

Privacy Compliant Internal Fraud Screening

Ulrich Flegel¹

Offenburg University of Applied Sciences
Department of Media and Information Engineering and Design
Badstr. 24, 77652 Offenburg, Germany
ulrich.flegel@fh-offenburg.de

Abstract

In the year 2009 several data privacy scandals have hit the headlines where major corporations had a legitimate need for detecting fraud conducted by their own employees, but chose inappropriate measures for data screening. This contribution presents architectures and pseudonymization technology for privacy compliant fraud screening or fraud detection, in order to reduce the number of undiscovered fraud cases and to reduce the time to discovery.

1 Introduction

During the first quarter of the year 2009 two major German corporations hit the headlines with large scale privacy law violations. Telekom conducted fraud screening by comparing employee account numbers and supplier account numbers to discover shell companies [Wel09]. Deutsche Bahn also compared employee telephone numbers and addresses with those of its suppliers [Wel09]. Shortly after, Airbus had to admit having conducted similar screenings [Reu09]. As a reaction to these scandals the German privacy law has been concretized, clarifying the scope of legal fraud detection measures. The scandals and the amendment of privacy law have created a demand for technology allowing for effective fraud screening in compliance with pertinent privacy law. This article presents architectures and pseudonymization technology allowing for automated fraud screening on pseudonymized data in compliance with privacy law in order to reduce the number of undiscovered fraud cases and to reduce the time to discovery.

Section 2 sets the scene by introducing a process that is prone to fraud and highlights fraud potential. Section 3 describes the state of the art process for detecting fraud in ERP systems and this fraud detection process is briefly assessed w.r.t. privacy law in Section 4, while Section 5 explicates the reconciliation of conflicting interests during this process. In Section 6 we propose an architecture for fraud screening on pseudonymized data, state the requirements for pseudonymization in this architecture and introduce a technical approach meeting the requirements. After assessing the improvement over the state of the art we conclude in Section 7.

2 Example Scenario

Private sector enterprises manage their business processes using *Enterprise Resource Planning* (ERP) systems. Business processes are implemented in an ERP system while accounting for inter-

¹ During authoring this contribution employed with SAP AG, Research Center Karlsruhe, Germany.

nal controls, such as the principle of *Segregation of Duties* (SoD). Internal controls aim at avoiding occupational fraud done by employees of the company. Major pain points for fraud are the purchasing and sales departments in a company. Therefore we use a purchasing business process as a running example. We restrict our description to the details that are necessary for understanding the remainder of this text.

2.1 Example Purchasing Process

When an employee needs to purchase an article for his business activity, she creates a respective *purchase requisition* (PR) in the ERP system (see Figure 1). Before approval by the employee's manager the purchase requisition is *blocked*, approval changes it into a *posted* PR. According to the posted PR the purchasing department determines a suitable supplier and creates a *purchase order* (PO) from the posted PR. Suppliers or vendors are entered, i.e. created, by a manager of the purchasing department into the ERP system.

Upon reception of the ordered article the employee of the company creates a respective *goods receipt* in the ERP system, such that the associated invoice may be approved. When an invoicing clerk of the company receives the aforementioned invoice from the supplier, he creates an appropriate *blocked invoice* document in the ERP system. If the invoiced amount exceeds a specified threshold a manager of the purchasing department checks the invoice against the PO and the goods receipt and approves the invoice, which is then converted into a *posted invoice*. Posted invoices are paid during the next *payment run*, which periodically occurs in an automated fashion.

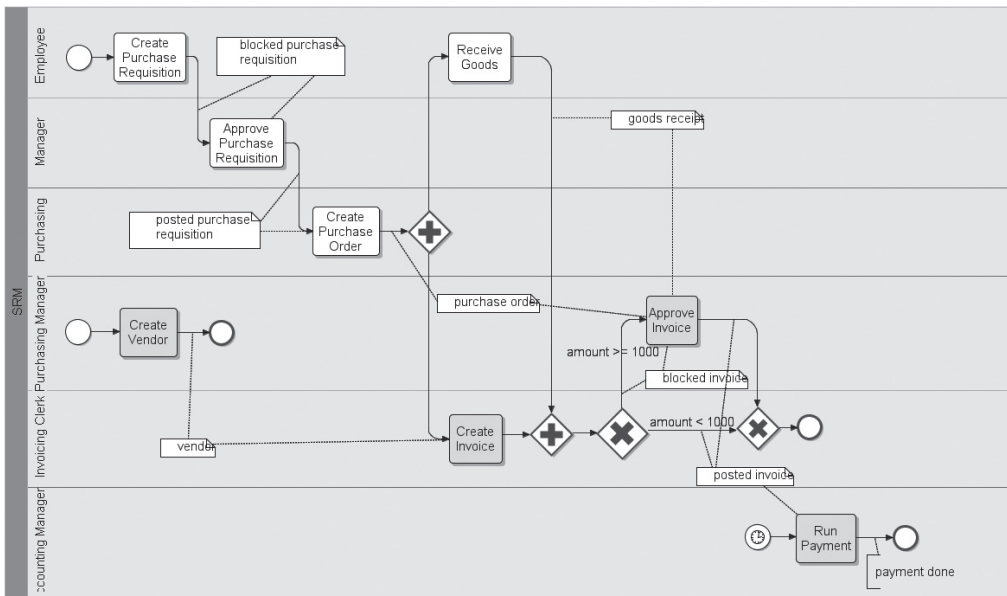


Figure 1: Simplified Purchasing Business Process in BPMN notation [GDW09]

This example process demonstrates the segregation of duties between different departments and roles for effective enforcement of internal controls. The involved employees use the same ERP

system during runtime (see Figure 2), which is hosted by the (internal) IT service provider of the company.

2.2 Example Fraud Scenarios

The internal controls arranged for in the business process implemented in the ERP impede possible fraud, but they cannot avoid fraud altogether. This results from the activity margin necessary to complete business process instances during daily business, and from the incompleteness of the process implementation. For example the business activity of external actors, i.e. suppliers, are usually only partially covered by the local ERP system of the purchasing company. Hence, in order to detect collusion between purchasing agents and suppliers there exist some indirect evidence, but the actual agreements are not documented in the ERP system. Examples for such indirect evidence are buyer-supplier pairs with higher than average early or advance payment of invoices.

3 Fraud Detection in ERP Systems

Evidence for fraud may be collected by inspecting the business transactions performed in an ERP system. ERP systems usually exhibit at least one *audit component* interface (see Figure 2) that allows for case-based extraction of business data from the ERP databank management system(s) (DBMS) for auditing purposes. This data necessarily is collected or generated during the course of the normal business activity within the ERP system. The audit component merely filters this data and exports it for the purpose of detecting fraud. Figure 2 depicts the state of the art architecture for detecting fraud in internal ERP systems.²

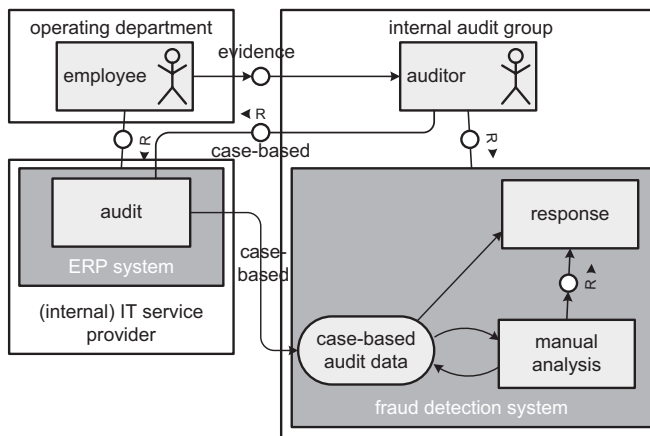


Figure 2: Conventional Realization of manual fraud detection

The fraud detection process within large enterprises works (with some variations) as described in the following. A member of the company’s internal audit group receives some confidential clue or evidence from some operating department employee (“whistleblower”), which justifies case-based extraction and analysis of audit data from the ERP system. The case-based analysis of

² In this article IT systems are modeled using *Fundamental Modeling Concepts* (FMC) [KGT06].

the extracted audit data aims at refuting or corroborating the suspicion implied by the initial evidence and is performed by internal auditors today mainly manually with some basic tool support [FVB10]. In addition to the extracted data the auditors leverage system-external information, e.g. from employee interrogation. In the case of a substantiated fraud a response is initiated, possibly comprising sanctions and criminal investigation. The actual response is out of the scope of this article.

3.1 Example Audit Data

The audit data extracted based on a specific case allows for comprehending the related business activity in its actual temporal order, including information on relevant documents, their owners and approvers. Documents relevant in the example purchasing process are *Purchase Requisitions*, *Purchase Requisition Approvals*, *Purchase Orders*, *Purchase Order Confirmations*, *Financial Documents*, *Invoices*, *Vendor Master Data*. The document field or attribute data extracted in the audit data is illustrated using purchase requisitions (PR) and the corresponding approvals as an example (fields/attributes with sensitive data w. r. t. privacy or business are underlined):

Purchase Requisition: PR Number (id), PR Item (id), Material (id), Short text (string), Requisition Tracking number (id), Quantity (integer), Unit (integer), Delivery Date (date), Material Group (id), Plant (id), Store Location (id), Pgr Requisnr. (user id), Fixed Vendor (id), Total Value (real), Currency (string), Unloading point (string), CoCode (id), Recipient (id), Cost Center (id), Profit Center (id), Name (string), Street (string), House number (string), Postal Code (integer), City (string), Country (string)

Purchase Requisition Approvals: PR Number (id), WF Instance (id), Approval Date 1(-7) (date), Approval Time 1(-7) (time), Approval Action 1(-7) (string), Approver 1(-7) (string)

4 Legal Assessment

A legal assessment of the implications of a given technical system is usually only possible for a given and specific legal system and law or regulation. A more general assessment across different legal systems or national laws is necessarily fuzzy and more geared towards general principles, such as Bizer's Seven Golden Rules for Data Protection [Biz07]. A comprehensive, more general assessment for international privacy law compliant fraud detection is bound to appear [FKM+10]. A detailed, legal assessment of fraud detection within the context of German privacy law has been published by Flegel, Raabe and Wacker [FRW09]. In the following that legal assumption is omitted and the results are summarized. While the results are based on German law, the general mind set should be transferable at least to national privacy law in European Union (EU) member countries, where directive 95/46/EC [95/95] harmonizes member country national privacy law.

ERP systems lawfully collect and store data for the purpose of executing the regular business processes of a company. Privacy law is involved when audit data is extracted by the ERP audit component and sent to the internal audit group for fraud detection. Privacy law already covers the case of investigating on a case-by-case basis for given clues for a criminal offense. However, this exemption does not cover a continuous extraction of audit data and screening for fraud evidence. The rationale of the lawmaker is to be understood in analogy to pertinent law about video sur-

veillance, where continuous surveillance for the purpose of exposing clues for criminal offenses would affect the employees of a company. As such, privacy law protects employees from a technical measure that continuously extracts clear text audit data from ERP systems for fraud screening.

5 Organizational Reconciliation of Conflicting Interests

Since fraud screening on continuously extracted audit data is prohibited by privacy law (cf. Section 4), a legal fraud detection facility needs to avoid audit data retention and rather focus on cases for which evidence justifies the suspicion that fraudulent activity is taking place. Hence, some piece of evidence is necessary and needs to be examined by the internal audit group, the works council and the responsible data protection officer in order to assess, if the given evidence justifies a fraud suspicion (refer to Section 3 for a description of the fraud detection process). Such evidence usually stems from whistle blowing or regular financial audits. As a result, the number of undiscovered fraud activity can be assumed to be high. The organizational reconciliation of the conflicting interests between employee privacy and fraud detection as described in Section 3 avoids a continuous surveillance. However, the control function of the internal audit group is restricted and leads to unreported fraud cases. Also does the organizational reconciliation not allow for timely results and response. In fact, even if a company employs measures for reducing fraud, the actual time before detection is 15 months (median) [oCFE06].

6 Towards Automated Fraud Screening

Both pain points, the long time before detecting a fraud and the possibly high number of undetected frauds could be mitigated, if business transactions could be continuously and comprehensively screened for fraud evidence in near real time. In such a system the internal audit group would not primarily rely on whistle blowing (cf. evidence arrow from operating department employee to internal audit group audit in Figure 2). Rather, audit data is continuously extracted from the ERP and automatically screened for fraud (see *analysis of evidence* at the IT service provider in Figure 3). Also the audit data in such a system may be analyzed for fraud evidence on the discretion of the internal audit group, i.e. not only for given fraud suspicions. Clearly, such a system would not be legal, if no additional measures are taken to comply with privacy law.

Compliance with privacy law can in this case be established by pseudonymizing the audit data that is continuously extracted and only making the pseudonymized audit data available for automated and/or manual analysis (see *pseudonymizer* in Figure 3). If this screening has produced a fraud suspicion based on the pseudonymized audit data, additional audit data may be needed to be analyzed for corroborating or refuting the suspicion. Since now a suspicion is given, the scope of the additional audit data is determined for the given case and may be extracted in clear text (see *case-based audit data* in Figure 3). As described in Section 3 the additional audit data extracted for the case must be relevant for the case, the scope must not be defined unnecessarily large. It may be necessary to disclose some pseudonyms in the pseudonymized audit data for the given suspicion in order to determine the exact scope for extracting clear text audit data. The technical realization of the pseudonymizer and the organizational processes must ensure that the original data behind pseudonyms can only be disclosed for audit data that leads to a fraud suspicion (see Section 6.1).

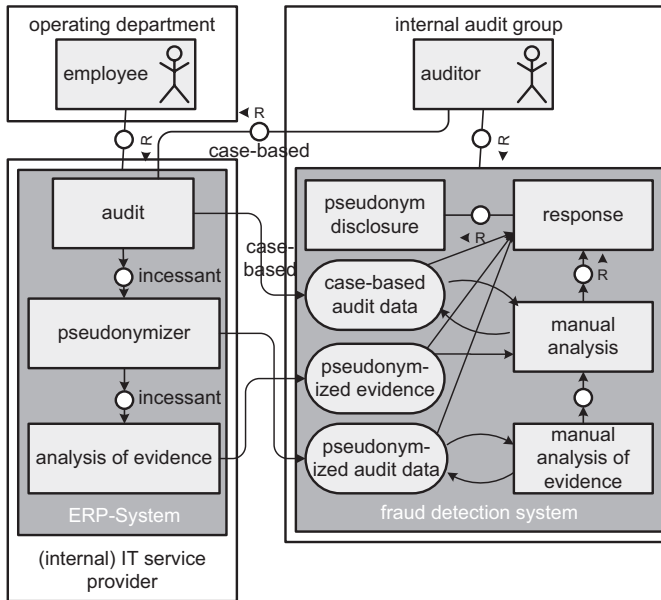


Figure 3: pseudonymized evidence screening for fraud detection

The proposed pseudonymization complements the rather random approaches to discovering fraud by whistle blowing and infrequent financial audits. Pseudonymization enables a comprehensive screening of all business transactions while respecting the confidentiality of sensitive data, such as personal data. This architecture design also allows for outsourcing of the fraud screening to a specialized external service provider, if we ensure that the service provider cannot disclose the pseudonyms.

6.1 Requirements for Audit Data Pseudonymization for Fraud Screening

Pseudonymization is a suitable measure to achieve privacy law compliance of continuous fraud screening, if the technical implementation meets several requirements:

1. **Confidentiality:** Pseudonymization is used to keep sensitive, e.g. personal, data confidential. This data must not be stored in a way that allows relating it to a natural person, which would allow bypassing the pseudonymization.
2. **Linkability:** Analyzing pseudonymized audit data for fraud suspicions, i.e. fraud screening, must be possible as effectively as for the unpseudonymized audit data. Technically the relationships between audit data features that are relevant for fraud screening need to be retained during pseudonymization.
3. **Technical purpose binding:** For fraud scenarios known in advance, the circumstances of pseudonym disclosure should also be agreed in advance and technically enforced. This allows for timely response to detected fraud evidence and avoids time-consuming decisions on the organizational level.
4. **Organizational purpose binding:** For fraud scenarios that are not known in advance, agreements for a technical purpose binding cannot be found in advance. Nevertheless,

it should be possible to disclose pseudonyms for new fraud scenarios subject to organizational purpose binding. The system must technically enforce the participation of the relevant stakeholders [Gem97], such as works council and/or data protection officer.

5. **Confidentiality of pseudonym mapping:** The mapping that relates the generated pseudonyms to the original data must be kept confidential.

6.2 Example Approach

In the following we assume that there is a process agreed by the works council and the data protection official according to which the internal audit group has identified and agreed collaboratively with the works council and with the data protection official, which audit data attributes are sensitive and/or personal data and need to be treated confidentially. In our running example the following attributes have been identified:

Purchase Requisition: *PR Number, PR Item, Short text, Requisition Tracking number, Fixed Vendor, Unloading point, Recipient, Name, Street, House number*

Purchase Requisition Approvals: *PR Number, WF Instance, Approver 1(-7)*

Accordingly the pseudonymizer (see architecture in Figure 3) shall replace these attributes with suitable pseudonyms, while respecting the requirements from Section 6.1. In the following we briefly demonstrate for the architecture described in Section 6 how a suitable technical approach for pseudonymization [Fle07] meets the requirements from Section 6.1.

6.2.1 Confidentiality and Linkability

Firstly the pseudonymizer replaces all sensitive and/or personal data attributes with pseudonyms and symmetrically encrypts each original data attribute with a randomly chosen key k of suitable length and unknown to a possible attacker³. This ensures confidentiality of the original data. Whenever during pseudonymization the linkability of data attributes needs to be retained, the same pseudonym is chosen for identical original data attribute values.

6.2.2 Technical Purpose Binding

Technical purpose binding aims at restricting pseudonym disclosure to the detection of a fraud suspicion. This requires that fraud suspicions be technically specified in advance. The activity of defining fraud suspicions needs to participate the works council and the data protection officer to ensure that the definitions actually comply with the definition of fraud suspicion given in privacy law.

A known fraud suspicion be technically characterized by n occurrences of specific types of business activity that are manifested in the audit data. From the key k that can be used to decrypt the encrypted original data then n or more shares are computed using an information theoretically secure secret sharing threshold scheme [Sha79]. The threshold scheme ensures that k may only be computed efficiently, if at least n shares have been generated. Each generated share is related to exactly one occurrence of aforementioned activities. Each time when an activity of a known fraud scenario is executed, the internal audit group receives the related share of key k . As soon as

³ In the current context we consider as attacker an entity that aims at disclosing the original data replaced by the pseudonyms.

the fraud suspicion has been completely substantiated, n shares have been generated and allow for efficient computation of key k . Using k the original data may be decrypted, i.e. disclosed. As a result, pseudonym disclosure is technically bound to the occurrence of fraud activity.

6.2.3 Organizational Purpose Binding

A similar approach is taken for organizational purpose binding of pseudonym disclosure. In advance the group of persons that need to participate in pseudonym disclosure decisions is determined, e.g. members of the internal audit group, of the works council and a data protection officer. Then a random key g is chosen secretly, and for each member of the decision group is generated a distinct share of g that is confidentially distributed to that member.

During runtime the random keys k are encrypted using g in a threshold cryptosystem [Gem97]. If for a new fraud scheme occurrence the group decides to disclose the pseudonyms, each group member uses her personal share of g to compute some share of k from the encrypted k . Combining the shares of k computed by the group members allows for recovering k and then to decrypt the original data.⁴

6.2.4 Confidentiality of Pseudonym Mapping

The extracted audit data must be pseudonymized before it is stored in secondary memory. Then the original data can only be accessed after pseudonym disclosure subject to technical or organizational purpose binding. Audit data extraction and pseudonymization are performed under the control of the IT service provider and out of the access reach of the internal audit group. The keys k are generated and stored in the primary memory of the pseudonymizer machine. Hence, pseudonymization can only be controlled via access to the primary memory of that machine or by modifying the configuration files of the pseudonymizer. State of the art technology cannot prevent both attacks, if conducted by administrators with system privileges. The group of system administrators for the pseudonymizer machine can be kept small, such that organizational controls are effective.

6.3 Revisited Assessment with Pseudonymization

The architecture depicted in Figure 3 together with the pseudonymization approach introduced in Section 6.2 implements the demarcation stipulated in pertinent privacy law insofar a comprehensive screening of clear text audit data is technically made infeasible and strongly hindered organizationally. A rededication for other purposes than fraud screening, e.g. behavior or performance monitoring, would require the collusion of the persons participating in organizational purpose binding and would create significant effort for disclosing all relevant pseudonyms. The data subject, i.e. the employee, can rely on the confidentiality of his personal data as long as she does not engage in fraudulent activity.

7 Conclusion

Enterprises rightfully collect and store data for the execution of their business processes. The recent fraud screening scandals were concerned with the fact that personal data of a large percent-

⁴ Note that the group does not recover g , but the individual k 's. Hence, g is not disclosed to any of the group members, which would allow bypassing technical enforcement of the organizational purpose binding.

age of the employees of enterprises were used as clear text data attributes for fraud screening. On the other, hand fraud screening facilitates the discovery of crimes. In analogy to the intention of law text on telephone and video surveillance, a comprehensive screening on clear text personal data puts the data subjects, i.e. the employees, under an undesired general suspicion. Hence, German law has been concretized to allow personal data for fraud analysis only, if evidence for a fraud suspicion already exists. As a result, comprehensive fraud screening on clear text personal data as a means to discover such fraud suspicion evidence is prohibited. [FRW09]

The presented architecture and pseudonymization approach replace a fraud screening on clear text personal data with an analysis on pseudonymized personal data. The pseudonyms may only be disclosed for discovered and a priori defined fraud suspicions in order to establish accountability. Disclosing pseudonyms is lawfully allowed, when a fraud suspicion already exists. Fraud screening on pseudonymized data may be used in compliance with privacy law in order to reduce the number of undiscovered fraud cases, while respecting the privacy of the employees not engaging in fraudulent activity. Such a measure can be implemented as a continuous process and may significantly reduce the time to discovery of fraud, since the internal audit group does not need to rely solely on whistleblowers and infrequent financial audits.

References

- [95/95] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, October 1995. http://europa.eu.int/eur-lex/en/lif/dat/1995/en_395L0046.html.
- [Biz07] Johann Bizer. Sieben goldene Regeln des Datenschutzes (in German). *Datenschutz und Datensicherheit*, 31(5):350–356, 2007.
- [FKM⁺10] Ulrich Flegel, Florian Kerschbaum, Philip Miseldine, Ganna Monakova, Richard Wacker, and Frank Leymann. *Insider Threats in Cybersecurity – And Beyond*, chapter Legally Sustainable Solutions for Privacy Issues in Collaborative Fraud Detection. *Advances in Information Security*. Springer, New York, 2010. To appear.
- [Fle07] Ulrich Flegel. *Privacy-Respecting Intrusion Detection*, volume 35 of *Advances in Information Security*. Springer, New York, 2007.
- [FRW09] Ulrich Flegel, Oliver Raabe, and Richard Wacker. Technischer Datenschutz für IDS und FDS durch Pseudonymisierung (in German). *Datenschutz und Datensicherheit (DuD)*, 33(12):735–741, December 2009.
- [FVB10] Ulrich Flegel, Julien Vayssière, and Gunter Bitz. *Insider Threats in Cybersecurity – And Beyond*, chapter A State of the Art Survey of Fraud Detection Technology. *Advances in Information Security*. Springer, New York, 2010. To appear.
- [GDW09] Alexander Grosskopf, Gero Decker, and Mathias Weske. *The Process: Business Process Modeling Using BPMN*. Meghan Kiffer, 2009.
- [Gem97] Peter Gemmel. An introduction to threshold cryptography. *Cryptobytes*, 2(3):7, 1997.
- [KGT06] Andreas Knöpfel, Bernhard Gröne, and Peter Tabeling. *Fundamental modeling concepts: Effective communication of IT systems*. Wiley, 2006.
- [oCFE06] Association of Certified Fraud Examiners. Report to the nation on occupational fraud and abuse, 2006.
- [Reu09] Reuters. German snooping scandal engulfs Airbus, April 2009.
- [Sha79] Adi Shamir. How to share a secret. *Communications of the ACM*, 22:612–613, 1979.
- [Wel09] Deutsche Welle. Spy scandal widens at German rail Deutsche Bahn, February 2009.