Automated Log Audits for Privacy Compliance Validation: A Literature Survey

Jenni Reuben, Leonardo Martucci, Simone Fischer-Hübner

Automated Log Audits for Privacy Compliance Validation: A Literature Survey

Jenni Reuben, Leonardo A. Martucci, and Simone Fischer-Hübner *

Karlstad University
651 88 Karlstad, Sweden
[firstname.lastname]@kau.se

Abstract. Log audits are the technical means to retrospectively reconstruct and analyze system activities for determining if the system is executed in accordance with the rules. This approach to compliance is referred to as compliance by detection. In the case of privacy adherence validation (or) privacy audits, the rules for compliance are less well defined and more contextual than in the case of traditional security audit. The aim of the paper is to understand the aims, techniques and challenges for realizing privacy compliance by detection. Using systematic literature review as the research tool we described the state-of-art privacy auditing approaches through taxonomies. We present two taxonomies, *i)* classified in terms of auditing techniques and *ii)* classified in terms of audit objectives. Following the observation gained from the state-of-the art we discuss challenges and suggest guidelines for utilizing log-based automated privacy audits.

Keywords: Log audit, privacy violation detection, privacy compliance, privacy audits, automation

1 Introduction

Compliance is a practice to ensure that the implemented practices and the execution of business processes of an organization is in accordance with regulations, legislations, industrial standards or agreed upon commercial contracts and policies. In particular, the practice to ensure the execution of business processes on personal data in accordance with enterprise policies thus data protection regulations is referred to as privacy compliance [17] or privacy audits.

Compliance is a perpetual and an essential practice for an organization, and it takes time, resources and cost [4]. As result of advancement in Information Technology (IT), the task of realizing compliance is increasingly automated [17, 20]. However, in an organizational setting automating the businesses processes

* This research was funded by A4Cloud and SmartSociety, two research projects of the Seventh Framework Programme for Research of the European Community under grant agreements no. 600854 and no. 317550.

to comply with regulatory requirements is not straight forward. A hierarchical approach is presented in [20] to facilitate compliance automation from the abstract regulatory constraints down to the concrete IT systems level. According to Sackmann et al. [20] there are two approaches to achieve compliance, one is a preventive approach and the other one is a detective approach. In the preventive approach the obligations (constraints derived from data protection laws in case of privacy compliance) are strictly enforced in the IT operations and thus non-compliant events are prevented from occurring. Whereas in the detective approach, activities of the system are monitored and validated against the compliance obligations. In this paper, we focus on the detective approach for achieving privacy compliance.

In the detective approach, compliance adherence is mainly validated through automated audits where the business process executions are recorded and retrospectively analyzed for policy conformance. This approach of realizing compliance demonstrates responsibility and thus accountability, provides flexibility in process executions, helps to validate future obligations, and deters policy violations. The key idea behind the automated audits, which automatically verifies the events that are registered in the logs against specifications is thus similar to intrusion detection methods. However in the context of privacy compliance, privacy principles and concepts, unlike security functions, are less well defined [10]. Hence, intrigued by the idea of privacy compliance verification we aim to synthesize the state-of-the-art auditing techniques that are used to realize privacy compliance by detection.

Research in automatic verification of privacy compliance has been ad hoc and diverse, ranging from well-evaluated proof-of-concepts to theoretical research ideas that propose approaches for validating the data handling practices of an organization. To the best of our knowledge, no scientific review has been done that describes the state-of-the-art in automated audits for privacy compliance validation. We reviewed existing streams of work in this field using the systematic literature review for summarizing the state-of-the-art. Major contributions of this paper are as follows:

- We systematically synthesize the existing technical knowledge of this research domain to provide a consolidated view of the state-of-the-art and explain its conceptual relationships. We present the consolidated view as a taxonomy, which is based on the auditing techniques. (Section 3).
- We qualitatively analyze the motivations for log-based automated privacy audits and derive a second taxonomy. which is based on privacy audit objectives (Section 4).
- Based upon the insights gained from the above two taxonomies, we conceptualize a set of initial guidelines for utilizing log-based automated privacy audits (Section 5).

2 Method

To provide an in-depth overview of the state-of-the-art, we used a systematic literature review process [18]. This process provides methodological rigor in terms of planning, conducting and reporting the literature review.

As the initial step, a research protocol was designed following [18]. The research protocol specifies the objectives of the review, the search strategy to identify relevant literature, and a strategy to classify the primary studies (the identified sources of this review) for evaluation. The following subsections describes each steps in detail.

2.1 Identification of Objects

A thorough and unbiased search for relevant literature is the essence of a systematic literature review. For the scientific literature sources, we used our university library's one-search service - EBSCO discovery service[1] mainly to ensure that disciplines other than information technology were not overlooked. Prior to the search, keywords, free texts and their synonyms were identified from the following objectives.

- What are the technical approaches that have been proposed to validate privacy compliance or to detect privacy non-compliance?
- What are the motivations for performing privacy compliance verification through log audits?
- Could privacy concerns be detected from log audits, if so what are the privacy concerns that can be validated?
- What are the limitations of the current approaches?

Table 1 lists the search strings used for this study. The identified search strings were combined appropriately during the search to locate as many relevant articles as possible. The search process was stopped when the different combination of the search strings were not returning any new results.

The search was performed during week 51 in 2014 and yielded 1723 results in total to be evaluated for relevance.

2.2 Selection of Primary Studies

The articles returned by the search were examined for relevance from the title, abstract, and subject headings following these exclusion criteria

[1] EBSCO discovery service is a gateway to access various scholarly databases (with and without full text) in many different disciplines and it also includes electronic resources like e-magazines, e-books and etc. The complete list of resources indexed by the discovery service can be found at [1]. The updated version of the list is not available yet, hence recently added database such as ACM is missing in the out dated list.

Table 1. Search strings

Phrases (Including free texts)	Synonyms
privacy concerns	
privacy analysis	privacy audit
privacy compliance	privacy enforcement
privacy non compliance	privacy violation
access logs	
audit trails	

– Exclude if the focus of the proposed approach is clearly not a privacy-compliance by detection.
– Exclude if the focus of the paper is not technical but relates to constitutions, legislation and regulations.

As a result a total of 25 articles were identified and imported to EndNote[2] - a bibliography management tool. Next, we performed a backward (i.e searching the citations of the identified articles) and a forward search (i.e locating the papers that cites the identified articles) on the identified articles to ensure fullest coverage of related work. All the relevant literature resulted from the forward and backward search was further narrowed following a second set of inclusion and exclusion criteria, which are listed below:

– Include the most completed and related version, if there exist several related studies by the same authors or by the same research group.
– Exclude short papers where the technical description of the solutions were not described.
– Exclude if the privacy-compliance auditing was *discussed* in parts of the proposed prototype but clearly was not the main focus, because in most of these cases the verification part is not automated.
– Exclude if the privacy-compliance verification is performed by other means than the log audits.

This step resulted in a reduction in the number of the primary studies. The total number of relevant literature identified and reviewed for this work are 14, full texts of the 14 articles were located and imported to EndNote for further processing.

2.3 Quality Assessment, Classification and Synthesis

As the next step in the process, all the primary studies were carefully analyzed. The aim of the study is to bring together existing but divergent research for

[2] `http://endnote.com/`, accessed March 11, 2016.

two purposes; 1) to give a snapshot of the field revealing the design foundations, which will help future researchers in this field to position their research and 2) to analyze the privacy requirements that are accounted for in the state-of-the-art thus indicating the uncovered privacy requirements. Classification systems and the resulting classification are proven to be an effective strategy for deconstructing the foundations of a domain [19]. The taxonomy[3] focuses on the characteristics of this particular research area and its conceptual relationships rather on the evolutionary relationships. Therefore a broad set of characteristics are extracted from the primary studies to inductively analyze the conceptual relationships. This follows though not rigorously[4] the taxonomy development framework proposed by Nickerson et al. [19], which is based on a three-level indicator model suggested by Bailey et al. [5]. We used particularly the empirical-to-conceptual approach, where the characteristics of each object under study are observed and analyzed for similarities in order to form categories.

After identifying the purpose of the taxonomy, which is to provide a scientific understanding of the state-of-the-art, the meta-characteristics [19] were specified to drive the development of a taxonomy. The meta-characteristic identified for the taxonomy that is presented in Section 3 is technical design. Hence the characteristics to be extracted from each primary study are algorithms, inputs and outputs of the algorithms, their performance, scalability and technical limitations. In addition, design principles such as the aim of the solution, system settings, users of the system were extracted, which are helpful to deduce the nature of each category. These extracted characteristics are later qualitatively analyzed for the taxonomy presented in Section 4.

3 A Taxonomy of Audit Techniques

Following the approach described in the Subsection 2.3, seven to eight different characteristics are extracted and observed from each primary study. Some of the characteristics are not clearly determined from the primary studies, hence related abstract notions are identified, simplified and analyzed. In order to develop a taxonomy of technical approaches to privacy-compliance verification, the main criterion that drives the taxonomy development is identified as detection (verification) algorithm, given the events and the privacy specifications. As mentioned earlier, checking the system activities registered in the logs for detecting non compliance is thus similar to Intrusion Detection Systems (IDS), which is evident from the detection algorithms used in the primary studies. The technical architecture of these detection algorithms are analyzed to derive functional dimensions. Within the 14 technical designs, 6 of the solutions specify patterns for detection prior to the verification phase, where five in six solutions define

[3] In this paper the words classification and taxonomy are interchangeably used to mean the result of the classification process.

[4] For example, the cycles of our taxonomy revision process was intuitive, there was no ending conditions such as objective and subjective endings were specified to terminate the iteration.

patterns for compliance and one solution defines patterns for non-compliance. The remaining 8 solutions train a classifier to learn the compliant state, which is further used to detect outliers. Evidently, this follows the two primary types of IDS techniques, namely misuse-based IDS and anomaly-based IDS. Further sub dimensions of the taxonomy are derived based on the verifying characteristics of the detection algorithms. Figure 1 depicts the resulting classification, which consists of 2 dimensions of the technical designs observed from the existing privacy-compliance by detection solutions.

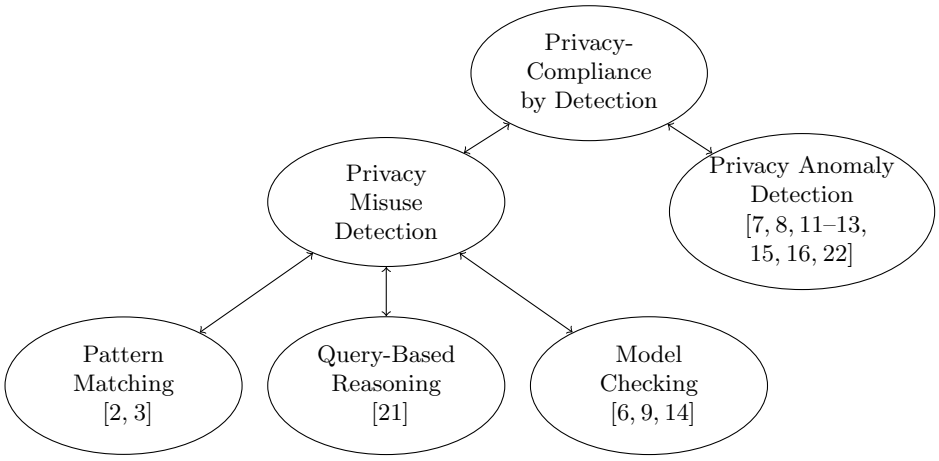

Fig. 1. Technical taxonomy of privacy-compliance by detection solutions.

3.1 Privacy Misuse Detection

As mentioned earlier, the common characteristic of the technical architecture in this category of solutions is, the privacy misuse patterns (either compliant or non-compliant) that are used for detection are expressed explicitly. Generally these patterns are known in advance, and are looked up in the logs for a mismatch in case of compliant patterns or a match in case of non-compliant patterns to detect violations. Three different checking approaches are observed in the technical design of the primary studies within this category, namely, pattern matching, query-based reasoning, and model checking.

Pattern matching. The system settings of these detection algorithms are integrated IT systems offering web-enabled IT services, self-service identity management feature of an application, etc. Individuals release data items while interacting with these systems. The inherent characteristic of these systems is that individuals can specify privacy preferences that define conditions and

obligations for their data usage. Activities of the system, i.e every action on the data items, are recorded in a log file (audit trails) for later review. During the review (audit) the actions recorded in the logs are matched against the applicable rules in the privacy preferences.

There are two primary studies which fall under the pattern matching approach, Accorsi [2] and Accorsi et al. [3]. Accorsi [2] proposes a matching approach based on counterexamples. Prior to the verification step, potential violations are derived from the privacy preferences of individuals. These patterns are later used to find a match in the log view that corresponds to an individual. A violation is detected if a match is identified.

Accorsi et al. [3], propose a novel approach for pattern matching to improve the efficiency of the log audits. In the first step of the audit process, individualized audit trails are transformed into an action tree which is a hierarchical representation of activities based on the structure of the data items. In the second step, the algorithm sequentially processes the rules in the privacy preferences of individuals and searches for a node in the tree that corresponds to the data item referred to in the rule. When such a node is determined, the actions recorded in the node are checked for agreement with the applicable rule in the privacy preference. This process is iterated and at the end of each iteration, matched actions are pruned from the action tree. At the end of the audit, the nodes that remain constitute the sum of violations pertinent to the privacy preferences.

Query-based Reasoning. The system setting of this solution is very similar to the setting mentioned above, where the individuals can formulate privacy preferences for their data usage. Privacy related events are logged when subjects (processes, softwares, participants) of the system act upon the managed data items. Samavi et al. [21] propose a query based reasoning approach to derive obligations applicable to an access request and to verify the fulfillment of these obligations. They define two ontologies in their framework, namely L2TAP (Linked Data Log to Transparency, Accountability and Privacy) and SCIP (Simple Contextual Integrity Privacy). L2TAP ontology allows logging of events such as changes in the privacy preferences, and information related to access requests and access activities. This ontology provides provenance assertions (when and by whom), which are associated with the activity that triggers the change. SCIP ontology provides semantics for the events logged in L2TAP logs, and uses SPARL queries with RDFS reasoning support for auditing. The first step in the auditing process is to use SPARL queries to identify all the obligations linked to an individual's privacy preferences that match the access requests. In the next step the status (fulfilled, pending, violated) of the linked obligations are identified using SPARL queries to determine the compliance of the access request. In case of multiple obligations that are linked to an access request a series of SPARL queries are used to evaluate the combined effect of these obligations in order to determine the overall compliance of the access request.

Model checking. The system setting of this approach is enterprise computing solutions and the design objective is to validate compliance of the system in accordance with privacy regulations, or enterprise policies, or data protection directives. These specifications are expressed in logical forms, and the actual usage of personal data recorded in the logs are also represented as models. Given this, a model-checking algorithm verifies whether a given formula is satisfied by the model. The auditing algorithms of three primary studies use the model-checking approach. Garg et al. [14] formalize all the 84 disclosure-related clauses of the Health Insurance Portability and Accountability Act (HIPAA)[5] Privacy Rule using first-order logic. Additionally, in their framework the audit trails are expressed as partial structures. During the audit, the auditing algorithm iteratively checks for a violation, given the policy and the audit trails. It is iterative because sometimes information contained in the logs is insufficient for determining whether a policy is satisfied or not (for e.g in case of future obligations), hence the algorithm runs iteratively when more information is available. Banescu et al. in [6] formalize the business processes of an enterprise. During the audit, the modeled system events which are generated during the execution of the business process are replayed over the business process model to identify violations, which are later quantified. The events are classified during this process to identify the type of violation in regard to the business process model. Butin et al. [9] adopt a similar approach; they express parts of the EU Data Protection Directive [6]and the audit trails in logical form. For the verification of the state of the logs, its compliance properties are derived and defined from the formalized EU Data Protection Directive. A log trace (audit trails) is compliant if it satisfies all its compliance properties. The audit framework proposed by Butin et al. operates on "privacy friendly" logs and they mathematically proves the correctness of the log analysis that verifies the compliance to the data protection obligations.

3.2 Privacy Anomaly Detection

The common characteristic of the technical architecture of the solutions presented in this category is that the detection is anomaly-based. The system setting of these solutions are similar to system setting of privacy misuse detection solutions, but the design objective is to detect unexpected attempts. The anomaly-based detection systems identify events that deviate from an abnormal profile in the case of supervised method, or from the other items in a dataset in the case of unsupervised method. A combination of statistical, machine learning and data mining techniques are observed in eight primary studies irrespective of the anomaly detection method. Venter et al. [22], Bhattacharya et al. [7], Boxwala et al. [8], Gupta [15] and Heatley et al. [16] use the supervised anomaly detection method. The training set required to train the classifier is built manually for two

[5] `http://www.hhs.gov/ocr/privacy/`, accessed on March 11, 2016

[6] `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML`, accessed on March 11, 2016

of the solutions but Bhattacharya et al. [7] use an association rule data mining technique to generate patterns of violations for the training set. Similarly Heatley et al. [16] and Gupta [15] use statistical methods like decision tree induction technique and Latent Dirichlet Allocation (LDA) respectively to build the training set, which contains the "normal" and the "abnormal" behavior. Given the model (training data set), the classifier using statistical techniques such as k-nearest neighbor, Logistic Regression (LR) and Support Vector Machine (SVM) predicts the probability for the access request to be an outlier.

However, Chen et al. [11] present an unsupervised anomaly detection method for system settings where building a clearly labeled training set is difficult. In their approach, the first step is to aggregate the features in the audit trails into community structures using a combination of statistical methods such as graph-based modeling and dimensionality detection. In the second step the deviation probability of the elements from its closest neighbor is determined using the k-nearest neighbor algorithm. Fabbri et al. [13] describe an approach similar to that of Chen et al. for forming aggregated social structures but use the social graph to explain clearly the context of the events in an audit trail. The basis for this is that these explanations are later reviewed by the privacy officers to detect violations. Similarly Duffy [12] defines a framework that combines anomaly detection algorithms such as the one proposed by Chen et al. and two other algorithms (Patient-Flow Anomaly Detection Systems (PFADS) and Specialized Network Anomaly Detection (SNAD)) to provide contextual explanations thus allowing patients and privacy officers to have better understanding of the access logs.

3.3 Implications of Intrusion Detection Approach to Privacy Audits

As illustrated and depicted in Figure 1, the approaches to validate privacy compliance follows intrusion detection techniques and hence inherit the underlying limitations yet bring intrinsic challenges. For instance, misuse detection algorithms match specific patterns from the audit trails for detecting misuse, under the assumption that all the patterns for validation are known in advance. However relying on well-known patterns degrade the purpose of the validation algorithms. Moreover, the patterns for privacy invasion are difficult to define primarily because privacy invasive incidents are mostly context dependent.

Furthermore, the expressiveness of the compliance requirements as patterns is constrained by the limitations of the formal languages, for instance, purpose(s) need to be explicitly expressed for each data item. In addition, human intervention is generally required for deciding policy violations, and for providing additional semantics besides the information registered in the logs.

Anomaly detection algorithms on the other hand are not concerned with known notable events but with the abnormal system or user behavior. However, in order to detect outliers (abnormality), the classifier in the case of supervised learning method requires a fair amount of training data set to model what is normal for each subject (could be a user, a process, and a software) in a system in order to decide on how probable the activity of a given subject is abnormal. In

the case of unsupervised learning though, it is important to model appropriate clusters for detecting outliers that is abnormal from rest of the members in a cluster. Nevertheless, similar to well-known patterns the signals of privacy invasions are challenging to discern because under a slightly different context a system or user behavior would have been legitimate. This may result in many false positives or many false negatives offsetting the purpose of the automation process.

Furthermore, in the literature review we observed that there are more privacy anomaly detection solutions than the privacy misuse detection solutions. This is because, in the case of anomaly-based detection the concept of normality is difficult, if not impossible to transfer from one system setting to another system setting. Although all the solutions in this category target health-care systems, a slightly different condition such as collaborative health-care systems require different training techniques or clustering for modeling expected behaviors.

4 A Taxonomy of Audit Objectives

In Section 3 we presented a taxonomy based on the auditing techniques that validate adherence to the privacy compliance requirements. We observe that the utilization of different auditing techniques are highly influenced by the audit objectives. Hence, in addition to the taxonomy of the technical approaches (see Figure 1), in this section we attempt to conceptualize the state-of-the-art in terms of compliance objectives. Whereby, we provide an analysis of the motivational trends for privacy compliance. There are three compliance objectives deduced from the state-of-the-art for engaging in a log-based automated privacy audits, they are;

- Audit for ex-post obligations
- Audit for permitted exceptions
- Audit for access legitimacy

Audit for ex-post obligations Privacy compliance requirements, which cannot be checked by a policy engine beforehand but are determined after the fact, are realized using automated log audits. Temporal Obligations such as "Delete the stored personal attributes after 2 years" and perpetual obligations such as "Send notice to the end-users every time there is a change in the enterprise policy" indicate after-the-fact compliance actions. To automatically realize these requirements, data handling processes are persistently monitored in the form of audit trails for subsequent actions. The granularity level of the audit trails greatly impact the achievable degree of automation of these requirements.

Audit for permitted exceptions In the health care applications, IT processes such as access control are often lenient to allow for emergency. Herein, patient care and safety outweigh the permission restrictions. However, unauthorized

access to patient information arise from the misuse of "break glass policies" may jeopardize the personal integrity of the patients. Therefore, these access exceptions must be supervised regularly in order to handle the risk of misuse. Audit for permitted exceptions helps to both detect and deter violation of these exceptions.

Audit for access legitimacy In service based applications, end-users are allowed to set access preferences to certain data objects that are under the control of the service providers or information fiduciaries such as identity management providers. However, regulation such as the proposed EU General Data Protection Regulation (GDPR)[7] and agreed upon end-user agreements mandate service providers' accountability, i.e. require the service providers to demonstrate the exercised responsibilities instead of just promising that the end-user preferences will be honored. To automate this requirement, related actions of the system are recorded in the logs, compared and verified using the log audits for compliance to the conditions arising from the end-users' privacy access preferences. Further, the log audits that validate compliance to laws and regulations also include access and disclosure compliance. Therefore, the solutions that provide mechanisms for validating compliance in accordance with the laws, and regulations are also included in this category.

Figure 2 shows the categorizations of the primary studies in accordance with the audit objectives listed above.

Taxonomy of audit objectives point out how well each objectives are covered (or not covered yet) by the state-of-the-art. Further, Figure 3 illustrates the mapping of the privacy audit techniques to the privacy audit objectives.

Pattern matching, and query-based reasoning algorithms provide solutions to determine if the permissions set by the end-users are indeed respected. From Section 3.1, we note that the model checking algorithms present auditing solutions to model the observable state of a system and validate if the system is compliant as expected. Hence, model checking solutions also address accountability requirement for access and disclosure activities.

Further, from our literature review, we observe that privacy anomaly detection mechanisms provide auditing solutions for supervising the risk integrated in the enforceable policies (for e.g "break glass"). It can be noted that anomaly detection algorithms could as well suitable for access legitimacy audit. However, as mentioned earlier the inherent limitations of anomaly detection algorithms such as high number of false positives and impractical conditions for transferring the state of normality makes the anomaly detection algorithms less reliable. Nonetheless, the practical usefulness of anomaly detection for other privacy audit objectives may be a new potential research direction to investigate further.

[7] `http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf`, accessed on March 11, 2016

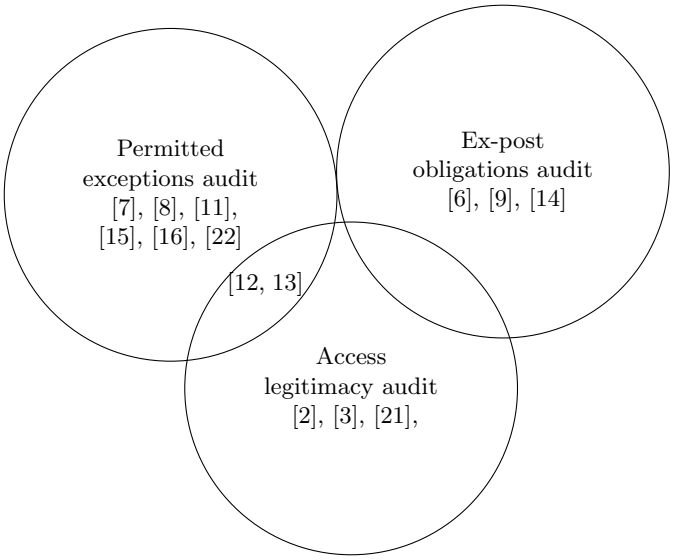

Fig. 2. Three audits objectives in the context of privacy

5 Guidelines

Compliance by detection approaches are long established IT processes in an organizational setting. However, in the context of privacy compliance requirements there is no one size fit all solution. In this section, we tentatively provide initial guidelines based upon our analysis of the state-of-the-art for aiding implementation of responsible log-based automated privacy audits.

- Any organization providing service to consumers usually processes information or data. An assessment procedure to classify the type of data is an essential point of departure. The classes of data such as private, public, sensitive or non-sensitive are very much context-dependent, i.e. the context of the application domain and various features of the service are need to be considered.
- Depending on the classification of the data and the application domain, applicable regulations, standards, and laws must be reviewed in order to derive the privacy compliance requirements.
- Classify the privacy-compliance by detection requirements into afore mentioned audit objectives. It is important to note that the list of objectives presented in Section 4 are the privacy audit directions inferred from the state-of-the-art. There could be different objectives for employing privacy audits depending on the class of the data, classification domain, and jurisdiction.
- Based on the identified privacy audit objectives, employ the respective algorithms as depicted in figure 3.

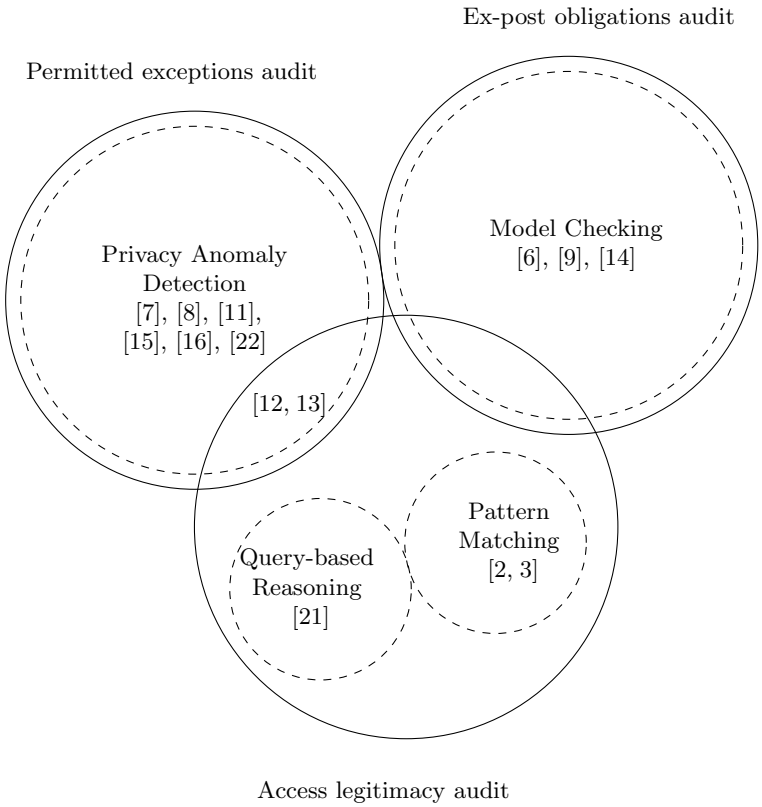

Fig. 3. Privacy audit objectives and respective techniques

6 Conclusions

We systematically reviewed and classified the existing privacy compliance validation solutions in relation to the auditing techniques and the audit objectives. In this paper we attempt to deconstruct the theoretical foundations of the state-of-the-art in order to explain the conceptual relationships of the existing solutions. Nonetheless, we recognize that the presented taxonomies are itself artifacts hence subject to evaluation. We attempt to conceptualize the presented taxonomies for mutually exclusivity and explanatory, to ensure they are extendible when more objects show up in the future.

Privacy compliance validation plays a very important role for success of any business. Despite the recognition of the importance of compliant to privacy requirements, the state-of-the-art has not matured proportionately with the growing privacy expectations. Furthermore, the feasibility of automating the privacy expectations in order to demonstrate compliance is practically restricted by two factors. First, the degree of expressiveness of regulations, laws, standards, and user-laid preferences in a machine-readable form. Second, the level of granu-

larity of the information logged in the audit trails, and the impractical conditions to automatically log certain system activities such as "informed consent". Hence, as a next step, in our future work, we are interested in studying a privacy-aware log design for privacy compliance validation.

Acknowledgments. This work benefits from the invaluable comments, inputs from Rose-Mharie Åhlfeldt and from the anonymous reviewers.

References

1. Databases in onesearch, `http://www.kau.se/sites/default/files/Dokument/subpage/2012/06/databaser_i_onesearch_pdf_18607.pdf`
2. Accorsi, R.: Automated Privacy Audits to Complement the Notion of Control for Identity Management, vol. 261, pp. 39–48. Springer (2008)
3. Accorsi, R., Stocker, T.: Automated Privacy Audits Based on Pruning of Log Data. In: 12^{th} Enterprise Distributed Object Computing Conf. Workshops. IEEE (2008)
4. Bace, J., Rozwell, C.: Understanding the Components of Compliance. Research G00137902, Gartner (2006)
5. Bailey, K.: A Three-Level Measurement Model. Quality and Quantity 18(3), 225–245 (1984)
6. Banescu, S., Petkovi, M., Zannone, N.: Measuring Privacy Compliance Using Fitness Metrics, LNCS, vol. 7481, pp. 114–119. Springer Berlin Heidelberg (2012)
7. Bhattacharya, J., Dass, R., Kapoor, V., Chakraborti, D., Gupta, S.: Privdam: Privacy Violation Detection and Monitoring Using Data Mining (2005)
8. Boxwala, A., Kim, J., Grillo, J., Ohno-Machado, L.: Using statistical and machine learning to help institutions detect suspicious access to electronic health records. Journal of the American Medical Informatics Association 18(4), 498–505 (2011)
9. Butin, D., Le Métayer, D.: Log Analysis for Data Protection Accountability (Extended Version). Research report no. 8432, INRIA (2013)
10. Cavoukian, A.: The Security-privacy Paradox: Issues, Misconceptions and Strategies. Tech. rep., Information Privacy Commission/Ontario and Deloitte & Touche (2003)
11. Chen, Y., Malin, B.: Detection of Anomalous Insiders in Collaborative Environments via Relational Analysis of Access Logs. In: Proc. of the 1^{st} ACM Conf. on Data and Application Security and Privacy (CODASPY). pp. 63–74. ACM (21–23 Feb 2011)
12. Duffy, E.: Facilitating Patient and Administrator Analyses of Electronic Health Record Accesses. Master thesis, University of Illinois at Urbana-Champaign (2013)
13. Fabbri, D., LeFevre, K.: Explanation-based Auditing. Proc. of the VLDB Endowment 5(1), 1–12 (2011)
14. Garg, D., Jia, L., Datta, A.: Policy Auditing Over Incomplete Logs: Theory, Implementation and Applications. In: Proc. of the 18^{th} ACM Conf. on Computer and Communications Security (CCS). pp. 151–162. ACM (2011)
15. Gupta, S.: Modeling and Detecting Anomalous Topic Access in EMR Audit logs. Master thesis, University of Illinois at Urbana-Champaign (2013)
16. Heatley, S., Otto, J.: Data Mining Computer Audit Logs to Detect Computer Misuse. Int. J. of Intelligent Systems in Accounting, Finance & Management 7(3) (1998)

17. Kahmer, M., Gilliot, M., Muller, G.: Automating privacy compliance with expdt. In: E-Commerce Technology and the 5th IEEE Conference on Enterprise Computing, E-Commerce and E-Services, 10th IEEE International Conference on E-Commerce Technology. pp. 87–94 (July 2008)
18. Kitchenham, B.: Procedures for performing systematic reviews. Joint technical report 0400011t.1, Keele University and Empirical Software Engineering National ICT Australia Ltd (2004)
19. Nickerson, R., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. Eur J Inf Syst 22(3), 336–359 (2013)
20. Sackmann, S., Kähmer, M., Gilliot, M., Lowis, L.: A classification model for automating compliance. In: 10th IEEE International Conference on E-Commerce Technology (CEC 2008) / 5th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (EEE 2008). pp. 79–86 (2008)
21. Samavi, R., Consens, M.P.: L2TAP+SCIP: an Audit-based Privacy Framework Leveraging Linked Data. In: 8th Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing. pp. 719–726. ICST / IEEE (2012)
22. Venter, H.S., Olivier, M.S., Eloff, J.H.: Pids: A Privacy Intrusion Detection System. Internet Research 14(5), 360–365 (2004)