TRANSACTIONS ON DATA PRIVACY 11 (2018) 239–277

EPIC: a Methodology for Evaluating Privacy Violation Risk in Cybersecurity Systems

Sergio Mascetti*, Nadia Metoui*, Andrea Lanzi*, Claudio Bettini*

*Università degli Studi di Milano, Department of Computer Science

Email:{sergio.mascetti, nadia.metoui, andrea.lanzi, claudio.bettini}@unimi.it

Received 26 September 2017; received in revised form 2 February 2018 and 26 May 2018; accepted 27 June 2018

Abstract. Cybersecurity Systems (CSSs) play a fundamental role in guaranteeing data confidentiality, integrity, and availability. However, while processing data, CSSs can intentionally or unintentionally expose personal information to people that can misuse them. For this reason, privacy implications of a CSS should be carefully evaluated. This is a challenging task mainly because modern CSSs have complex architectures and components. Moreover, data processed by CSSs can be exposed to different actors, both internal and external to the organization. This contribution presents a methodology, called *EPIC*, that is specifically designed to evaluate privacy violation risks in cybersecurity systems. Differently, from other general purpose guidelines, *EPIC* is an operational methodology aimed at guiding security and privacy experts with step-by-step instructions from modeling data exposure in the CSS to the systematical identification of privacy threats and evaluation of their associated privacy violation risk. This contribution also shows the application of the *EPIC* methodology to the use case of a large academic organization CSS protecting over $15,000$ hosts.

Keywords. cybersecurity System; Privacy violation risk; Privacy impact assessment.

1 Introduction

Privacy policy makers and data protection authorities all over the world are considering the impact on privacy of the large amount of identifiable sensitive data that are being collected and processed by public and private organizations. This is mainly the result of the adoption of new technologies like mobile and pervasive systems, social networks, and big data analytics, but also the evolution of technologies applied in surveillance and cybersecurity systems. An example of regulation activity motivated by these concerns is the EU General Data Protection Regulation, adopted in May 2016 [37]. While regulations differ in different countries, some general principles are shared; for example, user informed consent remains a pillar, and de-identification, despite the limits of anonymization techniques, is still considered a way to avoid or at least mitigate privacy violation risk [17]. Another shared recommendation to organizations deploying complex automated processes handling large amounts of personal data is to systematically and thoroughly analyze how the process affects the privacy of the individuals involved and evaluate the risks in order to

identify appropriate mitigation actions. This analysis is often called Privacy Impact Assessment (PIA) and it is in some cases a legal obligation as a necessary component in a *privacy by design* approach. However, its value goes beyond the design phase since it is also highly valuable when evaluating the compliance of already existing systems as well as when comparing the privacy risks of alternative systems.

Several documents exist guiding the experts in privacy impact assessments, but they usually consist of high-level guidelines instead of step-by-step instructions, partly motivated by the fact that they are sector independent. Indeed, the importance of designing sectorial PIA methodologies emerge in recent documents by EU data protection authorities [16]. While the interest is currently mostly focused on sectors like healthcare, e-commerce, finance, and insurance, less attention is paid to cybersecurity systems. These systems handle large amounts of sensitive information as, for example, the data obtained by monitoring employees PCs, mobiles and in general the whole organization network traffic [16]. In the last decade, cybersecurity systems have been increasing their strategic role for the protection of the IT infrastructure of industries and organizations. The wide adoption of digital technologies to control even critical infrastructure and the extension of organizational IT systems to include mobile and IoT devices have increased the attack surface and the impact that cyber attacks can have. This led to a significant increase in the complexity of cybersecurity systems in terms of components, architecture, amount of data being analyzed, and personnel involved in managing the systems.

The role of CSS with respect to privacy is twofold. On one side, CSS are an essential tool to prevent privacy violation, e.g., by avoiding unauthorized access to data. On the other hand, CSS often process a large amount of personal data, e.g., by monitoring network traffic, and hence they can pose a privacy threat. In general, privacy leaks from CSS can lead to discrimination in the workplace affecting both the relationships among colleagues and between the employee and the management, including effects on professional career. Privacy leaks from CSS can also affect external subjects, e.g., customers, with effects similar to the ones resulting from the release of private data through different channels. Among many others, we report some examples of possible problems arising from data leaked in CSS.

- Blackmailing. Alice works in the security team and has access to the organization's email logs; she finds evidence that her colleague Bob has an extramarital affair. Alice may then blackmail Bob threatening him to pass this information to his wife.
- Discrimination. During a routine inspection on the firewall, Carl discovers that his colleague David frequently accessed an event listing site known to be popular among gay people. Carl shares this information with his boss who on this basis discriminates David at the workplace.
- Identity theft. Eve works for a company that offers a web service. A file uploaded by a user is marked as 'suspicious' by the anti-virus and is then sent to Eve for investigation. The file actually contains Frank's ID card, credit card, and the SSN. Eve uses this information to impersonate Frank to gain financial advantage.

While in the above examples there is a direct damage to the person whose privacy is violated, there is also an indirect impact on the organization running the CSS, which is responsible for properly handling private data.

An accurate evaluation of privacy violation risks in a cybersecurity system is important for at least three reasons:

a) it identifies the gaps with respect to the applicable regulation so that appropriate remediation actions can be taken to achieve compliance;
b) it shows the responsibility of personnel like security, system, and network administrators in terms of personal data access, suggesting role-specific training and screening;
c) it highlights data collection practices that may make employees worry about their privacy and as a result, it can be an incentive for them to circumvent some of the cybersecurity mechanisms.

This paper presents the *EPIC* (Evaluating Privacy violation rIsk in Cybersecurity systems) methodology, that is composed of four steps and guides a privacy expert, with the collaboration of security experts from the organization running the system, to the identification of the main privacy threats, and to the assignment of a privacy violation risk value to each of them. Despite EPIC supports modeling of many aspects related to personal data handling, it is not intended as a complete PIA methodology, but rather an auxiliary tool specialized for the type of data exposures that can occur in a CSS. The proposed methodology supports both qualitative and quantitative risk values, the latter being preferable when it is possible to quantitatively assess how much a privacy threat would impact on the organization, for example in terms of monetary loss. The resulting evaluation can be used to prioritize mitigation actions to achieve legal compliance as explained in point a) above. Since training, and more generally trust, in a specific personnel role, is not considered until mitigation task prioritization, the evaluation is useful for point b) as well. Finally, our methodology can be used to compare different cybersecurity systems in terms of privacy implications, and possibly to design new cybersecurity systems that can effectively combine built-in privacy preserving features with protection from cyber attacks, addressing also point c) above. The methodology is illustrated through a running example and then applied in a use case considering the actual cybersecurity system of a large academic organization managing over 15,000 hosts.

The paper is structured as follows. We present an overview of the related work in privacy and security risk assessment in Section 2. In Section 3 we describe our privacy violation risk evaluation methodology and explain its three first steps. Section 4 is dedicated to the fourth step of the methodology dealing with the assignment of risk values and prioritizing mitigation actions, and Section 5 to the application of the methodology to the selected use case. We will conclude with a discussion in Section 6.

2 Related Work

A lot of research has been conducted in the last decades on various aspects of privacy, including the identification of privacy threats related to the use of technology, mitigation techniques, and methods to evaluate the risk of privacy violations. Considering this last point, most contributions proposing methodologies to analyze privacy threats mainly focus on general personal data collected as part of different applications, including e-health, geo-location apps, social networks, finance, and marketing. To the best of our knowledge, the only work in the literature that analyses the problem of privacy violations in cybersecurity systems is a survey paper by Toch et *al.* [39]. The survey proposes a new categorization of cybersecurity systems that help the privacy analysts to identify the personal data that these systems may expose to unauthorized parties. Our work builds on this categorization but takes the proposed analysis to a deeper and more operational level with the main goal of evaluating and comparing the risk of the identified privacy threats. Our methodology

considers also aspects like the adversary's knowledge, the capability to access the data, the amount of data leaked, the number of users involved, and other factors that determine the impact of a privacy threat. With respect to the survey that considered also cybersecurity systems for new ecosystems like mobile and IoT systems, we focus on organizational cybersecurity systems and test our proposed methods in a case study involving the CSS of a large organization.

In this section, we present the related work in three research directions: formal methods to measure privacy (Section 2.1), methodologies to assess security threats (Section 2.2) and methodologies to assess privacy threats (Session 2.3).

2.1 Privacy metrics

Various privacy metrics have been proposed in the literature to estimate the likelihood of an adversary of learning a private sensitive information when getting access to a given dataset (i.e., obtaining the identity of an individual and associated sensitive information). For example, since anonymity prevents privacy violations, several metrics have been proposed to quantify the level of anonymity of a dataset [6, 27, 24]. Extensions of these metrics have been proposed to evaluate anonymity in different data sharing contexts including location-based service requests [2]. However, their value is somehow limited by the problem of evaluating the adversary's knowledge which can determine which information can actually re-identify individuals. When identification cannot be successfully prevented, various sensitive data obfuscation techniques and related privacy metrics have been proposed. Some metrics measure the distortion or generalization applied to the data, and hence the probability of the adversary to infer the actual sensitive information. Other metrics are based on the notion of *indistinguishability* with differential privacy metrics [14] being an example. A quite comprehensive list of the privacy metrics that have been proposed in the literature can be found in [40]. Finally, there are valuable attempts to provide guidance in the application of privacy enhancing technologies (PET), often related to the above-mentioned metrics[1][15].

Some of these metrics (and related PETs) may be applied also in the context of cybersecurity systems; For example, some anonymity metrics may be used to evaluate how anonymous is a dataset of security alert logs, and some differential privacy notions may be used to measure the probability of privacy leak in releasing a statistically perturbed Web site access log. However, none of them in isolation seems appropriate to measure the general privacy violation risks involved in running a cybersecurity system. This is partly due to the fact that the validity of these metrics is dependent on specific assumptions on the considered data sharing model while typical cybersecurity systems have many different components that process and store data, complex architecture and data flows, and data access by users with different roles. This complexity calls for a principled but more high-level approach to privacy threat assessment.

2.2 Security threat assessment methodologies

Before considering privacy assessment methodologies we briefly report some methodologies adopted for security risk assessment since this is a related and more established field of investigation. Security threat analysis is a common step in the secure software development life-cycle. In the literature, we find several tools and methodologies such as the OCTAVE

[1] https://www.privacypatterns.org/

method [4], ISRAM [22], and the Common Vulnerability Scoring System (CVSS) [29] only to cite a few. Among the most widely used, the STRIDE model was proposed by Microsoft [19] as a security threat identification process, used to assist engineers to consider security aspects during the development of a software product. This process starts by analyzing the information flow within a system and then modeling system's components using Data Flow Diagrams (DFD); a list of possible security threats is identified for each of the components. STRIDE classifies security threats into six categories (Spoofing, Tampering with data, Repudiation, Information disclosure, Denial of service, and Elevation of privileges). This model-based analysis has inspired the methodology that we are proposing. Indeed, we extend the DFD notation to better model the system components and focus on the privacy threat identification for each component. STRIDE is often used with the threat evaluation model DREAD to assess security risks [35]. DREAD proposes to rate security threats by computing a score based on five criteria (Damage, Reproducibility, Exploitability, Affected users, and Discoverability). This score implicitly expresses the likelihood and severity aspects of a security threat. A similar approach is proposed in our methodology for privacy violation risk assessment.

2.3 Privacy threat assessment methodologies

The first approaches to privacy assessment were mostly in the form of checklists with the goal of demonstrating legal compliance[7]. Privacy impact assessment (PIA) methodologies emerged later-on to refine these approaches. Several definitions have been given to PIA (see [21, 34, 20]). David Wright in [41] defines PIA as a methodology for assessing the impacts on privacy of a project, and for taking remediation actions to avoid or minimize negative impacts. Several governmental bodies such as the CNIL (France), NIST (USA), ICO (UK) and the EU Art.29 Working Party have proposed various PIA methodologies [8, 30, 20, 16]. These guidelines, although very useful to understand the goals of the assessment, do not guide an organization through the specific steps that should be performed. Among the works that contribute in this direction, Oetzel and Spiekermann present a seven steps methodology to support a complete PIA analysis and systematically match the threats and the appropriate countermeasure [32]. However, their approach only considers the impact of a privacy threat and not the probability of occurrence of the threat, which may lead to an incorrect overall risk estimation. Another aspect that has a relevant impact on the effectiveness of the guidelines is their specialization for a given sector. The methodologies mentioned above are designed for a generic privacy assessment, and consequently they may not be straightforwardly implemented when addressing the problem in a specific context. Indeed, the development of sector-specific PIAs is mentioned among the priorities in recent EU recommendations [16]. We found very few sector-specific approaches, among which a PIA framework for RFID based applications [11], and a PIA template for smart-grid and smart-metering systems [36].

While EPIC at its current stage is not a complete PIA methodology, it is quite comprehensive in considering many aspects related to personal data handling. Regarding the analysis of type of data and its contribution in risk evaluation, it should be noted that a CSS can monitor basically whatever goes through the organization network and possibly also all the operations performed on the computers; differently from an information system where it is possible to analyze specific types of data based on a database schema, we find mostly unstructured data in the logs of a CSS. Hence, we adopt the categorization of type of data proposed in [39], distinguishing, for example, mail and HTTP headers from their associated body content, and file names from file content. This categorization is then used for the

evaluation of the data re-identification power, as well as its sensitivity. Note that unstructured data, like email body content, could contain any kind of information, hence it should be considered as potentially very sensitive. The computation of a global privacy violation risk guided by EPIC also includes other aspects of data handling like consent, retention period, and intervenability. They are considered in Step 4 of the methodology as part of the compliance impact evaluation.

Besides PIA, other privacy assessment approaches adopted a requirement engineering perspective to promote the privacy by design principles [12, 31, 26, 13]. Among them, the closest to our proposal, despite not being specific to cybersecurity systems, is probably LINDDUN, a privacy threat analysis framework for software-based systems proposed by Deng et al. [13] and based on the STRIDE model [19]. Privacy threats in LINDDUN are identified through potential misuse scenarios (i.e., scenarios in which an adversary can violate privacy requirements upon accessing the data). Unfortunately, the processes of identification and analysis of misuse scenarios are not specified by the methodology but rely on the expertise of the analysts. LINDDUN does not provide a risk evaluation support either. On the contrary, in our approach, we consider as a threat any data disclosure that can reveal sensitive information about a respondent. Our methodology is specialized for cybersecurity systems and hence the identification of threats is well guided by security and privacy factors (e.g., adversaries' capabilities and knowledge, types of exposed data). We also propose a domain-specific risk assessment model evaluating the likelihood and severity of a threat.

3 Methodology

3.1 Overview

The EPIC methodology is organized in four steps as illustrated in Figure 1. The whole process requires the participation of a team, involving members with different expertise, namely privacy, and security, as well as personnel of the organization in which the CSS is deployed. Security experts of the team have a major role in Step 2 while privacy experts take the lead in Step 3 and Step 4. Step 1 (modeling the CSS) and Step 2 (identifying data exposures) require the collaboration of personnel of the organization in which the CSS is deployed. Indeed, information about the actual configuration of the CSS, the processes involved, as well as about the structure of the organization including users, system, network and security personnel must be acquired. In the following we use the term **expert** to refer to a person that contribute to the analysis following the methodology.

3.2 EPIC First Step: Model the cybersecurity system

The first step of the methodology aims to model the specific CSS under investigation. This step is particularly relevant for two reasons. First, we can expect that some of the experts involved in the privacy threat modeling process do not have the required knowledge about the system. For example, privacy experts are not expected to know which are the components of the CSS, how data flow in the system and which actors are involved. Second, an explicit system description helps the experts to collaborate and prevents misunderstandings among them. In our use-case, this step was completed by members of our team supported by system and security administrators from the institution running the CSS. Modeling a CSS as part of Step 1 must include the following aspects.

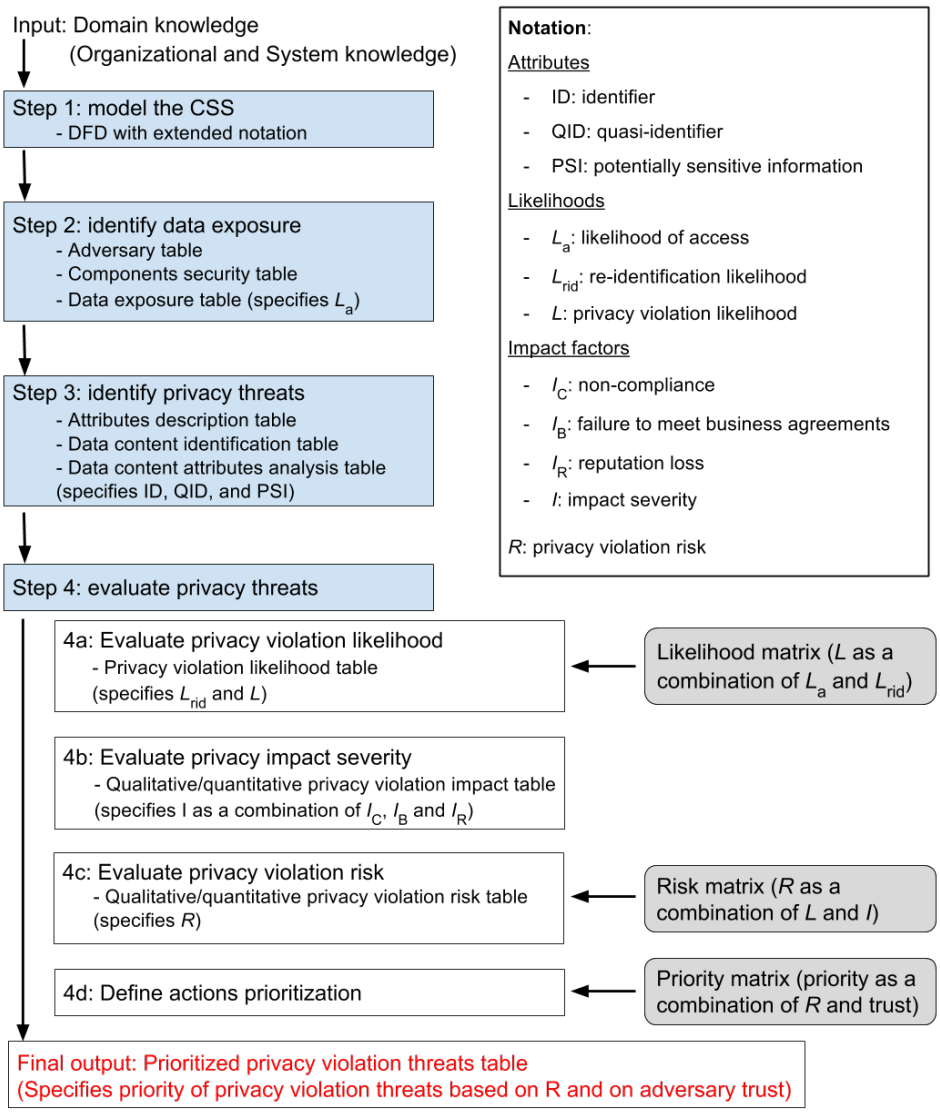

Figure 1: Methodology organization in four steps.

- System aspects: overall architecture and control processes.
- Data aspects: data flow and data storage.
- Functional aspects: users, roles, and functional processes.

A well-known formalism to represent data and functional aspects is Data Flow Diagram (DFD) [5]. This formalism allows us to represent four types of elements (see Figure 2): data flow is denoted with a full arrow, entities are denoted with a rectangle, storage with parallel line segments and functional processes (i.e., processes implementing the main system functionalities) with a circle. Note that a double circle is used to represent a complex

process i.e., a single component that represents several functional processes.

In this contribution, we extend DFD (and we call it **DFD+**) to also account for system aspects and hence to better detect situations in which data is exposed to an actor. We introduce four additional graphical symbols (see examples in Figure 2); a box represents a hardware component, an arrow with a small circle represents a physical channel connecting hardware components, a dashed arrow represents control flow and a dashed circle represents a control process that implements IT controls such as maintenance and security. In Example 1 we illustrate DFD+ and its use in CSS modeling as required by Step 1.

Example 1. Figure 2 describes an application level firewall. Data flows from the source entity *Network* to the destination entity *Security administrator*. Channel *C1* shows how data flows from *Network* to the *Firewall* hardware component. *C1* is marked as a physical channel and it is associated with a label (*Network Traffic*) that represents the type of data; in this case, it is the portion of network traffic that should be checked by the CSS. The logical destination of *C1* is the *Traffic Filtering* process. Upon detecting a security threat, this process sends the threat description to the data storage *DS1*. Note the different representation of *C2* with respect to *C1* due to the fact that *C2* is a logical channel.

From *DS1* data flows through the physical channel *C3* to another hardware component, *Remote Console*, where threat reports are organized for visualization by process *P2*. Then, *P2* sends this information through physical channel *C4* to the *security administrator* who is the destination entity and the main actor interacting with the CSS.

In this diagram we also model a secondary actor *system administrator* interacting with the hardware machine hosting the CSS (*Firewall*). The aim of the interaction is *Administration and Maintenance* and indeed *CP1* is marked with a dashed circle representing a control process. Similarly, the dashed arrows represent a control flow. Another control process (*CP2*) allows the *security administrator* to manage data storage *DS1*.

3.3 EPIC Second Step: Identify Data Exposure

The aim of the second step is to systematically identify all possible data exposures, i.e., situations in which data is disclosed to a potential adversary. A **data exposure** (or **exposure** for short) is identified by the component that is leaking data and by the adversary that can access that data; it is also characterized by other attributes that we specify in this section. **Component** refers to channels, processes and data storages identified in Step 1.

The term **adversary** refers to an actor identified in Step 1 as a subject normally interacting with the CSS or other people, which can either be external adversaries (e.g., a hacker violating a machine and accessing a data storage) or internal ones (e.g., a network administrator or other employees). An **adversaries table** (Table 1) containing a list of adversaries, each associated with a brief description, needs to be identified at the beginning of this step. In Table 1 we report this list considering our running example.

While the organization management and owner, in principle, may also be considered as adversaries, they usually do not have direct access to the system and the risk of them violating privacy can be easily evaluated by combining the risks computed for the operators that have direct access, since they are the ones that can take order from them. Moreover, the risk assessment is performed on their behalf and in their interest. This is similar to IT security threat modeling: system owners are usually not considered as potential attackers of their own system.

Step 2 also requires, for each component specified in the model, to identify the set of adversaries that can acquire data from that component. More specifically the aim is to

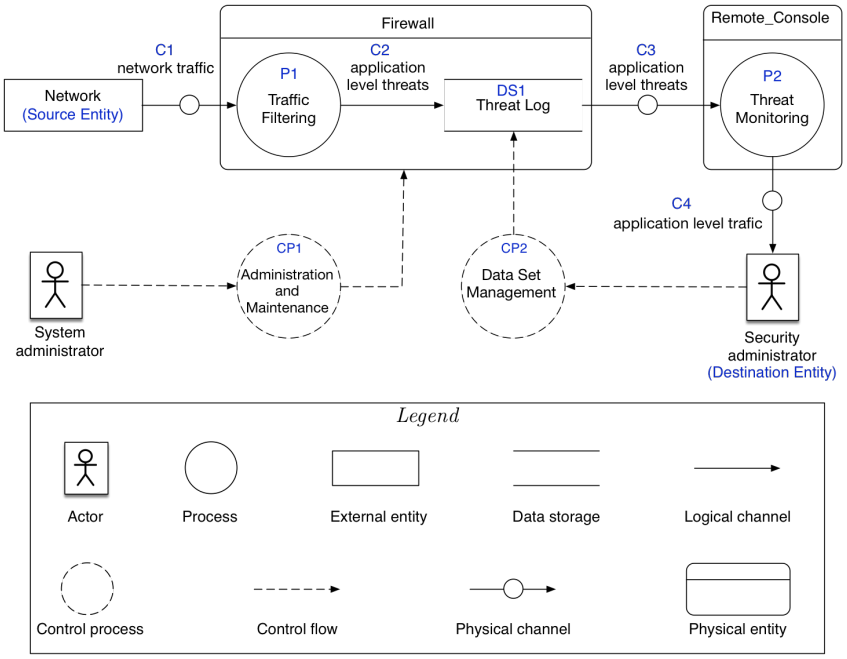

Figure 2: CSS modeling with DFD+ (running example).

Table 1: Adversaries (running example)

Adversary	**Description**
Security administrator	Their main tasks are to perform monitoring and investigation as well as the maintenance and configuration of the data storage (e.g. add, modify roles and privileges).
System administrator	Their tasks include maintenance of the system hosting the *Firewall* (e.g., troubleshooting, installing updates software/firmware)
Network administrator	Their main task is to ensure the correct functioning of the network (routing, DNS, etc.).
Other internal adversaries	Individuals attempting a nonauthorized access from inside the organization network.
External adversaries	Individuals attempting a non authorized access from outside the organization network.

identify the adversaries that:

- can access data transiting along a channel (either logical or physical) ;
- can read data from a data storage;
- can obtain data from a process, for example by observing the process output or altering the process behavior.

Clearly, different adversaries need different efforts to obtain data from a component. For example in the DFD depicted in Figure 2, the *Security Administrator* has the credentials to

access data storage *DS1*, hence the effort is negligible. Vice versa, an external adversary needs to violate a number of security systems, and resources are required to accomplish this task (economical, computational, knowledge). In principle, external adversaries may also obtain data from internal adversaries, and more generally adversaries may collude with each other. However, as in security threat analysis, we first assume that adversaries do not collude. More precisely, in our methodology we do not model the *effects* of collusion between parties but we do model the *likelihood* of collusion by means of the trust in an internal adversary. Indeed, the likelihood that an internal actor shares data with other adversaries (either internal or external) is related to the organization policies, legal agreements, and in general to the level of trust in that actor, which we model in the following (see Section 4.4).

We model the difference in the effort required to obtain data from a component through the **likelihood of access** (L_a) parameter, that, intuitively, is inversely proportional to the effort required to access to the component. The likelihood of access only takes into account the technical difficulties that a given adversary has to face to access a component; it does not depend on the willingness of the adversary to maliciously access that component or, in other words, the *trust* we have on the specific person or in personnel acting under a specific role (e.g., network administrators). These aspects are considered in Step 4.

We use the following five values for the likelihood of access:

- *Negligible*: it is technically very difficult for the adversary to access the component and it is highly unlikely that access can be obtained with a reasonable effort;
- *Low*: it is technically difficult for the adversary to access the component and a significant effort is required;
- *Medium*: it is technically possible for the adversary to access the component, but this requires moderate effort;
- *High*: it is technically easy for the adversary to access the component with a limited effort;
- *Authorized*: the adversary is authorized to access the component, hence no effort is required.

The likelihood of access depends on the security mechanisms (e.g., access control, encryption) implemented to protect that component. For this reason, for each component, we list the security mechanisms, together with their details, including, for example, which users are authorized to access by an access control system. This is called the **components security table** (see for example Table 2).

It is also clear that different exposures have different magnitudes and results in leaking different amount of data. To estimate the **exposure magnitude** different approaches should be used, depending on the type of component.

- *Exposure magnitude in data storage*. The amount of information incoming in the data storage, as well as the retention period of this information, can help to estimate the exposure magnitude. For example, if we know that approximately $1,000$ logs are recorded in a data storage daily and that retention period is 30 days, we can conclude that the data storage contains about $30,000$ logs.
- *Exposure magnitude in channels*. When data is exposed through a channel, we should take into account the data throughput (how much data is transmitted in the unit of

Table 2: Components security (running example)

Component	Authorized users	Security Certified	Security mechanisms
DS1	Security administrator	YES	Encryption, access control, authentication, firewall, NIDS
C3	None	YES	Firewall, NIDS, private network
P2	Security administrator	NO	Access control, authentication, firewall, NIDS

time) along the channel and an estimation of how long the adversary can listen to the channel.

- *Exposure magnitude in processes*. Similarly to channels, we should take into account how much data the adversary can access. This may depend on how long the adversary can access the process.

The results of Step 2 are reported in the **data exposures table** (for example Table 3) that lists, for each combination of components and adversaries, the likelihood of accessing data from that component by that adversary together with the exposure magnitude. Example 2 illustrates an instance of this process and the result is shown in Table 3 where a brief motivation is also reported for each row. These notes are very important to communicate with collaborators on the analysis (e.g., security expert) and they are also useful if the analysis has to be repeated again in the future. The motivation field is used in most of the other tables we present in our methodology, especially when the assessment relies on the expert's subjective judgment.

Note that the two leftmost columns of Table 3 are derived from previous tables (i.e., adversaries table and components security table) while the four columns on the right include new content. Henceforth we use the following notation: a double line (like between "Adversary" and "Exp." in Table 3) distinguishes previous content (on the left) from new one (on the right).

At the end of Step 2 all exposures with a *negligible* likelihood of access are cleared (e.g., those highlighted in Table 3), while the remaining ones are further investigated in Step 3.

Example 2. This example continues from Example 1 and presents the components security and data exposures tables for three components: *DS1*, *C3*, and *P2*.

From the CSS model, we know that the *security administrator* can access *DS1* and we report this information in the components security table (Table 2). In this example, it is relevant to know that the security of *DS1* has been certified, which means that a specific auditing, possibly including penetration attacks, has been performed. We report this information in the table. Finally, we list the security mechanism adopted to protect *DS1*: encryption, access control, authentication, firewall and NIDS. No user is authorized to access channel *C3*, whose security has been certified and that is protected by firewall, NIDS, and a private network. Finally, the *security administrator* can access *P2*, whose security has not been certified. This component is protected by access control, authentication, firewall and NIDS.

Based on the results of the components security table, we now show how to create the data exposures table considering four adversaries: *security administrator*, *system administrator*, *network administrator* and *external adversary*. The result is reported in Table 3.

Since the *security administrator* has access to the data storage *DS1*, the likelihood of access is reported as *authorized*. Instead, the *system administrator* is not authorized to access *DS1* but

Table 3: Data exposures (running example)

Cp.	Adversary	Exp.	L_a	Exp. Magn.	Motivation
DS1: Threat Log	Security admin.	$Exp1$	Authorized	Important $\approx 100k$ rec	Administrator of the DS (see DFD)
	System admin.	$Exp2$	Medium	*Same as above*	Can access machine but data is Encrypted
	Network admin.	$Exp3$	Negligible	*Same as above*	A network admin. has to elude the network protection bypass authentication and AC mechanisms and the data is encrypted
	Ext. adversary	$Exp4$	Negligible	*Same as above*	The adversary has to elude the network protection, bypass authentication and AC mechanisms, and the data is encrypted
C3: application level threats	Security admin.	$Exp5$	Negligible	Limited $\approx 20k$ rec	Need to bypass network protection
	System admin.	$Exp6$	High	*Same as above*	Can compromise the machine hosting the Firewall and listen to channel *C3*
	Network admin.	$Exp7$	High	*Same as above*	Have access to the Network equipment and can listen to channel *C3*
	Ext. adversary	$Exp8$	Negligible	Very limited $\leq 5k$ rec	The adversary has to elude the network protection, bypass authentication and AC mechanisms
P2: Threat Monitoring	Security admin.	$Exp9$	Authorized	Limited $\approx 30k$ rec	Can observe the output of process P2
	System admin.	$Exp10$	Low	*Same as above*	Should not be able to access, but security has not been tested
	Network admin.	$Exp11$	Low	*Same as above*	*Same as above*
	Ext. adversary	$Exp12$	Low	*Same as above*	*Same as above*

Table 4: Attributes description (running example)

Name	Description	Domain	Example
IP(out-dst)	The destination IP address of outgoing traffic	IP addresses	216.58.205.195
IP(in-src)	The source IP address of incoming traffic	IP addresses	192.30.253.112
IP(in-dst)	The destination IP address of incoming traffic	IP addresses	132.133.56.45
File	A file being transmitted	String of bytes	

has access to the physical machine hosting this component. By cracking data encryption (note in the components security table that *DS1* does implement encryption), the *system administrator* can obtain data from *DS1*, hence we associated this a *medium* likelihood of access. The effort required by the *external adversary* is even higher, as he needs to elude the security protections of the network (firewall, NIDS) to gain access to the machine hosting *DS1*, then bypass the authorization and access control mechanisms and decrypt the data. These security mechanisms have been certified (as reported in Table 2) and hence the likelihood of access by the *external adversary* is marked as *negligible*. The likelihood of access by a *network administrator* is also *negligible*. Indeed, since *DS1* is well configured and security tested, this adversary has to elude all the security mechanisms and make a considerable effort in order to gain access to data from *DS1*.

Regarding *C3*, no user is authorized to access. Since the component's security is certified, we can assign *negligible* likelihood of access to *external adversary*. In this case, the *security administrator* needs basically the same effort as an external adversary to access *C3*, so it is also marked as *negligible*. The same does not hold for the *system administrator*, who administers the firewall machine and hence can listen to channel *C3* with *high* likelihood of access. The *network administrator* has access to the network equipment and can attempt to listen to channel *C3*, thus the likelihood of access is considered *high*.

Considering the list of security mechanisms protecting process *P3*, an unauthorized access attempt from either *system administrator*, *network administrator*, or *external adversary* is very unlikely; however, since these mechanisms were not certified we assign *low* (instead of *negligible*) likelihood of access to these adversaries for *P3*. The likelihood of access for the *security administrator* is *authorized* as he is allowed to observe the output of *P3* as part of his security monitoring tasks.

3.4 EPIC Third Step: Identify Privacy Threats

The objective of Step 3 is to determine whether data leaked in each exposure identified in Step 2 can potentially lead to a privacy violation. In order to assess this, we need to take into account what type of data is actually exposed. A given component can expose heterogeneous data. For instance, *DS1* in Example 2 exposes some log records that only contain the IP address of a user as well as others that also include the file being transmitted by that user. Another example is reported in Figure 3, showing the user interface of an application-level firewall (PAN-OS 6.1). The upper part of the figure shows results from security threat detection based on URLs filtering while the lower part report results from threat detection based on file filtering and the two tables have different attributes.

We refer to each log record type being exposed as a **data content**, each composed by a set of attributes. The **attributes description table** (for example Table 4). lists all attributes exposed in each data content and reports their name, description, domain and some example

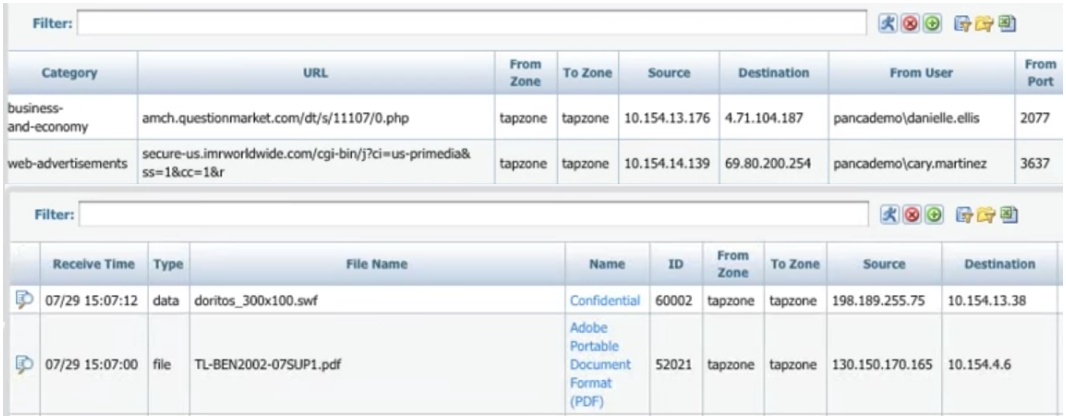

Figure 3: PAN-OS 6.1 interface to the logs (from Palo Alto Networks live community video tutorials)

Table 5: Data content identification (running example)

Exposure				Data content
Exposure	**Component**	**Adversary**	L_a	
$Exp1$	DS1. threat log	Security administrator	Authorized	$dc1$: IP(out-dst)
				$dc2$: IP(in-src), IP(in-dst)
				$dc3$: IP(in-src), IP(in-dst), File
$Exp7$	C3. application level threats	Network administrator	High	$dc1$: IP(out-dst)
				$dc2$: IP(in-src), IP(in-dst)
				$dc3$: IP(in-src), IP(in-dst), File
$Exp12$	P2. threat monitoring	Ext. adversary	Low	$dc1$: IP(out-dst)
				$dc2$: IP(in-src), IP(in-dst)
				$dc3$: IP(in-src), IP(in-dst), File

values. Table 4 shows the attributes description table for our running example.

We then associate each exposure (i.e., component and adversary) with the data contents it exposes. This is reported in the **data content identification table** (for example Table 5). that presents, for each pair of component and adversary derived from the data exposure table, the likelihood of access (as previously evaluated) and the list of data contents exposed by that component to that adversary. Table 5 shows an example reporting some selected exposures from Table 3. Note that in Table 5 each data content is exposed by each considered component to each considered adversary. This is not always the case as it can happen that two components expose different data contents and that a component exposes different data contents to different adversaries.

We then evaluate whether a combination of exposure and data content represents a privacy threat by analyzing how the adversary can discover the association between a sensitive information and an identified respondent. This is clearly related to the semantics of the data being exposed and on the knowledge accessible to the adversary. We first classify the attributes according to the following definitions.

- **Potentially Sensitive Information (PSI):** attribute or set of attributes that can be considered as sensitive. I.e., the combined values of the attributes in each of these sets

Table 6: Data content attributes analysis (running example)

Data content	ID		QID			PSI	
	Attribute	**Mtv.**	**Attribute**	**Bg. Kowledge**	**Mtv.**	**Attribute**	**Mtv.**
$dc1$: IP(out-dst)	None	...	None	None	...	IP(out-dst)	...
$dc2$: IP(in-src), IP(in-dst)	None	...	IP(in-dst)	List associating IP-addresses with user-names	...	IP(in-src)	...
$dc3$: IP(in-src), IP(in-dst), file	None	...	IP(in-dst), file	List associating IP-addresses with user-names	...	IP(in-src), file	...

reveal sensitive information about the data respondent.

- **Identifier (ID):** attribute or set of attributes that uniquely identifies a respondent in a data-set.

- **Quasi-Identifier (QID):** attribute or set of attributes that, combined with other information (including adversary's background knowledge), can be used to identify the respondent in a data-set (or to restrict the set of candidate respondents).

The recognition of QIDs and the related assumptions about background knowledge, also required by most anonymization techniques, is one of the most difficult tasks in privacy protection [3]; however, it becomes more feasible when considering a restricted domain with specific types of data content and adversaries, like the one we are considering. Table 6 shows an example of the **data content attributes analysis table** that reports the attributes classification for each data content and also describes the expected adversary's background knowledge. The privacy expert is also expected to motivate or comment the classification of each attribute. Note that for sake of brevity we will not report the motivations in the tables ("Mtv." columns) but we report them in the text.

Example 3. In Table 6, the attribute *IP(out-dst)*, contained in data content $dc1$, is classified as a PSI attribute . In fact, *IP(out-dst)* is the destination IP address of outgoing traffic/request (see Table 4); this address can reveal sensitive information about the respondent who sent the request e.g., in case of HTTP traffic this attribute will reveal the domain name of the web page visited by the respondent. $dc1$ contains neither ID attributes nor QID attributes because *IP(out-dst)* does not provide any information about the data respondent in the organization that initiated the communication.

Data content $dc2$ contains no ID attributes and a QID attribute *IP(in-dst)* that refers to the destination IP address of incoming traffic (see Table 4). It is the IP address of a respondent receiving a request or most likely an answer to a request. *IP(in-dst)* can be used to re-identify a respondent if the adversary has background knowledge allowing them to associate an IP address with a user-name. $dc2$ also contains the PSI attribute *IP(in-src)*. Similarly to *IP(out-dst)*, *IP(in-src)* indicates the IP address of a machine answering to a respondent's request that could be the domain name of a privacy-sensitive website that the respondent is visiting.

Data content $dc3$ and $dc2$ have two attributes in common: *IP(in-dst)* classified as QID and *IP(in-src)* classified as PSI. $dc3$ contains, in addition, the attribute *file* classified as QID because it might contain information that can be used to re-identify a respondent e.g., name

and surname. *file* is also considered as a PSI attribute since files are very likely to reveal sensitive information about the respondents health, a purchase, financial information.

Each combination of exposure and data content is considered a **privacy threat** if that data content contains PSI attributes and at least an ID attribute or a QID attribute. For example, the combination of Exposure *Exp 1* and $dc2$ (see Table 5) is a privacy threat (if the adversary has the necessary background information), because $dc2$ contains IP(in-src), which is a QID and IP(in-dst), which is a PSI.

If for a given combination of exposure and data content, that data content has no ID nor QID attributes or if it has no PSI attributes, that combination can be cleared as it is not a privacy threat. For example, $\{Exp1, dc1\}$, $\{Exp7, dc1\}$, and $\{Exp12, dc1\}$, highlighted in Table 5, are cleared. In fact $dc1$ (as shown in Table 6) is composed solely by IP addresses of external machines and contains no ID or QID attributes.

4 EPIC Fourth Step: Evaluate and prioritize privacy threat risk

In this section, we describe the fourth step of our methodology aimed at measuring the risk of each privacy threat identified in Step 3. Following a common approach in the field of IT security, we compute the privacy violation risk as the combination of likelihood of occurrence of a privacy violation $\mathcal{L}$ and its impact $\mathcal{I}$. In the following we first describe how to measure privacy violation likelihood (Section 4.1), its impact (Section 4.2) and then we show how to measure risk (Section 4.3). Finally, we show how to prioritize risk mitigation actions (Section 4.4).

4.1 Privacy violation likelihood

The **privacy violation likelihood** represents the likelihood that the privacy of any respondent is violated due to the disclosure of a given data content in given data exposure. It depends on two factors: the likelihood of access (specified for each data exposure in the third step) and the likelihood that, from the exposed information, the adversary can successfully complete the privacy attack.

In order to complete a privacy attack, the adversary needs to associate the sensitive information with the respondent's identity. While in general, this association task may not be trivial, in the domain that we are considering sensitive attributes most of the time appear in data logs together with identifying or quasi-identifying information (e.g., IP, MAC address, UID). Since in this step, we are only considering data contents that contain PSI (the others have been cleared in Step 3), the likelihood of successfully completing the privacy attack corresponds to the **re-identification likelihood** i.e., the likelihood that the data respondent is re-identified.

We define this likelihood with a qualitative scale, established mainly by analyzing the ID and QID set of attributes identified in the data content in the previous step and evaluating which background knowledge the considered adversary may actually have. We provide the following guidelines and examples to assign re-identification likelihood values (c is the data content):

- *Certain*. Data respondents' identity is explicitly reported in c. Consider, for example, a company that assigns to each employee an email address in the form *name.surname*

and assume that each record in c contains the senders' email address for outgoing email. In this case, each log record in c is explicitly identified.

- *High*. The adversary can discover the data respondents' identity because (i) the explicit identity is part of many records in c or (ii) c contains quasi-identifying information and the adversary has access to the background information that allows him, with limited effort, to re-identify the respondents. As an example for case (i), consider a company in which users can choose their email addresses; c contains the senders' email address of outgoing emails. In most of the cases, the email address will be in the form *name.surname*, so the data respondent can be often identified. As an example for case (ii), consider that c contains the source IP address of outgoing HTTP connections, the adversary is the network administrator and he has background information to map an IP address to the corresponding user's name.

- *Medium*. The adversary can discover the data respondents' identity because (i) the explicit identity is seldom part of c or (ii) c contains quasi-identifying information and the adversary can use it, together with background information so that, sometimes and possibly with an effort, he can re-identify the respondent. As an example for case (i), consider that c contains the name of a file being transmitted; it is possible, though rare, that the file name contains the sender's identity like in the case of a file named *name_surname_CV*. As an example for case (ii) consider that c includes the timestamp of outgoing HTTP connection; the adversary has access to the physical entrance/exit logs for the building, so he can infer when a person was in the building, and hence, in some cases, he can find the identity of the data respondent or at least restrict the set of possible respondents to a few individuals.

- *Low*. Explicit identity is not part of c but c contains quasi-identifiers that the adversary can seldom or with a significant effort exploit to discover the respondent's identity. Consider this example: c contains the source IP address of outgoing HTTP connection. The adversary is the system administrator that, generally, does not know the association between IP addresses and employees identities. However, when a system administrator is asked for help desk support, he can become aware of a static IP address associated with a given employee, hence being able to re-identify the data respondent.

- *Negligible*. Explicit identity is not part of c and any quasi-identifying information in c, if any, can only be used to re-identify a respondent by using background information that is unlikely to be available to the adversary. Consider the case in which c contains the source IP address of outgoing HTTP connections. An external adversary does not know which user is associated with each IP address, so he cannot re-identify data respondents, especially if the address is dynamic or masked by a gateway.

The qualitative values for re-identification likelihood and likelihood of access are combined to obtain a qualitative value for the privacy violation likelihood, which is measured with a 5-values scale from *negligible* to *very-high*. Table 7 shows how to compute privacy violation likelihood given re-identification likelihood and the likelihood of access. The intuition behind Table 7 is that the two input likelihoods are combined with an operation similar to a product. For example, if one of the two input likelihoods is *negligible* (this is intuitively analogous to a zero probability), then the output likelihood is also *negligible*.

The **privacy violation likelihood table** (see for example Table 8) lists all privacy threats and for each of them it reports the *likelihood of access* (L_a) (derived from Step 3), the *re-*

Table 7: Likelihood matrix defining privacy violation likelihood as a combination of likelihood of access and re-identification likelihood

Re-identification likelihood	Negligible	Low	Medium	High	Very-High
Certain	Negligible	Medium	High	Very-High	Very-High
High	Negligible	Low	Medium	High	Very-High
Medium	Negligible	Low	Medium	Medium	High
Low	Negligible	Low	Low	Low	Medium
Negligible	Negligible	Negligible	Negligible	Negligible	Negligible
	Negligible	**Low**	**Medium**	**High**	**Authorized**

Likelihood of Access

identification likelihood (L_{rid}), that is evaluated according to the five qualitative values defined above, the motivations behind this evaluation and, finally, the value of the *privacy violation likelihood* (L) , which is computed according to the **likelihood matrix** (see Table 7).

Example 4. Table 8 is the privacy violation likelihood table for the privacy threats identified in the running example in Section 3.4.

The likelihood of access reported in this table was computed in Step 2 (see Table 3). Values for the re-identification likelihood were defined according to the following reasoning. Let's first consider data content $dc2$, including the IP addresses that an adversary can use to re-identify a respondent if he can associate it with the user name (either directly or, for example, by first associating the IP address to the office number and then to the user name). As observed above, *security administrator* can know this association in some cases, so the re-identification likelihood is *medium*. The *network administrator* has access to the full list associating IP-addressed and user names, so the re-identification likelihood is *high*. Finally, *external adversary* cannot associate the IP-address to the user name, so in this case, the re-identification likelihood is *negligible*.

Let's now consider data content $dc3$. Also, in this case, the IP-address is part of the data content, so, for each adversary, the re-identification likelihood is at least as high as with $dc2$. However, $dc3$ also contains a file (i.e., file name, file content, etc.) that can sometimes be an explicit identifier or a quasi-identifier. For the *security administrator*, who is an internal adversary, the file can often identify the user. For example the *security administrator* can re-identify the user even if the file is a document signed with the first name only; this is possible because the *security administrator* knows that there is only one person with that name, or because, from the context, the adversary recognizes the file as coming from a given office, where there is a single person with that name. For this reason, the re-identification likelihood is set to *high* for *security administrator*. Instead, *external adversary* can only re-identify the issuer when the full name is reported in the file and, in some cases, this might not even be enough, for example for very common full names. For this reason, the re-identification likelihood is set to *medium* for this adversary.

4.2 Privacy violation impact severity

A privacy violation has a negative impact on the responsible organization. We model this by assigning an **impact severity** (I) value to each privacy threat. The value depends on three impact factors, defined in the following (Section 4.2.1). Impact severity can be assessed both qualitatively (Section 4.2.2) and quantitatively (Section 4.2.3).

Table 8: Privacy violation likelihood (running example)

Exposure	**Data content**	L_a	L_{rid}		L
			Value	**Mtv.**	
$Exp1$: DS1	$dc2$	Authorized	Medium	...	High
Sec. Admin.	$dc3$	Authorized	High	...	Very-High
$Exp2$: C3	$dc2$	High	High	...	High
Net. Admin.	$dc3$	High	High	...	High
$Exp3$: P2	$dc2$	Low	Negligible	...	Negligible
Ext. Adversary	$dc3$	Low	Medium	...	Low

4.2.1 Impact Factors.

To provide an impact severity assessment with as much accuracy as possible we first need to identify the consequences of a privacy violation, that we call *impact factors*. They are summarized in the following list:

- *Non-compliance* (I_C). If data content is exposed in a non-compliant way, then the organization might incur a certain cost in the form of e.g., non-compliance fines, respondents compensation for loss of their privacy, remediation measures to address the privacy issues that led to the unlawful leakage. For example, aspects that should be taken into account to evaluate this impact factor include: whether the respondent was informed or provided a consent for data processing, if the data was retained for a period longer than prescribed, and even general intervenability rights like the possibility of data subjects to rectify or delete their data.

- *Failure to meet business agreements* ($\mathcal{I}_B$). The organization might have agreements with end-users or other organizations that imply penalties in case of privacy violations. For example, privacy protection could be part of a service level agreement and the service provider may be subject to specific penalties in case of privacy loss.

- *Reputation Loss* ($\mathcal{I}_R$). A privacy violation can have an impact on the organization reputation, that is a commercially valuable asset. Indeed, reputation loss can "erode the ability of businesses to successfully retain their markets, maximize shareholders value, raise finance and manage debts, and remain independent" [23].

In the following, we discuss how to assign a qualitative or quantitative value to each factor. In both cases, there are three aspects that should be taken into account and that we collectively call **violation magnitude**.

i) The effect of the privacy violation on the respondent. While the effect of the privacy violation on the respondent does not have a direct impact on the organization, it is relevant for the evaluation of the three impact factors listed above. For example, if the privacy violation discloses a person's sexual orientation and this results in the person being sentenced (homosexuality is still illegal in some countries), then the reputation loss for the organization will be higher than in the case of a privacy violation that has limited impact on the data respondent. Indeed, as suggested in [9], a privacy violation can be assessed in terms of the physical, material and moral damage inflicted on the respondent.

ii) The number of respondents. It can be assessed based on the exposure magnitude (see Section 3.3) and an estimation of how exposed data is distributed among individuals.

iii) Nature of respondents. There are some categories whose privacy should be particularly protected (e.g., minors, social minorities) or individuals for whom a privacy violation can have worse effects than for others (e.g., a politician, a CEO).

By considering these three aspects the expert assigns a qualitative value to the violation magnitude in the scale: *Very limited*, *Limited*, *Medium*, *Important* and *Very important*.

When the experts assign values to the impact factors and to the violation magnitude, they should also keep track of the motivations behind the assigned values. This is important for two reasons. First, in order to ease future updates of the privacy evaluation risk. Second, in order to make it possible to intervene during the remediation phase. For example a high value may be assigned to I_C for improperly following the regulation with respect to retention period, intervenability or consent. The specific reason should be annotated so that targeted remediation actions can be taken if required by considering the resulting privacy violation risk for the considered threat.

4.2.2 Impact Severity: Qualitative Assessment.

With this form of assessment a *privacy expert* and an *organization representative* jointly evaluate the severity of each impact factor for each privacy threat and assign a qualitative severity level to each factor on a 5-levels severity scale (*Low*, *Med-low*, *Med*, *Med-high* and *High*). This evaluation takes into consideration different aspects for each of the three factors. For example, the non-compliance severity will depend on the measures the organizations deployed in order to be compliant with the regulation or the lack of these measures. It also depends on the violation magnitude; indeed, in case a compensation to the violation victims is required, the non-compliance severity will scale linearly with the number of respondents affected. The reputation loss impact may depend on the adversary, on the data handled by the organization, on insufficient organizational and technical control, and most importantly by the number of individuals affected. Indeed, reputation loss is likely to scale with the privacy violation magnitude, not only in terms of number of respondents affected, but also in terms of the nature of these respondents (e.g., a privacy violation for a social minority, a celebrity or a political figure will certainly have more reputation impact than other leakages). Finally, the impact of non-fulfillment of business agreements depends on the kind of data leaked and on the agreements themselves. An example business agreement may be an SLA (service level agreement) with a cloud provider. SLAs usually specify a minimum level of data security and privacy. In case of failure to meet those requirements, penalty fees should be paid to the client as compensation.

After evaluating the impact factors, impact severity is computed as the maximum severity level of the three factors: $I = \max(I_C, \mathcal{I}_B, I_R)$. In fact, the five severity levels intuitively represent significantly different range of values (possibly even different orders of magnitude). Thus, the overall impact severity will most likely preserve the range of values of the highest severity among the considered impact factors.

The results are reported in the **qualitative privacy violation impact table** (see for example Table 9) that reports, for each privacy threat, the violation magnitude, the qualitative values of each impact factor and the resulting qualitative impact severity.

Example 5. Table 9 reports the qualitative privacy violation impact table for a subset of the privacy threats reported in Example 4.

In the first row, impact severity is *low*. Indeed, in threat *Th1* users are informed that the IP addresses (both local and remote) are collected for security purposes and might be processed by the *security administrator*. For this reason, and because several measures were

Table 9: Qualitative privacy violation impact (running example)

Exposure	**Data content**	**Th**	**Violation magn.**		I_C		I_B		I_R		I
			Value	**Mtv.**	**Value**	**Mtv.**	**Value**	**Mtv.**	**Value**	**Mtv.**	
$Exp1$: DS1 Sec. Admin	$dc2$	$Th1$	Important	...	Low	...	Low	...	Low	...	**Low**
$Exp7$: C3 Net. Admin	$dc2$	$Th2$	Very-Limited	...	Med-high	...	Low	...	Med-low	...	**Med-high**
$Exp7$: C3 Net. Admin	$dc3$	$Th3$	Very-Limited	...	High	...	Low	...	High	...	**High**
$Exp12$: P2 Ext. Adver.	$dc3$	$Th4$	Limited	...	High	...	Low	...	Med-high	...	**High**

taken to avoid privacy violations, the non-compliance impact factor (I_C) is evaluated as *low*. I_B is *low* because the organization has no business agreements to fulfill. $\mathcal{I}_R$ is also *low* because the impact of this violation on reputation is minimal since a *security administrator* is somehow expected to access information about user IP addresses.

Impact severity of $Th2$ is *med-high*. Considering $Th2$, non-compliance impact factor is quite severe because respondents are not informed that the adversary can access exposed data (actually, *network administrator* is not expected to access exposed data). However, violation magnitude is very limited, because there are few respondents for the exposed data. This mitigates I_C that is evaluated as *med-high*. I_B is *low* because the organization has no business agreements. The impact on organization's reputation I is estimated *med-low* because the violation magnitude is *very limited* and exposed data does not contain particularly sensitive information (see Table 6 for $dc2$).

Threat $Th3$ is similar to $Th2$ with the difference that in this case, the adversary can also access files, which in turn can contain any type of data, including those particularly protected by existing regulations, e.g., health-related information. For this reason both I_C and I_R are *high*, and consequently impact severity is also *high*.

In threat $Th4$ a non-authorized person (i.e., an *external adversary*) has access to $cd3$ that includes files. Hence, similarly to $Th3$, I_C is *high* and consequently impact severity is *high*.

4.2.3 Impact Severity: Quantitative Assessment.

Another approach to assess impact severity is to quantitatively estimate the economic cost deriving from a privacy violation. We consider the same three factors as in the qualitative approach but in this case, we associate each of them with an estimation of the economic loss.

For example, *non-compliance* cost includes: (i) the fines that the organization has to pay, (ii) the cost of remediation actions (both organizational and technical), and (iii) the compensation to pay to each affected respondent times the number of respondents.

The *reputation loss* costs are caused by the loss of trust and the degradation of the relationship between the organization and its partners, employees, investors, customers and potential future customers. It can be reflected on several levels e.g., turnover of existing customers, diminished customer acquisition, cumulative abnormal stock returns, decline of equity value [25]. It can also include the costs of efforts to control the incident disclosure and reputation repair.

The *failure to meet business agreements* cost depends on the existing business agreements and their nature.

In the case of a quantitative assessment, we compute impact severity of a privacy threat as the sum of the costs associated to each impact factor: $I = I_C + I_R + I_B$.

The results are reported in the **quantitative privacy violation impact table** that is analogous to the qualitative privacy violation impact table (Table 9) with the only differences that impact factors and impact severity are reported as quantitative values.

4.3 Privacy violation risk

As mentioned in the beginning of this section, **privacy violation risk** depends on the privacy violation likelihood and impact severity. If impact severity is assessed quantitatively, then we can compute a quantitative privacy violation risk. Otherwise, we provide a qualitative privacy violation risk assessment.

4.3.1 Qualitative Evaluation

Table 10: Risk matrix defining qualitative privacy violation risk as a combination of privacy violation likelihood and impact severity.

Impact severity						
	High	Low	Medium	High	High	High
	Med-High	Low	Medium	Medium	High	High
	Med.	Low	Low	Medium	Medium	High
	Med-Low	Low	Low	Low	Medium	Medium
	Low	Low	Low	Low	Low	Medium
		Negligible	**Low**	**Medium**	**High**	**Very-High**

Privacy violation likelihood

We define the **qualitative privacy violation risk** with three levels: *low*, *medium* and *high*. We combine privacy violation likelihood and impact severity levels according to the **risk matrix** (see Table 10). The idea behind Table 10 is that when privacy violation likelihood is negligible, then we can exclude that the adversary can successfully complete the attack, so the risk is low. If the privacy violation likelihood is *low*, then risk is obtained by decreasing the value of the impact severity (e.g., impact severity *high* results in a *medium* risk). Similarly, if the privacy violation likelihood is *medium*, then risk is obtained by slightly decreasing the value of the impact severity (e.g., *medium-high* impact severity results in a *medium* risk but *high* impact severity results in *high* risk). A *high* value of privacy violation likelihood implies that the values of impact severity map to the same value of risk, with the exception of *medium-low* and *medium-high* that are "rounded up" to *medium* and *high* risk

Table 11: Qualitative privacy violation risk (running example)

Th	**Exposure**	**Data content**	L	I	R
$Th1$	Exp1: DS1, Sec. admin.	$dc2$	High	Low	Low
$Th2$	Exp2: C3, Net. admin.	$dc2$	High	Med-low	Medium
$Th3$	Exp2: C3, Net. admin.	$dc4$	High	High	High
$Th4$	Exp3: P2, Ext. adversary	$dc4$	Low	High	Medium

values, respectively. Finally, a *very-high* privacy violation likelihood results in risk values that are higher than those of the impact severity (e.g., *medium* impact severity maps to *high* risk).

The **qualitative privacy violation risk table** (see for example Table 11) reports, for each privacy threat, the values of privacy violation likelihood and impact severity (that were previously computed), together with the qualitative risk value that is computed based on Table 10.

4.3.2 Quantitative Evaluation

In the quantitative approach, we need to convert the qualitative measure of privacy violation likelihood into a numerical value. We propose the following association: *Very-High*$= 1$, *High*$= 0.75$, *Medium*$= 0.5$, *Low*$= 0.25$ and *Negligible*$= 0$. Then, for each privacy threat, we compute the quantitative privacy violation risk R as the product of the privacy violation likelihood and of impact severity: $R = L \cdot I$

The results are then reported in the **quantitative privacy violation risk table** that is analogous to the qualitative privacy violation risk table (Table 11) with the difference that quantitative values are reported for the privacy violation likelihood, for impact severity and risk.

4.4 Risk mitigation actions prioritization

In the fourth step, after assessing the risk values, we are now interested in defining in which order the privacy threats should be addressed with mitigation actions. We model this order with a **priority** value, a scale of integer values from 1 to 12 where 1 represents the highest priority.

The priority of a privacy threat depends on two factors: its privacy violation risk and the trustworthiness of the adversary involved in that privacy threat. Several definitions of trust have been proposed in the literature (see [28] for a survey). In this paper we consider the **trust** in an adversary as the organization's level of confidence about the actor not attempting to gain non-authorized data access or misusing the data to violate privacy. This level should be assessed by taking into consideration several aspects, including legal agreements, specific training on handling personal data, personal characteristics (such as morality, skills, and behavior [18]), and organizational procedures (e.g., motivational practices and reward systems). Regarding legal agreements, note that employees with access to the system usually have to sign such agreements as part of their contract. In EPIC, the knowledge about these agreements is part of the domain knowledge acquired as input for the whole methodology (see Figure 1).

Human factors are receiving increasing attention in the security field. Indeed actors trust assessment is often included in risk management processes. Some approaches discuss the trust level as a part of the risk computation [38] whereas others use this level as an independent indicator to balance the risk at the decision making stage [1]. It has been observed that the first approach tends to underestimate or hide the risks involving insider threats [10]. Actually, insiders have a big potential to create threats intentionally (by attempting malicious actions) or unintentionally (through lack of experience and awareness). For this reason the EPIC methodology adopts the second approach, and we do not consider adversary trustworthiness as a factor in the evaluation of privacy violation risk. The trustworthiness is rather used to define a priority value.

This approach has a twofold effect. On one side, it provides an effective priority classification of threats to act upon. On the other side, it provides an explicit classification of risk that also takes into account the adversaries' trustworthiness. This risk estimation will be useful in the process of deciding and designing what kind of training an actor should have in preparation to fill a high-risk position and what kind of profiles to select when hiring.

We consider the following four levels of trust.

- *Fully trusted:* Adversaries are fully trusted if they are trained to deal with personal data at the CSS level. Their activities with data are monitored by logging mechanisms and they are accountable for any personal data leakage. They often have very high privileges allowing them full access to data.

- *Trusted:* Trusted adversaries are also trained to deal with personal data and their activity is monitored. However, they have less responsibility in case of privacy leakage and have restricted access to the sensitive data.

- *Moderately trusted:* Actors are moderately trusted if they are trusted at the organization level, however, they are not specifically trained to deal with sensitive and personal information at the CSS level. These actors have often high privileges (e.g., administration privileges). They are responsible and accountable for any abuse of their privileges.

- *Untrusted:* Adversaries are considered as untrusted if they have no training on how to deal with private information and no authorizations to access the data.

We propose to use the priority distribution defined by the **priority matrix** (see Table 12) to combine privacy violation risk and adversary's trustworthiness in order obtain each threat priority. This matrix is designed to give more weight to the risk than to the trust. Priority of threats with the same risk level decreases (i.e., gets higher values) conversely to the trust level. In most of the cases, a privacy threat with a lower privacy violation risk than another is associated with a lower priority, with some exceptions. For example, a privacy threat with *medium* risk and *untrusted* adversary is associated with a priority higher than a privacy threat with *high* risk and *fully trusted* adversary.

The results of this procedure are reported in the **prioritized privacy threats table** (see for example Table 13) that indicates, for each privacy threat, its associated privacy violation risk (previously computed), the adversary trust, and the resulting priority value.

Adversaries that are not fully trusted may also be at risk of sharing data with external adversaries or colluding with other adversaries. While dealing with collusion is not explicitly taken into account by EPIC, the likelihood of this scenario can be reduced by remediation actions that include specific legal obligations, and organizational measures like preventing

Table 12: Priority matrix defining priority as a combination of privacy violation risk and adversary trust

		Adversary trust			
		Untrusted	Moderately Trusted	Trusted	Fully Trusted
Privacy violation risk	High	1	2	3	5
	Medium	4	6	7	9
	Low	8	10	11	12

the use of personal external storage or the use of any personal device in the CSS control room.

Table 13: Prioritized privacy violation threats (running example).

Th.	**Exposure**	**Data content**	R	**Adversary trust**		**Priority**
				Value	**Mtv.**	
$Th1$	$Exp1$: DS1, Sec. admin.	$dc2$	Low	Fully Trusted		12
$Th2$	$Exp2$: C3, Net. admin.	$dc2$	Medium	Moderately trusted		6
$Th3$	$Exp2$: C3, Net. admin.	$dc3$	High	Moderately trusted		2
$Th4$	$Exp3$: P2, Ext. adversary	$dc3$	Medium	Untrusted		4

Example 6. Table 13 illustrates the priority of threats considered in the previous section. The first threat $Th1$ has a *low* risk level. The adversary is the *security administrator* that is *fully trusted* to access and process the data content $dc2$ because they are highly trained to deal with personal data and assume high responsibilities for any potential leakage or misuse of this data. For these reasons, this threat has the lowest possible priority (12).

In the second and third rows (i.e., $Th2$ and $Th3$) the adversary is *moderately trusted*. *Network administrators* are trusted within the organization, but they are not authorized to access $dc2$ or $dc3$ nor specifically trained to deal with any private information collected by the CSS. Consequently, $Th2$ and $Th3$ have priority levels 6 and 2, respectively.

In the last threat the adversary is external and hence *untrusted*. Since risk is *medium*, according to Table 13 priority is 4.

5 Case Study

In this section, we show how to apply the proposed methodology to a real case study represented by the cybersecurity system protecting the network of an academic institution including over 15,000 hosts.

5.1 Case Study description

Figure 4 depicts the architecture of a university campus network along with its cybersecurity Systems (CCSs). From here on we will refer to the ensemble of these cyber security systems as *UCSS* (i.e., University cybersecurity System). The university network is divided into different network segments located in three geographic areas and connected among themselves by four main routing devices (R1, . . . , R4).

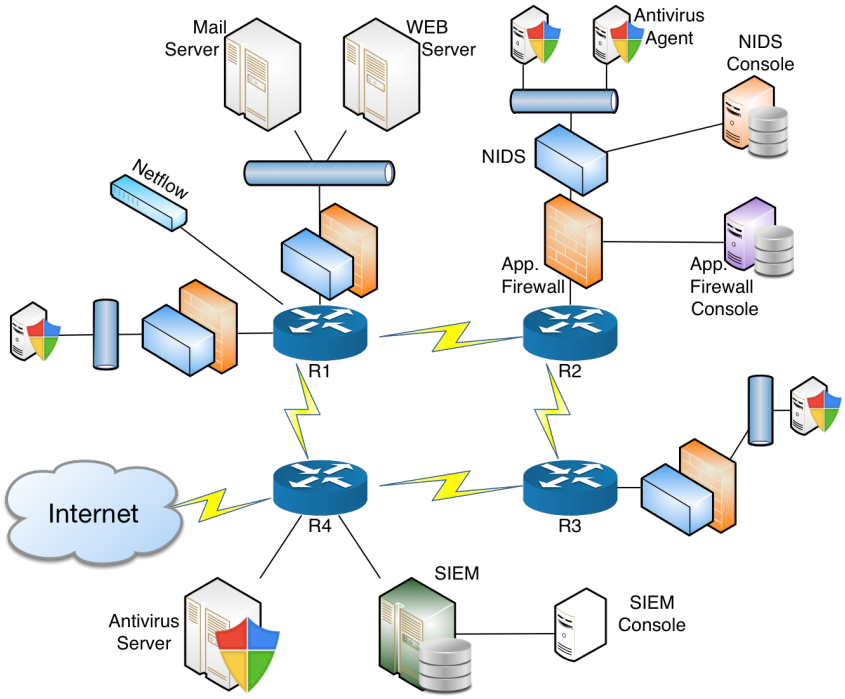

Figure 4: Architecture of the University's cybersecurity System

This network is protected in total by six types of cybersecurity systems: (1) Netflow network collector, (2) Network Intrusion Detection Systems (NIDS), (3) application-level firewall, (4) Security Information and Event Management System (SIEM), (5) cloud antivirus and (6) security mechanisms built into the routers and in particular (i) Firewall at IP level and (ii) Virtual Private Network at datalink level.

To identify and assess the privacy impact of UCSS, we run the four steps of the EPIC methodology for each of these six cybersecurity systems. In the following, we report the most important results of the analysis focusing on three of them: cloud-based antivirus, application level firewall and SIEM.

The Cloud antivirus is based on a technology that uses a lightweight software component on the protected host while offloading the majority of data analysis to the antivirus provider's infrastructure. The goal of the software agent is to identify suspicious files and send them to the network cloud where multiple antiviruses and behavioral detection engines are applied simultaneously to improve the detection rate. Cloud antivirus can also use a "retrospective detection" where the cloud detection engine rescan all files already checked when a new threat is identified. Such technique can improve the detection speed.

Firewall at application level is used to detect threats such as web attacks, exploitation techniques, malware infections, etc. (Figure 4 shows a single firewall connected to R2 but there is actually a firewall for each router). To this end, the firewall is able to process a large spectrum of data types such as: executables, PDFs, emails, multimedia files, etc. The firewall can be also configured to decrypt SSL traffic going to any external websites and it acts as a forward proxy. Like the other cybersecurity devices, the firewall is also equipped with a remote console to allow the security team to monitor the security events and investigate

threats.

The events and threats collected by Netflow, NIDS, application level firewall and routers are sent to the SIEM for further analysis, which is considered the mastermind of UCSS. Thanks to its capabilities such as data aggregation, event correlation, and advanced forensic analytics, this system provides a view on the big picture of potential attacks running under the network and that the other CSS cannot detect separately. The SIEM has a remote console allowing interaction between the system and human agents.

5.2 Applying EPIC's first step: Model UCSS

In this step, we use DFD+ to describe the system, data and UCSS functional aspects.

Cloud Antivirus. Figure 5 describes the data flow in the cloud antivirus. The source entity is an antivirus agent installed on a user's machine. It collects and sends suspicious files to an antivirus server through the physical channel *C1*. This data is processed in *P1* to detect potential threats and an action notification is sent back to the agent (e.g., to quarantine, to deletion, to consider as a false positive etc.). The detected threats are then sent through *C2* and stored in data storage *DS1* from where they can be accessed by a security administrator from process *P2* through the physical channel *C4*.

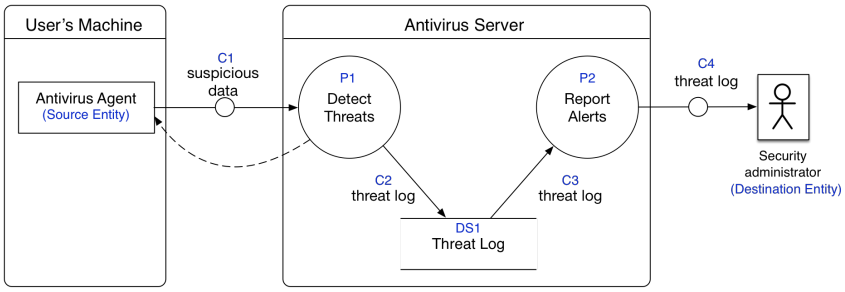

Figure 5: Modeling UCSS Antivirus component with DFD+

Application level firewall. Network traffic flowing through the network is filtered by the firewall, as shown in Figure 6. Process *P1* filters, at application level, two kinds of traffic: threats and traffic events. Threats are detected when network traffic matches threat patterns and specific information about each threat is sent through *C2* and stored in data storage *DS1*. Traffic events are security events, which are not considered as a threat but should be monitored in order to prevent security issues (e.g., a failed login is not considered a threat unless the number of attempts exceeds a given threshold). Traffic events are stored in *DS2*. The remaining traffic (normal traffic) is not logged. Threat logs can be accessed by two actors namely security administrator and security operator via process *P2* hosted on a remote console. Threat logs transit from the firewall to the remote console through *C2.1*, and from the console to the actors machines through *C2.1.1* or *C2.1.2*. The same actors can access log events via *P3* through *C3.1* and *C3.1.1* or *C3.1.2*. A security administrator has direct access to the data storage *DS1* and *DS2* to fulfill several management tasks, and a system administrator has access to the firewall machine to perform system administration and maintenance tasks.

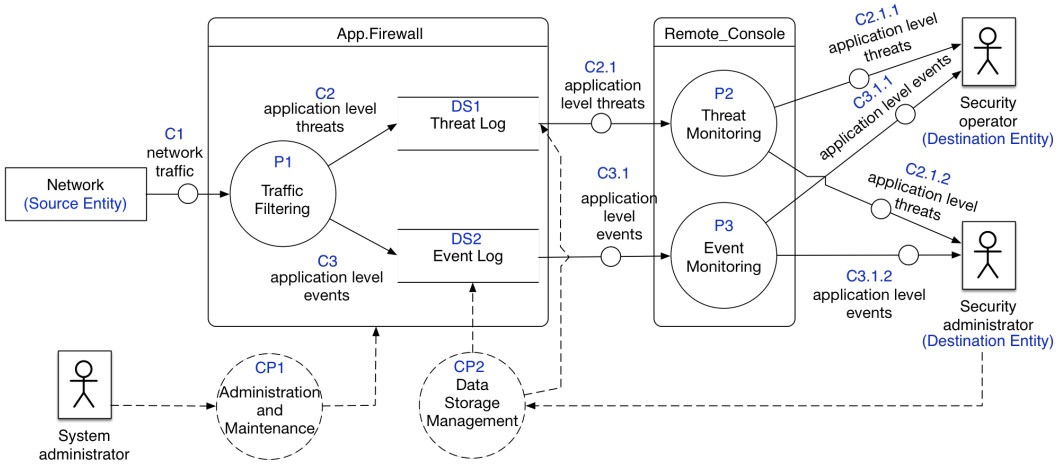

Figure 6: Modeling UCSS Firewall component with DFD+

SIEM. The SIEM receives security logs from the other cybersecurity systems deployed in UCSS (Figure 7). Logs data from Netflow, router firewall, application level firewall and NIDS flow through channels *C1.1*, *C1.2*, *C1.3* and *C1.4*, respectively. Data is fed to several processes (e.g., threat and anomaly detection) that we summarize in the complex process *P1*. The threats identified in *P1* are sent through *C2* and stored in *DS1*. The security administrator and the security operator can perform several investigation operations via process *P2* hosted in a remote SIEM console (this also involves channels *C3*, *C3.1* and *C3.2*). The system administrator has access to the SIEM machine for administration and maintenance purposes.

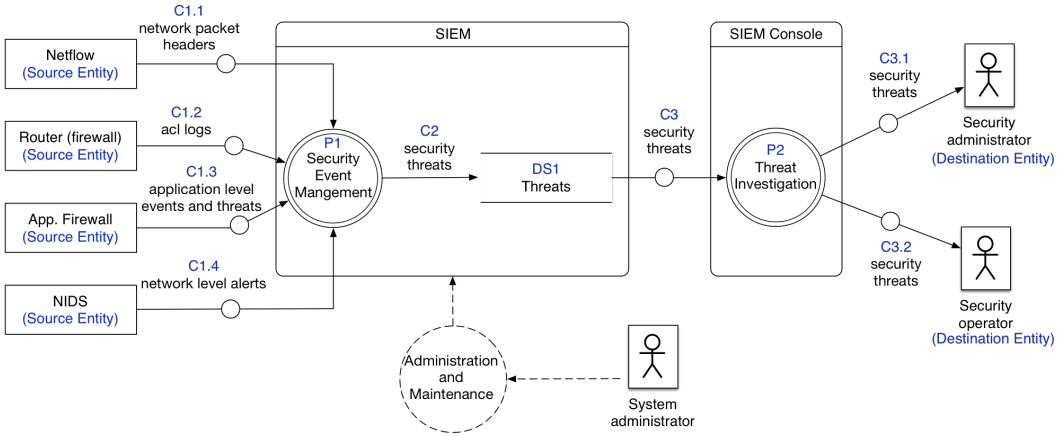

Figure 7: Modeling UCSS SIEM component with DFD+

5.3 Applying EPIC's second step: Identify data exposures in UCSS

The goal of this step is to identify and analyze exposures in each component identified in Step 1. For the sake of space, we report and comment only some of the results of this second step. We first define the list of potential adversaries, then we study the security mechanisms implemented in each component, and finally, we assess, for each exposure and adversary, the exposure magnitude and the likelihood of access.

Adversaries. We reported in Table 1 a reference list of adversaries for a cybersecurity environment that can be personalized for the specific CSS and organization being considered. In the UCSS use case, we added the following adversaries: a) *security operator*: in UCSS the security team is composed of a security administrator and four security operators; the operators have the same tasks but fewer privileges than the *security administrator*; b) *network user*: this adversary role applies to any individual with approved access to the university's network.

Security mechanisms. Table 14 lists only some of the components identified in Step 1, and for each of them, it reports the list of users that are authorized to access the component and the security mechanisms that protect the access.

Table 14: UCSS components security

Component	**Authorized users**	**Security mechanisms**
antivirus: C1	N/A	Private network, network protection (firewall, SIEM, NIDS)
Firewall: DS1	Security admin.	Authentication, access control, network protection (firewall, SIEM, NIDS)
Firewall: DS2	Security admin.	*Same as above*
SIEM: C1.3	N/A	Private network, network protection (firewall, SIEM, NIDS)
SIEM: P2	Sec. admin, Sec. Op.	*Same as above*

Data exposure. For each CSS, we considered all possible pairs of components and adversaries, we estimated the likelihood of access and the exposure magnitude. Table 15 reports some of these combinations. Consider, for example, the antivirus agent, which sends suspicious traffic from the local machine to the antivirus server through channel *C1* (see Figure 5). For this component we identify two adversaries that can gain access to data with non *negligible* likelihood: the *network administrator* and a *network user*. The former has a *high* likelihood of access as they have direct access to the router device and consequently to the non-encrypted traffic between agent and server. Although this adversary has the ability to listen to the channel for a long period of time, the magnitude of this exposure is quite *limited* due to the limited number of antivirus agents sending data through this channel (only suspicious data is transmitted). The *network user* can perform a Man-in-The-Middle (MiTM) attack on the local network by using techniques like ARP cache poisoning attack [33] whose

Table 15: UCSS Data exposures

Component	**Adversary**	**Exp.**	L_a	**Exposure magnitude**	**Motivation**
Antivirus: C1	Network admin.	*Exp1*	High	Medium	...
	Network user	*Exp2*	Low	Very limited	...
Firewall: DS1	Security admin.	*Exp3*	Authorized	Very important	...
	System admin.	*Exp4*	Medium	Very important	...
Firewall: DS2	Security admin.	*Exp5*	Authorized	Very important	...
	System admin.	*Exp6*	Medium	Very important	...
SIEM: C1.3	Network admin.	*Exp7*	High	Limited	...
SIEM: P2	Security admin.	*Exp8*	Authorized	Very important	...
	Security op.	*Exp9*	Authorized	Very important	...

main goal is to hijack the communication between two hosts. Such attacks are quite easy to accomplish, especially in LAN environment where the activation of these defensive mechanisms is not possible due to the heterogeneity of the network devices. Since the UCSS network implements some protection mechanism against those attacks (see Table 14), we estimate the likelihood of access by *network user* to be *low*. In this case, the magnitude of exposure is *very limited* because, in addition to the limited amount of transmitted data, this adversary can only listen to the channel for a limited amount of time.

For the SIEM there are a total of 77 combinations of components and adversaries(i.e., 7 adversaries and 11 components) and for the entire UCSS, the total pairs are about 350. Despite the high number, the overall effort of producing the data exposure table (Table 15) is still reasonable for three main reasons. First, given the results of the previous steps, the security expert can quickly assess the likelihood of access and the exposure magnitude. Second, from the component security table (Table 14) it is possible to automatically identify when the likelihood of access is *authorized*. Finally, when the likelihood of access is *negligible*, there is no need to assess the exposure magnitude.

5.4 Applying EPIC's third step: Identify privacy threats in UCSS

After identifying the exposures in the previous steps we now assess whether these exposures represent privacy threats. As explained in Section 3.4 we start by listing and describing the data attributes leaked in the exposures. The eight data attributes that are exposed in antivirus, firewall and SIEM components[2] are described Table 16.

As mentioned in Section 3.4 the data leaked in each exposure is composed by heterogeneous types of records (i.e., records with different attributes). For example, exposures from

[2]The exposures we identified actually leak other privacy neutral attributes (i.e., neither IDs, nor QIDs, nor PSIs), however for sake of brevity we don't report them.

Table 16: UCSS attributes description

Name	Description	Domain	Example values
IP_int	IP adress (source or destination) of a machin in the local network	IP Address	192.168.100.32
IP_ext	External IP adress (source or destination)	IP Address	8.8.8.8
URL	visited sites urls and parameters if any	URL	www.sitename.com/search?s=parameter
file_meta.	File name, size, author creation time etc.		name.pdf, 504kb, 2017-06-06 12:07:10
file	A file being transmitted	String of bytes	
email_header	Email Object, Sender and Reciver adresses	smtp header	from: to: date: subject: etc.
email_cont.	Email Object, Sender and Reciver adresses		
app. name	name of the application and protocol used	name, protocol etc.	Thunderbird 52.1.1, smtp

channel $C1$ of the antivirus (i.e., $Exp1$ and $Exp2$) leak records composed by *IP_ext*, *IP_int*, *file_meta*, and *file*. These exposures also leak another type of records composed by *IP_ext*, *IP_int*, and *email_headers*. In Table 17 [3] we call the first type of records data content $dc1.3$ and the second $dc1.6$. Other data contents are defined analogously.

As required by the EPIC methodology, we continue our analysis by classifying the attributes of each data content as identifying (ID), quasi-identifying (QID), or potentially sensitive information (PSI). When an attribute is classified as QID we indicate which background knowledge may lead to re-identification when joined with the attribute value. The result is reported in Table 17. As an example for interpreting the table, note that no attribute in $dc1.3$ is identifying, while there are three attributes that are quasi-identifiers *IP_int*, *file_meta*, and *file*. An adversary might be able to re-identify a respondent from the IP address *IP_int* and knowledge allowing to map this address with the respondent's identity. The *file_meta* can contain information about machines and applications. Adversaries might have knowledge about respondents machines/systems and use this knowledge to re-identify records with the attribute *file_meta*. The attribute *file* can contain identifying information such as respondents names and surnames. The attributes *file_meta* and *file* are also potentially sensitive information, along with *IP_ext*, which can disclose e.g., a site visited by a user.

We can conclude that all data contents reported in Table 17 should be further analyzed in Step 4 because their exposure is a privacy threat. Indeed, all data contents reported in Table 17 have at least one attribute marked as PSI and at least one attribute marked as ID or QID.

[3] Note that data contents should be identified separately for each exposure, however some exposures in our example share the same data contents e.g.,($Exp1$ and $Exp2$) or ($Exp5$ and $Exp6$). Thus we represent them together in Table 17.

Table 17: UCSS data content identification and attributes analysis

Exp.	Data content	ID		QID			PSI	
		Att.	**Mtv.**	**Att.**	**Bg. Knowledge**	**Mtv.**	**Att.**	**Mtv.**
$Exp1$ and $Exp2$ (Anti-virus. C1)	$dc1.3$	None	...	IP_int; file_meta; file	Mapping between IPs and user names, or knowledge about user machine/system	...	IP_ext; file_meta; file	...
	$dc1.5$	None	...	IP_int; file_meta	*same as above*	...	IP_ext; URL; file_meta	...
	$dc1.6$	None	...	IP_int; email_header	Mapping between IPs and user names, or email address and user names	...	email_header	...
$Exp5$ and $Exp6$ (Firewall. DS2)	$dc2.4$	email_ header	...	IP_int; email_cont	*contains an identifier*	...	IP_ext; email_header; email_cont	...
	$dc2.8$	None	...	IP_int; file_meta	Mapping between IPs and user names or other knowledge about user (e.g., HR)	...	file _meta	...
	$dc2.9$	None	...	IP_int; file_meta; file	*same as above*	...	file_meta; file	...
$Exp8$ and $Exp9$ (SIEM. P2)	$dc3.1$	None	...	IP_int	Mapping between IPs and respondent's identity	...	location	...
	$dc3.6$	None	...	IP_int	*same as above*	...	URL, http_content	...
	$dc3.9$	None	...	IP_int	*same as above*	...	application name	...

5.5 Applying EPIC's fourth step: Evaluate and prioritize privacy threat risk in UCSS

EPIC's fourth step aims at evaluating and prioritizing the privacy threats identified in Step 3. The results of this step for a selected set of threats are reported in Tables 18, 19, 20, and 21. In the following, we provide some details on how these values were obtained.

5.5.1 Evaluating privacy threats in UCSS

As defined by EPIC, privacy violation risk (Table 20) is evaluated as the combination of privacy violation likelihood (see Table 18) and impact severity (see Table 19).

Privacy violation risk is *high* for threat $Th1$ where a *network administrator* can access data content $dc1.3$ (i.e., *IP_int*, *IP_ext*, *files_meta* and *file*) from channel *C1* that transfers suspicious data from the antivirus agent installed in the end-user machine to the antivirus server. As shown in Table 18, privacy violation likelihood is *high* because this adversary has *high* likelihood of accessing the data and of re-identifying respondents from their IP address. The violation magnitude is *medium* since the exposure magnitude is *medium* (see Table 15) and the violation can affect only a fraction of respondents. Impact severity (Table 19) is *high*, despite the medium violation magnitude, because the non-compliance impact is *high* and the impact on reputation is *med-high*[4]. The non-compliance impact is *high* since the adver-

[4]In our use case since the academic organization does not have any business agreements, the impact factor I_B

Table 18: UCSS privacy violation likelihood

Exposure	**Data content**	**Th.**	L_a	L_{rid}		L
				Value	**Mtv.**	
$Exp1$ Antivirus. C1: Network admin.	$dc1.3$	$Th1$	High	High	...	**High**
$Exp2$ Antivirus. C1: Network user	$dc1.3$	$Th2$	Low	High	...	**Low**
$Exp3$ Firewall. DS2: Security admin.	$dc2.4$	$Th3$	Authorized	Certain	...	**Very-high**
$Exp4$ Firewall. DS2: System admin.	$dc2.4$	$Th4$	Medium	Certain	...	**High**
$Exp5$ SIEM. P3: Security operator	$dc3.9$	$Th5$	Authorized	Low	...	**Medium**

Table 19: UCSS qualitative privacy violation impact

Th	**Exposure**	**Data content**	**Violation magn.**		I_C		I_R		I
			Value	**Mtv.**	**Value**	**Mtv.**	**Value**	**Mtv.**	
$Th1$	$Exp1$ Antivirus. C1: Network Admin	$dc1.3$	Medium	...	High	...	Med-high	...	**High**
$Th2$	$Exp2$ Antivirus. C1: Network user	$dc1.3$	Very Limited	...	High	...	High	...	**High**
$Th3$	$Exp3$ Firewall. DS2: Security Admin	$dc2.4$	Limited	...	High	...	High	...	**High**
$Th4$	$Exp4$ Firewall. DS2: System Admin	$dc2.4$	Limited	...	High	...	High	...	**High**
$Th5$	$Exp5$ SIEM.P3: Security Operator	$dc3.9$	Important	...	Low	...	Low	...	**Low**

Table 20: UCSS qualitative privacy violation risk

Th.	**Exposure**	**Data content**	L	I	R
$Th1$	$Exp1$ Antivirus. C1: Network admin.	$dc1.3$	High	High	**High**
$Th2$	$Exp2$ Antivirus. C1: Network user	$dc1.3$	Low	High	**Medium**
$Th3$	$Exp3$ Firewall. DS2: Security admin.	$dc2.4$	Very-high	High	**High**
$Th4$	$Exp4$ Firewall. DS2: System admin.	$dc2.4$	High	High	**High**
$Th5$	$Exp5$ SIEM. P3: Security operator	$dc3.9$	Medium	Low	**Low**

sary is not allowed to access this kind of data and the respondents (i.e., data owners) are not aware of and did not give their consent for this access. This threat involves potentially

is low

very sensitive information (i.e., file content) and leaking this information can seriously affect the organization reputation. However, the magnitude of violation is *medium* thus the reputation impact factor is *med-high*.

In the second threat in Table 20, $Th2$, the same data content $dc1.3$ is exposed to a *network user* with a *medium* privacy violation risk. Unlike the *network administrator* this adversary can only sniff data from a limited number of respondents belonging to the same Ethernet segment where the adversary is connected The privacy violation likelihood is *low* (see Table 18) despite the *high* likelihood of re-identification as this adversary has *low* likelihood of access to the data. The re-identification likelihood is *high* because files exposed in $dc1.3$ can contain the identity of the respondent and the adversary may have background information about respondents (i.e., office colleagues). Despite the *very limited* violation magnitude the impact severity is *high* (Table 19) as both non-compliance and reputation impacts are *high*. Non-compliance impact is *high* because the adversary is not allowed to access this kind of data and the respondents are not aware of and did not give their consent for this access. Despite the very limited violation magnitude, the impact on reputation is also *high* because the violation is perpetrated by an internal and nonprivileged adversary and the leaked data is very sensitive.

Threats $Th3$ and $Th4$ (in Table 20) involve respectively the *security administrator* and *system administrator* accessing $dc2.4$ (i.e., *IP_int*, *IP_ext*, *email_header*, *email_cont*) from data storage *DS2* containing threat logs from the firewall. As shown in Table 18, the privacy violation likelihood is *very-high* for the security administrator and *high* for the system administrator. The two adversaries have very different likelihood of access but the same re-identification likelihood. The security administrator has *authorized* access to the data from *DS2* while the system administrator has physical access to the machine hosting *DS2* and can attempt to gain access to this data content with a *medium* likelihood of access. Both adversaries have a *certain* likelihood of re-identification as $dc2.4$ contains *email_headers* composed by the name and surname of the respondent. The exposure magnitude (assessed in Step 2, Table 15) and the violation magnitude (Table 19) are *limited*. Application level firewall processes a very big amount of data but the *email_cont* is very rarely collected and stored by a very limited number of rules. Despite the limited violation magnitude, impact severity (Table 19) is *high* for both adversaries. Non-compliance impact factor is *high* for a security administrator, although the respondents provided consent, a policy that allows the organization to systematically inspect the content of emails is in conflict with the regulation. Non-compliance impact factor is also *high* for the *system administrator*, since he is not supposed to access that data. In addition, the organization should respect higher security requirements when processing very sensitive information such as emails and email headers. Reputation loss impact factor is *high* for both adversaries due to the sensitivity of email content. Hence, based on privacy violation likelihood and Impact severity, both $Th3$ and $Th4$ have *high* privacy violation risk.

The last threat $Th5$ has *low* privacy violation risk (Table 20). $Th5$ involves the exposure of $dc3.9$ (*IP_int*, *IP_ext*, *application_name*) from the SIEM process *P2* to a security operator. The privacy violation likelihood is *medium*, as shown in Table 18. Although this adversary is *authorized* to access data from *P2*, the data content does not allow him to re-identify the respondents without the knowledge of the mapping between IP addresses and respondents identity. A security operator in UCSS is not very likely to have this kind of information, and for this reason, the re-identification likelihood is *Low*. The exposure magnitude (assessed in Step 2, Table 15) is *important* since the SIEM collects big amount of data corresponding to this data content. Impact severity instead is *low* because both non-compliance and reputation loss impact factors are *low*. In fact, the access to data is compliant for this adversary and

Table 21: UCSS prioritized privacy violation threats

Th.	Exposure	Data content	R	Adversary trust		Priority
				Value	Mtv.	
$Th1$	$Exp1$ Antivirus. C1:Network admin.	$dc1.3$	High	Moderately trusted	...	2
$Th2$	$Exp2$ Antivirus. C1: Network user	$dc1.3$	Medium	Untrusted	...	4
$Th3$	$Exp3$ Firewall. DS2: Security admin.	$dc2.4$	High	Fully trusted	...	5
$Th4$	$Exp4$ Firewall. DS2: System admin.	$dc2.4$	High	Moderately trusted	...	2
$Th5$	$Exp5$ SIEM.P3: Security operator	$dc3.9$	Low	Trusted	...	11

$dc3.9$ (mainly the application name) is not very sensitive and contains seldom any sensitive information.

The complete table reporting qualitative privacy violation risk (like Table 20) for the entire UCSS system contains several hundred threats. The great majority of these threats emerge in two CSSs (SIEM and Firewall) while fewer threats emerge in each of the other CSSs. For example, we identified 24 threats in the antivirus CSS, 8 with *high* risk, 13 *medium*, and 3 *low*.

The identification of these threats in UCSS has a high value for the academic institution, not only to better understand the privacy implications of the deployed CSS and possibly mitigate the threats but also to comply with regulation. For example, the new EU General Data Protection Regulation[5] (GDPR) requires to keep detailed "Records of personal data processing activities" (article 30), and the EPIC's threat analysis was an excellent tool to isolate this information for the CSS.

5.5.2 Prioritizing risk mitigation actions in UCSS

We now follow the EPIC methodology in assigning a priority to each threat evaluated in Table 20 by considering the risk value and the adversary's trustworthiness.

$Th1$ and $Th4$ are the threats with the highest priority (priority equals 2) among the threats reported in Table 21. $Th1$ has a *high* privacy violation risk and the adversary (network administrator) is *moderately trusted*. In fact, this adversary is trusted within the organization, but not authorized to access data content $dc1.3$ exposed in $Th1$. In addition, network administrators have no training to deal with private information collected by the UCSS. Threat $Th4$ is also characterized by a *high* privacy violation risk and involves the system administrator as adversary. Similarly to the network administrator, this adversary is *moderately trusted* because trusted within the organization, but not authorized to access the data content $dc2.4$ exposed in $Th4$.

Threat $Th2$ follows in priority order (priority equals 4). $Th2$ contains the same data content than $Th1$ (i.e., data content $dc1.3$) and the data exposures at the origin of these threats (respectively $Exp1$ and $Exp2$) were identified for the channel *C1* of the antivirus. These

[5]Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

threats, however, involve different adversaries. The adversary in $Th2$, the network user, is *untrusted* because not authorized to access the data and not trained to manage personal data. Despite they are less trusted than network administrators (adversary in $Th1$), $Th2$ is assigned a lower priority than $Th1$ because its privacy violation risk (*medium*) is lower than the risk of $Th1$.

Threat $Th3$ is similar to $Th4$ (i.e., same component and same privacy violation risk) but has a lower priority due to the fact that the adversary (i.e., security administrator) is more trusted than the system administrator.

Finally, threat $Th5$ has the lowest priority since it involves a *trusted* adversary, the security operator, and it has a *low* privacy violation risk.

The prioritized list of privacy violation threats has a central role in guiding the mitigation actions. Considering the UCSS use case, this step of the EPIC methodology highlighted two benefits. First, it forced the trust analysis of the different actors considered as adversaries, identifying the higher reliability of security operators with respect to system and network administrators because of their different training and expertise. Second, considering only trust, *Th2* would be considered at highest priority, while considering only risk *Th1*, *Th3*, and *Th4* would be considered before *Th2*. Only the balanced evaluation of the combination of the two factors suggests the non-trivial priority order reported in Table 12.

6 Conclusions and Future Work

In this paper, we proposed EPIC, a methodology to identify and evaluate privacy violation threats resulting from the deployment of an organizational cybersecurity system.

The methodology guides a privacy expert, with the collaboration of the organization's security team, through four steps of analysis namely modeling the cybersecurity system, identifying data exposures, identifying privacy threats and evaluating and prioritizing these threats.

The privacy risk assessment resulting from the methodology can be used to compare cybersecurity systems in terms of privacy preservation. By considering the trustworthiness of the adversary together with the privacy violation risk, the methodology also provides a prioritization of the activities necessary to mitigate the risk of the identified privacy threats.

We refined and validated the methodology by applying it to the actual cybersecurity system of a large academic institution reporting some of the analysis and results in the paper.

Two contrasting needs emerged while designing the *EPIC* methodology: on one side, in order to increase the accuracy of privacy violation risk assessment, a larger number of aspects needs to be modeled and deep evaluations by privacy experts need to be performed. On the other side, the methodology should be practical: the experts should be able to apply it to real systems with a reasonable effort and time. Balancing these needs required us to omit some details or special cases that add complexity to the process, while not always affecting the evaluation result in the specific context of privacy in CSS. For example, in a first attempt to model privacy violation, we explicitly took into account "linking information" i.e., attributes that can associate several pieces of information to the same individual. Consider for instance a data log that reports a given sensitive information associated with pseudo-id 123; another log contains the association between pseudo-id 123 and respondent's identity. By accessing these two logs the adversary can violate the respondent's privacy through the "linking information" i.e., pseudo-id 123. EPIC does not explicitly provide guidance to the experts for analyzing this re-identifying method since, in our case study, this form of reasoning never disclosed additional privacy threats, while

adding complexity. Despite we believe that our use case is representative of a large class of CSS, there may be cases that require a more detailed analysis, including linking. Actually, linking information is captured by our formal model as a special case of quasi-identifier (see our definition of quasi-identifier) and can be considered in Steps 3 and 4 of EPIC. More generally, a technically deeper analysis on specific aspects can be conducted as a second phase assessment or as part of the remediation for particular privacy threats and system components.

A natural follow-up to this work would be to guide through the selection and implementation of privacy protection solutions, including organizational and legal interventions. Regarding technical solutions, despite there are several privacy enhancing techniques that could be applied in this domain, a careful evaluation is required specifically for preserving data quality and computational efficiency in order not to impact on security protection. Indeed, some existing privacy enhancing techniques have been shown to reliably protect privacy, however, they often severely affect the quality of data and come with a substantial computational overhead. For this reason, we plan to continue our investigation by analyzing how existing techniques could be effectively adapted and combined, and possibly design new methods.

Acknowledgments

The authors want to sincerely thank the anonymous reviewers for their feedback, which helped improving this paper. This work is partly supported by the Ministry of Foreign Affairs, Italy for the Italy-Israel joint research project *PACS: Privacy-aware cybersecurity*. We thank Erez Shmueli, Eran Toch, and Laura Radaelli for valuable comments on the proposed methodology.

References

[1] N. Baracaldo and J. Joshi. A trust-and-risk aware RBAC framework: Tackling insider threat. In *Proceedings of the 17th ACM Symposium on Access Control Models and Technologies*, SACMAT '12, pages 167–176, New York, NY, USA, 2012. ACM.

[2] C. Bettini, S. Mascetti, X. S. Wang, D. Freni, and S. Jajodia. Anonymity and historical-anonymity in location-based services. In *Privacy in Location-Based Applications: Research Issues and Emerging Trends*, pages 1–30. Springer, Berlin, Heidelberg, 2009.

[3] C. Bettini, X. S. Wang, and S. Jajodia. The role of quasi-identifiers in k-anonymity revisited. *CoRR*, abs/cs/0611035, 2006.

[4] R. Caralli, J. Stevens, L. Young, and W. Wilson. Introducing octave allegro: Improving the information security risk assessment process. Technical report CMU/SEI-2007-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2007.

[5] Y.-L. Chen et al. Data flow diagram. In *Modeling and Analysis of Enterprise and Information Systems*, pages 85–97. Springer, 2009.

[6] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. k-Anonymity. In T. Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*. Springer-Verlag, 2007.

[7] R. Clarke. Privacy impact assessment: Its origins and development. *Computer Law & Security Review*, 25(2):123 – 135, 2009.

[8] CNIL (Commission Nationale de l'Informatique et des Libertés). *PRIVACY IMPACT ASSESSMENT (PIA): Methodology (how to carry out a PIA)*, June 2015.

[9] CNIL (Commission Nationale de l'Informatique et des Libertés). *PRIVACY IMPACT ASSESSMENT (PIA): Tools (templates and knowledge bases)*, June 2015.

[10] C. Colwill. Human factors in information security: The insider threat - who can you trust these days? *Information Security Technical Report*, 14(4):186–196, Nov. 2009.

[11] CORDIS. Privacy and Data Protection Impact Assessment Framework for RFID Applications. Report, European Commission's Community Research and Development Information Service, 2011.

[12] S. J. De and D. Le Métayer. Priam: A privacy risk analysis methodology. In *International Workshop on Data Privacy Management*, pages 221–229. Springer, 2016.

[13] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[14] C. Dwork. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*, pages 1–19. Springer, 2008.

[15] ENISA. Privacy enhancing technologies: Evolution and state of the art. Report, European Union Agency For Network and Information Security, December 2016.

[16] European advisory body on data protection and privacy. Guidelines on data protection impact assessment (pia). Report, European Commission, Directorate General Justice and Consumers, April 2017.

[17] W. Hartzog and I. Rubinstein. The anonymization debate should be about risk, not perfection. *Commun. ACM*, 60(5):22–24, Apr. 2017.

[18] D. Henshel, M. Cains, B. Hoffman, and T. Kelley. Trust as a human factor in holistic cyber security risk assessment. *Procedia Manufacturing*, 3:1117 – 1124, 2015. 6th International Conference on Applied Human Factors and Ergonomics AHFE 2015.

[19] M. Howard and S. Lipner. *The Security Development Lifecycle*. Microsoft Press, 2006.

[20] Information Commissioner Office (ICO). Conducting privacy impact assessments code of practice. Report, Information Commissioner Office, 2014.

[21] ISO. Information technology - Security techniques - Guidelines for privacy impact assessment. Standard, International Organization for Standardization, Geneva, CH, June 2017.

[22] B. Karabacak and I. Sogukpinar. Isram: information security risk analysis method. *Computers & Security*, 24(2):147–159, 2005.

[23] J. Larkin. *Strategic reputation risk management*. Springer, 2002.

[24] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115. IEEE, 2007.

[25] P. I. LLC. 20161 cost of data breach study: Global analysis. Technical report, Benchmark research sponsored by IBM, June 2016.

[26] J. Luna, N. Suri, and I. Krontiris. Privacy-by-design based on quantitative threat modeling. In *Risk and Security of Internet and Systems (CRiSIS), 2012 7th International Conference on*, pages 1–8. IEEE, 2012.

[27] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1), Mar. 2007.

[28] D. H. McKnight and N. L. Chervany. Trust and distrust definitions: One bite at a time. In *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives*, pages 27–54. Springer, Berlin, Heidelberg, 2001.

[29] P. Mell, K. Scarfone, and S. Romanosky. Common vulnerability scoring system. *IEEE Security & Privacy*, 4(6), 2006.

[30] NIST. Privacy risk management for federal information systems. Internal Report, National Institute of Standards and Technology , 2017.

[31] N. Notario, A. Crespo, Y. S. Martn, J. M. D. Alamo, D. L. Mtayer, T. Antignac, A. Kung, I. Kroener, and D. Wright. Pripare: Integrating privacy best practices into a privacy engineering methodology. In *2015 IEEE Security and Privacy Workshops*, pages 151–158, May 2015.

[32] M. C. Oetzel and S. Spiekermann. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems*, 23(2):126–150, 2014.

[33] A. Ornaghi and M. Valleri. Ettercap, 2006.

[34] The Privacy Office, Department of Homeland Security, Washington, DC 20528. *Privacy Impact Assessments: The Privacy Office Official Guidance*, June 2010.

[35] A. Shostack. *Threat modeling: Designing for security*. John Wiley & Sons, 2014.

[36] The European Commission. Data protection impact assessment template for smart grid and smart metering systems. Report, The European Commission, 2014.

[37] The European Parliament and the Council of the European Union. Regulation (EU) 2016/679 of the European parliament and of the council. *Official Journal of the European Union*, April 2016.

[38] G. Theodorakopoulos and J. S. Baras. On trust models and trust evaluation metrics for ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 24(2):318–328, Feb 2006.

[39] E. Toch, C. Bettini, E. Shmueli, L. Radaelli, A. Lanzi, D. Riboni, and B. Lepri. The privacy implications of cyber security systems: a technological survey. *ACM Computing Surveys (CSUR)*, x(x):xx, To appear.

[40] I. Wagner and D. Eckhoff. Technical privacy metrics: a systematic survey. *arXiv preprint arXiv:1512.00327*, 2015.

[41] D. Wright. The state of the art in privacy impact assessment. *Computer Law & Security Review*, 28(1):54–61, 2012.