computing@tanet.edu.te.ua
www.tanet.edu.te.ua/computing

ISSN 1727-6209
International Scientific
Journal of Computing

ON PRIVACY CLASSIFICATION IN UBIQUITOUS COMPUTING SYSTEMS

Dan Cvrček [1) 2)], Václav Matyáš [1) 3)] and Marek Kumpošt [1)]

[1)] Masaryk University Brno, Faculty of Informatics
[2)] Brno University of Technology, Faculty of Information Technologies
[3)] Contact author: matyas@fi.muni.cz

Abstract: *Many papers and articles attempt to define or even quantify privacy, typically with a major focus on anonymity. A related research exercise in the area of evidence-based trust models for ubiquitous computing environments has given us an impulse to take a closer look at the definition(s) of privacy in the Common Criteria, which we then transcribed in a bit more formal manner. This led us to a further review of unlinkability, and revision of another semi-formal model allowing for expression of anonymity and unlinkability – the Freiburg Privacy Diamond. We propose new means of describing (obviously only observable) characteristics of a system to reflect the role of contexts for profiling – and linking – users with actions in a system. We believe this approach should allow for evaluating privacy in large data sets.*

Keywords: – *anonymity, Common Criteria, PATS, Freiburg Privacy Diamond, unlinkability.*

1. INTRODUCTION

This paper outlines the development of our appreciation of privacy concepts that started with a research exercise on data mining in evidential data for evidence-based reputation systems. Earlier versions of this work were presented at the 2004 IEEE Workshop on Privacy and Security Aspects of Data Mining, and at the 2005 NATO Advanced Research Workshop on Security and Embedded Systems.

The novel idea of evidence-based reputation (or trust) systems is that such systems do not rely on an objective knowledge of user identity [1, 2, 11]. One has instead to consider possible privacy infringements based on the use of data (evidence) about previous behaviour of entities in the systems. We provide a brief introduction to evidence-based trust/reputation systems, as well as to the privacy issues, addressing the common problem of many papers that narrow the considerations of privacy to anonymity only.

The paper is structured in the following way – remaining parts of this introductory section provide a brief overview of issues related to evidence-based systems, Common Criteria and Freiburg Privacy Diamond models, motivation for our research, and a simple example used to illustrate the use of privacy models. Section two then presents some of the Common Criteria concepts used in the following discussions, and also outlines the Common Criteria approach to privacy issues (families), together with a discussion of unlinkability – the most complex property/quality of privacy. The third section presents the Freiburg Privacy Diamond – a semi-formal model allowing for expression of anonymity and unlinkability, focussing on the mobile environment. Section four then examines the role of contexts in these two approaches to modelling privacy. This leads to the fifth section that proposes using contextual information to model systems for privacy evaluations, and presents non-existential definitions of the four Common Criteria privacy concepts. Section six concludes with an outline of related ideas and open issues.

1.1 Evidence-based trust/reputation

Evidence-based systems work basically with two sets of evidence (data describing interaction outcomes). The primary set contains evidence that is delivered (or selected from locally stored data) according to a given request content. That data is used for reputation evaluation to grant/reject access requests. Data in this first set may contain information from third parties representing evidence about behaviour collected by other nodes – recommenders.

The secondary set comprises data relevant to a local system. That data is used for self-assessment of

the local system security in various contexts (it may be a non-deterministic process in a certain sense). This set may be also referenced as derived or secondary data. Note that there may be an intersection between the two evidence sets with implications to privacy issues that we are investigating in related projects [4, 5].

The approach of reputation systems is rather probabilistic and this feature directly implies properties of security mechanisms that may be defined on top of such systems. The essential problem arises with recommendations that may be artificially created by distributed types of attacks (Sybil attack [7]) based on large number of nodes created just to gather enough evidence and achieve maximum reputation that would allow them to launch their attack(s).

1.2 A NOTE ON THE COMMON CRITERIA AND THE FREIBURG PRIVACY DIAMOND MODELS

This paper proposes formal definitions of existing Common Criteria concepts/areas of privacy and compares them with the Freiburg Privacy Diamond model (FPD) [18]. Recent research in anonymity systems [6, 10, 15] demonstrates that it is usually unfeasible to provide perfect anonymity and that implementations of privacy enhancing systems may provide only a certain level of privacy (anonymity, pseudonymity). This leads to definitions of several metrics that can quantify level of privacy achievable in a system, most often a (remailing) mix.

The Common Criteria class Privacy deals with aspects of privacy as outlined in their four families. Three of these families have a similar grounding with respect to entities (i.e., users or processes) whose privacy might be in danger. They are vulnerable to varying threats, which make them distinct from each other. These families are Unobservability, Anonymity, and Unlinkability. The fourth family – Pseudonymity – addresses somewhat different kind of threats.

1.3 MOTIVATION

While working on related issues [5], we became aware of the need to define the Common Criteria concepts (called families) dealing with privacy in a bit more precise fashion. As we were examining definitions of privacy concepts/families as stated in Common Criteria two negative facts emerged. First, the definitions are given in an existential manner, and secondly, not all aspects of user interactions relevant to privacy are covered. Both issues come from research carried out in the areas of side-channel analysis and security of system implementations, showing that it is not sufficient to take into account only the idealised principals and messages. It is also very important to consider the context, in/with which the interactions are undertaken. Information like physical and virtual (IP, MAC addresses) positions of users and computers, time, type of service invoked, size of messages, etc. allow to profile typical user behaviour and successfully deteriorate privacy of users in information systems.

We propose to introduce context information (side/covert channels, like physical and virtual location of users and computers, time, type of service invoked, size of messages, etc.) into the CC model and compare it with the FPD model that reflects only one very specific context information – location.

Our objectives for starting this work are as follows. Firstly, we want to provide a model that allows one to cover as many aspects of user interactions as is beneficial for improving quantification/measurement for different aspects of privacy; this model shall definitely provide for better reasoning/evaluation of privacy than Common Criteria and Freiburg Privacy Diamond models do. Secondly, and in a close relation to the first objective, we want to illustrate the deficiency of the Common Criteria treatment of privacy, and to provide a foundation that would assist in improving this treatment. Thirdly, with a long-term perspective, we aim to provide basis for partly or fully automated evaluation/measurement of privacy.

This paper does not address all aspects of data collection for privacy models, and neither does it suggest any means for improving the level of privacy protection.

1.4 A SIMPLE EXAMPLE

Let us present a trivial example that we use later in this paper to compare the formal models for privacy. The attacker attempts to determine which payment cards are used by a certain person with a particular card – she is interested in linking together all the cards of this person (identification of the particular person is not part of the attacker's goal at the moment). We assume the attacker is able to collect till receipts of shoppers from the same house or the same company. For this subset of supermarket clients we then do not mind a given receipt to show only a part of the payment card number.

There are three payment cards (with numbers *11*, *21*, *25*) used for three actual shoppings (visits of the supermarket resulting in payments – *A*, *B*, *C*), and there is also a set of typical baskets/shopping lists (*l*, *m*) in our simplistic example.

The attacker has a precise (100%) knowledge about connections between payment cards and shoppings, and an imprecise knowledge about classification of individual shoppings into typical "consumer group" baskets. This classification to "typical baskets" is usually done with some kind of a data-mining algorithm over actual shopping lists. Note that one could obviously achieve perfect knowledge should loyalty cards be used (and their numbers on the receipts), introduction of this has no qualitative impact to this example illustration in our model.

With just changing semantics, we may define a very similar example based on users of chat services connecting from a given Internet cafe. The categories would then be chat-room pseudonyms, chat sessions, and classification into groups based on interest (content) and/or language, with the attacker's goal of identifying pseudonyms used by one user in different chat sessions.

2. PRIVACY IN THE COMMON CRITERIA

2.1 THE STARTING POINT – MODEL

Since some of the discussions and proposals in this paper are based on the Common Criteria concepts, let us briefly present the related information. Relevant Common Criteria notions and concepts are as follows [17]:

Target of Evaluation (TOE) – An IT product or system and its associated administrator and user guidance documentation that is the subject of an evaluation.

TOE Security Functions (TSF) – A set consisting of all hardware, software and firmware of the TOE that must be relied upon for the correct enforcement of the TOE security policy.

TSF Scope of Control (TSC) – The set of interaction that can occur with or within TOE and are subject to the rules of the TOE security policy.

Subject – An entity within a TSC that causes operations to be performed.

Assets – Information or resources to be protected by the countermeasures of a TOE.

Object – An entity within a TSC that contains or receives information and upon which subject perform operations.

User – Any entity (human user or external IT entity) outside the TOE that interacts with TOE.

We can see (Fig. 1) that user does not access objects directly but through subjects – internal representation of herself inside TOE/TSC. This indirection is exploited for definition of pseudonymity as we will see later. Objects represent not only information but also services mediating access to TOE's resources. This abstract model does not directly cover communication like in (remailer) mixes as it explicitly describes only relations between users/subjects and resources of target information system.

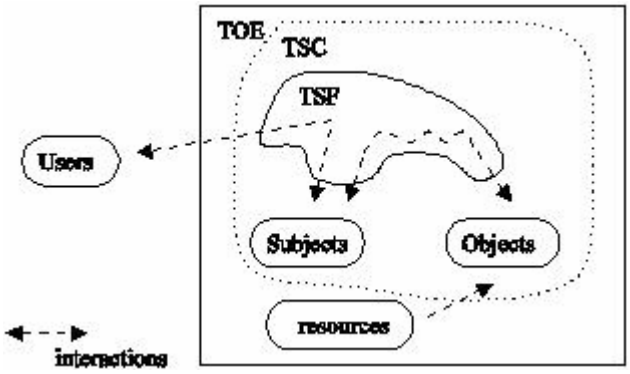

Fig. 1 – Common Criteria Model.

However, it is not difficult to extend the proposed formal definitions of major privacy concepts based on this model for communication models.

2.2 PRIVACY IN THE COMMON CRITERIA

Unobservability: *This family ensures that a user may use a resource or service without others, especially third parties, being able to observe that the resource or service is being used.* The protected asset in this case can be information about other users' communications, about access to and use of a certain resource or service, etc. Several countries, e.g. Germany, consider the assurance of communication unobservability as an essential part of the protection of constitutional rights. Threats of malicious observations (e.g., through Trojan Horses) and traffic analysis (by others than communicating parties) are best-known examples.

Anonymity: *This family ensures that a user may use a resource or service without disclosing the user identity. The requirements for Anonymity provide protection of the user identity. Anonymity is not intended to protect the subject identity.* Although it may be surprising to find a service of this nature in a Trusted Computing Environment, possible applications include enquiries of a confidential nature to public databases, etc. A protected asset is usually the identity of the requesting entity, but can also include information on the kind of requested

operation (and/or information) and aspects such as time and mode of use. The relevant threats are: disclosure of identity or leakage of information leading to disclosure of identity – often described as "usage profiling".

Unlinkability: *This family ensures that a user may make multiple uses of resources or services without others being able to link these uses together.* The protected assets are of the same as in Anonymity. Relevant threats can also be classed as "usage profiling".

Pseudonymity: *This family ensures that a user may use a resource or service without disclosing its user identity, but can still be accountable for that use.* Possible applications are usage and charging for phone services without disclosing identity, "anonymous" use of an electronic payment, etc. In addition to the Anonymity services, Pseudonymity provides methods for authorisation without identification (at all or directly to the resource or service provider).

2.3 PRIVACY FAMILIES REVISITED

Common Criteria privacy families are defined in an existential manner and any formal definition of them has to tackle a number of ambiguities. It is unrealistic to assume perfect/absolute privacy as demonstrated by several anonymity metrics, based on anonymity sets (number of users able to use a given resource/service in a given context) [12] or entropy assigned to a projection between service and user/subject identities (uncertainty about using a service) [15].

Can we introduce more formal definition of privacy notions and use them to define mutual relations? It is not easy, but the prospects of getting a clearer picture of mutual relations between different privacy aspects/qualities are encouraging.

Our proposal for the CC model privacy formalisation is based on the following graphical representation (Fig 2.). The set S represents observations of uses of services or resources, P_{ID} is equivalent of subjects and ID stands for users as defined in the CC. Sets U_S and U_{ID} are sets of all possible service use observations and identities, respectively – not only those relevant for a given system. By stating *with probability not significantly greater than* in the following definitions, we mean negligible difference (lower than ε) from a specified value [3]. Let A be any attacker with unbounded computing power.

Our formal transcription of existential definitions of CC privacy families is as follows.

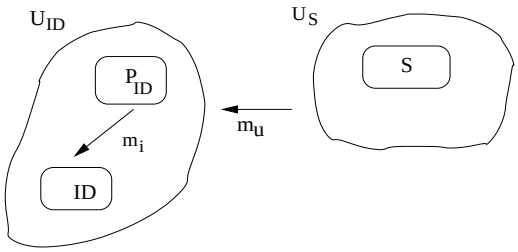

Fig. 2 – Schematics for the CC view of privacy.

Unobservability – there is a space of encodings (U_S) from which some elements are defined to encode use of service/resource (S). However A is not able to determine $\forall s \in S$ with a probability significantly greater than 1/2 whether a particular $s \in S$ or $s \in (U_S - S)$.

Anonymity – there is a probability mapping $m_u : S \to U_{ID}$. When

1. A knows the set ID – then $s \in S$, $u_{ID} \in ID$, she can only find $m_u(s) = u_{ID}$ with a probability not significantly greater than 1/|*ID*|.
2. A does not know anything about ID (particular elements or size) – then for $\forall u_{ID} \in U_{ID}$, she cannot even guess whether $u_{ID} \in ID$ with a probability significantly greater then 1/2. (The probability of finding $m_u(s) = u_{ID}$ would not be significantly greater then 0.

Unlinkability – let us assume there is a function $\delta : m \times S \times S \to [no, yes]$. This function determines whether two service uses were invoked by the same $u_{ID} \in U_{ID}$ or not. Parameter m stands for a function that maps service uses (S) into sets of identities U_{ID} (e.g. m_u from Fig 2.).

It is infeasible for A with any δ and any $s_1, s_2 \in S$, $s_1 \neq s_2$ to determine whether $m(s_1) = m(s_2)$ with a probability significantly greater than 1/2.

Pseudonymity – there exists and is known to A an unambiguous mapping $m_u(s) = u, \forall s \in S, u \in P_{ID}$. There also exists a mapping $m_i(u) = u_{ID}, \forall u \in P_{ID}$, $u_{ID} \in ID$, but is subject to strict conditions and is not known to A. When A

1. knows ID, she cannot determine u_{ID} with a probability significantly greater than 1/|*ID*|;
2. does not know ID, she can only guess with a probability not significantly greater than 1/2 whether $u_{ID} \in ID$.

These existential expressions can then be easily

turned into probabilistic ones that allow for expressing different qualitative levels of all these privacy concepts/families. This can be done simply by changing the "not significantly greater than" expression to "not greater than Δ ", where Δ is the given probability threshold.

2.4 THE UNLINKABLES

Unlinkability cannot be satisfied without other privacy families. It is now understood [13, 14] that the Common Criteria definition of unlinkability is not supporting some aspects of unlinkability in real systems, and a Common Criteria modification proposal in this manner is currently submitted. We point the reader to the fact that when pseudonymity is flawed, an attacker may obtain the ID of an actual user. The same holds when anonymity is breached.

Moreover, we are convinced that unlinkability may be a property of other privacy families. This comes straight from the formal unlinkability definition as stated above, where mapping m is the link binding the families together. Unlinkability should ensure that the particular family (or rather its implementation) does not contain side-channels (context information) that could be exploited by an attacker. We have found, in this context, two other meanings for unlinkability during our analysis. The first meaning is expressed in the following definition of unlinkable pseudonymity. It says that when a user employs two different pseudonyms, any A is not able to connect these two pseudonyms together.

Unlinkable pseudonymity – As for the definition of pseudonymity above in part 2.3, and also for any $s_1, s_2 \in S, \quad s_1 \neq s_2, \quad m_u(s_1) = u_1, \quad m_u(s_2) = u_2$ (where $u_1, u_2 \in P_{ID}$)

1. if A knows ID – she cannot find (with probability significantly greater than $1/|ID|$), whether $m_i(u_1) = m_i(u_2)$, or
2. A does not know ID – she cannot guess with a probability significantly greater than 1/4 whether $m_i(u_1) \times m_i(u_2)$ belong to $ID \times ID$, $ID \times \overline{ID}$, $\overline{ID} \times ID$, $\overline{ID} \times \overline{ID}$, respectively. $(\overline{ID} = U_{ID} - ID)$

The second semantics is built on the assumption that knowledge of several pieces of mutually related information is much more powerful than knowledge of just one piece of such information. When compared with the previous definition of unlinkable pseudonymity, the definition is now concerned with a property ensuring that there is no increase in the probability of correct identification of a given user when more information is available. The same reasoning lies behind the following definition of unlinkable anonymity.

Unlinkable anonymity – As for the definition of anonymity above in part 2.3, and

1. If A knows ID – she cannot find (with probability significantly greater than $1/|ID|$), such $s_1, s_2 \in S$, where $s_1 \neq s_2$, $m_u(s_1) = m_u(s_2)$.
2. A does not knot ID – with a probability not significantly greater than 1/4 whether $m_u(s_1) \times m_u(s_2)$ belong to $ID \times ID$, $ID \times \overline{ID}$, $\overline{ID} \times ID$, $\overline{ID} \times \overline{ID}$, respectively.

We can apply profiling when unlinkability is breached. Basically, unlinkability should ensure that the particular family (or its implementation) does not contain side-channels that could be used when several service invocations appear.

The example: The Fig. 3 depicts how CC models our example from part 1.4. It is obvious that there is no information about the context information for the basket (chat) contents. This implies that an attacker will not id any link between payment cards (pseudonyms) using this model, even though the link/connection exists. This shows that CC simply do not address contextual information.

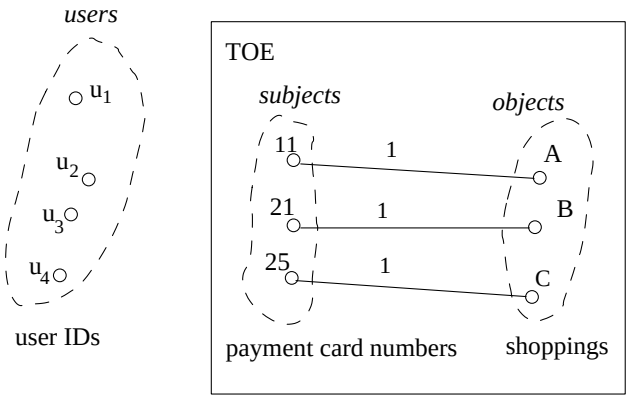

Fig. 3 – The example in the CC model.

3. FREIBURG PRIVACY DIAMOND

FPD is a semiformal anonymity (and partly also unlinkability) model by A. Zugenmaier et al. [18, 19]. The model originated from their research in the area of security in mobile environments. The model is graphically represented as a diamond with vertices User, Action, Device (alternatives for CC's user, service, and subject), and Location (Fig. 4). The main reason for introducing *location* as a category here is probably due to the overall focus of this model on mobile computing.

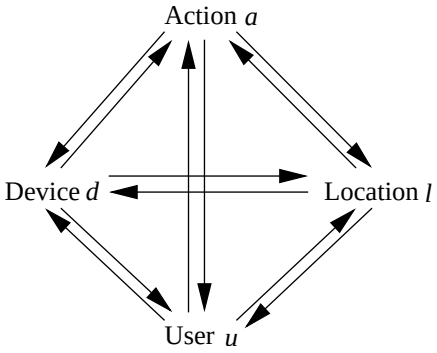

Fig. 4 – Freiburg Privacy Diamond.

Anonymity of a user *u* performing an action *a* is breached when there exists a connection between *a* and *u*. This may be achieved through any path in the diamond model. Let us recap basic definitions of the FPD model:

1. Any element x has got a type $type(x) \in \{User, Action, Device, Location\}$. Any two elements such as $x, y \in \{e \mid type(e) = User \vee Action \vee Device \vee Location\}$, $type(x) \neq type(y)$ are in relation R if the attacker has evidence connection x and y.
2. An action is anonymous if $U_R = \{u \mid type(u) = User \wedge (u,a) \in R\}$ is either empty or $|U_R| > t > 1$, where t is an anonymity threshold defining minimum acceptable size of anonymity set.
3. There is the transitivity rule saying that if $(x,y) \in R$ and $(y,z) \in R$, and $type(x) \neq type(z)$, then $(x,z) \in R$.
4. The union of all initial relation known to an attacker A defines his initial view $View_A$.
5. The transitive closure $\overline{View_A}$ of $View_A$ defines all the information an attacker A may infer from her initial view.

The book [18] also introduces three types of attacks with context information.

- Recognition attack – A realizes that several users $(x_i, type(x_i) = User)$ are in fact a single user.
- Linking attack – $(x,y) \in R$ and $(z,y) \in R$ are in the $\overline{View_A}$. When A is able to find just one pair $(y,x_i) \in R$ then she will know that $x_i = x$ and $(z,x) \in R$.
- Intersection attack – A knows anonymity sets for several actions. When she knows that a certain user is in all anonymity sets, she can apply intersections to reduce size of anonymity set and eventually identify the user.

Finally, the model assigns probabilities to edges in order to express attacker's certainty about existence of particular relations with some simple rules how to derive certainty for transitive relations.

The example: When attempting to model the example scenario (see part 1.4) in the FPD model, the attacker ends up with three diamonds for each service use (see Fig. 5). Here *user* and *location* represent domains with no particular values as there is no such information available. The attacker cannot find any intersection of the three diamonds – i.e., there is no attack as defined by the FPD model theory. This is obvious since the FPD model does not cover any other contextual information, only location and device.

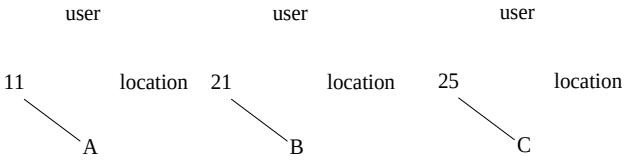

Fig. 5 – The example in the FPD model.

4. CONTEXTS IN THE TWO MODELS

Contexts and their roles are not reflected in the CC model. Considering Fig. 2, we see that the two vectors in question (m_i, m_u) are bound together through a pseudonym – subject in the CC language. Contexts may be assigned to any element of the model. *ID* represents physical entities and we may know their mobile phone locations, addresses, patterns of network usage, etc. P_{ID} – virtual IDs – can be characterised by previous transactions and possibly virtual locations (a virtual location may be in some cases very effectively mapped on a physical location). Elements of S may be further characterised by type, provider, etc.

The edges between sets (their elements) represent sessions taking place in the system. The information we may gather about them are highly dependent on actual implementation of the system and may comprise contextual information such as time, length, routing path, content, etc.

4.1 CONTEXTS IN FPD

The FPD model only briefly mentions context information but does not introduce any definition of it. The attacks based on context information do not say how to perform them but only defines changes in $\overline{View_A}$ when an attack is completed

Since the FPD model newly addressed the mobile computing environment, as opposed to the old-fashioned "static" environment, location had a very prominent role, as did the device to some extent. We have decided to treat these as "ordinary" context information, i.e. as any other additional information about the system that can link a user and an action (or more precisely, their identifiers).

5. CONEXT REVISITED – BASICS OF THE PATS (PRIVACY ACROSS-THE-STREET[1]) MODEL

We propose the following approach, inspired by the way location and device (descriptors) are represented in FPD.

We suggest all context information available to an attacker to be represented as vertices in a graph, where edges are weighed with the probability of the two incident vertices (contextual information, user and service IDs) to be related/connected. Those connections may be between any two vertices, and a path connecting a user ID and a service ID with a certain probability value of the path suggests a link between the service use and the user ID exists.

The graph reflects all knowledge of an attacker at a given time. Attackers with different knowledge will build different graphs for a system as will likely do the same attacker over some time.

What is not clear to us at the moment is the question whether pseudonyms should be treated differently from other contexts or not. Clearly they are more important in the model since their connection to users and actions defines level of pseudonymity achieved in the system. Yet at the moment we suggest all vertices to be treated equally, although we suspect that some of them might be more equal than others. :-)

5.1 OUTLINE OF THE GRAPH MODEL

We denote the set of all vertices by *V*, the set of all identifiers of service instances by *S*, and the set of all user IDs by *ID*. There are no edges between any pair of elements of *ID*, only indirect paths through a linking context, and the same applies to elements of *S*. There is also a function $W_{\max}$ calculating overall probability weight for a path in the graph, and therefore also a way to determine the highest value $W_{\max}(v_a, v_b)$ for a path between v_a and v_b. The value of any path is calculated as a multiplication of the weights (w) of all its individual edges, e.g. for the path $P = v_1, v_2, \ldots, v_i$ of i vertices of the graph, the value of the path P is $W(v_1, v_i) = w(v_1, v_2) \times w(v_2, v_3) \times \ldots w(v_{i-1}, v_i)$.

Unobservability (of service s_i) – a graph that *A* can build after observing a system at a given time does not include s_i at all.

Unlinkability (between two nodes v_1, v_2, at the level Δ) – a graph that *A* can build when observing the system at a given time has no path connecting v_1 with v_2 with the overall probability greater than Δ, i.e. provides $W(v_1, v_2) \le 1/|V| + \Delta$, where $v_1, v_2 \in V$.

Anonymity (of a user $u_{ID} \in ID$, at the level Δ) – then $\forall v \in V$, when *A*

1. knows the set *ID*, she can only find a path from v to u_{ID} with the weight not greater than $1/|ID| + \Delta$, such that $W_{\max}(v, u_{id}) \le 1/|ID| + \Delta$;
2. does not knot anything about *ID* (particular elements or size), she can only find a path from v to u_{ID} with the weight not greater than Δ, i.e. $W_{\max}(v, u_{id}) \le \Delta$.

Pseudonymity (of a subject/pseudonym $u \in P_{ID}$, at the level Δ) – there exists a path known to *A* from any $s \in S$ to u with a satisfactory value of $W_{\max}(s, u)$, but for *A* there is no knowledge of an edge from *u* to any $u_{ID} \in ID$ such that when *A*

1. knows *ID*, the path from *u* to any u_{ID} has weight not greater than $1/|ID| + \Delta$, i.e. $W_{\max}(u, u_{id}) \le 1/|ID| + \Delta$;
2. does not know anything about *ID* (particular elements or size), the path from *u* to u_{ID} has weight not greater than Δ, i.e. $W_{\max}(u, u_{id}) \le + \Delta$.

There are several proposals for formal framework for anonymity [8, 9] and unlinkability [16]. Frameworks introduced in these papers define typed systems with several defined categories like agents, type of agents, messages [9] or an inductive system based on modal logic of knowledge [8]. We believe that our proposal would be more flexible and would cover context information as an inherent part of the model thus opening interesting questions.

[1] Authors of this proposal work for different institutions located across the street.

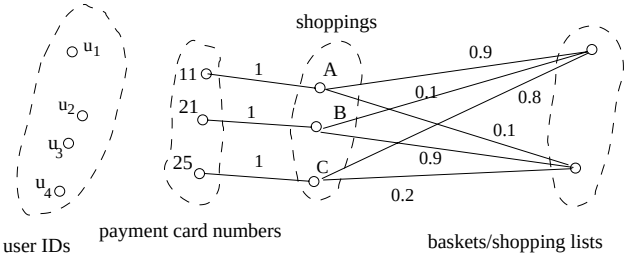

Fig. 6 – An example with a PATS model graph.

The example: Let us express our example from part 1.4 in the PATS model. Fig. 6 shows how the context information about typical basket contents is connected to actual instances of shoppings. As we are interested in connections between payment cards (pseudonyms), we are looking for paths (and their aggregate values) containing pairs of particular payment cards. Let us try to find paths between card 11 and the other two cards:

Path	Probabilities	Aggregate
$11 - A - l - B - 21$	1 * 0.9 * 0.1 * 1	0.09
$11 - A - l - C - 25$	1 * 0.9 * 0.8 * 1	0.72
$11 - A - m - B - 21$	1 * 0.1 * 0.9 * 1	0.09
$11 - A - m - C - 25$	1 * 0.1 * 0.2 * 1	0.02
…	…	…

Fig. 7 – Paths connecting payment card 11 with the other two cards.

These are the shortest (and highest value) paths only. The attacker may deduce (with a high probability) that payment cards 11 and 25 belong to the same person, though she does not know who this person is. According to our definitions, unlinkable pseudonymity is breached.

6. CONCLUSIONS AND OPEN ISSUES

This paper points out that contexts provide (or perhaps we can even say that they produce) side-channels that are not covered neither by the Common Criteria Privacy Class, nor by the Freiburg Privacy Diamond model. We also believe that contexts in general are not well reflected in other current research attempts to quantify the levels (and deterioration) of privacy. A simplistic introduction of pseudonyms will not guarantee perfect privacy, and we need to have some means to quantify what levels of privacy is needed and/or achievable for specific scenarios. There are two solutions for protection against side-channels: hiding and so-called anonymizing. Hiding is what anonymizing networks utilise – they combine number of messages together, thus creating satisfactory anonymity set. Anonymizing (or rather more often in practice pseudonymizing) requires creation of layers that cloak the identity of the protected entity. Common Criteria use this concept when defining pseudonymity that still enforces accountability of users, but hides/shades their identity.

One particularly interesting issue relates to the Common Criteria definition of unlinkability, as empirically reviewed by Rannenberg and Iachello [13, 14] and more formally specified above in section 2.4, is whether the unlinkable "items" in question should only be operations (service invocations) or whether other kinds of unlinkability should also be considered. We have provided a supporting evidence for a substantial revision of unlinkability specifications, while leaving the actual revision as an item for the future research.

We also provide our basic PATS model that is not so limited in the coverage of selected aspects of user interactions and therefore allows for better quantification/measurement of different aspects of privacy. This proposal, unlike the CC or FPD models, introduces a computational model (based on graph theory). One of the problems we are currently examining is atomicity for the vertices, i.e. contextual information. We currently review various approaches to this problem, being aware that the issue of atomicity has a critical impact on the possibility of graph normalisation and therefore also for the provision of the critical properties of completeness and soundness. This work in progress includes the issue of edge dependence, for it is clear that the edges are not completely independent. We can mark sets of nodes from distinct kinds of context (e.g., pseudonyms, IP addresses used in connections from the same provider) – let us call them *domains*. Then we can address additional graph properties, e.g., such that for all pairs of domains D_1, D_2 all sums of probabilities from any node in D_1 to all nodes in D_2 are not higher then a given value, typically 1.

The PATS approach allows for two definitions of anonymity, a weaker one considering a weight of the entire path from $u_{ID} \in ID$ to s_i can be added to the stronger one above that considers the intermediate edges from u_{ID} only (to any other vertex – contextual information – that would then be identifiable).

Another interesting issue is the role of time that has a two-fold role – firstly, it can be a contextual information (time of an action invoked by a certain subject, i.e. three mutually connected vertices). Secondly, the probabilistic weights of edges in a graph change with time, as do the sets of vertices

and edges as such. Obviously, the contextual role of time may be reflected by the latter view – time of an action invoked by a certain subject is denoted by existence of vertices describing action and subject identifiers, connected by an edge with weight 1, at the given time.

7. ACKNOWLEDGEMENTS

Thanks go to Andrei Serjantov and Alf Zugenmaier for their opinions and links to some interesting references in the area of privacy, to Flaminia Luccio and others for valuable discussions of the PATS graph model details.

8. REFERENCES

[1] A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *Hawaii International Conference on System Sciences 33*, pages 1769–1777. ACM, 2000.

[2] J. Bacon, K. Moody, J. Bates, R. Hayton, C. Ma, A. McNeil, O. Seidel, and M. Spiteri. Generic support for distributed applications. *IEEE Computer,* pages 68–76, March 2000.

[3] M. Bellare. A note on negligible functions. Technical Report CS97-529, Department of Computer Science and Engineering, UCSD, 1997.

[4] V. Vahill et al. Using trust for secure collaboration in uncertain environments. *IEEE Pervasive Computing Magazine*, 2003 (July-September):52–61.

[5] D. Cvrček and V. Matyáš. Pseudonymity in the light of evidence-based trust. In *Proc. of the 12th Workshop on Security Protocols*, LNCS (forthcoming), Cambridge, UK, April 2004. Springer-Verlag.

[6] C. Diaz, S. Seys, J. Claessend, and B. Preneel. Towards measuring anonymity. In R. Dingledine and P. Syverson, editors, *Proceeding of Privacy Enhancing Technologies Workshop* (PET 2002), LNSC 2482. Springer-Verlag, April 2002.

[7] J. Douceur. The Sybil attack. In *1st International Workshop on Peer-to-Peer Systems* (IPTPS`02), LNCS 2429, pages 251–260. Springer-Verlag, 2002.

[8] J. Y. Halpern and K. O`nelil. Anonymity and information hiding in multiagent systems. In *Proceedings of the 16th IEEE Computer Security Foundations Workshop*, pages 75–88, 2003.

[9] D. Hughes and V. Shmatikov. Information hiding, anonymity and privacy: A modular approach. *Journal of Computer Security, special issue on selected papers of WITS 2002*, 12(1):3–36, 2004.

[10] D. Kesdogan, D. Agrawal, and S. Penz. Limits of anonymity in open environments. In F. Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002),* LNCS 2578. Springer-Verlag, October 2002.

[11] M. Kinateder and S. Pearson. A privacy-enhanced peer-to-peer reputation system. In *Proceedings of the 4th International Conference on Electronic Commerce and Web Technologies, EC-Web 2003*, LNCS 2738, pages 206–215, Prague, Czech Republic, September 2003. Springer-Verlag.

[12] A. Pfitzmann and M. Köhntopp. Anonymity, unobservability and pseudonymity – a proposal for terminology. In *Designing Privacy Enhancing Technologies: Proceedings of the International Workshop on the Design Issues in Anonymity and Observability*, LNCS 2009, pages 1–9. Springer-Verlag, 2000.

[13] K. Rannenberg and G. Iachello. Protection profiles for remailer mixes. In *International workshop on Designing privacy enhancing technologies: design issues in anonymity and unobservability*, LNCS 2009, pages 181–230, Berkley, California, 2002. Springer-Verlag.

[14] K. Rannenberg and G. Iachello. Protection profiles for remailer mixes – do the new evaluation criteria help? In *16th Annual Computer Security Applications Conference (AC-SAC`00)*, pages 107–118. IEEE, December 2000.

[15] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In *Privacy Enhancing Technologies (PET)*, LNCS 2482, pages 41–53. Springer-Verlag, April 2002.

[16] S. Steinbrecher and S. Köpsell. Modelling unlinkability. In R. Dingledine, editor, *Privacy Enhancing Technologies (PET),* LNCS 2760, pages 32–47. Springer-Verlag, 2003.

[17] The Common Criteria Project Sponsoring Organisations. *Common Criteria for Information Technology Security Evaluation – part 2, version 2.1.* August 1999.

[18] A. Zugenmaier. *Anonymity for Users of Mobile Devices through Location Addressing.* RHOMBOS-Verlag, ISBN 3-930894-96-3, Berlin, 2003.

[19] A. Zugenmaier, M. Kreutzer, and G. Müller. The Freiburg Privacy Diamond: An attacker model for a mobile computing environment. In *Kommunikation in Verteilten Systemen (KiVS) `03*, Leipzig, 2003.

Ing. Daniel Cvrček, Ph.D. *is currently employed by Brno University of Technology and Masaryk University Brno. He is doing research in information system security and applied cryptography. The current interests cover reputation systems, privacy issues, key management schemes, sensor networks, and hardware security. He finished his PhD. study in 2001 and was a research associate at Cambridge University, UK in 2003-2004. During the years, he was also involved in Czech Digital Signature Law, or implementation of the first Czech certification authority, or research projects for Czech National Security Agency. He is also taking an active role in establishing an independent university security laboratory – BUSLab.*

Dr. Václav (Vashek) Matyáš Jr. *is an Associate Professor at the Masaryk University Brno, Czech Republic. His research interests relate to applied cryptography and security, with over forty peer-reviewed publications, including two books. He also worked as a Visiting Researcher with Microsoft Research Cambridge, Visiting Lecturer with University College Dublin, as a co-founder and CEO of ecom-monitor.com, was an Associate Director with Ubilab, UBS AG; Royal Society Postdoctoral Fellow with the Cambridge University Computer Lab; and Director, Technology and Security, of a London-based CA Uptime Commerce. Vashek is an Editorial Board member of Data Security Management (Czech security journal), edited the Computer and Communications Security Reviews, and also worked on the development of Common Criteria and with ISO/IEC JTC1 SC27.*

Mgr. Marek Kumpošt *is an Ph.D. student at the Masaryk University Brno, Faculty of Informatics, Czech Republic. He is interested in computer security and computer networks. He is a member of BUSLab group www.buslab.cz. The areas of his Ph.D. research are reputation systems, trust building and privacy protection. He worked as a security advisor with AEC Ltd. in Czech Republic.*