A Model for Reasoning About the Privacy Impact of Composite Service Execution in Pervasive Computing

Roberto Speicys Cardoso and Valérie Issarny

Abstract Service composition is a fundamental feature of pervasive computing middleware. It enables users to leverage available computing power by using existing services as building blocks for creating new composite services. In open and dynamic environments, service composition must be flexible enough to admit realization by different executable workflows that have similar functionalities but that present different partitions of tasks among available services. This flexibility, however, raises new privacy issues e.g., a single service performing all tasks of a workflow has access to more data than different services executing parts of the workflow.

In this paper we propose a model that enables users to reason about the impact on privacy of executing a composite service. The model is based on an extension of Fuzzy Cognitive Maps, and considers the impact of the composition as a whole according to the partition of tasks. We introduce our extension called Fuzzy Cognitive Maps with Causality Feedback, describe how they can be used to model the relationship among different personal data and the privacy impact of their disclosure, and give an example of how the model can be applied to a composition scenario.

1 Introduction

In service-oriented pervasive computing environments, accessible resources and applications are modeled as services that users may combine to obtain new functionalities. There lies the greatest potential of service-oriented pervasive computing: clients can create innovative and unexpected applications by simply combining available services. Those composite services can be later reused to create yet more novel services, effectively enabling the user to leverage the existing pervasive computing power. Service composition is a key middleware functionality to realize

Roberto Speicys Cardoso · Valérie Issarny
INRIA Paris-Rocquencourt, 11 domaine de Voluceau, Le Chesnay Cedex France
e-mail: firstname.second_name@inria.fr

this scenario. Composition mechanisms must overcome challenges such as service heterogeneity and user mobility to enable effortless creation of services that are reliable, secure and that respect user-defined quality constraints. Many works proposed methods to improve service composition flexibility and to increase the probability of finding corresponding executable processes on dynamic environments [21, 19, 3].

An often neglected aspect of service composition is its impact over user privacy. In today's information society, personal information has become a valuable asset and many companies exist whose single purpose is to collect, combine and analyze personal data, threatening civil liberties and the citizen's right to privacy [16]. Pervasive computing environments are a valuable source of personal information since individuals are expected to use pervasive services to perform daily tasks and to generate a great quantity of virtual imprints that could be used to infer a number of intimate traits and beliefs. As a result, pervasive environments may become a notable target for privacy attacks and users need tools and models that allow them to identify risky situations and to avoid such attacks.

Companies offer identification services that explore data correlation to infer personal user details from apparently harmless data. For instance, even though one's gender, zip code and date of birth may not reveal much information individually, researchers affirm that between 63% [9] and 87% [23] of the American population can be uniquely identified when combining those three attributes. Some private data are also naturally conflicting and can expose crucial information when correlated. As an example, a customer may not feel comfortable to share the list of medicines he buys with his bank, since knowledge of this data could influence his bank's rates and terms for a personal loan. Such empirical studies and perceptions evidence the risks to privacy posed by entities that correlate different sources of personal data. When executing composite services and disclosing personal data to different providers, users must be aware of the privacy risks posed by their execution to decide whether or not to execute the service or to search for less invasive alternatives.

Existing research on privacy-enhanced workflow execution focuses mainly on more static and closed environments where workflows hardly ever change. Some solutions assume environments under a single administrative domain and propose mechanisms for specification of separation of duties in control flows [4, 13], while others enable specification of access control constraints on the data flow [6]. Pervasive computing environments, however, are more dynamic and hardly ever the exact user-defined workflow can be constructed with available services. A more suitable approach is based on a middleware component responsible for service composition, that separates or blends tasks of the original abstract workflow and creates different but functionally equivalent executable workflows using existing services. Specification of privacy constraints should be done independent of the composition workflow to cope with the dynamics of the environment.

In this paper, we introduce a model to reason about privacy when composing services in pervasive environments. The privacy consequences of disclosing personal data are modeled as an extension of Fuzzy Cognitive Maps (FCMs) [14]. This model can be later used to select the least privacy invasive executable workflow of a composite service and the privacy impact of reducing or increasing the precision of

disclosed personal data. In Sect. 2 we give a brief introduction to Fuzzy Cognitive Maps and present our extension that allows the specification of FCMs with causality feedback. After that, in Sect. 3, we define how to model the privacy impact of independently disclosing atomic and composed personal information. Section 4 describes how to create Fuzzy Cognitive Maps with Causality Feedback that model the relationship among different personal information required for an executable workflow and how the model can be used to measure the privacy impact of its execution. In Sect. 5 we introduce an example scenario and show how to apply our model. Section 6 compares this work to related research and finally in Sect. 7 we draw some conclusions and discuss future work.

2 Fuzzy Cognitive Maps with Causality Feedback

In this section we quickly review the basics of Fuzzy Cognitive Maps, discuss its limitations and describe our extension called Fuzzy Cognitive Maps with Causality Feedback (FCM-CF). More details about FCMs can be found in [15].

2.1 Fuzzy Cognitive Maps

Fuzzy Cognitive Maps were proposed by Kosko [14] based on the work developed by Axelrod [2] on cognitive maps to model social knowledge. Cognitive maps are graphs that describe causality (edges) among varying concepts (nodes). A positive edge between two concepts means that increasing the first concept will result in increasing the second and conversely a negative edge indicates that increasing the first concept will cause a decrease on the second.

Kosko noticed that causality generally happens in degrees, and the traditional model of cognitive maps, where edges can assume only one value in $\{+,-\}$, was too rigid to represent real-world reasoning. To overcome this limitation he introduced fuzziness in cognitive maps, allowing the representation of causality with different levels (such as *a little*, *much* and *a lot*). Formally, an FCM is a fuzzy graph containing N nodes $C_1,...,C_N$ representing concepts, and N^2 edges $W_{i,j}$ representing causality between concepts C_i and C_j. Concepts C_i are fuzzy sets and can assume values in the interval $[0,1]$ while edges $W_{i,j}$ can represent positive or negative causality between concepts i and j receiving values in the interval $[-1,1]$. The initial state of a FCM contains the initial values of each concept, $S(0) = \{C_1(0),...,C_N(0)\}$. Figure 1 shows a FCM with 5 concepts (denoted by circles) and 8 edges (with causality values in boxes).

Fuzzy Cognitive Maps can be used to answer *what-if* questions related to the concepts represented. Given an initial state $S(0)$ and a matrix W containing the values of all edges $W_{i,j}$ between concepts on the FCM (where $W_{i,j} = 0$ means that there are no edges between concepts i and j), state $S(t)$ can be obtained by the matrix

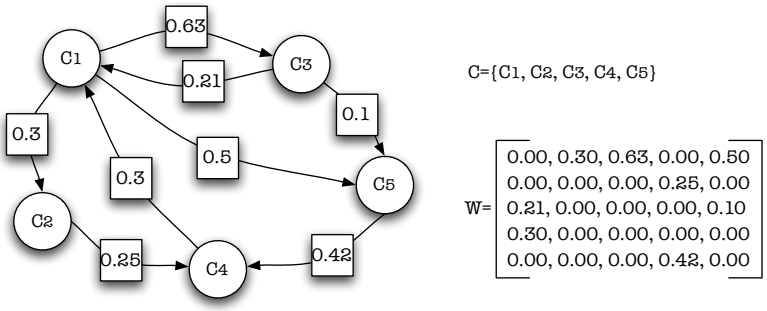

Fig. 1 A simple Fuzzy Cognitive Map with its set of concepts and its weight matrix

product $S(t-1) \times W$. To better represent learning, however, a sigmoid function is used to attenuate the new value of $C_i(t)$. Equation 1 is normally used to compute the next state of a concept on a FCM, where f is a sigmoid function such as arctangent. This operation is repeated until the map reaches a stable state (either $S(t) = S(t-1)$ or a cycle of states is detected).

$$C_i(t) = f(C_i(t-1) + \sum_{j=1}^{N} W_{j,i} C_j(t-1)) \tag{1}$$

FCMs proved to be a good abstraction to identify hidden patterns in causal relations and to simulate the global effects of changes in the values of some concepts. FCMs have been used to model causal relations in society [12], international politics [18], control engineering [22] and virtual worlds [7]. Researchers developed learning algorithms for FCMs that fine-tune edge weights to lead the FCM to certain pre-defined steady states [5]. However, FCMs have some weaknesses such as lack of support for time, conditional weights and non-linear weights [10].

2.2 Fuzzy Cognitive Maps with Causality Feedback

Despite their power, FCMs fail to capture an important notion of private information disclosure: when a personal attribute is disclosed to an entity, the privacy impact of disclosing other attributes to the same entity increases. There is a relation among concepts and causality that cannot be modeled by traditional FCMs. To enable the expression of such relations, we propose an extension to FCMs named Fuzzy Cognitive Maps with Causality Feedback, or FCM-CF.

A FCM-CF is the combination of two graphs. The first is a traditional FCM with N concepts $C_1, ..., C_N$ and N^2 edges $W_{i,j}$ representing causality between concepts C_i and C_j, where each C_i can assume values in the interval $[0,1]$ and each $W_{i,j}$ can receive values in the interval $[-1,1]$. The second graph represents causality feedback

and has two sets of nodes: *source* nodes and *destination* nodes. Source nodes are the concepts C on the FCM and destination nodes are the edges W on the FCM. Edges $F_{i,jk}$ on the causality feedback graph can assume a value in $[-1,1]$ and always connect a source node C_i to a destination node $W_{j,k}$. As a result, a causality feedback graph CF=$(C \cup W, F)$ can have at most N^3 edges. Figure 2 shows a simple FCM-CF graph, where solid lines are edges on FCM while dashed lines are edges on CF.

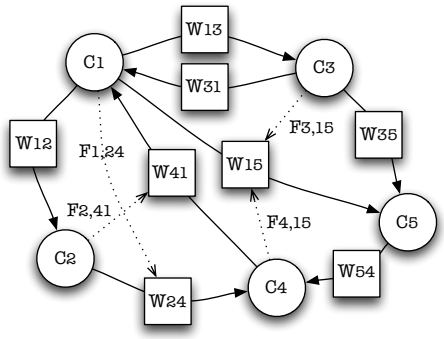

Fig. 2 An example of a Fuzzy Cognitive Map with Causality Feedback

The state of an FCM-CF at step t is composed of the values of its concepts and its weights on instant t, $S(t) = (C(t), W(t))$. Next state computation is a two-step process in FCM-CFs. Given a state $S(t)$, the values of $C(t)$ and $W(t)$ are fed into the CF graph to obtain the new matrix $W(t+1)$ that considers the causality increase on the FCM caused by the values of concepts. New values for W are computed using an equation similar to Eq. 1 as in traditional FCMs. After that, the new FCM= $(C(t), W(t+1))$ is used to compute $C(t+1)$. In summary, for each state containing the values of concepts and the weights on a FCM, first the edges of the FCM are updated according to the CF graph, and then the new concept values are obtained through the FCM.

3 Modeling the Privacy Impact of Information Disclosure

We propose a model based on FCM-CFs to enable individuals to measure the impact of disclosing personal information when executing composite services. Users first create an FCM-CF that represents the privacy impact of revealing each personal attribute to a different entity and afterwards they combine those graphs to model the effects of revealing subsets of data to the same entity. This final FCM-CF can be used to compare the privacy impacts of equivalent executable workflows featuring different service partitions or to simulate the effects on privacy when disclosing personal information with different levels of precision, as described in Sect. 4.

3.1 Characterizing Data Privacy Impact

There are two properties of private data that determine the privacy consequences of its disclosure: **identifiability** and **sensitivity**. Identifiability measures how easy it is to identify an individual after disclosure of a specific information; it is a metric related to the distribution of an attribute a with value v on a population U. We consider two metrics of identifiability: how much a user can be typically identified by the attribute a (I_a) and how much a particular user can be identified by a specific value v of attribute a (I_v). Definition of the value for I_v requires knowledge about the distribution of v on the population. When this distribution is unknown, I_a can be used as an approximation of I_v. Those values are expressed in the interval [0,1] where 0 means that the attribute a or the value v are common to all members of the population U and do not allow to identify the user, and 1 means that the attribute a or the value v are exclusive to the user and can identify him uniquely. The value of I is the value of I_a if $I_v = 0$, or I_v otherwise.

Sensitivity, on the other hand, is a metric associated with the user perception of privacy. It measures how comfortable the user is when disclosing a particular attribute a with value v. We also consider two metrics for sensitivity, namely, the attribute sensitivity (S_a) that measures how sensible it is for the user to disclose attribute a in general, and the value sensitivity (S_v) that quantifies how embarrassed the user is to disclose value v of attribute a. Those values are also expressed in the interval [0,1] where 0 means that the user does not see any issue when revealing attribute a or value v, and 1 means that the user is very embarrassed to reveal attribute a or value v. The data sensitivity S is also defined as S_a if $S_v = 0$, or S_v otherwise.

In many cases private information can be disclosed using different precision levels. For instance, when asked for his age, the user can disclose his exact age (e.g., 27 years old) or just his age range (between 20 and 30 years old). Each possibility represents different degrees of identifiability and sensitivity on the interval $[0,1]$. The values of I and S in this case may vary according to the precision of the information the user reveals. In other cases personal information can be decomposed into smaller components. As an example, a *complete name* contains a *first name* and a *last name*. The next section discusses how to combine information components to model the privacy impact on disclosing the complete information and how individuals can use the model to evaluate the different impacts on revealing subsets containing the information components.

3.2 Modeling Privacy Impact

The privacy impact of disclosing a specific personal information can be obtained from its identifiability and from its sensitivity as defined previously. Since sensitivity is a user-related metric, we give it a bigger weight than identifiability so that the user's perspective on privacy prevails. The sum of their weights does not have to be 1, the values only represent that identifiability has a positive effect slightly smaller

than average (0.4) and that sensibility has a positive effect a little bigger than average (0.6) on the privacy impact of disclosing the information. The FCM-CF representing the privacy disclosure impact of an atomic personal attribute is described by Fig. 3 (I), where *Ia* is the identifiability of information *a*, *Sa* is its sensitivity and *Pa* is the privacy impact when the information *a* is disclosed. If information *a* has different levels of detail – e.g., location can be expressed in GPS coordinates, street name or city – this diagram can be used to measure the consequences of disclosing data using different resolutions. The values of *Ia* and *Sa* are fuzzy and may assume different degrees according to data resolution. For instance, identifiability of location in GPS coordinates may be closer to 1 while identifiability of location as a city name is much smaller. For each possible data resolution *i* corresponds a pair $\{Ia_i, Sa_i\}$. Variables *Ia* and *Sa* can be replaced by values Ia_i and Sa_i on the map to compute the privacy impact Pa_i of disclosing variable *a* with resolution *i*.

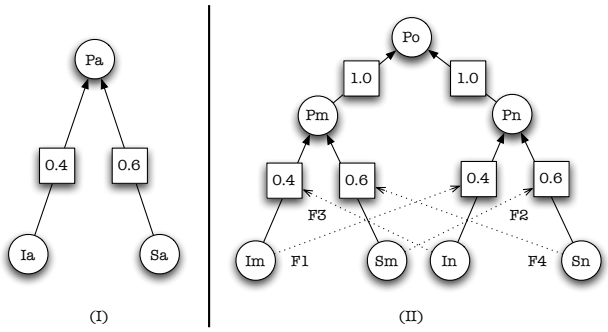

Fig. 3 An FCM-CF modeling the privacy impact of disclosing (I) an atomic personal information and (II) a composed personal information

Personal information can also represent a composition of other atomic information. Users can define the identifiability and sensitivity for each atomic attribute and create an FCM-CF that represents the impact caused by disclosure of all the attributes. Figure 3 (II) shows the FCM-CF of information *o*, which is composed by atomic informations *m* and *n*. This map shows how causality feedback is used to represent the relation between disclosure of different pieces of personal information. If information *m* is independent from *n*, or in other words, if knowledge of *m* does not affect knowledge of *n*, edges F_i have value 0 and the effect of disclosing *m* and *n* together (or *o*) is the same as disclosing each data separately. However, in many cases unveiling one personal information has consequences on further disclosures. For example, the privacy impact of disclosing a first name increases if the last name was already revealed to the same entity. Edges F_i can thus be used to represent the increase on privacy impact when both data are revealed. Their values are defined according to the strength of the relation between the information data types.

4 Privacy-Aware Selection of Pervasive Composite Services

After creating maps for all relevant personal information that the user may disclose, those maps can be combined to measure the privacy impact of service compositions. This section explains how the model is instantiated according to available service compositions and how the user can select the one that poses the smallest threat to his personal life.

4.1 Pervasive Service Composition

We assume a service-oriented pervasive computing environment, where services are described by their inputs, outputs and properties. Clients do not have any previous knowledge about the services available on a given environment, and find required functionalities by either discovering an appropriate service through the middleware **service discovery** mechanism or by composing existing services to obtain that functionality using the middleware **service composition** mechanism. Users are mobile and can play the roles of service providers, service consumers, or both.

In such open and dynamic environments, service composition must be flexible enough to take into account all services available at a given moment and to combine them in multiple ways to obtain the operations requested by the user. We suppose that clients describe service compositions by defining **abstract workflows**. An abstract workflow describes the process to obtain a certain functionality including its control and data flows. Abstract workflows can be built based on the user's past experiences creating composite services or using directions provided by other users or service providers. However, unlike a traditional workflow, an abstract workflow can be realized by different **executable workflows** [3]. The service composition mechanism may split an abstract task into different services, or merge different abstract tasks into a single executable service to obtain a composition that provides functionalities equivalent to the user-defined abstract workflow and that is achievable using available services.

Even though the process described by an abstract workflow and defined by its corresponding executable workflows are similar in terms of functionalities, the same is not true if we consider the user's perceived privacy impact. Although a user-defined abstract workflow may define that data about user's diseases and the user credit card number should be processed by different services, an executable workflow could merge those services in a composition that still provides the user-required functionality but that does not respect his privacy. As discussed in Sect. 1, data correlation is a big threat to privacy and users should be able to control which personal information can be accessed by a specific entity. However, executable workflows contain data flows that are not part of and cannot be predicted by the original abstract workflow, what makes specification of privacy constraints on abstract data flows not effective. These characteristics of service composition in open and dy-

namic environments require a flexible method for handling personal data disclosure that is not coupled to a particular workflow specification.

4.2 A Model for Reasoning about Privacy of Service Compositions

Ideally, the user wants to perform complex tasks without disclosing personal information to protect his privacy. However, privacy management involves a negotiation between user convenience and private information disclosure, and to obtain some specific outputs, certain private data may be necessary. A user who wants to buy a book and receive it at home will have to eventually disclose his home address. On the other hand, a service may ask the user to provide private data that is not required for the output, such as telephone, address and age to send an e-card[1]. Disclosure of personal data necessary to execute a task can be viewed as an acceptable risk, while transmission of unnecessary private data may be deemed inadmissible.

To model the compromise between revealing a private information and obtaining the composition functionality, we introduce the concept of **convenience**. Convenience is opposed to privacy, so the more convenient to the user it is to disclose a personal information, the smaller will be the privacy impact of its disclosure. Convenience has, thus, a weakening effect on the privacy consequences of personal data disclosure. Its value is fuzzy, and is related to the necessity of disclosing the information to obtain the desired output. It can be determined according to user experience (how many times that information was already requested to obtain the same output), recommendation (privacy activist groups may publish lists of abusive requests) or legislation (defining which categories of private data that merchants are allowed to require when providing a specific output).

Figure 4 shows an FCM-CF representing a service composition that requires three types of personal data. To model the privacy impact of the composition, a new concept P representing the total impact of executing the composition is created. The particular privacy impact of every information required to run the composition is connected to P with weight value 1. This is due to the fact that there is no attenuation on the disclosure privacy impact caused by their combination. Each information has its specific convenience factor C, that has an effect of -0.5 over its privacy impact. This weight corresponds to the notion that if an information has small identifiability, its disclosure does not trouble much the user and revealing it is highly convenient, then its privacy impact is close to 0.

Finally, the model requires weights for all dashed lines representing causality feedback. Identifiability of an information can only increase identifiability of another information, and sensitivity likewise. Causality feedback happens in service composition when related personal information is disclosed to the same service provider, and its value depends on the service composition being considered. If each data is disclosed to a different provider, the composition does not present causality

[1] www.netfuncards.com

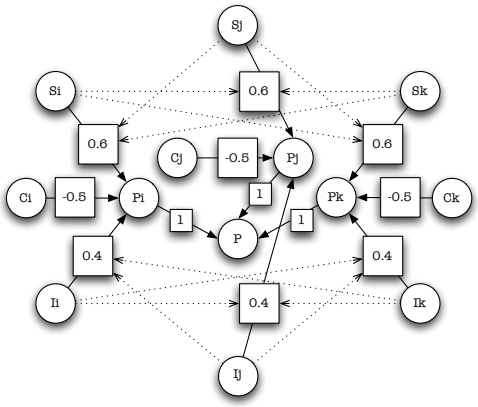

Fig. 4 An FCM-CF of a composition requiring three different atomic personal informations

feedback and the weight of all dashed arrows is 0. However, if the composition requires sending different information to the same provider, the CF edges relating those informations will have weights proportional to the privacy impact increase caused by their correlation. For instance, when the zip code and the gender of an individual are disclosed to the same entity, we can say that knowledge of the zip code greatly increases the identifiability of gender, while knowledge of the gender slightly increases the identifiability of zip code. Some service discovery protocols explicitly provide mechanisms for service provider description (such as the *bussinessEntity* field in UDDI [20]) and the model can use this information to define which causality feedback edges are active when evaluating a specific workflow. Other discovery protocols do not particularly support service provider specification, but this information could be extracted from data such as the service access point URL or external directories that categorize service providers.

Since only a few causality feedback edges are used to model the privacy impact of disclosing groups of personal data to the same provider, the resulting CF graph is very sparse which reduces the complexity of computing FCM-CF states. Also, since the resulting FCM-CF does not contain cycles, the map stabilizes after a few interactions. The value of outermost concepts never changes among interactions, so the edge values stabilize after one interaction. After two interactions the whole map reaches a stable state.

5 Application Example

To present how our model can be used to evaluate the privacy impact of executing different service compositions we introduce in this section a pervasive computing

service composition scenario that can be realized by two possible executable workflows.

William is passing his vacations at a city very distant from where he lives. Even though he suffers from a critical medical condition, due to advances in treatment he can lead an almost normal life. This morning, however, he is not feeling very well and he thinks it would be better to see a doctor. From his hotel room, he uses his smart phone to book a taxi that will take him to the nearest hospital that has a treatment center specialized on his condition. He uses a pervasive web service application provided by his personal doctor to help him to perform this task.

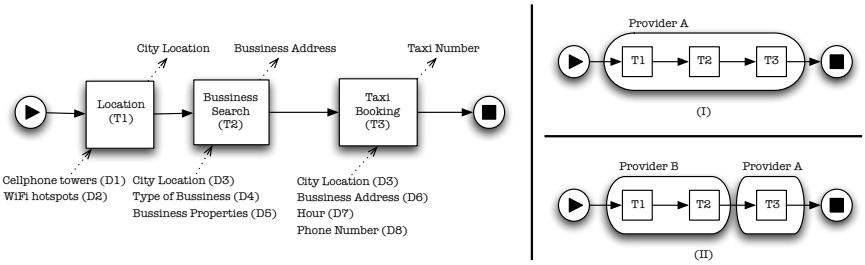

Fig. 5 Abstract workflow representing a composition and two possible executable workflows

The abstract composition workflow used by the application is depicted on the left side of Fig. 5. It uses the YAWL workflow notation [1] to represent the control flow, and incoming and outgoing dashed arrows to represent the data flow. It consists of three tasks: the first tries to locate the user based on information such as cellular phone towers and wireless hotspots nearby. The second task receives as input the user location and a type of business (in this case an hospital with a specific treatment center) and outputs the address of the closest establishment. Finally, the last task books a taxi based on the address of origin and destination, the time, and a phone number for confirmation. Figure 5 shows on the right side two possible executable workflows that provide the user-required functionality but present different privacy properties. Tasks inside the same capsule are executed by the same provider. Workflow (I) shows a situation where all the services are offered by the same entity, while workflow (II) represents an executable workflow where location-aware search is performed by one service provider and taxi reservation by another.

5.1 Building the Model

Two steps are necessary to model the privacy impact of composite service execution. First the user must create the model, and then instantiate it with information specific to the workflow. Model creation is independent of executable workflows and can be performed offline. It consists of defining the identifiability and sensitiv-

ity of each personal information that can be provided by the user, and the causality feedbacks that appear when different data is disclosed to the same entity. In this example, we will focus on the eight types of data required to execute the composition. Typical values are used for most of the data, except for the values "hospital" and the specific treatment center that have user-defined identifiability *medium* and *high*, and sensitivity *high* and *very high*. Typical values can be obtained from other users or from privacy activism groups, and can be related to particular inputs on service composition workflows by using ontologies (for instance to identify that *city location* is a type of *location*). Table 1 contains the values assigned to identifiability and sensitivity of each data.

Table 1 Definition of the identifiability and sensitivity of each personal data

Data	I_{a_i}	I_{v_i}	S_{a_i}	S_{v_i}
D_1	0.2	-	0.4	-
D_2	0.2	-	0.4	-
D_3	0.3	-	0.5	-
D_4	0.4	0.5	0.5	0.8
D_5	0.4	0.7	0.5	1.0
D_6	0.3	-	0.6	-
D_7	0.3	-	0.6	-
D_8	0.8	-	0.5	-

The weight of causality feedback edges can also be defined in advance. Afterwards, depending on the executable workflow under evaluation, those edges may receive the value 0 (for data disclosed to different entities) or the pre-defined value (for data disclosed to the same entity). Considering our scenario above, disclosure of the phone number greatly increases the identifiability of other personal data disclosed to the same entity. We define CF edges with value 0.7 from concept D_8 to the weights of all edges connecting data identifiability with its privacy impact.

5.2 Using the Model

The model above must be instantiated for each executable workflow under consideration. Model instantiation consists of the definition of convenience values C_i to each information according to the need to disclose it to obtain the desired output, and neutralization of CF edges connecting data that is not disclosed to the same entity on the executable workflow. In this example, all information required by the tasks is necessary to obtain the desired composition output, so all the convenience values are set to 1 in both model instances. Causality feedback edges, on the other hand, are different on both maps: the model for the first executable workflow will contain all causality feedback edges, since the same provider has access to the user's phone number and all other data the user provides, while on the second case the model

instance will only have CF edges between the identifiability I_8 of data D_8 and the identifiability of other data accessed by the same entity, namely I_3, I_6 and I_7. Figure 6 shows both model instances.

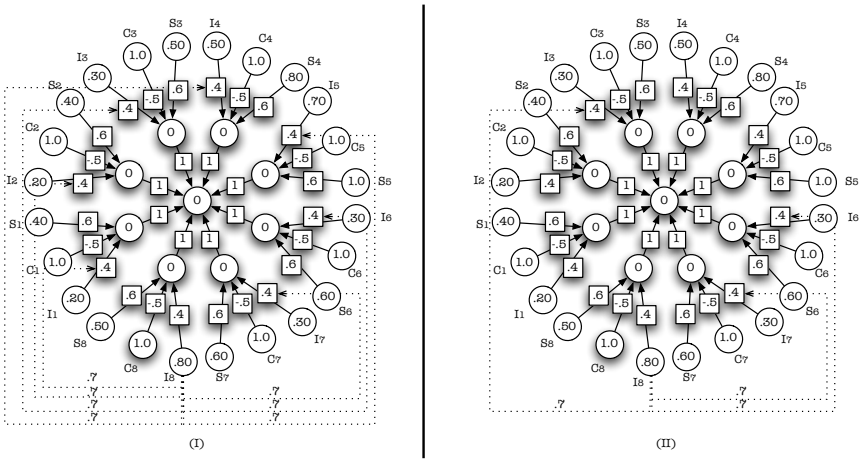

Fig. 6 Model instances corresponding to executable workflows (I) and (II)

Once values for all concepts are defined, we can iterate the FCM-CFs until they reach a stable state to compare the privacy impact of each executable workflow. The iteration begins by computing the new weights of all causality edges that are influenced by CF edges as explained in Sect. 2. In model (I) this results in changing the weight of all edges connecting the identifiability of an information to its privacy impact from 0.4 to 0.76. In model (II), only edges coming out from concepts I_3, I_6 and I_7 suffer causality feedback, so only their weights are updated to 0.76 while the others remain at 0.4. After updating the weights, the new state of the FCM-CF can be computed as in traditional FCMs. After two iterations the models reach a stable state, and the value of the total privacy impact concept (the innermost node of the graph) is 0.98 for executable workflow (I) and 0.71 for executable workflow (II). As a result, the user can compare the privacy impact of executing each workflow and select the second workflow since it is less privacy invasive.

6 Related Work

Many extensions to Fuzzy Cognitive Maps exist on the literature. The one that comes closer to represent the type of causality relations that we identified in personal information disclosure are Extended Fuzzy Cognitive Maps (E-FCMs) [10]. E-FCMs provide time support, weights $W_{i,j}$ can be any function on concept C_i (and

no longer only a linear function as in traditional FCMs) and weights can be conditional (if C_m then $W_{i,j} = x$). However, E-FCMs do not provide features to express situations where concept C_k gradually affects the weight of $W_{i,j}$ as in FCM-CFs.

The idea of using information identifiability and sensitivity as metrics for privacy is shared by [11]. They propose a model to generate policies for attribute disclosure in access control based on factors such as attribute identifiability and sensitivity, inference rules and public knowledge. In order to compare different requests for personal information, they organize sensitivity on a lattice such that the sensitivity of revealing two attributes is always bigger than the sensitivity of revealing a single attribute. As a consequence, revealing data $\{D_1, D_2\}$ is always more sensitive than revealing only $\{D_1\}$ or only $\{D_2\}$, but $\{D_1, D_2\}$ and $\{D_1, D_3\}$ are incomparable since they are at the same level of the lattice. We believe that our model is better adapted to deal with real world situations where the privacy impact of disclosing groups of data must be compared e.g., the impact of disclosing {*first name, gender*} is smaller than the impact of disclosing {*first name, last name*}.

The results of research on workflow access control specification and separation of duty can help to increase the privacy of users executing composite services. The mechanism proposed by [6] enables specification of tuples <source, destination, message> on the abstract workflow that represent allowed messages between the source domain and the destination domain. In pervasive computing, however, a task on the abstract workflow may correspond to a composition of services on the executable workflow or a group of tasks on the abstract workflow may correspond to a single service on the executable workflow. Consequently, the source and destination of messages specified on the abstract workflow may be different than the ones found on the actual executable workflow. Specification of rules that account for all possible executable workflows originating from a service composition may be tiresome and error-prone. Definition of separation of duties for task execution [4, 13] may also help the user to avoid services to access conflicting data. Still, such solutions usually require that roles are well-defined among workflow participants and this is not always possible in open environments.

Finally, Falcone et. al. proposed the use of FCMs for modeling user trust decisions [8]. They categorize concepts that influence trust judgment into internal and external factors, and for each factor they identify four concepts that are sources of causality, namely, direct experience, categorization, reasoning and reputation. Their approach is complementary to ours since both models could be combined to enable users to compare executable compositions in terms of privacy and trust.

7 Conclusion and Future Work

In this paper we present a method for creating a model that allows users to reason about the privacy impacts of disclosing personal information. This model is specially valuable for users composing services in open and dynamic environments such as pervasive computing. In those environments, abstracts workflows defined in

service compositions can be realized by functionally similar executable workflows whose control and data flow may be different from the ones the user specified. Tasks in the abstract workflow may be executed by a composition of services, or groups of abstract tasks may be realized by a single service, according to existing services on the environment. In those cases, users must be able to compare the privacy impact of possible executable workflows and select the one that presents the smallest risk.

Our model is based on an extension of FCMs called Fuzzy Cognitive Maps with Causality Feedback, or FCM-CFs for short. With FCM-CFs it is possible to model situations where an increase on a concept may strengthen the causality between two other concepts, which is common in personal data disclosure scenarios. We describe how FCM-CFs can be used to model the privacy impact of revealing a personal information based on its identifiability and its sensitivity. The user can later connect those models to each other, and according to available executable workflows, define weights to relevant edges and assign convenience values to information disclosure, to evaluate the privacy impact of their execution. Based on the model results, users can then select the least invasive executable workflow or simulate the impact of disclosing personal data using different resolutions.

Even though this work was inspired by pervasive computing scenarios where the user needs to protect his personal data from malicious services, the model can also be used in situations where a service that stores personal information must protect its data from malicious users performing consecutive accesses, e.g. a database server. In this case, client queries to the database could be represented by a workflow and the model could be used to measure the privacy impact of executing the complete query, complementing other approaches to database privacy protection such as the ones surveyed by [24]. We plan to investigate in the future how to successfully use our model in such scenarios. Also, we intend to expand the model to include other factors that are relevant to perform privacy-aware selection of executable workflows e.g., the service reputation on respecting the client's privacy or the trust that the service will use private data for the right purposes. We also plan to evaluate our model on real privacy-sensitive composition scenarios (such as hospitals) and to create techniques that enable fine-tuning of the weights defined based on known results, possibly using FCM learning algorithms [17, 5].

Acknowledgements This work is part of the IST PLASTIC project and has been funded by the European Commission, FP6 contract number 026955.

References

1. van der Aalst, W.M.P., ter Hofstede, A.H.M.: YAWL: Yet Another Workflow Language. Information Systems **30**(4) (2005)
2. Axelrod, R.: Structure of Decision: The Cognitive Maps of Political Elites. Princeton University Press (1976)
3. Ben Mokhtar, S., Georgantas, N., Issarny, V.: COCOA: COnversation-based Service COmposition in PervAsive Computing Environments. In: ICPS'06: Proceedings of the IEEE Interna-

tional Conference on Pervasive Services (2006)
4. Botha, R.A., Eloff, J.H.P.: Separation of Duties for Access Control Enforcement in Workflow Environments. IBM Systems Journal **40**(3) (2001)
5. Carlsson, C., Fullér, R.: Adaptive Fuzzy Cognitive Maps for Hyperknowledge Representation in Strategy Formation Process. In: Proceedings of International Panel Conference on Soft and Intelligent Computing (1996)
6. Chafle, G., Chandra, S., Mann, V., Nanda, M.G.: Orchestrating Composite Web Services under Data Flow Constraints. In: ICWS '05: Proceedings of the IEEE International Conference on Web Services (2005)
7. Dickerson, J.A., Kosko, B.: Virtual Worlds as Fuzzy Cognitive Maps. In: Proceedings of the IEEE Virtual Reality Annual International Symposium (1993)
8. Falcone, R., Pezzulo, G., Castelfranchi, C.: A Fuzzy Approach to a Belief-Based Trust Computation. In: Trust, Reputation, and Security: Theories and Practice (2003)
9. Golle, P.: Revisiting the Uniqueness of Simple Demographics in the U.S. Population. In: WPES '06: Proceedings of the 5th ACM Workshop on Privacy in Electronic Society (2006)
10. Hagiwara, M.: Extended Fuzzy Cognitive Maps. In: IEEE International Conference on Fuzzy Systems (1992)
11. Irwin, K., Yu, T.: An Identifiability-Based Access Control Model for Privacy Protection in Open Systems. In: WPES '04: Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society (2004)
12. Kang, I., Lee, S., Choi, J.: Using Fuzzy Cognitive Maps for the Relationship Management in Airline Service. Expert Systems with Applications **26**(4) (2004)
13. Knorr, K., Stormer, H.: Modeling and Analyzing Separation of Duties in Workflow Environments. In: Sec '01: Proceedings of the 16th International Conference on Information Security: Trusted Information (2001)
14. Kosko, B.: Fuzzy Cognitive Maps. International Journal of Man-Machine Studies **24**(1) (1986)
15. Kosko, B.: Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Prentice-Hall International Editions (1991)
16. O'Harrow, Jr., R.: No Place to Hide. Free Press (2005)
17. Parsopoulos, K.E., Papageorgiou, E.I., Groumpos, P.P., Vrahatis, M.N.: A First Study of Fuzzy Cognitive Maps Learning Using Particle Swarm Optimization. In: CEC '03: The 2003 Congress on Evolutionary Computation (2003)
18. Perusich, K.: Fuzzy Cognitive Maps for Policy Analysis. In: Proceedings of the International Symposium on Technology and Society Technical Expertise and Public Decisions (1996)
19. Rajasekaran, P., Miller, J., Verma, K., Sheth, A.: Enhancing Web Services Description and Discovery to Facilitate Composition. In: SWSWPC '04: Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition (2004)
20. Singh, M.P., Huhns, M.N.: Service-Oriented Computing - Semantics, Processes, Agents. John Wiley and Sons (2005)
21. Sirin, E., Hendler, J., Parsia, B.: Semi-automatic Composition of Web Services Using Semantic Descriptions. In: Web Services: Modeling, Architecture and Infrastructure Workshop in conjunction with ICEIS 2003 (2003)
22. Stylios, C.D., Groumpos, P.P.: Fuzzy Cognitive Maps in Modeling Supervisory Control Systems. Journal of Intelligent and Fuzzy Systems **8**(2) (2000)
23. Sweeney, L.: Uniqueness of Simple Demographics in the U.S. Population. Tech. Rep. LIDAP-WP4, Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA (2000)
24. Verykios, V.S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y., Theodoridis, Y.: State-of-the-art in Privacy Preserving Data Mining. ACM SIGMOD Record **33**(1) (2004)