TOWARDS DATA PROTECTION COMPLIANCE*

Nicola Zannone
Eindhoven University of Technology, Eindhoven, The Netherlands
n.zannone@tue.nl

Milan Petković
Philips Research, Eindhoven, The Netherlands
Eindhoven University of Technology, Eindhoven, The Netherlands
milan.petkovic@philips.com

Sandro Etalle
Eindhoven University of Technology, Eindhoven, The Netherlands
University of Twente, Enschede, The Netherlands
s.etalle@tue.nl

Keywords: Privacy, Auditing, Distributed Systems, Accountability, Legal requirements

Abstract: Privacy and data protection are fundamental issues nowadays for every organization. This paper calls for the development of methods, techniques and infrastructure to allow the deployment of privacy-aware IT systems, in which humans are integral part of the organizational processes and accountable for their possible misconduct. In particular, we discuss the challenges to be addressed in order to improve organizations privacy practices, as well as the approach to ensure compliance with legal requirements and increasing efficiency.

1 Data Protection

The ability to protect sensitive information is becoming a critical success factor for an increasingly large number of organizations, because of market pressure and legal constraints on data processing. Concerning the market side, we witness that users are taking into greater consideration the security and privacy practices of organizations before they subscribe to a certain service. Concerning the legal side, there exist a number of laws and regulations in place (e.g., Directive 95/46/EC, Privacy Act, HIPAA, etc.) that put stringent requirements on the collection, processing and disclosure of personal data. Organizations handling personal data have to implement such requirements in their business procedures.

The protection of sensitive information is often implemented by access control (Samarati and di Vimercati, 2001) and usage control (Park and Sandhu, 2004) systems. Here, protecting data implies guaranteeing that data are disclosed solely under specific conditions to specific users (the legitimate recipient), and that specific *obligations* are fulfilled after the data have been accessed. When specifically addressing *privacy protection* there is an additional requirement that has to be taken into account, the so called purpose specification: *"personal data shall be collected for specified, lawful and legitimate purposes and not processed in ways that are incompatible with the purposes for which data have been collected"* (Guarda and Zannone, 2009).

Nowadays, a number of frameworks exist for the specification and enforcement of purpose in access control (Backes et al., 2004; Byun and Li, 2008). However, these frameworks cannot cope with a number of increasingly important requirements for modern organizations. In the last years, organizations have become more fluid and decentralized, with a bigger interplay of physical and digital aspects and thinner security barriers. To mention a few aspects that have an impact on policy compliance:

- Humans play a fundamental role in organization processes. However, only some human tasks are under the control of the IT system. Consequently, the system cannot capture all actions performed by users and, therefore, it has to make security decisions on the basis of a partial knowledge.
- Organizations have to adapt quickly to changes in structure, businesses, and environmental conditions. Organizations thus have to deal increasingly often with unexpected situations.
- Outsourcing is nowadays a common business practice to reduce costs. Outsourcing has, however, a strong impact on the data protection requirements of an organization: personal data are

*This work has been partially funded by the EU-IST-IP-216287 TAS^3 project.

often disclosed to an external supplier over whom the organization may not have direct control.

None of the access control and privacy enforcement platforms existing today can cope with these aspects. The fundamental reason is that all these frameworks are preventative, in the sense that data are disclosed only if the purpose in the access request matches the purpose for which the data have been collected. Preventative frameworks have two main drawbacks: (1) they do not allow one to determine if data are actually processed accordingly to the intended (specified) purpose once being disclosed, and (2) they are too rigid to deal with exceptions.

Therefore, we need a new radically more flexible approach for the specification and enforcement of data protection policies to cope with the aforementioned circumstances. The aim of this paper is to discuss the main challenges that have to be faced and to propose research directions for the realization of such an approach. In particular, we propose:

- a novel purpose representation model that connects the intended purpose of data to the business activities performed by an organization;
- a-posteriori mechanisms for determining if data are used in accordance with the specified purpose;
- methods for analyzing user behavior and the purpose of data usage when audit-logs are partial;
- metrics for identifying and measuring the privacy risks of infringements;
- a novel infrastructure that combines a-priori and a-posteriori controls to support data protection compliance in distributed systems.

2 Challenges

Ensuring compliance to data protection policies that include purpose specification while detecting infringements (and allowing them when appropriate) requires addressing a number of challenges.

Challenge 1 *How can we verify that information has been used to achieve the specified purpose?*

Data protection is usually addressed by augmenting access control models with purpose, obligations, and conditions (Backes et al., 2004; Byun and Li, 2008; Hilty et al., 2005; Karjoth et al., 2002). In existing purpose-based frameworks (Backes et al., 2004; Byun and Li, 2008), the purpose is treated as a label attached to data which specifies their intended use. Access requests are evaluated not only against the identity of the access requester, the data to be accessed, and the action to be performed, but also against the *purpose* for which access is requested. If the purpose for which data are requested matches the intended purpose associated with the data, then the access is granted. For example, in healthcare a doctor can specify treatment or clinical research as a purpose when he requests the data, using the Cross-Enterprise Security and Privacy Authorization (XSPA) profile of XACML that defines a purpose attribute and a corresponding coded value set. However, this and the other proposals rely on the fact that the data requester (1) has specified the purpose correctly and legally and (2) will process data in accordance to the specified purpose.

Challenge 2 *How can we verify if the user behavior is compliant with data protection policies if some human activities cannot be observed?*

Humans have a key role in the business activities of an organization. Organizational procedures usually describe human activities next to IT activities. However, some of human tasks cannot be IT observable (e.g., a physician discussing patient data with a colleague over the phone for second opinion). The compliance with the purpose thus cannot be checked. Adopting mechanisms like video surveillance to monitor user behavior makes the system too intrusive and consequently can encounter social resistance.

Challenge 3 *How can we ensure policy compliance while dealing with unexpected situations?*

Policy enforcement is carried out by access/usage control (Park and Sandhu, 2004; Samarati and di Vimercati, 2001), digital rights management (DRM) (Rosenblatt et al., 2001), or trust management (Chapin et al., 2008). They share a preventive nature: they permit actions authorized in the policy and just stop the process if something unexpected happens (Hamlen et al., 2006; Ligatti et al., 2009). In many application domains like healthcare, users may take actions that deviate from the normal behavior to respond to the urgency. In such domains, exceptions are dealt with using the so called break-the-glass protocol. This protocol takes priority over the running enforcement mechanism allowing the user to perform the required action. This makes existing enforcement mechanisms inadequate for the enforcement of data protection policies. Auditing has been used to analyze the user behavior: the system logs users' actions and then the logs are analyzed manually to identify possible misconduct. However, analyzing every exception can be time consuming and costly. Some steps towards systems which allow a-posteriori policy compliance are already made (Cederquist et al., 2007). However, existing proposals do not verify if data are used for the intended purpose.

Challenge 4 *How can we ensure the compliance with data protection policies in modern distributed systems*

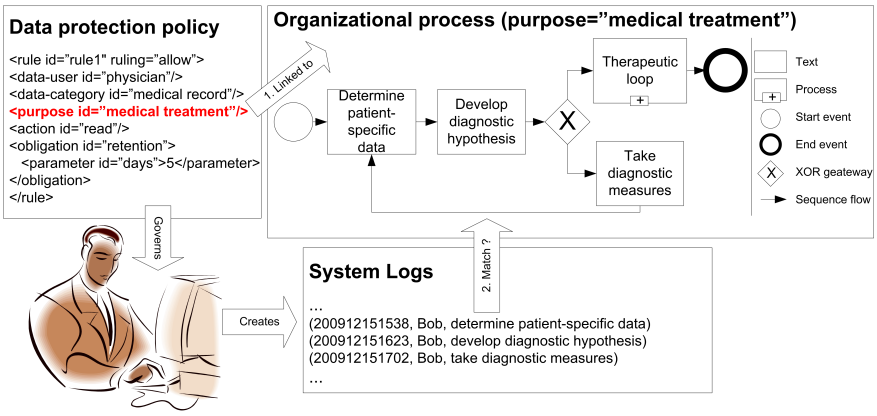

Figure 1: Linking organizational processes to data protection policies

and when the data processing is outsourced?

Once information has been outsourced, it is no more under the direct control of the organization. Existing compliance approaches applicable to distributed systems can be classified in three categories: (1) solutions that impose the responsibility of providing evidences to the data recipient, (2) solutions that allow the data controller to monitor data recipient's IT system, (3) solutions that limit possible executions to prevent policy infringements using mechanisms like DRM. None of these approaches is flexible enough to address the requirements of modern distributed systems mentioned in the introduction. The first approach relies on the assumption that the data recipient provided all the evidence necessary to prove his correct behavior. This implies a trust relationship between the data controller and the data recipient. The second approach is invasive and may not be accepted by the data recipient. The last approach is too restrictive to support exceptions.

3 Approach

We propose a method to verify the compliance of user behavior with data protection policies. The proposed approach uses and interlinks the following three components (Fig. 1):

1. *Data protection policies* define who can access the data and for which purpose. Policies can also specify the actions that should be taken once an authorization is granted (i.e., obligations).
2. *System logs* record the sequence of actions performed by users (i.e., the actual user behavior).
3. *Organizational processes* describe the business processes and procedures of an organization. Organizational processes represent the actions that users are expected to take in order to accomplish a certain goal.

The basic idea is to link the purpose specified in data protection policies to the goal of organizational processes (arrow 1 in Fig. 1). Intuitively, if a user requires access to some data for a certain purpose, we can determine whether the data are used for that purpose by verifying if the system log corresponds to a valid execution of the process associated with the purpose (arrow 2 in Fig. 1).

Organizations often describe their processes and procedures using graphical languages like BPMN. Unfortunately, those languages are informal and leave room for ambiguity (Dijkman et al., 2008). Moreover, they are not suitable for formal analysis. To address these issues, we can take advantage of methods that provide a translation of BPMN into formal frameworks for the analysis of business processes (Dijkman et al., 2008; Prandi et al., 2008). This allows the development of tools that automatically detect the infringement of data protection policies.

Besides providing effective and tool-supported methods to verify the compliance with data protection policies, defining the purpose using business process models has a number of advantages: (1) organizations can reuse the knowledge they already have without further efforts; and (2) it allows for the management of purpose and data protection policies.

The proposed approach alone may not be sufficient when we consider the human component in organizational processes. Procedures may contain human actions that cannot be logged by the IT system. Because of missing information about user behavior we cannot ensure that data have been processed wrt the intended purpose. Therefore, we have to extend our initial approach with a method that verifies if system logs correspond to a valid sequence of observable actions of the workflow. In addition, qualitative rea-

soning on business processes (Prandi et al., 2008) can be used to obtain evidence of possible misbehavior.

Many application domains like healthcare require dealing with exceptions. For instance, a physician can take actions that diverge from the procedures adopted by the hospital to face emergency situations. Preventing such actions may be critical for the life of the patient. To address this problem, we need to define acceptable infringements and metrics for assessing privacy risks by performing real-time risk analysis. This make it possible to narrow down the number of situations where rightful data usage is under investigation by considering only situations whose privacy risks cannot be tolerated by the data subject.

Outsourcing is becoming a common business practice of many organizations. This business strategy is adopted to reduce costs, but it involves transfer of business activities and data to an external supplier. Suppose that the hospital outsources some therapeutic activities to a subcontractor together with the data needed for its execution. The hospital is no more in control of such data. Therefore, it needs evidence that the subcontractor has used the information only for providing therapeutic treatments. To address those requirements for data protection policy compliance in distributed systems, we need an infrastructure that supports different security mechanisms as well as infringement management. The infrastructure should allow seamless interoperability of different enforcement mechanism, such as DRM, and on the other hand deterring security mechanism such as audit logic. This also requires real-time risk analysis models that make it possible to define at run-time flexible boundaries between preventive and deterring mechanisms (i.e., dynamically select the most proper mechanism for a certain situation based on the risks it involves).

4 Conclusion

This paper has defined the basis for the development of more trustworthy and privacy-aware IT systems. In particular, this work intends

- *support organizations in ensuring compliance with data protection policies* as the ability of identifying policy infringements will provide a concrete way to ensure organizations that they are meeting their privacy promises;
- *make users more accountable for their actions* as user behavior will be analyzed and possible misconducts detected;
- *be more flexible to deal with exceptions* as the a-posteriori policy compliance would make it possible to continue operations and account users for eventual misconduct afterwards;
- *provide more usable and scalable tools* which will automate many operations that are currently performed by humans.

REFERENCES

Backes, M., Karjoth, G., Bagga, W., and Schunter, M. (2004). Efficient comparison of enterprise privacy policies. In *Proc. of SAC'04*, pages 375–382. ACM.

Byun, J.-W. and Li, N. (2008). Purpose based access control for privacy protection in relational database systems. *VLDBJ*, 17(4):603–619.

Cederquist, J. G., Corin, R. J., Dekker, M. A. C., Etalle, S., den Hartog, J. I., and Lenzini, G. (2007). Audit-based compliance control. *International Journal of Information Security*, 6(2-3):133–151.

Chapin, P. C., Skalka, C., and Wang, X. S. (2008). Authorization in trust management: Features and foundations. *ACM Comput. Surv.*, 40(3):1–48.

Dijkman, R. M., Dumas, M., and Ouyang, C. (2008). Semantics and analysis of business process models in BPMN. *Information and Software Technology*, 50(12):1281–1294.

Guarda, P. and Zannone, N. (2009). Towards the Development of Privacy-Aware Systems. *Information and Software Technology*, 51(2):337–350.

Hamlen, K. W., Morrisett, G., and Schneider, F. B. (2006). Computability classes for enforcement mechanisms. *ACM Trans. Program. Lang. Syst.*, 28(1):175–205.

Hilty, M., Basin, D. A., and Pretschner, A. (2005). On Obligations. In *Proc. of ESORICS'05*, LNCS 3679, pages 98–117. Springer.

Karjoth, G., Schunter, M., and Waidner, M. (2002). Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data. In *Proc. of PET'02*, LNCS 2482, pages 69–84. Springer.

Ligatti, J., Bauer, L., and Walker, D. (2009). Run-time enforcement of nonsafety policies. *TISSEC*, 12(3):1–41.

Park, J. and Sandhu, R. (2004). The UCON_{ABC} usage control model. *TISSEC*, 7(1):128–174.

Prandi, D., Quaglia, P., and Zannone, N. (2008). Formal analysis of BPMN via a translation into COWS. In *Proc. of COORDINATION 2008*, LNCS 5052, pages 249–263. Springer.

Rosenblatt, W., Mooney, S., and Trippe, W. (2001). *Digital Rights Management: Business and Technology*. John Wiley & Sons, Inc., New York, NY, USA.

Samarati, P. and di Vimercati, S. D. C. (2001). Access Control: Policies, Models, and Mechanisms. In *FOSAD 2001/2002*, LNCS 2946, pages 137–196. Springer.