Adding Privacy Protection to Distributed Software Systems

George O. M. Yee
Computer Research Lab, Aptusinnova Inc., Ottawa, Canada
Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada

Keywords: Distributed, Software, Development, System, Privacy, Risks, Mitigation.

Abstract: Distributed software systems are increasingly found in many aspects of our lives, as can be seen in social media, international online banking, and international commerce (e.g. Internet shopping). This widespread involvement of software in our lives has led to the need to protect privacy, as the use of the software often requires us to input our personal or private information. A first step to protecting privacy is to identify the risks to privacy found in the software system. Once the risks are known, measures can be put in place to mitigate the risks. This is best done at the early stages of software development due to the heavy costs of making changes after the software is deployed. This paper proposes a two-pronged approach, consisting of privacy risk identification followed by risk mitigation, for adding privacy protection to distributed software. The paper illustrates the approach with examples.

1 INTRODUCTION

Many distributed software systems targeting consumers have appeared in recent years, accompanying the rapid growth of the Internet. Such systems are available for banking, shopping, travel reservations, learning, healthcare, and even government online (e.g. the European Union). Most of these systems require a consumer's personal information in one form or another, leading to concerns over privacy.

Various approaches have been used to protect personal information, including data anonymization (Iyengar, 2002) and pseudonym technology (Song et al., 2006). Other approaches for privacy protection include treating privacy protection as an access problem and then bringing the tools of access control to bear for privacy control (Adams and Barbieri, 2006). However, these approaches presume to know where and what protection is needed. They presume that some sort of analysis has been done that answers the question of "where" and "what" with respect to privacy risks. Without such answers, the effectiveness of the protection comes into question. For example, protection against house break-ins is ineffective if the owner only secures the front door without securing other vulnerable spots such as windows (where and what). In the same way, privacy risk identification considering "where" and "what" is essential to effective privacy protection. The author's earlier work (Yee, 2016) proposed a visualization method for identifying privacy risks, based on this notion of "where" and "what". This paper extends that work by adding risk prioritization and mitigation.

This paper is organized as follows. Section 2 defines privacy, privacy preferences, and privacy risks. Section 3 presents a summary of privacy risk identification from (Yee, 2016). Section 4 describes the application of measures to mitigate the privacy risks. Section 5 presents an application example. Section 6 discusses related work. Finally, Section 7 presents conclusions.

2 PRIVACY

As defined by (Goldberg et al., 1997), privacy refers to the ability of individuals to *control* the collection, retention, and distribution of information about themselves. We add "purpose" to this definition. To see that "purpose" is needed, consider, for example, that one may agree to give out one's email address for the purpose of friends to send email but not for the purpose of spammers to send spam. This definition also suggests that "personal information", "private information" or "private data" is any information that can be linked to a person;

Yee, G.
Adding Privacy Protection to Distributed Software Systems.
DOI: 10.5220/0006434903510358
In *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - Volume 4: SECRYPT*, pages 351-358
ISBN: 978-989-758-259-2
Copyright © 2017 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

otherwise, the information would not be "about" the person. Thus, another term for private information is "personally identifiable information (PII)". These terms are used interchangeably in this paper. In addition, controlling the "collection" of information implies controlling *who* collects *what* information. Controlling the "retention" of information is really about controlling the *retention time* of information, i.e. how long the information can be retained before being destroyed. Controlling the "distribution" of information is controlling to which other parties the information can be *disclosed-to*.

A user's *privacy preference* expresses the user's desired control over a) *PII* - what the item of personal information is, b) *collector* - who can collect it, c) *purpose* - the purpose for collecting it, d) *retention time* - the amount of time the information is kept, and e) *disclosed-to* - which other parties the information can be disclosed-to. A *privacy risk* is the potential occurrence of any action or circumstance that will result in a violation of any of the components PII, collector, purpose, retention time, and disclosed-to in a user's privacy preference.

3 PRIVACY RISK IDENTIFICATION

As mentioned above, the method for privacy risk identification was previously presented as (Yee, 2016) and we summarize it here. The method may be applied to distributed software systems having the following common characteristics:

a) The software system requires the user's personal information in order to carry out its function.
b) The software system is distributed, i.e. modules of the system operate in different locations.
c) The software system may transmit the information (e.g., move it from one place to another within the system), store the information (e.g., store the information in a data base), and use the information to carry out its function.

The method is based on the notion that the *location* of personal information gives rise to privacy risks and consists of the following steps: i) determining all the possible locations in the software system where the user's personal information could reside, and ii) evaluating at each of these locations the possible ways in which the user's privacy preferences could be violated.

Step i) is accomplished by modeling the software system in terms of a Personal Information Map (PIM), using the notational elements in Table 1.

Table 1: Elements of a PIM.

Element		Description
Use Circle	○	Identifies where PII is used. Labeled with a letter together with a description of the use in a legend.
Data Store	□	Identifies where PII is stored. Labeled with a letter together with a description of the data store in a legend.
Same Physical Platform	⬚	Identifies use circles and data stores that execute on the same computing platform.
PII Data Flo w	⟶	Identifies the movement of PII from one location to another. Labeled with a number together with a description of the data in a legend.
Non-PII Data Flow	⇢	Identifies the movement of SD from one location to another. Labeled with a number together with a description of the data in a legend.
Legend		Descriptions corresponding to the letters or numbers with which the above elements were labeled.

Physically separate units, as delineated by dashed rectangles, allow the identification of risks for any data flow between them. Circles or squares not enclosed by a dashed rectangle are understood to be already physically separate units. Figure 1 shows the PIM for the software system of an online seller of merchandise (e.g. Amazon.com) that has modules developed and running in both the United States and Canada. The system requires the user's name, address, merchandise selection, and credit card number. These are considered as three personal information items where name and address together are considered as one item. Figure 1 also shows three non-personal information flows (4, 5, 6). The dashed rectangle enclosing A, B, and C indicates that A, B, and C all run on the same physical computing platform.

Step ii) is accomplished by inspecting the PIM resulting from step i). For each location (flow arrow, storage square, and use circle) and each personal information item, enumerate the possible ways in which a privacy preference component may be violated. This may be achieved by asking risk questions for each component (see Table 2), and drawing conclusions based on security and systems knowledge, as well as experience. Record the results in a Privacy Risks Table containing two columns: the left column for records of the form "(PII_1, PII_2, …/ locations)" and the right column containing the corresponding privacy risks. Table 3 illustrates this step for the online seller of Figure 1.

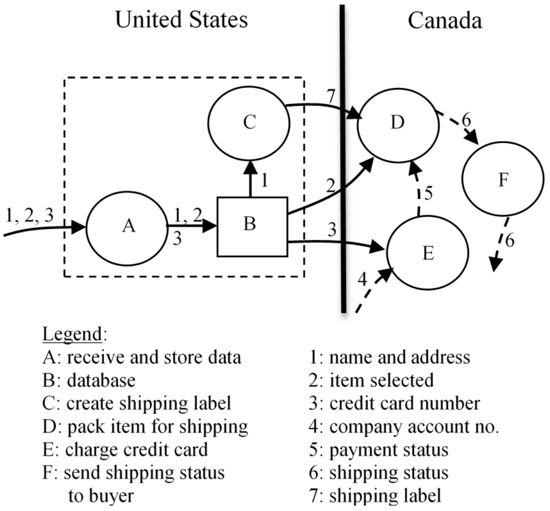


Figure 1: PIM for an online seller of merchandise.

Table 2: Risk questions.

Component	Risk Questions
PII	How can the user be asked for other PII, either intentionally or inadvertently?
collector	How can the PII be received by an unintended collector either in addition to or in place of the intended collector?
purpose	How can the PII be used for other purposes?
retention time	How can the PII retention time be violated?
disclose-to	How can the PII be disclosed either intentionally or inadvertently to an unintended recipient?

Table 3: Partial privacy risks table corresponding to Fig. 1.

(PIIs / locations)	Privacy Risks
(1, 2, 3 / path into A); (2 / path into D); (3 / path into E); (7 / path into D)	Man-in-the-middle attack violates *collector*, *purpose*, and *disclose-to*
(1, 2, 3 / path into A)	User could be asked for personal information that violates *PII*
(1, 2, 3 / A); (1 / C); (2, 7 / D); (3 / E)	Trojan horse or hacker attack on A, C, D, and E violates *collector*, *purpose*, and *disclose-to*
(1, 2, 3 / B)	SQL attack on B violates *collector*, *purpose*, and *disclose-to*
(1, 2, 3 / B)	Information in B could be kept past the *retention time*

4 PRIVACY RISKS MITIGATION

Privacy risks mitigation involves the application of security safeguards (e.g. encryption) to mitigate the risks. Due to the shortage of company resources, e.g. time and money, not all risks will be mitigated. In this case, it is necessary to choose a subset of risks for mitigation.

4.1 Selecting Privacy Risks for Mitigation

Intuitively, one would want to mitigate risks that are highly probable to be realized, and that once realized, would result in very costly damages. In other words, due to budget constraints, we feel that we can ignore the risks that tend not to be realized and even if realized would cause very little damage. Determining which risks to mitigate may be assisted though weighting the risks according to criteria.

(Salter et al., 1998) proposed a method for applying weights to various forms of attacks in order to determine if a particular attack would be probable. They focused on three aspects of an attack, namely “risk”, “access”, and “cost”, where “risk” is risk to the safety of the attacker, “access” is the ease with which the attacker can access the system under attack, and “cost” is the monetary cost to the attacker to mount the attack. To avoid confusion between “risk” to the safety of the attacker and “risk” to privacy, we use “safety” for “risk” to the safety of the attacker. The weight values are simply “L”, “M”, and “H” for Low, Medium, and High, respectively. These attack aspects can be represented using a 3-tuple, as [safety, access, cost] and so [H, M, L] would be an instance of the weights. For example, consider a physical attack such as a mugging incident in a park. In this case, the risk to the safety of the attacker would be high (the person being mugged could be an undercover police officer), the attacker’s ease of access would be high (people stroll through the park all the time), and the attacker’s cost would be low (not much needed to mount the attack). Thus, this attack has the weights [H, H, L].

In this work, we add a fourth aspect of an attack, namely the resulting damages from the attack. Thus, we use the 4-tuple [safety, access, cost, damages] with the same weight values L, M, and H. Hence we would definitely want to defend against privacy risks leading to attacks with weights [L, H, L, H]. We feel that we can ignore privacy risks having attacks with weights [H, L, H, L]. In reality, there is a spectrum of weights between these two boundaries, where a decision to defend or ignore may not be clear, and ultimately a judgment, perhaps based on other factors, may be needed. For example, it is not clear whether or not a privacy risk with associated weights [L, L, H, H] should be ignored, and one would decide to defend if one believes that no matter

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

how improbable the attack, the resulting damages must never be allowed to occur.

The uncertainty of deciding which risks to mitigate using the weights may be remedied through the use of a Mitigation Policy. This policy would identify the 4-tuples of weights whose associated risks are to be mitigated. For example, the policy might state that risks with associated 4-tuples [L, *, *, H] and [L, *, *, M] are to be mitigated, where "*" indicates possibilities L, M, and H. This policy may stipulate the mitigation of more risks or fewer risks, depending on the perceived level of attacker activity over a past period of time (e.g., 6 months). The perceived level of attacker activity could be based on news of attacks from a security newsletter source such as SANS Newsletters (https://www.sans.org/newsletters/). As an example, suppose it is perceived that the level of attacker activity over the previous 6 months is high. Then the above policy might be changed from "mitigate risks with associated 4-tuples [L, *, *, H] and [L, *, *, M]" to "mitigate risks with associated 4-tuples [L, *, *, *] and [L, *, *, *], thus permitting an increase in the number of risks mitigated, as a reflection of the high level of attacker activity. Of course, it is assumed here that management could be persuaded to increase the financial budget for mitigation.

4.2 Method for Privacy Risk Mitigation

1. Apply weights to the privacy risks using the procedure described in Section 4.1 above. Develop a Mitigation Policy that can be used as a basis for selecting the 4-tuples of weights whose associated risks are to be mitigated. Select the risks for mitigation based on this policy.
2. Apply security measures (e.g. encryption) to mitigate the privacy risks.

Applying this method for mitigation to the example of Section 3 gives Table 4, containing the weights (Step 1) and mitigations (Step 2) corresponding to the risks identified in Table 3.

The weights in Table 4 were assigned as follows. For the man-in-the-middle attack, the risks to the attacker's safety is low since he or she is attacking at a distance; the access is high since it's the Internet; the cost is low as not much equipment is needed; the damages would be high since the attacker could post the private information leading to heavy damages to the company's reputation. Similar considerations apply to the weight assigned to the Trojan horse or hacker attack. For the SQL attack on B, accessibility was assigned as low and cost as high because improvements to the database user interface were recently carried out to guard against SQL attacks. The risk of the user being asked for information violating PII and the risk of information kept past the retention time were considered as potential accidents caused by the company itself. Therefore, the risk to safety, the accessibility, and the costs were deemed to be low, high, and low respectively. The resulting damages were considered to be medium because the accidents would likely be quickly discovered through auditing and remedied. The privacy risks in Table 4 that are labeled as "not considered for mitigation" were so labeled as a result of a Mitigation Policy that states "only mitigate risks with weights [*, *, L, H]".

Table 4: Weights and mitigations for risks in Table 3.

(PIIs / locations)	Privacy Risks	Weights (Step 1)	Mitigations (Step 2)
(1, 2, 3 / path into A); (2 / path into D); (3 / path into E); (7 / path into D)	Man-in-the-middle attack violates *collector*, *purpose*, and *disclose-to*	[L, H, L, H]	Use SSL for (1, 2, 3 / path into A); use encryption with electronic signatures for the rest.
(1, 2, 3 / path into A)	User could be asked for personal information that violates *PII*	[L, H, L, M]	Not considered for mitigation
(1, 2, 3 / A); (1 / C); (2, 7 / D); (3 / E)	Trojan horse or hacker attack on A, C, D, and E violates *collector*, *purpose*, and *disclose-to*	[L, H, L, H]	Use a combination firewall and intrusion protection system in front of A.
(1, 2, 3 / B)	SQL attack on B violates *collector*, *purpose*, and *disclose-to*	[L, L, H, H]	Not considered for mitigation
(1, 2, 3 / B)	Information in B could be kept past the *retention time*	[L, H, L, M]	Not considered for mitigation

5 APPLICATION EXAMPLE

Consider an airline reservation system called AccuReserve offered by a Canadian airline with headquarters in Toronto, Canada. AccuReserve (fictitious) is composed of the globally developed and distributed modules shown in Table 5, along with their private information requirements.

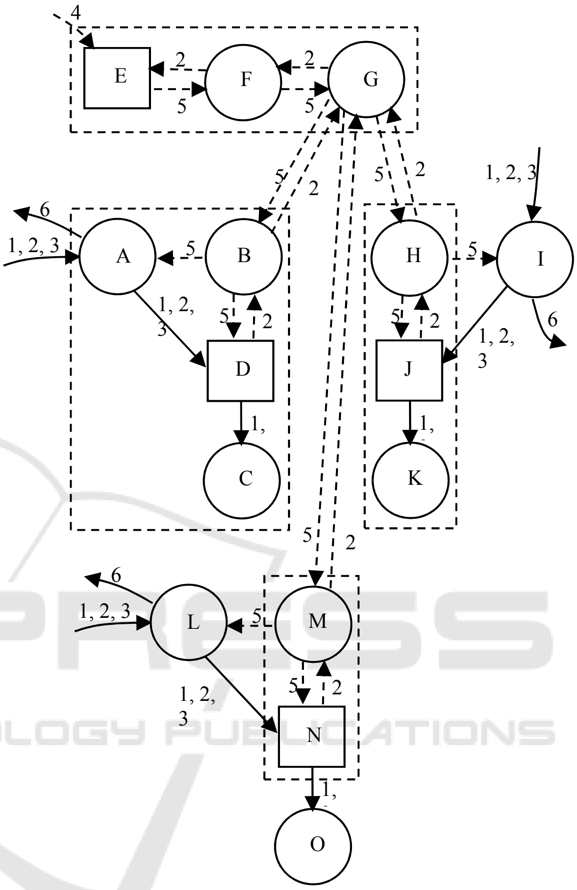
The country-specific modules (modules other than Main) were developed and runs in their respective countries to fulfill a requirement by the respective governments to have databases of their citizens' personal information reside within their respective countries. Given this requirement, it was decided to develop the user interface within each country, to take advantage of local expertise in customizing the interface to client behavior, in order to improve user friendliness and efficiency.

Table 5: AccuReserve modules and private data required.

Module	Function	Private Information Required
Main – developed and runs in Canada	Maintains central database of aircraft routes and seat availability; also acts as the client interface and client database for Canada	Identification details (name, address, phone, citizenship, passport number), flight details requested (originating city, destination city, date of departure, one-way ticket or return ticket, date of return (if return ticket)), payment details (name, credit card number, credit card expiry date)
Mod-US – developed and runs in the US	Client interface and database for the US	Same as for Main
Mod-EU – developed and runs in the EU (Germany)	Client interface and database for the EU	Same as for Main

5.1 Identification of Privacy Risks

Draw the PIM for AccuReserve. The resulting PIM is shown in Figure 2, and was achieved by drawing the PIM for each module (Main, Mod-US, and Mod-EU) and then linking the modules together with communication links. Main is made up of the master database of available flights together with the user interface for Canada. The user interfaces all have the same PIM structure, which is not surprising since they all serve the same purpose. However, the components that are responsible for communicating with the user (represented by A, L, and I in the PIM) are specialized within each country to maximize user friendliness and efficiency. These components are not necessarily distinguishable at the PIM level. Recall that the PIM is not a software design or architecture diagram and does not necessarily show all the software components in the system. An interesting aspect of this PIM is that it reflects the fact that the software design is already privacy friendly in the sense that private information does not flow outside the country of origin. Thus, "flight details requested" and "flight details assigned" (data items 2 and 5) are non-private, being associated with the user by way of an identification hash (considered part of each data item), except for 2 when it is associated with the identification details of 1 (on the path leading into and out of A, I, and L).



Legend:

A, I, L: receive and store data
B, H, M: communicate w. Main
C, K, O: charge credit card
D, E, J, N: databases
F: flight availability manager
G: communicate with countries

1: identification details
2: flight details requested
3: payment details
4: flight availability updates
5: flight details assigned
6: travel itinerary

Figure 2: PIM for AccuReserve; Main consists of E, F, G, A, B, D, and C; Mod-US consists of L, M, N, and O; Mod-EU consists of H, I, J, and K.

Enumerate Privacy Risks at Private Information Locations. Table 6 gives a partial Privacy Risk Table for locations in Figure 2 that have interesting or serious privacy risks. The theft of personal information means that the information is under the control of an unintended party. Clearly, this can violate the corresponding privacy preference or preferences in terms of violating *collector*, *purpose*, *retention time*, and *disclose-to*. The risk of personal

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

information theft arises so often that it is convenient to call it *CPRD-risk*, from the first letters of collector, purpose, retention time, and disclose-to. The risks in Table 6 were obtained as follows. For the first and second rows, it was noticed that the personal information flows through transmission paths connecting physically distinct units. The risk questions of Table 2 were then considered, leading to possible man-in-the-middle attacks that give rise to CPRD-risk. Notice that "(1, 2, 3 / path between A and D)" is excluded because A and D both run on the same platform (so the path is not very accessible to attack). For the third row, violations of PII are always possible unless strict controls are in place against it. For the fourth row, it was observed that private data are input to information use processes A, I, L, C, K, O. The risk questions of Table 2 were again considered, leading to possible Trojan horse or hacker attacks that again give rise to CPRD-risk. For the fifth row, it was noticed that private data are stored in databases. Once again the risk questions were considered, leading to possible SQL attacks against the databases, giving rise to CPRD-risk. For the second to last row, it was noticed that private information stored in databases could be subject to insider attacks. Finally, for the last row, it was observed that the private data stored in the databases could be kept past their retention times. In each of these rows, knowledge of the system (private data locations) and knowledge of information security (possible attacks) were needed to identify the risks. It should be noted that the links between G and B, G and M, and G and H are also vulnerable to man-in-the-middle attacks, but these attacks would not be privacy attacks, since these links are not used for private information. Non-privacy attacks are outside the scope of this paper.

Table 6: Privacy risks table corresponding to Fig. 2.

(PIIs / locations)	Privacy Risks

(1, 2, 3 / path into A);
(1, 2, 3 / path into I);
(1, 2, 3 / path into L)
	Man-in-the-middle attacks lead to CPRD-risk.

(1, 2, 3 / path between I and J); (1, 2, 3 / path between L and N); (1, 3 / path between N and O)	Man-in-the-middle attacks lead to CPRD-risk.
(1, 2, 3 / path into A); (1, 2, 3 / path into I); (1, 2, 3 / path into L)	The user could be asked for personal information that violates PII (i.e. asked for PII other than 1, 2, 3).
(1, 2, 3 / A, I, L); (1, 3 / C, K, O)	Trojan horse, or hacker attacks on the personal information use circles lead to CPRD-risk.
(1, 2, 3 / D, J, N)	Potential SQL attacks on D, J, and N lead to CPRD-risk.
(1, 2, 3 / D, J, N)	Potential insider attack steals private information from D, J, and N resulting in CPRD-risk.
(1, 2, 3 / D, J, N)	Private information in D, J, and N could be kept past the retention time.

5.2 Mitigation of Privacy Risks

Applying the method for privacy risk mitigation described in Section 4.2 produces Table 7, containing the weights (Step 1) and mitigations (Step 2) corresponding to the risks identified in Table 6.

The weights in Table 7 were assigned as follows. A weight of [L, H, L, H] was assigned to the first row after the same considerations as that described in Section 4.2 for man-in-the-middle attacks. A weight of [M, M, L, H] was assigned to the second row since the paths in this row are relatively short (connecting components in the same module), leading to greater risk for the attacker (greater risk of being seen) and lower accessibility (fewer places to access the link). A weight of [L, H, L, M] was assigned to the third and last rows out of the same considerations as in Section 4.2, for the risk of the user being asked for information that violates PII and the risk of private information kept past the retention time. A weight of [L, H, L, H] was assigned to the Trojan horse or hacker attack in the fourth row and the SQL attacks in the fifth row since the attacker could operate from a distance with easy access through the Internet and with relatively low costs. A weight of [L, H, L, H] was assigned to the risk of an insider attack in the sixth row since an insider can hide in plain sight, has high access by virtue of being an insider, and carry out the attack at zero cost to herself.

The privacy risks in Table 7 were prioritized using the Mitigation Policy "only mitigate risks with weights [L, *, *, H]".

6 RELATED WORK

This section primarily concerns related works involving privacy risk prioritization and mitigation. Please consult (Yee, 2016) for related works on privacy risk identification.

In terms of risk prioritization, no references were found that deals directly with the prioritization of privacy risks. However, abundant work exists on the assessment of security risks, which is closely related to prioritizing privacy risks. (Alizadeh and Zannone, 2016) present a risk-based framework that facilitates the analysis of business process executions. The framework detects non-conforming process behaviors and ranks them according to criticality, which is determined by the execution's impact on organizational goals. The criticality ranking enables a security analyst to prioritize the most severe

SCIENCE AND TECHNOLOGY PUBLICATIONS

incidents. (Jorgensen et al., 2015) propose decomposing risk associated with a mobile application into several risk types that are more easily understood by the application's users and that a mid-level risk summary be presented that is made up of the dimensions of personal information privacy, monetary risk, device availability/stability risk, and data integrity risk.

Table 7: Weights and mitigations for risks in Table 6.

(PIIs / locations)	Privacy Risks	Weights (Step 1)	Mitigations (Step 2)
(1, 2, 3 / path into A); (1, 2, 3 / path into I); (1, 2, 3 / path into L)	Man-in-the-middle attacks lead to CPRD-risk.	[L, H, L, H]	Use SSL for (1, 2, 3 / path into A); 1, 2, 3 / path into I); (1, 2, 3 / path into L)
(1, 2, 3 / path between I and J); (1, 2, 3 / path between L and N); (1, 3 / path between N and O)	Man-in-the-middle attacks lead to CPRD-risk.	[M, M, L, H]	Not considered for mitigation
(1, 2, 3 / path into A); (1, 2, 3 / path into I); (1, 2, 3 / path into L)	The user could be asked for personal information that violates PII (i.e. asked for PII other than 1, 2, 3).	[L, H, L, M]	Not considered for mitigation
(1, 2, 3 / A, I, L); (1, 3 / C, K, O)	Trojan horse, or hacker attacks on the personal information use circles lead to CPRD-risk.	[L, H, L, H]	For each module, constrain all incoming traffic to a firewall with intrusion protection
(1, 2, 3 / D, J, N)	Potential SQL attacks on D, J, and N lead to CPRD-risk.	[L, H, L, H]	Use strong encryption on the databases D, J, and N.
(1, 2, 3 / D, J, N)	Potential insider attack steals private information from D, J, and N resulting in CPRD-risk.	[L, H, L, H]	Strengthen screening of potential employees; deal with employee grievances fairly
(1, 2, 3 / D, J, N)	Private information in D, J, and N could be kept past the retention time.	[L, H, L, M]	Not considered for mitigation

Their work suggests that privacy risk prioritization may be facilitated by decomposing the risks into more easily understandable categories or dimensions (as also done in this work). (Islam et al., 2016) present a framework for threat analysis and risk assessment of automotive embedded systems to systematically tackle security risks and determine security impact levels. The latter serve to prioritize the severity of the risks. The framework aligns with several industrial standards.

In terms of privacy risk mitigation, no other work similar to this work was found. (Kandappu et al., 2013) examine crowd-sourced survey platforms and show how easily user privacy can be compromised by collating information from multiple surveys. They propose a new crowd-sourcing platform that allows users to control their privacy loss using at-source obfuscation. (Lucas and Borisov, 2008) address the privacy risks of social networking websites where the providers of such websites can observe and collect the information that users transmit through the network. They propose to mitigate these risks through the implementation of a new architecture for protecting the information transmitted, using encryption. These authors have implemented the architecture using a prototype Facebook application and claim that their implementation strikes a balance between protecting privacy and maintaining Facebook's usability. (Oladimeji et al., 2011) look at healthcare delivery through ubiquitous computing and suggest that new techniques are needed to deal with the concerns for security and privacy within such delivery. They propose a goal-centric and policy-driven framework for deriving security and privacy risk mitigation strategies in ubiquitous health information interchange. These authors employ scenario analysis and goal-oriented techniques to model security and privacy objectives, threats, and mitigation strategies in the form of safeguards or countermeasures.

7 CONCLUSIONS AND FUTURE WORK

This work has proposed a straightforward approach for adding privacy protection to distributed software systems, consisting of two parts. The first part identifies the privacy risks, and the second part mitigates the risks, both parts focusing attention on locations that contain PII.

Some of the strengths of the method include: a) provides a structured way to identify privacy risks, and b) prioritizes the risks to be mitigated, to account for budgetary constraints.

Some weaknesses of the method are: a) drawing

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

the PIM is a manual task, prone to error, and b) the assignment of weights for selecting the risks to be mitigated is subjective and dependent on the expertise and experience of the person or persons doing the assignment. Weakness a) can be addressed by building tools for automatically drawing the PIM. Similar tools already exist for rendering a software architecture diagram from the reverse engineering of code, e.g., (Nanthaamornphong et al., 2013). Furthermore, automated analysis of the PIM should be feasible by using a rules engine to automate the enumeration of privacy risks, based on machine understanding of the graphical notation in the PIM. Weakness b) may be overcome by replacing the weights with probabilities of risk realization, calculated objectively with an algorithm that takes account of factors such as past attacks, existing threats, the state of attack technology, and existing defenses of the software system. Note that we have already proposed to partially incorporate the impact of past attacks, in terms of enlarging the number of risks to be mitigated (see Section 4.1) based on a high level of past attacker activity.

Future work includes resolving weaknesses a) and b), and validating the effectiveness of the approach in industrial settings.

REFERENCES

Adams, C. and Barbieri, K., 2006. Privacy Enforcement in E-Services Environments. Chapter in *Privacy Protection for E-Services*. Edited by G. Yee. Idea Group, Inc.

Alizadeh, M. and Zannone, N., 2016. Risk-based Analysis of Business Process Executions. Proceedings of the *6th ACM Conference on Data and Application Security and Privacy (CODASPY'16)*, pp. 130-132.

Goldberg, I., Wagner, D., and Brewer, E., 1997. Privacy-Enhancing Technologies for the Internet. Proceedings of *IEEE COMPCON'97,* pp. 103-109.

Islam, M., Lautenbach, A., Sandberg, C., and Olovsson, T., 2016. A Risk Assessment Framework for Automotive Embedded Systems. Proceedings of the *2nd ACM International Workshop on Cyber-Physical System Security (CPSS'16)*, pp. 3-14.

Iyengar, V. S., 2002. Transforming Data to Satisfy Privacy Constraints. Proceedings of the *8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02*), Edmonton, Alberta, pp. 279-288.

Jorgensen, Z., Chen, J., Gates, C., Li, N., Proctor, R. and Yu, T., 2015. Dimensions of Risk in Mobile Applications: A User Study. Proceedings of the *5th ACM Conference on Data and Application Security and Privacy (CODASPY'15)*, pp. 49-60.

Kandappu, T., Sivaraman, V., Friedman, A., Boreli, R., 2013. Exposing and Mitigating Privacy Loss in Crowdsourced Survey Platforms. Proceedings of the *2013 Workshop on Student Workhop*, pp. 13-16.

Lucas, M. and Borisov, N., 2008. flyByNight: Mitigating the Privacy Risks of Social Networking. Proceedings of the *7th ACM Workshop on Privacy in the Electronic Society (WPES '08)*, pp. 1-8.

Nanthaamornphong, A., Morris, K. and Filippone, S., 2013. Extracting UML Class Diagrams from Object-Oriented Fortran: ForUML. Proceedings of the *1st International Workshop on Software Engineering for High Performance Computing in Computational Science and Engineering (SE-HPCCSE'13),* pp. 9-16.

Oladimeji, E. A., Chung, L., Jung, H. T. and Kim, J., 2011. Managing Security and Privacy in Ubiquitous eHealth Information Interchange. Proceedings of the *5th International Conference on Ubiquitous Information Management and Communication (ICUIMC '11)*. Article No. 26.

Salter, C., Saydjari, O. S., Schneier, B. and Wallner, J., 1998. Toward A Secure System Engineering Methodology. Proceedings of the *New Security Paradigms Workshop*, pp. 2-10.

Song, R., Korba, L., and Yee, G., 2006. Pseudonym Technology for E-Services. Chapter in *Privacy Protection for E-Services*. Edited by G. Yee. Idea Group, Inc.

Yee, G., 2016. Visualization of Privacy Risks in Software Systems. Proceedings of the *Tenth International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2016)*, pp. 289-294.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS