAC addresses in use, a list of IPv6 addresses and their respective Home Addresses; in a common index scenario, only one ID-Pointer would suffice, along with the information necessary in both cases, eliminating several steps of mapping and translation between identifiers, such as CoA and HoA.

Also, it provides an easier interaction with user profiles. In previous architectures, the user profile is not necessarily the user's identity, and in the same reasoning the users identity does not contain network profile information, such as QoS in an Information Card [109]. With the proposed scheme both can coexist on the identity layer, providing information such as resource authorization and QoS information, which are important in a 4G scenario, coupled with the user identity, enabling user-centric architectures. Each entity can retrieve this information easily, whether working with Link Layer or Application Layer information, providing that they have the necessary access credentials to the IDManager.

Current architectures can be complex systems, making them hard to improve since there are many design constraints. But, pre-existing constraints, such as user centric multi-device environments, must be respected. Complexity is therefore inherent. The proposed design can make innovation simpler by reducing the system complexity. Simpler is better in the sense that we can assume simple access to a large set of information, provided by an abstraction layer, valid at each point of the network stack. Network architectures can evolve by turning the focus to truly user-centric paradigms, taking advantage of the provided integration, which fails in current systems. This also means that the abstraction layer can be summarized in concise APIs, making it simpler to build on top of. Application developers can easily create

user-centric software since the architecture itself provides the metaphors and handles required by today's business processes.

The cost of such an integrated view is mainly the resolution of the ID-Pointer at each network element: it must be resolved into an IDManager, provided by the Realm. This can be done through functions such as reverse DNS queries, or using more evolved mechanisms such as Distributed Hash Tables to locate the correct IDManager. Nonetheless, the cost of resolution can be minimized through caching processes, or optimized through information deduction (e.g. if the A4C receives a pre-configured HoA, it can safely infer the Realm by looking at the address).

The presented architecture requires a cross-layer introduction of identifiers as a major privacy tool. While the benefits are several, legacy support of current protocols is a major concern, especially if we consider that user (critical) information must be distributed across such environments, endangering privacy. Taking this legacy requirement into account, we presented integration means that take advantage of the existing identifier spaces in each protocol, bringing to a minimum the impact on each protocol level. The implicit disclosure of the identifiers enables a non-ID enabled node to communicate seamlessly with an ID enabled node, at the cost of neglecting the Identity properties in the communication.

6.7 Conclusion

When analyzing the vertical and horizontal dimensions of privacy that originated from the privacy model, we acknowledge the recurrence of different architectural aspects used to provide privacy. Properly analyzing these concepts led us towards defining identity as a simplifying tool towards user-centric privacy architectures.

In this chapter, we tried to synthesize the common aspects of the different proposed approaches into architectural drivers that increase user privacy. The outcome of this process resulted in four different aspects that take advantage of identity as a tool towards defining an architectural support for privacy. Based on these four different aspects, we proposed an architectural instantiation for each, that tried to highlight the important aspects of combining the proposed architectural drivers. The different instantiations were aligned with the VID approach, increasing the strength of each proposal towards a strong link to the vertical privacy control layer, taking advantage of identity concepts through an IdM framework, as well as network pseudonyms stemming from VNS.

The first of the discussed architectural drivers focused on providing privacy through a strong identity control layer. Particularly, this dealt with re-using the VID framework from Chap. 4, but with a clear focus on integrating an IdM solution into the vertical model, in this case SAML. The advantage of this approach was showing that a concrete IdM solution can provide an adequate instantiation for the VID concept, and in turn the different issues concerning the network can be addressed. Particularly, we showed how a SAML based approach can help regulating the usage of different VNS's. Furthermore, we showed how to address specific instantiation problems, such as selecting and using different identities, using a concrete example, resulting from the VID generic approach.

From an architectural perspective, this reinforced the notion that the privacy features of using pseudonyms can be pragmatically integrated in the network, using an IdM solution as the control layer, promoting a concrete instantiation to the VID paradigm. This completed the different pieces of the vertical puzzle, and showed that, while not all IdM solutions can be

directly mapped onto the VID/VNS solutions, relying on cross-layer pseudonyms and vertical strategies preserves the user's privacy under almost all conditions. We also observed that SAML provides the sought after user-centricity on the application layer, and empowers all network layers with policy-based negotiation mechanisms that can now operate cross-layer, as discussed in Chap. 5.

Using the vertical layer for privacy management, it is possible to control the different VNS instances on a device towards the network. However, the control aspects of the vertical layer seemed to enable more features than simply managing the user's pseudonyms. The connection to IdM showed that it is possible to have a central layer, aware of user information and therefore capable of maintaining privacy. Starting from this assumption, we explored the privacy gains of using an identity control layer to guide different aspects of network operation, resulting in the architectural driver of privacy through an identity control layer. This conclusion led to the instantiation of an architecture that attempts to clearly separate control and execution, where one part is highly dependent on user-related information (control) and the other is highly operational (execution). Pragmatically, we explored a use-case centered around mobility management, where the control was outsourced towards the IdM plane, as a vertical privacy-aware decision plane, while the execution is left to individual protocols. This proposal indicated that the richness of user-centric information stemming from an identity layer can positively impact the decisions taken over different network procedures, making them privacy-aware and even distributed. It also showed that, from a purely operational perspective, the treatment of mobility protocols as plug-ins can lead to a situation where it is not necessary to adopt a one-size-fits-all protocol, but rather a customized selection that adapts to the user's needs.

While the previous instantiation concerned splitting decision and execution for the sake of outsourcing control to identity mechanisms, it showed that it is also possible to address privacy issues by focusing on identity as an operational driver in the network, our third discussed architectural driver. By taking an already identity focused protocol, HIP, we instantiated a network-oriented privacy solution, extracting direct benefit from the identity properties of HIP that create a new namespace. Using the identity focused separation layer, we were able to leverage HIP's mobility mechanisms to the advantage of the user, using the separation to address specific privacy issues, such as location privacy. This solution was also complemented from a vertical perspective, establishing a relationship between HIP's notion of identity and IdM systems. This opened the door to a straightforward introduction of the combined VID and VNS ecosystem into different network protocols, providing a secure and privacy-aware coupling between technologies. Using protocols that already provide a notion of identity simplifies the overall integration process, requiring only an alignment with the VID framework, given that the overall approach is naturally aligned with the PRIVED model.

The previous proposals required explicit network modifications, as means to reach an integration with the vertical control layer. However, to provide privacy, we required an intrinsic solution that is able to establish an identity focus on the different interactions in NGN. This represents the last architectural driver, that was instantiated by providing an identification solution that inherits properties from the VID solution to provide a tight identity relationship that can be used in all the different network protocols.

Using a relationship to the identity vertical layer can define a privacy baseline that can be used in most network interactions. Accordingly, we introduced an architecture that used carefully crafted identifiers, that fit into existing protocols, as means towards establishing the required identity connection. This connection provided a two-fold advantage: first, it

provided the mechanisms towards defining a more privacy-aware environment, since it is possible to re-use the privacy tools coming from the identity solutions; second, it allowed defining mechanisms to deal with the growing complexity of managing the different assortment of protocols that needs to be contemplated in order to maintain privacy. As described by the PRIVED model, if all the different protocols are not considered in a privacy solution, it is possible for one (or several) to undermine the entire effort of the privacy proposals. From this perspective, the VIDID integrated into the architectural instantiation created the possibility of redesigning current protocols to enable a clear reference to identity and to user-centric information. This not only supports but encourages pseudonyms that can integrate with the different technologies, and it also enables an unparalleled view of the user across different network components, which were previously disconnected from the information plane. Through the use of implicit identity references, we enabled protocols to communicate using the same "language" and dealing with the same "objects", which results not only in a privacy increase, but also on a complete user-centric paradigm for the network.

Chapter 7

Conclusion

> If you do not change direction, you may end up where you are heading.
>
> **Lao Tzu**

With such a wide problem space and a comparable solution spectrum, it is important to keep the proposed contributions in perspective, as well as understanding what can be done as follow-up work on the different aspects of this Thesis. This final chapter presents the most important conclusions on the explored privacy topics, while revisiting the initial hypothesis under the light of the presented proposals. We present how the different aspects come together to form a consistent view on privacy, and also how the Thesis results can be a starting point for future and emerging networking paradigms.

7.1 Results and Achievements

Throughout the different chapters we presented several complementary solutions that build on each other to promote a more effective privacy environment for the network user. We started by presenting a model that complies with the outlined privacy definitions, attempting to shed some light on the complexities of network-related privacy aspects. This enabled a vertical approach to privacy that tried to address the cross-layer problems caused mostly by information correlation. The result focused on a privacy protecting model, leading to a vertical architecture that places identity as a main tool for architecture design, specially centered on user privacy. The properties of the vertical framework also highlighted several shortcomings on the different layers, motivating the analysis of specific layer problems using the proposed privacy model. This horizontal focus led to several presented solutions, covering the different layers in the OSI stack, with particular attention on several network-based threats, which were the primary concerns.

The combination of the different privacy aspects and proposals contributed to a paradigm shift that relies on a vertical layer for information control and user-centric privacy management. This role was assumed by identity, which became a tool towards providing privacy in the network, and allowed the definition of several architectural drivers for designing privacy-aware architectures. The identification of these architectural features transformed identity, and Identity Management, into a privacy-driving feature in the network, rather than an application layer-only Privacy Enhancing Technology (PET).

Going back to the initial objectives, it is important to understand that the different contributions were only proposed as means of pursuing the privacy hypothesis presented in Chap. 1. Therefore, in this chapter we analyze and discuss the hypothesis in the light of the work described between chapters 3 through 6.

Nevertheless, a PhD thesis should offer more than a compiled package of results, aligned with a proposed theory and hypothesis. It should contribute to future solutions by opening different (research) paths, that others may follow. To comply with this vision, we present several directions that complement the proposed work, along with insights stemming from the different privacy aspects that can be taken into other domains. Therefore, as a concrete result of the discussions presented throughout the different chapters, we promote an outlook on privacy, along with a vision that focuses mostly on evolving network, architectures and paradigms based on the Thesis' results.

In the following sections, we first go through the different aspects that frame the hypothesis and attempt to address it, using the conclusions to promote an outlook on privacy and a possible evolution path for network paradigms.

7.1.1 Understanding Network Privacy

Concerning the network, usually confusing privacy concepts are made clearer through the proposed definition and subsequent framing discussed in this Thesis (Sec. 3.2), which is an important contribution that should be highlighted. This motivation led to the several contributions that try to simplify network related privacy issues, through tangible and pragmatic definitions. We started with a lexicon that frames our privacy discussion on the network, trying to avoid, when possible, confusions with orthogonal or concurrent definitions. These definitions also served as a starting point for a pragmatic model that attempted to handle privacy from a network perspective, using three recurring concepts: Information Set, Event

and Relationship. These concepts enabled the description of several network-based privacy threats, addressing both basic and complex network interactions, through a pragmatic approach that can relate multiple network protocols through events and relationships. This event-focused approach defines the user as the central object of the Information Set, making it the single most important entity in the privacy equation. Also, the different model components allow mixing simple network observation events with complex correlation events, defined as relationships, that can only be made through extensive information collection and pattern observation. These properties make the proposed model suitable for a diverse range of use-cases and network conditions.

In order to prevent the model from becoming exceedingly conceptual and loosing its potential impact, we presented a network instantiation, going through several layers and the network stack model present in Chap. 3. The instantiation tries to build an effective bridge between theory and practice, contributing to the clarification of network privacy concepts.

The conclusions, beyond the resulting models and threats, showed that there is a vertical and horizontal dimension to privacy. In this context, vertical privacy deals with the traversal aspects of privacy, and the cross layer threats that compromise security and privacy at different layers. Conversely, the horizontal aspects focus on technology specific privacy aspects that stem from particular network properties, especially identifiers and network mechanisms. In our approaches we tackled both, showing how can their privacy issues could be addressed or improved.

7.1.2 Vertical Dimension

An important conclusion of the privacy review done in Chap. 2, and consequent privacy model presented in Sec. 3.3, was that network privacy is not limited to a single layer. In fact, a single layer can compromise others, creating a clear dependency between the different layers in the network stack. As a direct consequence, this understanding led to a vertical notion of privacy that resulted in a cross-layer conceptual approach defining virtual personae, modeling how the user interacts with the world, covered mostly in Chap. 4. The Virtual Identity is a digital construction of the user that has repercussions on how the user is perceived and handled in a digital environment. This concept enables a vertical approach that resembles how users interact with the different services, outlining the tools which enable the construction of a privacy framework. The VID framework presents a new way of dealing with users in network-intensive environments, using IdM concepts to address network privacy concerns.

The VID model, presented in Sec. 4.2, was designed to have an impact on all the network layers that required user interactions, creating several requirements to support multiple identities belonging to the same user on a single terminal. Given that privacy is a cross-layer issue, as concluded by analyzing the different correlation threats, it must be provided across the different network layers. We turned to the idea of pseudonymity to address cross-layer threats and to enable the coexistence of several VIDs for a single user. We discussed the notion of pseudonymity in the network, all of its benefits and drawbacks, and how well it matched against the proposed privacy model in terms of information set. Studying network pseudonymity, as discussed in Sec. 4.4, as well its relationship to VID and Information Set, provided enough insight to define an instantiation of cross-layer pseudonyms as Virtual Network Stacks (Sec. 4.5). VNS uses a per-identity virtual device metaphor to support different information sets that encompass the network stack. A prototype implementation was integrated into working test-beds and demonstrators, showing the feasibility of the approach in

different contexts.

This proved that there is the space and the means to build a vertical layer that retains control capabilities over the network and that, with the proper semantics, can support the VID concept with multiple pseudonyms on the network. This contributes to an improved privacy environment, and allows us to focus on specialized network privacy issues that exist on the different layers.

7.1.3 Horizontal Dimension

Given that several privacy problems stem from the lower layers in the network, we tried to provide a system that deals with identifier related information, outlined by the PRIVED model, and used in the VID approach. These tools can handle different threats, such as identifier correlation and identification problems, resulting in user tracking, recognition or location pinpointing (among other threats), which can result from different aspects of the lower layers. In Chap. 5, we attempted to identify these threats on each layer, resulting in a multi-layer study that addressed mostly link and network layer issues, with some considerations of privacy aspects on the transport and application layers.

At the link layer, we proposed a solution that uses a novel transport mechanism to support several privacy features, presented in Sec. 5.2. The goal of the solution is to make every packet visible on the network indistinguishable, especially regarding information that can jeopardize privacy. Therefore, it should not contain any noticeable information bits, except for the intended destination. This was achieved with the concept of identifying the communication channel by its encryption key, and not by explicit identifiers like sender and receiver. To test the different system properties of the system, we implemented a simulation scenario (using NS-2) around the proposed solution, evaluating its feasibility and potential performance impacts. The result was a comprehensive study, yielding several interesting observations of the link layer behavior, which met the original requirements.

We also addressed the network layer, where the most important issues revolved around user identification and location information contained in unique identifiers. After a study of network layer functions, we proposed a solution to handle the issues at the routing level, entitled Waypoint Routing, described in Sec. 5.3. By inserting Waypoint Routers in the communication path, which are routers that mask the sender (following the tradition of MixNets), which perform hop-by-hop (between WPR) routing to protect user privacy. Due to the changing addresses on the packet, it is virtually impossible for any single node along the path to determine the original packet source. The novelty of the solution resided in the proposed lightweight mechanisms that allow core routers to efficiently perform privacy-providing mechanisms. Consequently, privacy can be provided as a value added service by the operator, rather than a peer-to-peer end-user system. We presented this novel solution along with a study on how inefficient privacy techniques can impact routing, crippling its performance.

Exploring the remaining layers led to the conclusion that transport and application layer also require privacy considerations, outlined in Sec. 5.4, even though different from the two previous analyzed cases.

Focusing on the transport layer, we identified that it suffers from mostly the same issues as the network layer, given their tight identifier reutilization (IP addresses). The direct result is that, a solution that addresses the network layer aspects can directly benefit the transport layer, as they both share identifiers. The only exception to this scenario is when locator-

identifier split solutions are employed, which we address further in Sec. 6.5 as a result from architectural privacy design in mobility protocols, and also application ports that do not pose severe privacy threats, but can still be addressed in the scope of pseudonymity,

The application layer provides a very different model than the other layers, as it deals with user information and attributes. In this scenario, where many different protocols deal directly with user information and identification, it is necessary to consider a different approach. While addressing all of the application layer mechanisms is mostly out of scope of the addressed network interactions in this Thesis, the application layer can provide mechanisms for privacy that we already used in the vertical layer. IdM solutions provide privacy aware application protocols that enable the definition of user-centric privacy technologies. While these solutions focus mostly on service interactions, they can address most specific privacy issues on the application. We extend this by identifying the important technologies that accomplish these features, and by defining cross-layer considerations, which is then carried over to Chap. 6, as we instantiate a specific application layer solution, SAML, in a cross-layer privacy environment provided by Virtual Identities and Virtual Network Stacks.

7.1.4 Architectural Drivers

For the most part, the proposed solutions incorporate privacy as a core architectural component, often resorting to user-centric mechanisms. Considering privacy as an integral part of the network, led to several architectural features that can be used to define new privacy-aware network solutions. As presented in Chap. 6, these privacy features stem mostly from the vertical control layer, which can consolidate privacy aspects in network related technologies and mechanisms, resorting to identity concepts as the main tool to promote privacy-aware network solutions.

These architectural feature were used as means to instantiate several architectures, most of which rely on identity centric mechanisms, embodied by IdM-based solutions. By linking privacy with identity we are providing more than simply security enhancements. More importantly, we are linking privacy with the concept of identity. This is beneficial because it relates to social concepts of privacy, that always deal with the concept of "self". In return, using identity concepts in the network leads to a user-centric environment with scalable control for user (private) information. Through these assumptions, we defined several architectural drivers, that enable privacy as a core feature of most network interactions.

A proposed method to integrate privacy in the network was defined by using a vertical enabler that can instantiate the VID layer. This control layer, which we define by using a SAML-based solution in Sec. 6.3, promoted privacy by using identity as the vertical privacy enabler. The IdM solution provided a VID framework realization, that focused on SAML as the main protocol used for the vertical information coordination, including the user's pseudonyms (as VNS).

While we presented some strategies that enable the mapping of SAML identities and pseudonyms to the network, the control features that are privacy-aware by definition can further serve to control information flow on the network. Therefore, we not only used identity as a driving technology for the vertical layer, but we also encouraged using the vertical layer for further network control. This was realized by providing a simple architecture that separated control from execution, as instantiated in Sec. 6.4), using mobility as an example technology. This approach focuses on three aspects: i) consolidating all user-centric information on the IdM infrastructure, ii) decoupling that information from mobility processes, and iii) express-

ing mobility through common semantics that are capable of turning mobility protocols into action only mechanisms. The real benefit of this proposal is that once mechanisms are truly decoupled, control can be made user-centric in a privacy-aware context, as proposed by a vertical layer. In fact, this separation is not necessarily just from a control/execution perspective. We also explored such a separation between layers, in our case, between the network and transport layer, showing how simple identity concepts and information decoupling can provide privacy oriented operation drivers on the network.

Finally, as a result of the requirements for integrating identity as a tool for network privacy, we proposed a solution capable of instantiating identity centric operations in several different aspects of typical network operation. In Sec. 6.6, we used an ID-Pointer that mimics the definition of the VIDID from the VID model, to provide an identity focus for the different network protocols. This promoted a simpler control mechanism for the different privacy features in the network, as well as means of simplifying the utilization of user-related information in low level network operations. Not only does this simplify network privacy, but it also provides more adequate handles for integrating cross-layer pseudonymity in the network.

The main result of the architectural privacy drivers can be highlighted as the different architectures that provide clear instantiations of the different proposed concepts. The nature of these concepts proved very important towards enabling privacy as a core network feature to the use of a support technology, such as identity, that adequately deals with user information. In most network level protocols, this user bias is not something that can be easily achieved without the support of adequate identifiers (VIDIDs), as well as technologies capable of controlling the user's information set using identity concepts (VID), well integrated with NGN networks.

7.1.5 Reviewing the Hypothesis

We started with a question revolving around the privacy threats that originate from the different identifiers present in heterogeneous mobile networks, and potential consequences towards the user. More importantly, we set out to discover if, when found, these threats can be properly mitigated. As the scientific method dictates, we postulated an hypothesis stating that network level threats can be caused by the identifiers present on different protocols and layers, which present a unique way to undermine privacy, resulting in reproducible breaches.

The conclusion, six chapters later, is that the hypothesis was validated in the light of the results, but with complicated ramifications demanding more than a binary response. The discovered vertical and horizontal dimensions of privacy showed that on the network, identifier relationships are corrosive to privacy and can provide consistent means to undermine it. The vertical aspects create constant privacy threats, reaching several layers and compromising isolated privacy efforts. Similarly, the horizontal aspects show that each layer poses unique threats yielding private information, covering different aspects of user privacy. The validity of the hypothesis started to become evident with the PRIVED model, and was corroborated by the following solutions, both vertical and horizontal, which showed how containment spaces, either separating vertical interactions using personae, or horizontal separations constraining identifiers, prove to enhance privacy. Each solution reinforces one of these two aspects, and shows that there can be recurring approaches towards privacy, which can be implemented in future network solutions.

Concerning the more abstract theme of network privacy, we observed that the network is not necessarily a privacy jeopardizing environment. In fact, there are even some scenarios

where network technologies improve on our privacy, allowing anonymous interactions, where the user can be protected by pseudonyms and alias identifiers. Unfortunately, this is not the case with most of the analyzed network protocols, to which we must be alert, that can compromise all privacy efforts. Currently, the best place to start compromising privacy, is indeed in the network, where protocols make consistent use of identification paradigms to carry out their objectives. In this context, identifiers as the way we use them, proved to be particularly vulnerable to correlation and linking threats, defining an environment where a single flaw can contribute to undermining user privacy. This shows that we must always look at the network as a whole, considering a vertical paradigm, but always mindful that specific layer identifier threats exist.

To support this broader vision on privacy, we acted on the proposed privacy models and concepts, leading to several novel solutions presented through the Thesis, which we believe contributed positively to the privacy landscape. By focusing on network privacy, the identified threats and proposed solutions bring us to a final conclusion regarding the hypothesis: the different articulations of privacy solutions, along with the model that relies on two distinct axes, revealed the validity of the hypothesis by defining identifiers as a major threat in the network, concerning their internal properties as well as the correlation mechanisms they provide. These findings corroborate the initial postulated hypothesis that identifiers are in fact a major threat on the network.

7.2 Future Outlook on Privacy

Throughout the presented work there was an effort to provide tools that enable us to increase our understanding of privacy in the network. As a result, we tried to elaborate on a core model that focuses on the most pressing network threats, using identifiers and identification, but that still has room for improvement. There are several ways in which we can conceive a roadmap that enables us to further pursue some of the privacy topics studied here. The most immediate subjects relate to constructions of the privacy model and also orthogonal privacy issues, as discussed below.

7.2.1 Measuring Privacy

While there is always great debate about user privacy, there is always a gap, which we struggled with many times, dealing with how privacy can be perceived by users and services, and even to some extent, measured. Because providing privacy can not be directly comparable across different situations, several discussions can become superficial. Consequently, one of the primary ways to improve the work presented here, and to improve the privacy field in general, is to provide mechanisms that enable any type of privacy measurement. This quantification is important to enable the user and providers to perceive privacy, both in terms of effective protection, and of network related measures. It is always complex to debate privacy vs performance trade-offs and gains, without a clear evaluation of the privacy gains/losses, because there is no direct metric or comparison means.

There are some proposals on how to address this particular issue, using the concept of entropy [132] from Shannon's Information Theory, which is mostly applicable in anonymity schemes [36]. While anonymity is not privacy, it can still be the foundation for many types of information centric measurement mechanisms, that deal with randomized information (as argued within the PRIVED model, where the uniqueness or not of identifiers can be a cause for

discussion). Accordingly, a direct evolution of the proposed model is to enable an arithmetic proposition for the quantification of privacy, which can be done using entropy as a starting point.

7.2.2 Building on the Privacy Model

Beyond the conceptual parametrization of the model, which directly implies a quantification system that enables measuring privacy, there is a need for experimental validation of the privacy model. While we have done several validations on the efficiency of privacy protecting mechanisms, there is a need for an experimental tool that directly maps the PRIVED model as an attack tool, thus demonstrating the flexibility of the proposed approach. In practice, what we are suggesting is a privacy scanner that builds on network events and creates an information set handler that consumes network information. While this has been done in several distinct tools that can be used for specific threats or layers, there is no single tool that follows the complete PRIVED approach. The efforts for building such a tool are already planed as a continuation of the work shown in this Thesis.

The implementation of a network scanner using the PRIVED methodology should result in the consolidation of the conceptual model. The tool could be used to explore complex relationships, especially the evidence collecting model and the relationship factor. This could be considered as another potential branch of the proposed work, which would provide a pragmatic approach towards determining the effectiveness of the model.

While some of the work that deals with complex correlation mechanisms has been proposed as future work, there is still a need to verify the model in such conditions. In Chap. 3 we introduced several possible techniques to deal with complex correlations mechanisms, such as using Bayesian inference for relationship establishment, which are still not explored in the scope of the PRIVED model. From these, an attractive research direction is to consider probable relationships, resulting from evidences correlated against each other to determined information leaks, outlining potential privacy threats. This can become an important approach to privacy, when properly validated through real test-cases, and supported by our practical approach to privacy in the network using the PRIVED model.

7.2.3 Improving Orthogonal Conditions

Most of the discussed improvements so far relate to the privacy model. However, the envisioned network environment stemming from the different explained proposals shows that, once there is a more meaningful layered separation supported by privacy aware identifiers, there can be a more meaningful exploration of layered privacy, and resolution mechanisms. In fact, identifier issues, multiple identifiers and resolution has been one of the implicit technologies, for which we introduced some considerations, but that requires a Thesis on its own. This is precisely how it is being pursued, within the scope of another PhD Thesis, as means of using identifiers and resolution to power several of the privacy and identification related technologies and proposals, and has already been the object of a MsC Thesis.

As we proposed in Chap. 5, there is now space for layered improvements, in order to complete and complement the privacy mechanisms of the vertical approach. Taking this into account, there is already ongoing work to improve our proposed link layer solution. Because of the unique mechanisms of that particular proposal, there are no distinguishable packets on the link. This would cause an attacker to turn to traffic patterns. The natural evolution

is to provide traffic pattern concealment, as part of the overall strategy of hiding link layer information, thus providing privacy to the user.

On the network layer, and due to the framing of the discussion around lightweight privacy mechanism, a bigger discussion considering routing mechanisms can arise: how to relate Waypoint Routing relationships with TOR; and to provide a distributed discovery mechanism for WP routers and Privacy Services. We believe that it is possible to design a system that suits both Waypoint Routing and TOR, which implies privacy metrics and discovery systems for privacy providers, provided either by the network operator or by end-user, with the discussed benefits and drawbacks. Another idea that is yet to be explored, is to study Waypoint Routing mechanisms as control schemes in the network, creating a better privacy environment that relies on key properties derived from the Waypoint mechanisms. This can be done to enhance the privacy control layer, while using unobtrusive routing mechanisms, thus protecting user privacy and establishing a better trade-off for privacy.

There are also some aspects of the proposed solutions that would benefit from implementation and real wold deployment, but that could require the work of another thesis, relating it to privacy metric systems as mentioned above. This also serves as a general conclusion, because several solutions presented here require implementation and deployment for validation. As they mature, these ideas are becoming ready to be incorporated into more widespread contexts, and yield tangible privacy benefits.

7.3 Evolving Paradigms

Working on privacy implies an in-depth inspection of network mechanisms, focusing on several operations and content. At this point, it is safe to assume that privacy is not simply about network operational aspects, but also about what information is conveyed and how. This implies dealing with the relationships between layers, and performance or security trade-offs, specially concerning the user. These different vectors play important roles regarding with privacy in the network. Therefore, exploring complex privacy issues can provide insight not only on the current state of user privacy in the network, but also on fundamental operational aspects of the network. There are lessons that can be extracted and applied generically to any networked environment or solution. The gained knowledge can assist ongoing efforts towards future solutions that try to rethink current networking paradigms, such as clean slate designs and Future Internet approaches.

Applying the knowledge gained throughout the Thesis on the different network mechanisms, identifiers and user-centric design, makes it easier to step out of our zone of comfort, and equate solutions that not only consider privacy aspects, but allow an evolution into new paradigms. In this section we explore how using privacy as a core future technology, especially focusing on aspects that stem from the generalization of the proposed privacy solution, can provide a path towards evolving network paradigms.

7.3.1 A step into future architecture

What we have witnessed in the evolution of Internet-based network paradigms, is that the layered model has self-imposed limitations. Even though the initial concepts are still valid (independent layers that build upon each other to provide aggregated value), in some cases the design has become diluted through several compromises. There are clear limitations to this model due to the fact that layers became very dependent on each other, specially considering

identification aspects. This is emphasized in privacy discussions because information specific to a single layer, in most cases, cannot be replaced, as it would impact adjacent layers. This shows that while the network stack is very sound in theory, identifier and layer functionality re-usage has lead to several potential violations of the layered model approach.

Another recurring idea that appears throughout the Thesis is that we turn to identity for control. If we take a step back, we can see that this happens due to the lack of any other layer or tool that provides control primitives over network and user information, which creates a more effective privacy environment. A potential cause for this phenomena is the lack of reusable control semantics. Reusable semantics enable the network to benefit from a common control structure. This is what we attempted to introduce as we explored derivative of first, a VID solution that creates a controllable network structure, and later, an IdM approach which provides the control infrastructure.

One of the subtle changes that is implicitly (and sometimes explicitly) mentioned relies on the fact that we shift the communication paradigm towards a model where the user increasingly becomes the endpoint of the communication, rather than the device. While the user identity has been substantial clarified, on the network side, it can still be complicated to grasp a proper management structure for the network aspects. While our first attempt to embed identity information into network protocols (Sec. 6.6.1) proved interesting, it can be seen as a first step towards a new model.

Based on the identifier concepts (that provide relationships towards a control layer, e.g. identity) and explicit layer separation, we can define a session-based approach towards networking. This approach is supported by the fact that most protocols and digital systems already provide the notion of a session (even though sometimes in different formats). This is highlighted in Fig. 7.1, which defines the connection between two endpoints. Further considerations reveal that a session based approach to networking can in fact be a powerful abstraction to drive network innovation, because it can be used from a pure vertical perspective, as an aggregation of session (a recursive definition), and a horizontal session, the basic unit that drives protocols.

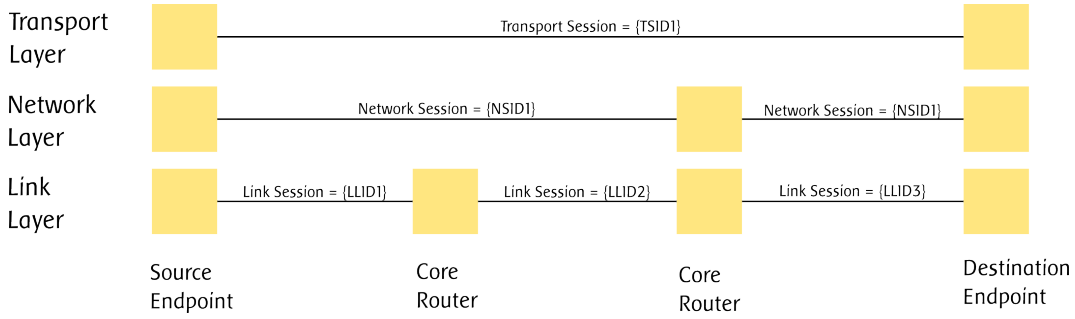


Figure 7.1: Different session and corresponding identifiers on the network.

Our first work in this direction was “Mobility Aware Paths: The identity connection” [94], were we proposed the idea of building identity controlled structures, named paths, that enable the conversion of identification solutions through resolution mechanisms, on different layers. This reinforces the user as the communication endpoint, leading to the idea that communication can become user-centric, and provide control semantics for the network. When the identity moves from just an aid in the network to the end-point of the communication

protocols, we are faced with a new communication paradigm.

By acknowledging the transformation of endpoints, we continue to extract common aspects of the network communication, starting from the idea that the network operates based on session semantics. This enables flexible communication, either through a path or a session, allowing bigger steps towards future architectures. Using sessions as a base concept, or better yet, as an evolution of paths which only deal with network connectivity, we have a driver for horizontal and vertical relationships in the network. Furthermore, sessions can be related to the user and, at the same time, to the simpler abstraction mechanisms, such as network packets. What is intended with sessions is a concept that promotes both a horizontal approach, using targeted layer sessions, and a vertical glue that aggregates the layered sessions - a meta-session. The layered approach can be traced back to the original plan for the OSI or TCP/IP models. The session concept can be reused in a system wide view of isolated layers, which present individual sessions that are related in the scope of a higher level session, defining a consistent abstraction model.

The main advantage of the presented model is that it enables a clear separation between layers and protocols, making it possible to introduce a concise control layer. Following the proposed ideas in this Thesis, this layer can revolve around identity, providing the missing user-centric components for Future Internet architectures. The Future Internet aspects are introduced because achieving the layer separation requires re-engineering existing paradigms, which would lead to a new Internet design. But, because this approach draws on well-known concepts, such as layer separation and isolation (something which was originally intended in Internet's layered design approach), it can present a smoother evolution path than a clean slate design. Also, the user-centric control layer (which can be IdM or any evolution of this user-centric paradigm) enables us to outline integration mechanisms similar to the concepts proposed by HIP (Sec. 6.5.2 and Sec. 6.5.4). The basic methodology would be to provide separation layers, with independent identifiers that relate to IdM, which already proved interesting in the solutions discussed earlier. The paradigm shift implies refocusing the roles of each layer, cutting any superfluous or redundant functions. Once that is completed, it is possible to design abstraction mechanisms, like those proposed by the Generic Mobility architecture (Sec. 6.4): the mobility abstraction layer provided means for clear control and execution, while still reusing existing protocols and enabling the integration of new solutions. This concept can be extended to the different network layers.

We try to introduce these ideas by relying on vertical information sets that use identity and session and the driving concepts. By separating the different layers, it is possible to rediscover the nature of the layered network model, with meaningful implications on concepts like privacy and mobility, which suffer from the close layer relationships and dependencies. However, this approach has a requirement that can be seen as a key element in future designs: it needs a complex naming and addressing system that is agile enough to support a networking paradigm built upon user and session references.

The outlined approach represents a model that requires the use of several shim layers on the network stack, as show in Fig. 7.2. This is an evolution of concepts like HIP [113] and LNA [11], but where the introduced identifiers are handled by the naming and resolution layers, which should be coordinated by IdM systems with strong privacy requirements.

This provides an evolution path towards future networks, that still takes advantage of today's protocols. The main issue that needs to be considered is the definition of a model to aggregate the proposed concepts and establish the required layer separation. Nevertheless, the implicit conclusion of this discussion, which is an indirect conclusion from studying how

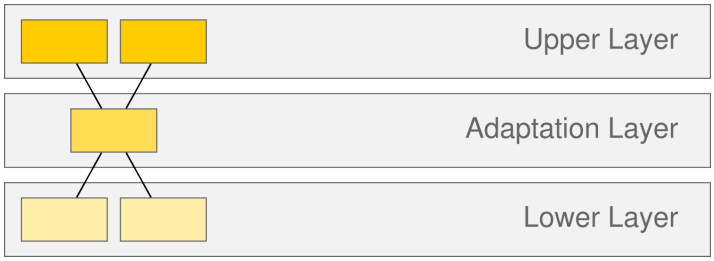


Figure 7.2: Wedge layer design.

privacy on one layer can affect others, is that a Future Internet architecture can be built using vertical and horizontal separation. While this may appear to be a daunting task, it does not need to be done in a single step. By using a modular approach, it can become an evolving effort, similar to what happened with the current Internet design.

7.3.2 A step into the real world

After several iterations with IdM technology and user-centric paradigms, we realized that this is a digital concept that brings both security (specially privacy) and personalization. In fact, several different systems can benefit from the identity information relationship, since it allows customizable environments. The companion concepts are secure authentication and authorization mechanisms, introduced by IdM, which are enhancements that can be integrated into current systems. This would allow the creation of a reusable component that is usually a complicated piece of software and architecture. Therefore, we can look at the IdM system as an architectural component that can serve as a privacy-aware AAA platform, especially in mobile environments where users become very agile and require further security and privacy features.

Using our devices in real world transactions, either for purchase, discounts, or social interactions, is an already established trend. While we turn to our devices for these transactions, we are not taking full advantage of their capabilities. We are mostly using them as transport mechanisms for authentication mechanisms, that can sometimes be easily forged and have limited security. This is incoherent with increasing security and privacy requirements. But, if we introduce an authentication and authorization component that has a user representation, we can provide a platform that can enhance real world interactions using electronic devices. This approach enables exploring IdM as the driving technology for security and privacy in real world interactions, where the most interesting applications reside in different scenarios like e-ticketing, domotics, or any type of device interactions, promoted by the user.

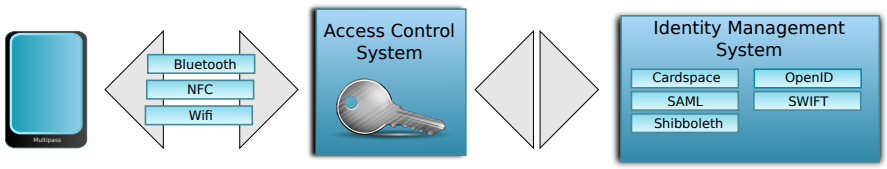


Figure 7.3: High level Multipass architecture using IdM for M2M a U2M interactions.

These concepts were applied in a spinoff effort derived from the presented Thesis, the Multipass project, which fosters the use of the IdM backend for a symbiotic user/device relationship. It enables a scenario where physical Machine-to-Machine (M2M) and User-to-Machine (U2M) interactions are more secure and privacy aware. The overall approach is shown in Fig. 7.3, where we can see the IdM backend directly interacting with "real-world" access control mechanisms, that are engaged with the user over wireless technologies. To support this ecosystem, the project built a transaction oriented architecture that turns to IdM for identity oriented control and validation of all transactions that happen between a user (which is translated in his device) and a provider. This enables stepping out of the normal applications of digital systems, venturing into a space that is in demand for security and privacy features, such as true user authentication (rather than device) and support for mutual authentication mechanisms, outlining a new evolution of digital interactions.

7.3.3 A step into the cloud

A technology that is quickly shaping into one of the most prominent computing trends is Cloud computing, or simply Clouds. The new paradigm is built around services provided to the end user, where even the network can be regarded as a service. Users create their environments on cloud services, where they store their files and most important backups, run publishing services (e.g. blogs or websites), and basically run every desired service. Currently, most aspects of domestic (the average user) computing have a Cloud counterpart, many running on an affordable (often free) offering model. While it requires little technology knowledge from users, these new cloud services introduce unprecedented service conditions in availability, cost, and (mobile distributed) access.

However, because users have to surrender control over their data to services, privacy is something that is severely lacking in cloud computing. It is almost impossible to determine where the data is, who accesses it, how safe it is, and how can it be deleted (if it actually can be deleted). These features, which were taken for granted in our ordinary use of personal computers, now elude us because there is no physical access to the servers running the Cloud, which might not even be all on the same geographical location or country. As such, in distributed systems that can serve millions of users, a single security breach can jeopardize many users, rather than just a few when the same breach occurs on a home computer. Consequently, the same principles that make Clouds attractive (make a service available to all devices, anywhere in the world, for a massive number of users) also raise several new problems. In fact, the business model is the first to undermine privacy, because free applications and services can generate profit by selling private user information, even if anonymized.

The above mentioned issues are not new in the scope of this Thesis. Privacy throughout the other network layers suffers from similar privacy problems, where once information reaches the wire, it leaves the user's grasp. A potential solution can be including Cloud technology as one of the layers discussed throughout previous chapters. Because IdM is already becoming an increasing trend in the Cloud, through lightweight services that provide SSO, like OpenID [122] based services as provided by Google, and now Twitter and Facebook using OAuth [116], we can still shape identity to be one of the primary privacy drivers. This is possible through a paradigm shift to user-centric clouds, making the cloud also about the user, and not only about the infrastructure and services. By leveraging Identity Management, there is a path where the user takes the cloud concepts and turns them to his advantage. The traditional separations on cloud services as Infrastructure-as-a-Service (IaaS), Platform-as-a-

Service (PaaS) and Software-as-a-Service (SaaS) then can be structured according to slightly different lines, as shown in Fig. 7.4.

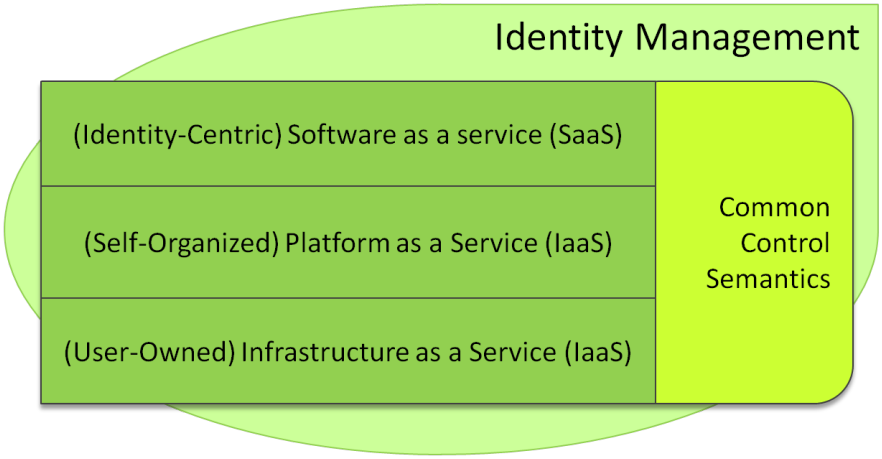


Figure 7.4: Identity as Core component of Cloud technology.

A cloud system that uses user-centric technology can put the user back in the driver seat of his data, providing the required data control, information security and privacy that must come with all modern systems. This can be achieved by designing a distributed system that explores the heterogeneous user-owned infrastructure, while providing privacy, security and authentication. Such a system could provide Operating System like semantics to access all the existing services and API's, harmonizing application access. However, the current cloud model does not consistently provide such features, and while they could be implemented separately - enhancing current cloud technologies - they would not benefit from user-centric characteristics beyond the fragmented model that exists today.

This presents identity as a kernel cloud technology in a distributed environment, where linking activities to specific users is essential. In this scenario, informations become distributed, using an infrastructure with multiple ownerships. The result is a user-centric service layer built around user data and resources, that is capable of managing a pool of distributed resources (the cloud). Every service running in the cloud will have the notion of user, which is in fact the currency (user information) for service access, through user-dependent credentials. By aggregating this information on the IdM system, it is possible to have a broad view over the Cloud resources available to each user, and to provide this view to self-organizing software. From this point on, the IdM layer becomes a mechanism capable of aggregating user resources, scattered through the cloud, in a coherent architecture. The omni-presence of identity and user information justifies that IdM should be a kernel component of the cloud ecosystem: it is the entity capable of making cloud services available to end users, controlling the available information and preserving privacy while still creating a better user experience.

The proposed concept can be summarized as wrapping traditional Cloud organization (SaaS/PaaS/IaaS) in modified logical layers. These should be managed by a top level control layer, built with IdM technologies, and by a lower layer consisting of standardized interfaces (which can be either achieved through a well-defined middleware). Each one of the blocks proposed in Fig. 7.4 can be considered as a separate sub-system, and its implementation may lead to user-centric clouds with different characteristics. The intended outcome of using IdM

as the driving technology is giving users the control over services and cloud information. By providing IdM as the glue layer between services and their interaction, we enable intrinsic user-centric mechanisms, especially concerning authentication and access management, along with privacy control of sensitive user information. We proposed this concept in [16], where the user-centric mechanisms are provided by IdM, and the lower layer semantics are provided through a distributed Operating System approach.

In this context, IdM allows the creation of long-lived trust relations between digital entities, and the creation of a reputation system allowing for distinct access control service provision taking in consideration provided resources and past behavior. We can describe the IdM subsystem as a set of services, which must have well known interfaces, in order to be reused throughout the entire architecture.

The key feature of this future vision is that, even in futuristic scenarios that involve current and upcoming technologies, it is possible to use the concepts presented in this Thesis to leverage metaphors and systems that provide added privacy, and integrate the new key aspects into user-centric operations that value the primary subject of privacy - the user.

7.4 Final Thoughts

We have looked at different aspects of privacy, and how they can work in a networked environment. We considered this specially important as we continue to move our interactions into the digital plane. But, privacy is and will continue to be a hard topic to tackle, given the surrounding confusion, and even antagonism, towards the concepts at stake.

When discussing privacy, there is always a balance to be struck: either between privacy and performance, privacy and usability, and as the social trends are now showing, between privacy and full disclosure. Unfortunately, in these discussions, there are still those who do not care about privacy. This can be a consequence of privacy suffering from a delayed cause-effect, but also of an ever growing desire for public interaction. This can lead to the discussion of whether privacy is really needed, especially in computer networks, which usually require a performance or usability compromise. However, we believe that this argument is moot because it does not exist unless the underlying architecture, framework or environment actually provides privacy. Privacy must come first, and by design. It is only when we have privacy-aware systems that we can decide whether or not to forfeit privacy, as a compromise for improved network conditions. Regardless of the incentives, if there is no privacy protection there is no privacy leverage, and the information is already jeopardized. For any of these discussions to be meaningful, the information balance must tip towards the user, which should have control over his information, rather than network entities and services (thus supporting the mentioned concepts of asymmetric knowledge).

The balance surrounding privacy is especially important in technical terms, because in order to provide privacy, we need privacy-aware networks and computing systems. To achieve this goal we need a model, architectural drivers and targeted solutions that relate the concepts of user and network privacy, all of which are conclusions supported by the work presented throughout the Thesis.

It is important to understand that we can only discuss whether or not privacy fits all situations and conditions, or when do we want or need to revoke it, if the user already enjoys the bias of privacy. If the user was not already protected, it is not necessary to decide to retain or revoke the privacy of past actions, since this was already imposed by the system.

From a technical perspective, to enable a choice between both conditions, either preserving or revoking privacy, requires technologies that support the concept of privacy. The direct conclusion is that privacy-aware (network) systems are a precondition towards allowing us the choice of upholding our Human Right for Privacy.

Bibliography

[1] Julien Abeillé, Rui Aguiar, Telemaco Melia, Ignacio Soto, and Patrick Stupar. Mobisplit: a scalable approach to emerging mobility networks. In *MobiArch '06: Proceedings of first ACM/IEEE international workshop on Mobility in the evolving internet architecture*, pages 17–22, New York, NY, USA, 2006. ACM.

[2] Alessandro Acquisti. Nudging privacy: The behavioral economics of personal information. *IEEE Security and Privacy*, 7:82–85, 2009.

[3] R.L. Aguiar, A. Sarma, D. Bijwaard, L. Marchetti, and P. Pacyna. Pervasiveness in a competitive multi-operator environment:the daidalos project. *Communications Magazine, IEEE*, 45(10):22–26, October 2007.

[4] Rui Aguiar, Hans Einsiedler, and R. Karrer. Daidalos: The operators vision of the next generation internet. *Infocom*, 45:23–29, April 2006.

[5] Liberty Alliance. Liberty alliance id-ff 1.2 specifications. `http://projectliberty.org/resource_center/specifications/liberty_alliance_id_ff_1_2_specifications/`. Last Checked: December, 2010.

[6] Xiangdong An, D. Jutla, and N. Cercone. A bayesian network approach to detecting privacy intrusion. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops*, pages 73 –76, December 2006.

[7] Frederik Armknecht, Joao Girao, Alfredo Matos, and Rui Aguiar. Method for establishing a secret key between two nodes in a communication network. Patent No. DE 102006036165, European Patent Office, June 2008. Also published as: US 2010008508, WO 2008014958, JP 2009545264, EP 2047631, CN 101496340.

[8] Frederik Armknecht, Joao Girao, Alfredo Matos, and Rui Aguiar. Method for protecting location information in wireless communication networks. Patent No. DE 102006036164, European Patent Office, February 2008. Also published as: WO 2008014971.

[9] Frederik Armknecht, Joao Girao, Alfredo Matos, and Rui L. Aguiar. Who said that? privacy at link layer. In *26th Annual IEEE Conference on Computer Communications*, Anchorage, Alaska, USA, May 2007. INFOCOM 2007. Minisymposium.

[10] T. Aura. Cryptographically generated addresses (cga). RFC 3972 (Experimental), March 2005.

[11] Hari Balakrishnan, Karthik Lakshminarayanan, Sylvia Ratnasamy, Scott Shenker, Ion Stoica, and Michael Walfish. A layered naming architecture for the internet. *SIGCOMM Comput. Commun. Rev.*, 34:343–352, August 2004.

[12] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield. Xen and the art of virtualization. In *SOSP '03: Proceedings of the nineteenth ACM symposium on Operating systems principles*, pages 164–177, New York, NY, USA, 2003. ACM.

[13] Marc Barisch. Modelling the impact of virtual identities on communication infrastructures. In *Proceedings of the 5th ACM workshop on Digital identity management*, DIM '09, pages 45–52, New York, NY, USA, 2009. ACM.

[14] Marc Barisch and Alfredo Matos. Integrating user identity management systems with the host identity protocol. In *The Fourteenth IEEE Symposium on Computers and Communications*, Sousse, Tunisia, July 2009. ISCC '08.

[15] Marc Barisch, Elena Torroglosa, Mario Lischka, Rodolphe Marques, Ronald Marx, Alfredo Matos, Alejandro Perez, and Dirk Scheuermann. Security and privacy enablers for future identity management systems. In *Future Network and Mobile Summit*, Florence, Italy, June 2010. MS'10.

[16] João Barraca, Alfredo Matos, and Rui Aguiar. User centric community clouds. *Wireless Personal Communications*, 58:31–48, May 2011.

[17] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2:46–55, January 2003.

[18] T. Berners-Lee, R. Fielding, and L. Masinter. Uniform resource identifiers (uri): Generic syntax. RFC 3986 (Proposed Standard), January 2005.

[19] Kim Cameron. The laws of identity. `http://msdn.microsoft.com/en-us/library/ms996456`, May 2005. Last checked: December, 2010.

[20] Scott Cantor, John Kemp, Rob Philpott, and Eve Maler. Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0, March 2005.

[21] David Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, Jan 1985.

[22] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24:84–90, February 1981.

[23] Jie Cheng, Russell Greiner, Jonathan Kelly, David Bell, and Weiru Liu. Learning bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, 137(1-2):43–90, 2002.

[24] David Clark, Robert Braden, Aaron Falk, and Venkata Pingali. Fara: reorganizing the addressing architecture. In *FDNA '03: Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture*, pages 313–321, New York, NY, USA, 2003. ACM Press.

[25] Gregory F. Cooper and Edward Herskovits. A bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992. 10.1023/A:1022649401552.

[26] Cherita Corbett, Raheem Beyah, and John Copeland. A passive approach to wireless NIC identification. In *IEEE International Conference on Communications*, June 2006.

[27] Jason Cornwell, Ian Fette, Gary Hsieh, Madhu Prabaker, Jinghai Rao, Karen Tang, Kami Vaniea, Lujo Bauer, Lorrie Cranor, Jason Hong, Bruce McLaren, Mike Reiter, and Norman Sadeh. User-controllable security and privacy for pervasive computing. In *Proceedings of the Eighth IEEE Workshop on Mobile Computing Systems and Applications*, pages 14–19. IEEE, March 2007.

[28] Daniel Corujo, Alfredo Matos, Rui Aguiar, Julien Abeille, and Telemaco Melia. Problem statement on common interfaces for local mobility management. Internet Draft, March 2007. Expired.

[29] European Council. Directive 95/46/ec of the european parliament and of the council of 24 octover 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, 1995. L281/31 - L281/39.

[30] European Council. Directive 97/66/ec of the european parliament and of the council of 17 of december 1997 - the data protection telecommunications directive. *Official Journal of the European Communities*, 1997. L281/31 - L281/39.

[31] European Council. Directive 2002/58/ec of the european parliament and of the council of 12 july 2002 concerning the processing of personal data and the protection of privacy in the electronic communications secto (directive on privacy and electronic communications). *Official Journal of the European Communities*, 2002. L201/37 - L201/47.

[32] European Council. Directive 2006/24/ec of the european parliament and of the council of 15 march 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications and amending directive 2002/58/ec. *Official Journal of the European Communities*, 2006. L105/54.

[33] IST FP6 Integrated Project Daidalos. Designing advanced network interfaces for the delivery and administration of location independent, optimised personal services. url: http://www.ist-daidalos.org.

[34] T Dalenius. Finding a needle in a haystack - or identifying anonymous census record. *Journal of Official Statistics*, 3(2):329–336, 1986.

[35] D.E. Denning. An intrusion-detection model. *IEEE Transactions on Software Engineering*, 13:222–232, 1987.

[36] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Proceedings of the 2nd international conference on Privacy enhancing technologies*, PET'02, pages 54–68. Springer-Verlag, 2003.

[37] R Dingledine, N Mathewson, and P Syverson. Tor: The second-generation onion router. *Proceedings of the 13th conference on USENIX Security Symposium-Volume 13*, page 21, 2004.

[38] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, August 2004.

[39] Matt Duckham and Lars Kulik. Simulation of obfuscation and negotiation for location privacy. *Spatial Information Theory*, 3693:31–48, September 2005.

[40] E. Hammer-Lahav (Editor). The oauth 1.0 protocol. RFC 5849 (Informational), April 2010.

[41] Sri Gundavelli (editor), Kent Leung, Vijay Devarapalli, Kuntal Chowdhury, and Basavaraj Patil. Proxy mobile ipv6. RFC 5213 (Proposed Standard), August 2009.

[42] A. Escudero, M. Hedenfalk, and P. Heselius. Location privacy in mobile internet - an extension to freedom network. *Internet Society Conference (INET2001)*, June 2001.

[43] A. Escudero and G.Q. Maguire Jr. Role(s) of a proxy in location based services. *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. (PIMRC2002)*, 3:1252–1257, September 2002.

[44] Antonio Escudero. Location privacy in ipv6: Tracking binding updates. *Tutorial at Interactive Distributed Multimedia Systems (IDMS2001)*, September 2001.

[45] Alberto Escudero-Pascual. *Privacy in the next generation Internet: Data protection in the context of European Union policy*. PhD thesis, Royal Institute of Technology - KTH / IMI, Sweden, December 2002.

[46] FIDIS. Future of identity in the information society. *IST-EU Network of Excelence*, 2005. http://www.fidis.net.

[47] Organisation for Economic Cooperation and Development (OECD). The guidelines on the protection of privacy and transborder flows of personal data. *OECD Council Recomendation*, 1980.

[48] Helsinki Instititute for Information Technology. Infrastructure for hip. `http://infrahip.hiit.fi/`.

[49] New European Schemes for Signatures, Integrity, and Encryption NESSIE. Performance of optimized implementations of the nessie primitives, version 2.0, IST-1999-12324, 2003.

[50] The Convention for the Protection of Human Rights and Fundamental Freedoms. The European Convention on Human Rights (ECHR). Council of Europe, 1953.

[51] Michael J. Freedman and Robert Morris. Tarzan: a peer-to-peer anonymizing network layer. In *Proceedings of the 9th ACM conference on Computer and communications security*, CCS '02, pages 193–206, New York, NY, USA, 2002. ACM.

[52] Xiaoming Fu, H. Schulzrinne, A. Bader, C. Hogrefe, D. Kappler, G. Karagiannis, H. Tschofenig, and Van den Bosch S. Nsis: a new extensible ip signaling protocol suite. In *IEEE Communications Magazine*, volume 43, pages 133–141. IEEE, 2005.

[53] Joao Girao, Bernd Lamparter, Marco Liebsch, and Telemaco Melia. A practical approach to provide communication privacy. In *IEEE International Conference on Communications*, Istanbul, Turkey, June 2006. ICC2006.

[54] Ian Glazer and Bob Blakley. Identity and privacy strategies: In-depth research overview. *Burton Group Report*, January 2009.

[55] David Goldschlag, Michael Reed, and Paul Syverson. Onion Routing for Anonymous and Private Internet Connections. In ACM, editor, *Communications - ACM*, volume 42, pages 39–41. Springer-Verlag, LNCS 2009, 1999. ISSN 0001-0782.

[56] Diogo Gomes and Rui Aguiar. Privacy through Virtual Hoarding. In *IEEE Globecom*, pages 1–6, San Francisco, USA, November 2006. IEEE.

[57] Diogo Gomes, Alfredo Matos, Emanuel Fonseca, and Rui Aguiar. Deploying and testing a ngn testbed : Ist daidalos testbed. In *Open NGN and IMS Testbeds Workshop at TRIDENTCOM 2009*, Washington, USA, April 2009. ONIT '09.

[58] Paul Graham. A plan for spam. `http://www.paulgraham.com/spam.html`, 2002. Last checked: October, 2010.

[59] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*, MobiSys '03, pages 31–42, New York, NY, USA, 2003. ACM.

[60] W. Haddad. Privacy for Mobile and Multi-homed Nodes: Formalizing the Threat Model. Internet Draft (Work in Progress), February 2005.

[61] P. R. Halmos. *Naive Set Theory, ISBN 0387-90092-6*. Springer, 1974.

[62] Ruth Halperin and James Backhouse. A roadmap for research on identity in the information society. *Identity in the Information Society*, 1:71–87, 2008.

[63] M. Hansen and M. Vanfleteren (KULeuven) H. Krasemann (ICPP)(Editors), Reviewers: P. Keller (Swisscom). "privacy and identity management for europe - prime white paper". *WP 15.1*, 2005. http://www.prime-project.eu.org.

[64] T. Heer and S. Varjonen. HIP Certificates. http://www.ietf.org/ internet-drafts/draft-ietf-hip-cert-00.txt, October 2008.

[65] D Hughesa and V Shmatikovb. Information hiding, anonymity and privacy: a modular approach. *Journal of Computer Security*, 12(1):3–36, 2004.

[66] IEEE Std 802.11-2007. IEEE Standard for Local and metropolitan area networks - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Technical report, IEEE Computer Society, 12 2007. Revision of IEEE Std 802.11-1999.

[67] IEEE Std 802.11i-2004. IEEE Standard for Local and metropolitan area networks - Amendment 6: Medium Access Control (MAC) Security Enhancements. Technical report, IEEE Computer Society, 07 2004.

[68] IEEE Std 802.21-2008. IEEE Standard for Local and metropolitan area networks - Part 21: Media Independent Handover, Jan 2009.

[69] Internet2. Internet2 middleware initiative. `http://www.internet2.edu/`. Last checked: December, 2010.

[70] Internet2 Middleware Initiative. Shibboleth. `http://shibboleth.internet2.edu/`. Last checked: December, 2010.

[71] IST Daidalos - Phase 2. Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimised personal Services (DAIDALOS). FP6-2005 Contract 02694. `http://www.ist-daidalos.org`.

[72] IST SWIFT. Secure Widespread Identity for Federated Telecommunications (SWIFT). EU FP7-2008 Contract 215832. `http://www.ist-swift.eu`.

[73] ITU-T Y.2011. Next Generation Networks - Frameworks and functional architecture models, General principles and general reference model for Next Generation Networks, 10 2004.

[74] P. Jayaraman, R. Lopez, M. Y. Ohba, A. Parthasarathy, and Yegin. Protocol for carrying authentication for network access framework. RFC 5193 (Proposed Standard), May 2008.

[75] Vitor Jesus, Susana Sargento, Daniel Corujo, Nuno Senica, Miguel Almeida, and Rui Aguiar. Mobility with qos support for multi-interface terminals: Combined user and network approach. *IEEE Symposium on Computers and Communications (ISCC'07)*, pages 325–332, July 2007.

[76] Joao Girao (Ed.). Swift, deliverable 203, first draft of the identity-driven architecture and identity framework. Technical report, IST SWIFT, 2008.

[77] David Johnson, Charles Perkins, and Jari Arkko. Mobility Support in IPv6 (MIPv6). RFC 3775 (Proposed Standard), june 2004.

[78] A.K. Jones and R.S. Sielken. Computer system intrusion detection: A survey. *Technical Report*, 2000.

[79] Kantara Initiative. Kantara Initiative: Shaping the Future of Global Identity. `http://kantarainitiative.org/`. Last checked: December, 2010.

[80] D. Kesdogan and C. Palmer. Technical challenges of network anonymity. *Computer Communications*, 29(3):306 – 324, 2006.

[81] Ponnurangam Kumaraguru and Lorrie Faith Cranor. Privacy indexes: a survey of westin's studies. In *School of Computer Science Technical Report*, page 22, Pittsburgh, USA, 2005. School of Computer Science, Carnegie Mellon University.

[82] David C. Kurtz. *Foundations of Abstract Mathematics*. McGraw-Hill International Editions, New York, USA, 1992.

[83] J. Laganier, T. Koponen, and L. Eggert. Host identity protocol (hip) registration extension. RFC 5203 (Experimental), April 2008.

[84] Ninghui Li, Tiancheng Li, and S Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. *Proceedings of IEEE International Conference on Data Engineering*, pages 106–115, 2007.

[85] M. Liebsch and L. Eggert. Host Identity Protocol (HIP) Rendezvous Mechanisms. Internet Draft (Work in Progress), July 2004.

[86] Linux Kernel. Linux containers - network namespace. `http://lxc.sourceforge.net`, 2008.

[87] Mario Lischka, Yukiko Endo, and Manuel Sánchez Cuenca. Deductive policies with xacml. In *Proceedings of the 2009 ACM workshop on Secure web services*, SWS '09, pages 37–44, New York, NY, USA, 2009. ACM.

[88] Teresa F. Lunt. Aggregation and inference: Facts and fallacies. *IEEE Symposium on Security and Privacy*, 0:100–102, 1989.

[89] Teresa F. Lunt. Aggregation and inference: Facts and fallacies. *Security and Privacy, IEEE Symposium on*, 0:102, 1989.

[90] Gabriel López, Cánovas, Antonio F. Gómez-Skarmeta, and Joao Girao. A swift take on identity management. *Computer*, 42(5):58–65, 2009.

[91] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1):3, 2007.

[92] Bradley Malin. k-unlinkability: A privacy protection model for distributed data. *Data Knowl. Eng.*, 64(1):294–311, 2008.

[93] Rodolphe Marques, Rui Ferreira, and Alfredo Matos. Cross layer privacy support for identity management. In *Future Network and Mobile Summit*, Florence, Italy, June 2010. MS'10.

[94] Alfredo Matos and Rui Aguiar. Mobility aware paths: The identity connection. In *Special Sessions of the 11th International Symposium on Wireless Personal Multimedia Communications*, Lapland, Finland, September 2008. WPMC'08.

[95] Alfredo Matos, Rui Ferreira, Susana Sargento, and Rui Aguiar. Virtual network stacks: From theory to practice. *Wiley Security and Communication Networks*, October 2011.

[96] Alfredo Matos, João Girão, Frederik Armnecht, and Rui Aguiar. Towards dependable networking: Secure location and privacy at link layer. *IEEE Wireless Communications Magazine*, 15(6), October 2008. Special Issue on Dependability Issues with Ubiquitous Wireless Access.

[97] Alfredo Matos, João Girão, Susana Sargento, and Rui Aguiar. Preserving privacy in mobile envinronments. In *Globecom '07*, pages 1971–1976, Washington D.C., USA, November 2007. Globecom2007.

[98] Alfredo Matos, Ricardo Pereira, and Joao Girao. Identity driven mobility architecture. In *Future Network and Mobile Summit*, Florence, Italy, June 2010. MS'10.

[99] Alfredo Matos, Justino Santos, Rui Aguiar, Joao Girao, and Marco Liebsch. Location privacy extensions for the host identity protocol. In *Revista do Departamento Electrónica e Telecomunicações*, volume 4, nº 8, Universidade de Aveiro, Portugal, 2007. DET.

[100] Alfredo Matos, Justino Santos, João Girão, Marco Liebsch, and Rui Aguiar. Hip privacy extensions - version 00. Internet Draft, August 2005. Expired.

[101] Alfredo Matos, Justino Santos, João Girão, Marco Liebsch, and Rui Aguiar. Hip privacy extensions - version 01 (revised). Internet Draft - Expired, March 2006. Expired.

[102] Alfredo Matos, Justino Santos, Susana Sargento, Rui Aguiar, Joao Girao, and Marco Liebsch. Hip location privacy framework. In *First ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture*, San Francisco, USA, December 2006. MobiArch2006.

[103] Alfredo Matos, Susana Sargento, and Rui Aguiar. Embedding identity in mobile environments. In *Second ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture*, Kyoto, Japan, October 2007. MobiArch2007.

[104] Alfredo Matos, Susana Sargento, and Rui Aguiar. Prived: A privacy model for heterogenous mobile networks. In *First International Workshop on Privacy Management in Mobile Applications (PriMo2011), Held at IFIPTM 2011*, Copenhagen, Denmark, June 2011.

[105] Alfredo Matos, Susana Sargento, and Rui Aguiar. Waypoint Routing: A Network Layer Privacy Framework. In *Globecom'2011*, Houston, Texas, USA, December 2011.

[106] Christoph P. Mayer. Security and privacy challenges in the internet of things. *Proceedings of KiVS Workshop on Global Sensor Networks GSN09*, 17, 2009.

[107] Alfred J. Menezes, Scott A. Vanstone, and Paul C. Van Oorschot. *Handbook of Applied Cryptography*. CRC Press, Inc., Boca Raton, FL, USA, 1996.

[108] Merriam-Webster. Merriam-webster online dictionary. `http://www.merriam-webster.com/`, July 2010. Last checked: July 2010.

[109] Microsoft Developer Network. Windows cardspace. `http://msdn.microsoft.com/en-us/library/ms733090(VS.90).aspx`, 2007. Last checked: December, 2010.

[110] Microsoft Developer Network (MSDN). Introduction to windows live id. `http://msdn.microsoft.com/en-us/library/bb288408.aspx`, February 2008. Last checked: December, 2010.

[111] Carlos Molina-Jimenez and Lindsay Marshall. True anonymity without mixes. *Internet Applications, IEEE Workshop on*, 0:32, 2001.

[112] R. Moskowitz. Host Identity Protocol Architecture. RFC 4423 (Proposed Standard), may 2006.

[113] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson. Host Identity Protocol. RFC 5201 (Experimental), April 2008.

[114] P. Nikander, J. Arkko, and B. Ohlman. Host Identity Indirection Infrastructure (Hi3). In *The Second Swedish National Computer Networking Workshop*, November 2004.

[115] ns 2. The network simulator 2. `http://www.isi.edu/nsnam/ns/`, as in June 2006.

[116] OAuth. `http://oauth.net/`. Last checked: December, 2010.

[117] T. Okagawa et al. Ip packet routing mechanism based on mobility management in a ip based network. *8th International Conference on Intelligence in next generation networks*, 2003.

[118] Judith S. Olson, Jonathan Grudin, and Eric Horvitz. A study of preferences for sharing and privacy. In *Extended abstracts on Human factors in computing systems*, CHI '05, Portland, OR, USA, April 2005.

[119] ITU-T Workshop on Ubiquitous Network Societies. Privacy and ubiquitous network societies. ITU-T, April 2005.

[120] A Pashalidis. Measuring the effectiveness and the fairness of relation hiding systems. *Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE*, pages 1387–1394, 2009.

[121] R. Koodli, Ed. Fast handovers for mobile ipv6. RFC 4068 (Experimental), July 2005.

[122] David Recordon, Johnny Bufu, Josh Hoyt, Brad Fitzpatrick, and Dick Hardt. OpenID Authentication 2.0, December 2007.

[123] Ricardo Azevedo (Ed.). Swift, deliverable 403, swift mobility architecture. Technical report, IST SWIFT, August 2009.

[124] Ronal Marx (Ed.). Swift, deliverable 302, specification of general identity-centric security model that supports user control of privacy. Technical report, IST SWIFT, January 2009.

[125] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. Sip: Session initiation protocol. RFC 3261 (Proposed Standard), June 2002.

[126] Rui Ferreira. Privacy and identity selection. In *Departamento Electronica, Telecomunicações e Informática - Tese Mestrado Integrado*, Aveiro, Portugal, June 2008.

[127] Rui Ferreira, Alfredo Matos, Susana Sargento, and Rui L. Aguiar. Enabling identity aware applications. In *Proc Conf. sobre Redes de Computadores - CRC*, Oeiras, Portugal, October 2009.

[128] Mehran Sahami, Susan Dumais, David Heckerman, and Eric Horvitz. A bayesian approach to filtering junk e-mail. *AAAI Workshop on Learning for Text Categorization*, July 1998.

[129] Amardeo Sarma, Alfredo Matos, João Girão, and Rui Aguiar. Virtual identity framework for telecom infrastructures. In *Wireless Personal Communications*, Netherlands, February 2008. Springer. ISSN 0929-6212.

[130] S. Schmidt et al. Turfnet: An Architecture for dynamically composable networks. *Proceedings in 1st IFIP TC6 WG6.6 Workshop on Autonomic Communication (WAC 2004)*, 2004.

[131] Bruce Schneier. Anonymity and the tor network. `http://www.schneier.com/blog/archives/2007/09/anonymity_and_t_1.html`, September 2008. Last checked: June, 2010.

[132] A Serjantov and G Danezis. Towards an information theoretic metric for anonymity. *Privacy Enhancing Technologies*, Jan 2003.

[133] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. *Privacy Enhancing Technologies*, 2482:259–263, 2003.

[134] Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[135] Hesham Soliman, Claude Castelluccia, Karim El Malki, and Ludovic Bellier. Hierarchical Mobile IPv6 Mobility Management (HMIPv6). RFC 4140 (Proposed Standard), august 2005.

[136] Daniel J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154:477, January 2006.

[137] Daniel J. Solove. *Understanding Privacy*. Harvard University Press, 2008.

[138] Lara Srivastava and Tim Kelly. digital.life. *ITU Internet Report*, December 2006.

[139] I. Stoica et al. Internet indirection infrastructure. *Proceedings in ACM SIGCOMM Conference (SIGCOMM'02)*, pages 73–88, August 2002.

[140] A. Stubblefield, J. Ioannidis, and A. Rubin. Using the Fluhrer, Mantin, and Shamir Attack to Break WEP. In *ATT Labs Technical Report - TD4ZCPZZ*, August 2001.

[141] L Sweeney. k-anonymity: A model for protecting privacy. *International Journal Of Uncertainty Fuzziness and Knowledge Based Systems*, 10(5):557–570, 2002.

[142] P Syverson, D Goldschlag, and M Reed. Onion routing for anonymous and private internet connections. *Communications of the ACM*, 42:39–41, 1999.

[143] The European Convention on Human Rights (ECHR). Article 8 - right to respect for private and family life. Council of Europe.

[144] T.Narten, E. Nordmark, and W. Simpson. Neighbor Discovery for IP Version 6 (IPv6). RFC 2461 (Proposed Standard), december 1998.

[145] Virtualbox. Virtualbox. `http://www.virtualbox.org/`, 2008.

[146] VMWare, Inc. Vmware workstation. `http://www.vmware.com`, 2010.

[147] Alan F. Westin. *Privacy and Freedom*. Atheneum, New York, USA, 1967.

[148] What Is My IP Address. How accurate is geolocation. `http://whatismyipaddress.com/geolocation-accuracy`. Last checked: June, 2010.

[149] Aukky Ylitalo and Pekka Nikander. Blind: A complete identity protection framework for end-points. *Security Protocols, Twelfth International Workshop, Cambridge*, April 2004.

[150] Alf Zugenmaier. The freiburg privacy diamond. In *Global Telecommunications Conference, Globecom'03*, volume 3, pages 1501–1505, San Francisco, USA, May 2003. Global Telecommunications Conference, Globecom'03.

[151] Alf Zugenmaier. Flasche - a mechanism providing anonymity for mobile users. In *Privacy Enhancing Technologies - 4th International Workshop*, pages 121–141, Toronto, Canada, May 2004. Privacy Enhancing Technologies - 4th International Workshop.