An Approach for Producing Privacy-Aware Reusable Business Process Fragments

Mohamed-Anis Zemni, Nejib Ben Hadj-Alouane
OASIS Laboratory
National School of Computer Sciences (ENSI)
Manouba, TUNISIA
mohamedaniszemni@gmail.com,
nejib_bha@yahoo.com

Moez Yeddes
CRISTAL Laboratory
National School of Computer Sciences (ENSI)
Manouba, TUNISIA
yeddes@yahoo.fr

Abstract—Nowadays, companies are more concerned to organize their business processes in the most flexible way, as to cope with the tough competition. Two major concerns are emerging, while dealing with the construction of new process functionalities: shortening the development periods and managing the privacy risks. An effective solution consists of reusing fragments of existing business processes. These fragments need to be declared as safe, however, from a privacy perspective. In this paper, we propose a comprehensive approach for decomposing existing processes in order to provide reusable fragments, while at the same time, preventing so-called sensitive association disclosure.

Keywords— Business Process, Fragmentation, Privacy-Preserving, Privacy-Aware, Activity Clustering.

I. INTRODUCTION

In business process management, providing good designs for business processes [1], mainly made of activities and data objects respectively linked with control flows and data flows, may improve the quality of the delivered services. This quality can be further improved by taking into account important privacy features. The reuse of existing processes or even only fragments of them into the new ones can increase both efficiency and productivity. Our work focuses on extracting the fragments from existing process models, till now, mostly carried out manually [2], rather than constructing fragments from scratch, since it appears to show a good potential geared towards maximizing productivity and efficiency gains. Existing approaches are either on enhancing process execution [3] or on securing the access to data that are rooted in between process activities [4]. To our best knowledge, none has worked on fragmentation for reuse purposes.

In addition, privacy and personal data preservation concerns should also be taken into account during the fragmentation task to prevent sensitive information leakage. Within business processes, personal data may be represented via data objects, and are routed in between the process activities by means of data flows. Important data privacy violations can occur by inferring information, through the construction of so-called sensitive associations of data belonging to published fragments of a same business process. Typically, these sensitive associations are not intended for sharing but for the purpose of performing key process functions. Addressing this problem within the context of business process fragmentation requires, among things, a good design of the fragments boundaries.

The fragmentation approach we detail in this paper is as depicted in figure (1). We first introduce a Privacy-preserving Constraints, in section (II), to protect the sensitive associations between process data objects, that the process owner, who has knowledge of the process artifacts, should expose in addition to the business process. We have also proposed a fragmentation technique based on Formal Concept Analysis (FCA) [5], in section (III), that we extend with the Privacy-preserving Constraints in order to functionally decompose the process into useful and privacy-aware fragments.

It should be noted, however, that focusing on the mechanisms for fragments reuse, is still an ongoing work, with promising results, but is not within the scope of the current paper due to space constraints.

II. PRIVACY-PRESERVING CONSTRAINTS

We have already proposed an initial approach [6] for introducing sensitive associations, in the business process domain. A sensitive association is the association between non-sensitive data objects that are safe when they are considered independently, but turn out to be critical when they are associated together (e.g. name/illness). Securing such associations is motivated as follows. Consider a fragment containing two distinct activities outputting two distinct data object sets that are involved in a sensitive association. The fragment may be subject to sensitive association leakage,
when integrated in a new process with some malicious activities equipped with inference mechanism. Indeed, the malicious activities can exploit the integrated fragment outputs to infer the sensitive association with pretty good confidence. In this work, we seek to improve on our privacy enforcing mechanism. For this, we propose a privacy-preserving constraint to avoid disclosing sensitive associations between data objects that are rooted between activities. Data objects are either activity inputs or activity outputs and are denoted by $E$. We denote by $E^* \subseteq E$ the set of data objects involved in the sensitive association called, critical data object set, and by $E \setminus E^* \subseteq E$ the set of neutral data objects. Activities that output some critical data objects in $E^*$ are called critical activities. The rest of activities are called neutral. Activities that input some critical data objects in $E^*$ are also neutral because they do not produce critical data objects but only receive them.

We propose the following privacy-preserving constraint (or privacy-constraint) on sensitive associations, to be considered while performing the fragmentation task, in order to ensure that distinct critical activities that output distinct critical data objects involved in the same sensitive association should never figure in the same cluster.

**Definition 1** (Privacy-preserving Constraints). $(E_i^∗ \stackrel{p}{\rightarrow} E_j^∗)$ is a privacy-preserving constraint to fix the sensitive association between the data objects $E_i^∗$ and $E_j^∗$. $C_N$ is a privacy-preserving constraint set such as $C_N = \{(E_i^∗ \stackrel{p}{\rightarrow} E_j^∗)|E_i^∗ \subseteq E^*, E_i^∗ \cap E_j^∗ = \phi\}$

Given a privacy-constraint set, $C_N$, we classify activities into groups, that we call privacy-groups and denote $G_{E_i^∗}$ ($E_i^∗$ involved in $C_N$) where $G_{E_i^∗}$ would enclose the activities that are not in conflict with $E_i^∗$, i.e. critical activities outputting some elements in $E_i^∗$ and neutral activities. Thus, composing activities of a same privacy-group does not reveal any sensitive association. Let us denote $G = \bigcup G_{E_i^∗}$ the set of privacy-groups.

**III. Privacy-aware Fragmentation**

In this work, we propose to provide useful and privacy-aware fragments. A business process fragment [7], [8] is a portion of a process intended for reuse purposes. It should contain no cycles, and a single control flow linking up two distinct activities. It is connected and can be composed with other fragments to build a complete and useful process. It is made of at least one activity, and of several control (dangling or not) and data flows.

The proposed fragmentation approach is intended to be carried out by an automatic fragmentation tool. To render the fragmentation automatic, we need to fix the features that the fragmentation technique is based on. [9] has defined a context as the set of relevant information, that are used to characterize the situation of an entity. Consequently, we propose to use the data objects that are used in the original process, such that, having the data objects considered as relevant information, we can guess the process context describing the functionalities. Therefore, by selecting a set of activities handling some data objects (e.g. “prescription”), it can be deduced, with quiet good confidence, the subprocess’s main functionalities (e.g. “prescription management”). Note that the functionality generating is beyond the scope of this paper.

We will use the FCA [5] mathematical formalism, that is based on contexts and concepts. It is commonly used to check the affinities between objects. A context is commonly described as a relationship between objects and attributes. A concept is a objects/attributes couple, where all the objects share the attribute set and the attributes are all shared by only the same set of objects. This is checked by the Galois correspondence functions. Readers are kindly referred to the existing literature for in-depth details.

We first translate the process model, into the FCA Formal Context where objects correspond to activities and data objects are attributes. The relation $R$ captures the data flows between activities and their corresponding data objects. The classical FCA, following this reasoning, even if it may provide useful fragments involved by the concept set, is not privacy-aware as it lacks privacy features. Indeed, some sensitive association may be revealed by the fragments. Therefore, we propose to incorporate the privacy-constraints $C_N$ in the FCA logic. More specifically, the fragmentation will be processed within each privacy-group in $G$, derived from the privacy-constraints. The following shows the definitions for the extended FCA formalism.

**Definition 2** (Privacy-Aware Formal Context). Given a business process model and a privacy-preserving constraint set $C_N$, a privacy-aware formal context is the tuple $C = (E, A, G, R)$:

- $E$ and $A$ are respectively data objects and activities.
- $G$ is the set of privacy-groups, derived from $C_N$.
- $R : E \times A \times G$ is the relation set. If an activity $a$ that is involved into $G_{E_i^∗}$, has a data object $e$, then the relation $(e, a, G_{E_i^∗}) \in R$.

We use the the Privacy-Aware Galois Correspondences to check the affinities between the activities according to the data objects.

**Definition 3** (Privacy-Aware Galois Correspondence). For each privacy-group $G_{E_i^∗} \in G$, the common data objects shared by all the activities in $A_i \subseteq A$ are given by $\Theta(A_i, G_{E_i^∗}) = \{e \in E \setminus (E^* \setminus E_k^*)|(e, a, G_{E_i^∗}) \in R, \forall a \in A_i\}$. Similarly, for each privacy-group $G_{E_i^∗} \in G$, the common activities sharing all data objects in $E_j^∗ \subseteq E \setminus (E^* \setminus E_k^*)$ are depicted by $\Delta(E_j, G_{E_i^∗}) = \{a \in A|(e, a, G_{E_i^∗}) \in R, \forall e \in E_j\}$.

From the Privacy-Aware Galois Correspondences we generate the Privacy-Aware Fragments.
Definition 4 (Privacy-Aware Formal Concept/Fragment). Given a privacy-aware formal context $C = (E, A, G, R)$, a privacy-aware formal concept is the tuple $\text{Con} = (E_j, A_i, G_{E_j}^i)$, where $E_j \subseteq E$ is a set of data objects shared by each $a \in A_i \subseteq A$ within $G_{E_j}^i$ with $\Theta(A_i, G_{E_j}^i) = E_j$, and $\Delta(E_j, G_{E_j}^i) = A_i$. $F = A_i$ is the corresponding privacy-aware fragment.

The following example performs the extended FCA.

Example. Consider a process $P$ made of a set of activities $A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\}$, and a set of data objects $E = \{o, d, n, q, c, i, b\}$. Let $C_N = \{(\{d\} \neq \{c\})\}$ be a privacy-constraint set, fixed by the process owner. This results into the privacy-group set $G = \{G_{\{d\}}, G_{\{c\}}\}$ where $G_{\{c\}}$ would enclose activities that are not in conflict with $\{d\}$ and $G_{\{c\}}$ would enclose activities that are not in conflict with $\{c\}$. $C = (E, A, G, R)$ (illustrated in Figure 2) is the privacy-aware formal context corresponding to the process. $R$ is the relation set capturing the data flows within each privacy-group. Data objects are columns, activities are rows, table 2a and table 2b respectively correspond to the privacy-groups $(G_{\{d\}}$ and $G_{\{c\}}$). Each $'x'$ corresponds to the relations capturing the data flows. Neutral data objects are represented with $'x'$ in the two tables. Critical data objects are represented with $'x'$ in the corresponding table. We assume that the privacy-constraints are correct and there are no activities using both $d$ and $c$ at the same time.

From the context $C$, we process the fragmentation task, within each privacy-group, to generate the Privacy-Aware Formal Concept set. Thus, we use the Privacy-Aware Galois Correspondence functions. For example, $\text{Con} = \{(\{n\}, \{a_6,a_8,a_{10}\}, G_{\{d\}})\}$ is a Privacy-aware formal concept according to the correspondence functions $\Theta$ and $\Delta$ and $F = \{a_6,a_8,a_{10}\}$ is its corresponding fragment.

Using the extended FCA formalism, we have generated privacy-aware fragments. Indeed, these latter are formed within the boundaries of each privacy-group. Each fragment describes a given functionality through the data objects depicted in the concept from which the fragment is derived. To consider them as useful, fragments should also be connected. A fragment is connected when there is a path between each two activities. This ensures the execution of all the activities when reusing a given fragment. Nevertheless, some fragments are disconnected. Such fragments cannot be reused in new processes and are considered as residues. Connected fragments are considered as useful. They contain no cycles and a single flow linking an activity to another one (i.e. ensured in the original process). Some flows are still dangling. Indeed, some target or source activities are missing. They will be fulfilled with activities when reusing the fragment to build a new process. Some techniques can be used to glue fragments into processes [8].

IV. CONCLUSION

We have proposed a framework for providing useful, well-formed and privacy-aware fragments that can be integrated in new processes. This approach can be further improved in future works from a semantic viewpoint by using keywords from activity description, instead of data objects. The results still be interesting and to our best knowledge, this is the first research on functional fragmentation while respecting privacy concerns.

REFERENCES


