PRIVED: A Privacy Model for Heterogenous Mobile Networks

Alfredo Matos[1,2], Susana Sargento[1], and Rui Aguiar[1]

[1] Instituto de Telecomunicações, Universidade de Aveiro, Aveiro, Portugal,
{alfredo.matos|susana|ruilaa}@ua.pt
[2] Caixa Mágica Software, Lisboa, Portugal
alfredo.matos@caixamagica.pt

Abstract. We propose a network-oriented privacy model (PRIVED) composed by a well defined information model, using events, information sets and relationships to define the conceptual privacy relationships that can occur in the network. We propose formal rules and a network instantiation, using linkability and correlation as the main tools for network applicability. We also use the model to determine the best approaches towards privacy protection in the network, resulting in a vertical/horizontal network privacy dichotomy.

Keywords: Privacy, model, network, linkability, correlation, events, information set, relationships.

1 Introduction

Privacy is a complex and multi-disciplinary concept that covers technological, legal, social and even philosophical issues. When considering individual user privacy in modern networks, we must isolate different views, each with its own set of threats, turning privacy in the network into a tractable problem. In this paper, the concept of network privacy refers to the ability of accessing the network without unwillingly revealing information about user or device. Therefore it becomes important to understand how information is leaked (or revealed) by accessing the network.

An approach to network privacy is to define a privacy model that covers the information related to the different network assets. It should identify what information jeopardizes privacy, so that a clear set of mitigation measures can be proposed. Different privacy models exist using different concepts, such as databases [5], anonymity [1, 6], network [8] or bayesian approaches [7]. However, most of them do not deal directly with the network, creating a gap. Network privacy requires a specific model, that identifies network based privacy threats. To close this gap, we propose a novel network oriented privacy model based on *Events* as a bridge between conceptual approaches and practical network applications, leading to *Relationships* and *Information Sets*. The relationships can be seen as the privacy threats, whereas the information sets represent the collected information about users, all of them using events as the basic operation. These events, which can be extracted from network observation (e.g. exchanged data packets) are at foundation of the presented work, providing a simple information model that uses resorts to event information extraction and correlation. The model is then adapted to the network, in Sec. 3, through the formalization of Linkage and Correlation as network observations, especially considering mobile networks, which are a key aspect in this respect, as they increase information correlation, besides increasing the means for tracking the location of end-points. Such formalizations provide the starting point for defining identifier based threats, as well as protection mechanisms. By defining the means for breaking relationships between events and containment barriers for the information sets, discussed in Sec. 4, these concepts can be used to define abstract privacy protection mechanisms for the network stack. We conclude by reviewing the contributions of proposed work in Sec. 5.

2 Privacy Model

Packet based networks can be characterized by information blocks, exchanged and stored across the network, depending on context and purpose. We define a network oriented model where information can be extracted

from these blocks, or even from relationships between different blocks, threatening privacy by attributing data to identifiable subjects - the users. Based on network protocol semantics, we assume that information can be split into data, a generic block that by itself can have no particular meaning towards a subject, and identifiers, which can uniquely identify or represent the subject to whom the data belongs, similar to *Quasi-Identifiers* [2], which can be used for privacy in databases [5]. Therefore, it is possible to separate privacy in the network into identifier threats and data mining threats. We focus on identifiers, which can be used to relate different information blocks (Figure 1). Once a relationship is established - which is where the privacy threat resides - a larger set is built, resulting in aggregated information. This view represents the notion of a user as the knowledge gained around a certain subject. By relying on concept that it is possible to gather information into specific sets, we propose a Privacy Event Driven (PRIVED) model, that uses the concepts of events, information sets and the relationships to model privacy breaches on the network.

$$e_x = (identifier, [subject,]information) \tag{1}$$

To understand how these relationships are established, where information stems from, and how to build and maintain a larger notion of relationships, we propose a three-fold approach that tries to model network behavior. We use discrete *Events* (1), e_x, which convey information blocks along with potential information for relationship establishment, leading to *Relationships*, $\varphi(e_x, e_y)$, and *Information Sets* (IS). IS are an aggregate set of information that is composed by different information blocks (or events) that share a relationship greater than a define threshold, t, as show by 2, whereas relationships are the links between events or sets that enable building the IS. Events matter because they can model the occurrence of information in the network. Also, 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.

$$S = \{e_1, e_2, e_3, \ldots, e_n\}, \varphi(e_x, e_y) \geq t \tag{2}$$

2.1 Events and Information Sets

$$\varphi_{(e_1,e_2)} \geq t \Rightarrow e_3(e_1, e_2, \varphi_{e_1,e_2}) \tag{3}$$

Using events, it is possible to link information and establish relationships that jeopardize user privacy through direct observations or probabilistic evidence gathering (taking into account different network conditions, information relevance and even previous observations), assuming that if if a relationship with value greater than the defined threshold, we define an event that relates them (3).

Event correlation also complements the construction of the information set, which can be done through linkage and correlation on the network. When an identifier refers to a subject or user, it allows building an information set around it. Therefore, if two identifiers relate to the same subject, we can expand the information set (4). This defines the correlation process, where identifiers become the common ground to threaten user privacy and can hold true for any type of data (e.g. addresses, location, bandwidth, services,

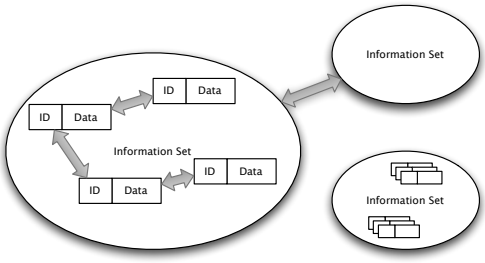


Fig. 1. Abstract information model

eye color, height, etc.). Similarly, when two different sets share an event, or there is a relationship between two events, they become the same expanded IS, as detailed in (5).

$$e_x \in s_k, e_y, \varphi(e_x, e_y) \Rightarrow e_y \in S_k \tag{4}$$

Part of the proposed model includes formal rules for building the information set using Set Theory [3], which is a generalization of network identifier based rules that we presented in [4]. We also express several cases where graph theory is superior for handling identifier relationships within sets, and how to use the model for expressing formal rules of existing work (thus incorporating the Freiburg Privacy Diamond [8]).

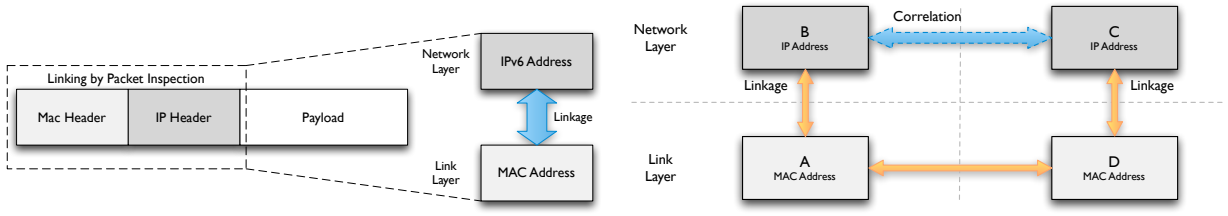
$$e_i \in S_i, e_j \in S_j : \varphi(e_i, e_j) \Rightarrow S_k = S_i \cup S_j \tag{5}$$

3 Privacy in the Network

$$e_x(e_{L2}, e_{L3}) \tag{6}$$

In network terms, events are data packets: each one has unique identifiers or circumstances, containing data (or payload) which can be linked or mined for more information. Therefore, the established relationships lead to linkage between independent (discrete) packet observations in the network. We can highlight several means of identifier driven correlation, such as time-based, contextual, or layered. The plethora of available identifiers, such as link, network or transport layer identifiers, create several threats, and can be aggregated into a network based information set, jeopardizing user privacy. Furthermore, the importance conferred to identifiers provides a strong connection between the model and mobile networks, given that the relationships between identifiers and network properties define several threats only important in mobile environments, such as location (provided by different identifier types). The mentioned relationships can follow two distinct approaches: relationships can be established based on a single event that conveys distinct information (identifiers), or multiple events related together through network techniques or inference.

This allows defining linkage according to two vectors: *Vertical Linkage* as the correlation of identifiers present in a single event, typically by observing multiple layer identifiers in one packet, and *Horizontal Linkage*, dealing with relationships between identifiers of the same nature (belonging to the same layer, horizontally on the network) and leading to relationships between multiple events. These definitions make a concrete difference in the network, directly mapping to Fig. 2(a), which represents linking between link layer identifiers and IP addresses shown in Fig. 2(a), formally described as an event that relates two other events (6). A similar process for identifiers in the same layer is shown in Fig. 2(b). These different relationships are specially important in mobile networks, where the means of correlation are supported by location properties and exchanged information, which provide more information for both vertical and horizontal linkage (e.g. events sharing similar locations can be bound together). The proposed network properties are consequences of the privacy model, using the event rules to define these types of relationships on the network, thus bridging the practical and theoretical approaches. In the proposed work, these rules are completed by a study on the different identifiers properties (e.g addressing structures) and their contributions to the IS.



(a) Cross Layer Linkage by by packet inspection.

(b) Same layer linkage by horizontal inference.

Fig. 2. Vertical and horizontal linkage on the network.

4 Network Privacy Protection

Understanding (and proposing) privacy models is only important when they contribute towards privacy preservation. So far we've isolated two different ways to undermine privacy: the establishment of relationships between events and the aggregation of those events into an IS. To protect privacy, it is necessary to break the relationships between events thus preventing them from being aggregated into an IS.

We can break relationships by protecting identifiers, avoiding any privacy leakage. Because most threats on network identifiers stem from the vertical or horizontal threats, we propose using vertical or horizontal barriers, promoting vertical separation. With a horizontal barrier (e.g. opaque tunnels like TLS or IPSec) between different layers, an identifier conveyed by lower layers is hidden and unavailable to any attacker on the network, thus providing a horizontal separation mechanism as a shield between layers. Alternatively, we can protect the user's privacy as a whole by preventing information from being aggregated in the IS. This can be done with vertical approaches that enable controlling the disclosed identifiers (e.g. pseudonymity), therefore defining boundaries and rules for the IS. Therefore, privacy mechanisms can be coupled with different techniques, such as identity, to reduce the size of IS by using transient identifiers (e.g. network pseudonyms) that loose relative importance and contribute to much smaller information sets, but require a vertical coordinated approach.

5 Conclusion

The PRIVED model introduces the concepts of Event, Relationship and Information Set. The IS frames user information, directed towards the network, while the events translate network interactions into data blocks, related through relationships that threaten privacy. However, the model is directly applicable to network paradigms, through the defined information model. By discussing how linking and correlation occurs on the network, we show the usefulness of the model, and derive how to provide network privacy. We establish horizontal and vertical relationships between network identifiers that result in privacy loss, along with conceptual approaches to mitigate such threats. One of the most important conclusions of identifiers and network as a whole, is that protocols can compromise above layers, thus jeopardizing privacy throughout the network stack using the vertical and horizontal conceptualizations. This promotes a two-fold approach where two dimensions present a conceptualization of privacy threats and solutions: the vertical aspects that are directly related to the user and must be handled vertically, and the horizontal aspects deal with each individual protocol or layer, and can be handled orthogonally (provided there is a vertical solution).

References

1. Chaum, D.L.: Untraceable electronic mail. Commun. ACM 24, 84–90 (February 1981)
2. Dalenius, T.: Finding a needle in a haystack. Journal of Official Statistics 3(2), 329–336 (1986)
3. Halmos, P.R.: Naive Set Theory, ISBN 0387-90092-6. Springer (1974)
4. Matos, A., Giro, J., Sargento, S., Aguiar, R.: Preserving privacy in mobile envinronments. In: Globecom '07. Globecom2007, Washington D.C., USA (November 2007)
5. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal Of Uncertainty Fuzziness and Knowledge Based Systems 10(5), 557–570 (2002)
6. Syverson, P., Goldschlag, D., Reed, M.: Onion routing for anonymous and private internet connections. Communications of the ACM 42, 39–41 (1999)
7. Xiangdong, An et al: A bayesian network approach to detecting privacy intrusion. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops pp. 73 –76 (December 2006)
8. Zugenmaier, A.: The freiburg privacy diamond. In: Global Telecommunications Conference, Globecom'03. vol. 3, pp. 1501–1505. Global Telecommunications Conference, Globecom'03, San Francisco, USA (May 2003)