Information Disclosure in Identity Management

all of us

Abstract

User Controlled Identity Management Systems have the goal to hinder the linkability between the different digital identities of a user. We perform a theoretical and an experimental study of the following information leakage problem: given a consistent view on the actions of a strong identity management system (e.g. *Idemix*) where k users pseudonymously issue and show some credentials, do these observations disclose sufficient information about the linkability of the digital identities?

We show that in theory, linking the different pseudonyms of a user is a NP-complete problem, by using first order logic. In addition, we evaluate practical instances of the problem, and show that there are non-negligible probabilities that despite full anonymization and use of an identity management system, pseudonyms are unambiguous linkable.

1 Introduction

A digital identity is a collection of information about a person that an organization may have for various purposes, e.g. marketing. It usually consists of a set of personal attributes, like hobbies, education, favorite books, combined with an identifier. This identifier could be the name of the person, a nickname, an email address, cookies etc.

Obtaining user data to build a digital identity is not always in line with the users. One of the most prominent demonstration of unwanted provisioning of digital identities is the overall increased amount of daily email Spam. Thus, it is a nearby idea to empower the user with a software, so that the user can manage her identities, the so called *Identity Management System*. The goal of such an identity management system for users interacting with organizations is to minimize the personal information without loosing functionality. Thus, the security requirements on such an Identity Management System are twofold. First, it has to hinder unwanted data aggregation by using data minimization, anonymity and pseudonymity techniques and second it has to ensure that the transferred information is accountable. The latter one is mainly important for the organizations in context of authentication, authorization and accounting (AAA) but also to offer a better service.

To comply with both requirements (i.e. anonymity and accountability) several cryptographic protocols (known as *anonymous credentials*) have been suggested [2]. Actually, anonymous credentials are the core part of security and privacy enhanced identity management systems. One of the most known example is the identity management system of the European project PRIME using the Idemix anonymous credentials.

However, designing a secure identity management system is a complex task, even though anonymous credentials are proven to be cryptographically secure. Indeed as for a lot of real life systems, the used cryptographic elements are the strongest part of it, but do not cover all aspects of Identity Management Systems. For instance, they do not cover the existence of information

prior to the application of the Identity Management System on one side or information about the person and their interaction on the other side. Hence, there is a need for extending the core model to get a deeper insight in identity management systems.

In this work we want to use the model of pseudonymous credentials with a focus on information flow that might be gained by observing the interactions of the participants. In particular, we are interested in the following information leakage problem: given a consistent view on the actions of a strong identity management system (e.g. *Idemix*) where k users pseudonymously issue and show some credentials, do these observations disclose sufficient information about the linkability of the digital identities?

There are only few publications considering the information flow of a system by observing the *interactions* of the system, as we will show in the next section. One of the recent suggestions is the consistent view attack. We follow these works and give a new theoretical description of the linkability problem. Thus, we are able to measure the information flow within the given system model. In particular we prove conditions when the system becomes insecure and show the impact of the observations on unlinkability. As a demonstration and extension of our theoretical work we provide experiments by the means of simulation. Here, we are able to show the effects of some parameters of the system determining the security of the whole system, and even demonstrate that in some situations the system fails to provide any protection.

We consider the same system as in Idemix [2], with two competing groups of participants: users and organizations. Both groups are acting within a given Identity Management solution, and we assume that all organizations are cooperating and form together the *attacker*. To avoid side channel information, we assume that all communication takes places via anonymizing network layers, thus efficiently blending out any other information, apart from the users' pseudonyms and the credential. Thus, there are also only two types of interactions: users can ask for a certain credential to be issued for one of their pseudonyms, or they can proof the possession of a credential. Note that this setup is identical with the assumptions in Idemix, thereby making our work directly applicable to the security of an Idemix system.

Our work shows that even given these strong assumptions, i.e. use of anonymous communication, unlinkable pseudonyms and strong cryptographic credentials, there is still considerable information leakage; sometimes it is possible to unambiguously link pseudonyms. Given that most the assumptions are too strong with respect to the real world, the amount of leaked information would be even larger in reality - thus leaving the users basically unprotected to attacking organizations despite the use of Identity Management Systems.

However, we will first precisely describe the discrete mathematical structure behind basic identity management systems and our model in the next steps, before the practical evaluation.

2   Related Works

A digital identity management should fulfill anonymity or pseudonymity, as well as accountability. On the one hand users should interact mostly anonymously with organizations and should be able to control the nature and amount of personal information disclosed to the organizations. On the other hand the organizations should be able to control the authenticity of the given information and in case of misuse the corresponding user should be identified (accountability requirement) [15].

The privacy-research community has suggested a number of cryptographic protocols that ba-

sically provide an information exchange between users and organizations with the desired properties [3, 4, 5, 7, 11, 2]. The system model consists of two major groups of players, the users and the organizations. Organizations know the users only by their pseudonyms (digital identifier). Different pseudonyms of the same user cannot be linked. An organization can issue a credential (digital certificate about the person) to a pseudonym. Users can later on prove possession of these credentials to other organizations by using different pseudonyms. Hence organizations can cryptographically check the possession of credentials (the links between pseudonyms and credentials) without revealing anything more than these facts.

However, Pashalidis and Meyer have introduced in [12] the idea of a so called *consistent view* attack: if the organizations accumulate all traffic[1] between them and the users, then it might be possible that the pseudonyms in the described system become linkable. In general they state the *linking problem* as following: given an anonymous credential system with k users and a consistent traffic observation ab initio, the task is to link uniquely all observed pseudonyms to k subsets representing the k users. Pashalidis and Meyer show in their work that the linking problem is a NP complete problem by reducing SAT to the linking problem. However they do not provide an algorithm to solve the consistent view problem itself.

Our work is highly related to two different areas of work: unlinkability theory and disclosure attacks.

Unlinkability Theory as given in [15] is more a framework about unlinkability and does not focus on evaluating unlinkability by determining the information flow of a given system. It rather describes and defines the unlinkability property: Let $P = \{p_1, \ldots, p_n\}$ be the set of items within a given system. For someone with full knowledge of the system some items of this set are related while others are not. We consider a notion of "is related" that forms an equivalence relation $\sim_{r(P)}$ on the set P. By this relation P is split in k ($1 \leq k \leq n$) pairwise disjoint equivalence classes $P_1, \ldots, P_k$. [15] defines unlinkability in probabilistic terms of items: For a random variable X let $Prob(p_i \sim_{r(P)} p_j) := Prob(X = (p_i \sim_{r(P)} p_j))$ denote the attacker's a posteriori probability that given two items p_i and p_j, X takes the value $(p_i \sim_{r(P)} p_j)$ (compare this also with [13]).

There are a number of works in the area of unlinkability with a different focus (for a general overview see [6]). They focus on the information leakage due to the given attributes (credentials), i.e. a number of personal attributes may identify a user in a given population or decrease the number of suspected persons (i.e. the anonymity set). Another track of attacks on unlinkability system are active attacks, see [6].

The second area of work investigates the evaluation of anonymity, which is highly related to unlinkability (see [13]). Disclosure attacks have many similarities to the consistent view attack as mentioned also by Pashalidis and Meyer but have to be extended and applied in the unlinkability area (compare also [12]).

Disclosure Attacks as e.g. in [10] evaluate the fundamental information gain that a passive attacker can obtain by traffic analysis in the area of anonymity [9]. These attacks have the focus on the information leakage (inference) inherent in the anonymity systems. They model the information leakage and its accumulation over time mostly in terms of probabilities (stochastic models) but also in information theoretic setting.

As in all disclosure attacks, we also assume that the attacker knows all available information in the system in order by draw conclusions. We use this information theoretical approach in

---

[1]The traffic consists of the complete issuing and showing process of credentials with different pseudonyms.

order to show basic security limits and underlying mathematical structures. As we will show in the next section, even the limited information available in systems with strong credential systems is sufficient to profile users under certain conditions.

3 Theoretical Evaluation

In this section we want to show, that it is possible to link the pseudonyms of the users of an Idemix system by passive observation of show and issuing events. These events appear, if a user issues or shows a credential to an organization according to the protocol of the Idemix system. The Idemix protocol even allows us to derive rules about sequences of events, so that the space of possible linkings between pseudonyms can be restricted. These rules will be defined in section 3.1. Thus this section will prove that Idemix is information theoretic insecure.

Basing on these rules and the observations, section 3.2 will introduce a formal description of the solution space in first order logic (FOL). The solution space describes all possible linkings of pseudonyms, which can be logically derived from the given information of the system. We will also contribute an algorithm to compute this FOL description.

Finally section 3.3 will prove that finding a unique linking between the pseudonyms of the users is NP-complete, by reducing the minimal vertex coloring problem to the Idemix game.

3.1 Our Model - The Idemix Game

Our Model is the same as the one used in [2]: it consists of two sets of participants, a set of k users and a set of organizations. Since we assume the organizations to be cooperating, we consider them as a single entity: the adversary. In this model the users have serial interactions with the adversary, where all events will be logged. Since all of the users' communication takes place over untraceable network layers, we assume that the adversary gains no additional information[2].

We model the system as the *Idemix game*: The adversary has m different types of credentials to issue. At the beginning of the game, no user is in possession of a credential. These can only be obtained from the adversary, where issuing a credential will get logged. If a user has obtained a credential (or several credentials) then she will eventually show this or a subset of the received credentials; as all organizations comprise the attacker, the show event will get logged, too. In order to issue or to show a credential, a user has to show one of her pseudonyms. The pseudonym sets Q_i for $i \in \{1, \ldots, k\}$ of the users are pairwise disjoint. We denote the set of all pseudonyms by the variable $P := \bigcup_{i:=1}^{k} Q_i$.

Following *basic rules* have to be followed by the game considered in this paper:

Rule 1 If the user has not obtained the credential from the adversary she can not show it.

Rule 2 If a user has issued a credential, she can not share it with other users.

Rule 3 Different users can get issued the same credential.

Rule 4 The users stay pseudonymous all the time, i.e. the adversary does not know by trivial means to whom a credential has been issued and neither who is showing a credential.

Rule 5 The user can show the same credential several times to the adversary.

Rule 6 No user is allowed to get the same credential issued more than once.

---

[2]We acknowledge that this assumption is not given in the real world: however we want to show that even in the absence of more information, the attacker is able to learn a lot of information.

The last rule corresponds to the real world implementation of the Idemix system, since it is not designed to allow the issuing of the same credential to the same user several times. It should also be noted, that the rules for the consistent view attack are the same apart from the absence of rule 6. This little difference in the Idemix game has a high impact on the structure of the problem, as it provides information which induce relations to graph theoretic problems. Rule 6 also effects that the results of the consistent view attack can not be devolved to the Idemix game. For example the NP proof strategy of the consistent view problem (by using the reduction from SAT) can not be applied to the Idemix game.

It is the task of the attacker, to unambiguously group all pseudonyms into k partitions for k users, given the log files from all transactions. If the attacker is able to do this, then he wins the game.

3.1.1 Observations of the Adversary

In order to define the structure of the communication events, which an adversary records, some terms are introduced in this section. At first we define the domain of the pseudonym set P, the credential set C and the event set E.

$$P := \{p_1, \ldots, p_n\}, \text{ where } p_i \text{ are pseudonyms}$$
$$C := \{1, \ldots, m\}, \text{ where each number is a distinct credential}$$
$$E := \{issue, show\}, \text{ where } issue, show \text{ are events}$$

An *observation* O is a three-tuple $(P \times C \times E)$. The *history* $\mathcal{H} := (O_1, \ldots, O_t)$ is a chronologically ordered tuple of observations, which the adversary collects from the beginning up to time t.

Example 3.1 *Let us consider an example for a history $\mathcal{H}$, which is supposed to comply to the basic rules. The adversary is assumed to collect observations in the following chronological order with a time line increasing from left to right:*

$$\mathcal{H} := ((p_1, 1, issue), (p_2, 1, show), (p_3, 2, issue), (p_4, 2, show), (p_5, 1, show),$$
$$(p_5, 2, show), (p_6, 3, issue), (p_7, 3, issue), (p_8, 1, show), (p_8, 3, show))$$

In this example, the pseudonym set P consists of the eight pseudonyms $\{p_1, \ldots, p_8\}$ and the credentials are $C := \{1, 2, 3\}$.

The first observed event is a user with the pseudonym p_1, who issues the credential 1. Then a user with another pseudonym p_2 shows the credential 1. Finally the last two observations register a user with the pseudonym p_8, who shows the credential 1 and 3.

The task of the adversary is to extract knowledge about the links between the pseudonyms from the given history with respect to the game definition. For example the pseudonym p_1 and p_2 must belong to the same user. The reason for this conclusion is rule 1. When considering the first two observations of $\mathcal{H}$, it can be noticed that p_1 is the only pseudonym, which has issued the credential 1 before the event of showing credential 1 in the second observation. Therefore, the user, who shows the credential 1 with the pseudonym p_2 must be the same one, who has issued the credential 1 with p_1 before.

5

3.2 Computing Solutions

This section defines a strategy to extract a FOL term from the given history and rules, which describes all solutions of the Idemix game. Formally a *solution* $\mathcal{S} := \{Q_1, \ldots, Q_i\}$ is a partition of the pseudonyms P participating in the show and issuing events with respect to the rules 1 to 6. We call each set $Q_j \in \mathcal{S}$ a *segment* of the partition $\mathcal{S}$, where $Q_j \subseteq P$. Each segment has the meaning, that all pseudonyms withing this segment belongs to the same user. A partition, which does not respect the history or the rule set is no solution.

3.2.1 Sub-history

At the first step, a more concise representation of the history is introduced. The idea is to separate $\mathcal{H}$ into sub history files with respect to the credential types. This splitting does not change the solution space of the problem. The credentials in $\mathcal{H}$ are 1,2,3, thus the resulting sub-histories are:

$$\mathcal{H}_1 := ((p_1, 1, issue), (p_2, 1, show), (p_5, 1, show), (p_8, 1, show))$$
$$\mathcal{H}_2 := ((p_3, 2, issue), (p_4, 2, show), (p_5, 2, show))$$
$$\mathcal{H}_3 := ((p_6, 3, issue), (p_7, 3, issue), (p_8, 3, show))$$

Note that the rules only enable the conclusion of links between pseudonyms, which issue or show the same credential. In particular rule 1 allows the derivation of pseudonym relations by their order in the issue/show events of the same credential. Therefore the *sub-histories* of $\mathcal{H}$ are defined to be the unique separation of the observations in the history by their credential types. Thereby each *sub-history* is a chronologically ordered[3] tuple containing all observations of a particular credential.

3.2.2 Solution of Sub-histories

A partition of the pseudonym set P must be consistent with the history $\mathcal{H}$ and thus consistent to each of the sub-histories $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$. Thereby a partition is called *consistent*, if it can be logically derived from the given history and the rule set. Pseudonyms p_i, p_j, which belong to the same segment of a partition are in the same equivalence class [4]. This equivalence relation is denoted by $p_i \sim p_j$, whereat $\sim$ is symmetric, reflexive and transitive. The equivalence class, to which a pseudonym p_i belongs is addressed by $[p_i]$ and the equation $[p_i] = [p_j]$ holds, if and only if $p_i \sim p_j$ holds. Using the rules, we can derive the following relation between the pseudonyms in each sub-history:

$$\begin{aligned}
\mathcal{H}_1 : \quad & (p_1 \sim p_2) \wedge (p_1 \sim p_5) \wedge (p_1 \sim p_8) && \Leftrightarrow ([p_1] = [p_2]) \wedge ([p_1] = [p_5]) \wedge ([p_1] = [p_8]) \\
\mathcal{H}_2 : \quad & (p_3 \sim p_4) \wedge (p_3 \sim p_5) && \Leftrightarrow ([p_3] = [p_4]) \wedge ([p_3] = [p_5]) \\
\mathcal{H}_3 : \quad & (p_6 \sim p_8 \vee p_7 \sim p_8) \wedge (p_6 \not\sim p_7) && \Leftrightarrow ([p_6] = [p_8] \vee [p_7] = [p_8]) \wedge ([p_6] \neq [p_7])
\end{aligned}$$

Note that rule 6 provides no gain of information in $\mathcal{H}_1$ and $\mathcal{H}_2$, as there is only one issue event in each of these cases. In contrast to this, two pseudonyms p_6 and p_7 issue the credential 3 in $\mathcal{H}_3$. Hence due to rule 6, the term $(p_6 \not\sim p_7)$ is added to expresses that these pseudonyms can not belong to the same equivalence class.

---

[3]The relative order of the observations in the sub-history is given by the order of the observations in $\mathcal{H}$.

[4]We use the term equivalence class and segment interchangeably

By the transitivity of $\sim$, it can be concluded, that the pseudonyms in $\mathcal{H}_1$ and $\mathcal{H}_2$ are in the same equivalence class. Let us call this class Q_1, where $p_1, p_2, p_3, p_4, p_5, p_8 \in Q_1$. The assertion of $\mathcal{H}_3$ requires, that either $[p_6] = Q_1$ in the case that $p_6 \sim p_8$, or that $[p_7] = Q_1$ in the case that $p_7 \sim p_8$. Observe that the "or" here is an exclusive or, as p_6 and p_7 cannot be in the same equivalence class by the term $(p_6 \not\sim p_7)$. Thus all possible solution of the history are:

$$\mathcal{S}_1 := \{\{p_1, p_2, p_3, p_4, p_5, p_6, p_8\}, \{p_7\}\} \qquad \mathcal{S}_2 := \{\{p_1, p_2, p_3, p_4, p_5, p_7, p_8\}, \{p_6\}\}$$

In this example, the sub-histories are also alternatively described by *first order logic* (FOL) terms by considering the relation between equivalence classes $[p_i]$ instead of the pseudonyms p_i. If the pseudonym classes $[p_i]$ are treated as variable names, which can be assigned an identifier corresponding to a segment Q_j, then the solutions of $\mathcal{H}$ are all variable assignments, so that $\mathcal{H}_1, \mathcal{H}_2$ and $\mathcal{H}_3$ are *true*. Those variable assignment are said to *satisfy* the FOL term, which describes $\mathcal{H}$. For example let the identifier of the first segment of $\mathcal{S}_1$ be 1 and of the second segment be 2. The assignment of all pseudonym classes $[p_1], [p_2], [p_3], [p_4], [p_5], [p_6], [p_8]$ with the number 1 and $[p_7]$ with the number 2, result in *true* for all FOL terms of the sub-histories. Thus the FOL terms in this example are a formal representation of the solution space of the history. Section 3.2.3 provides an algorithm to compute the FOL representation of a history.

3.2.3   Computing the CNF of a History

The description of the solutions by the relation $\sim$ and the logical connections $\wedge$ and $\vee$ can be easily transformed into an equivalent *conjunctive normal form* CNF expression as shown in section 3.2.2.

For simplicity, we will omit for the rest of the paper the equivalence class operation $[\cdot]$. Hence p_i now stands for the equivalence class $[p_i]$, unless it is otherwise stated. The following algorithm 1 computes the CNF representation of a given history $\mathcal{H}$.

For each of the n credentials $i \in \{1, \dots, n\}$, the algorithm repeats the following three phases:

1. Extract the sub-history $\mathcal{H}_i$ of credential i.

2. Represent the equivalence relations in $\mathcal{H}_i$ (rule 1, 2) by clauses cl.

3. Determine the non equivalence relations in $\mathcal{H}_i$ (rule 6).

The computation of the sub-history in phase 1 is not explicitly represented by the algorithm, as it was already treated in section 3.2.1. Phase 2, which are lines 4 to 10 determines the CNF cnf_i of the sub-history $\mathcal{H}_i$ with respect to the rules 1 to 5. Phase 3 corresponds to line 12, which applies the new rule 6 and adds the constrains that pseudonyms, which issue the same credential must belong to disjoint equivalence classes. This is expressed by the conjunction:

$$\bigwedge_{(p_{i_u}, i, issue), (p_{i_v}, i, issue) \text{ in } \mathcal{H}_i, u \neq v} (p_{i_u} \neq p_{i_v})$$

Note that the *cnf* obtained from the history $\mathcal{H}$ contains information about the equivalence and non equivalence relations between the pseudonyms, but it does not cover pseudonyms, for which no relation information exists. These are pseudonyms, which do not appear in the scope of any show events. In order to incorporate this, line 14 covers by the set P_{assign} all pseudonyms

within the scope of show events, i.e. all pseudonyms, which are in *cnf*. These pseudonyms are also called *kernel-pseudonyms*. Thus the set

$$P_{free} := P \setminus P_{assign}$$

embodies those pseudonyms, which can be assigned to any equivalence classes, i.e. those classes resulting from *cnf*, or new classes. Thereby P is the set of all pseudonyms appearing in the history given to the algorithm. Hence the solutions of $\mathcal{H}$ are the equivalence classes of *cnf* combined with P_{free}. We call a partition derived from *cnf* a *kernel-partition*, or a *kernel-solution*. Thus kernel-partitions refer to partitions of P_{assign}, while the plain term solution or partition (without the term kernel) references a partition of P. The set of all *kernel-partitions* constitute the *kernel* of *cnf*. If $P_{free} = \emptyset$, then the kernel of *cnf* is also the set of all solutions of the history $\mathcal{H}$.

Algorithm 1 HistoryToFOL

 1: **procedure** HISTORYTOFOL($\mathcal{H}$)
 2:      $cnf := true$
 3:      **for** $i := 1$ to n **do**
 4:          $cnf_i := true$
 5:          **for** $j := $ (first appear. of *show* in $\mathcal{H}_i$) to (last appear. in $\mathcal{H}_i$) **do**
 6:              $cl := false$
 7:              **for** each observation $(p_{i_l}, i, issue)$ in $\mathcal{H}_i$ with index $i_l < j$ **do**
 8:                  $cl := cl \vee (p_{i_l} = p_{i_j})$
 9:              **end for**
10:              $cnf_i := cnf_i \wedge cl$
11:          **end for**
12:          $cnf := cnf \wedge cnf_i \bigwedge\limits_{(p_{i_u},i,issue),(p_{i_v},i,issue) \text{ in } \mathcal{H}_i, u \neq v} (p_{i_u} \neq p_{i_v})$
13:      **end for**
14:      $P_{assign} := \{p_i \mid p_i \text{ in } cnf\}$
15:      **return** cnf, P_{assign}
16: **end procedure**

An example for the case of $P_{free} \neq \emptyset$ is given by a history consisting of the following two sub-histories:

$$\mathcal{H}_1 := ((p_1, 1, issue), (p_2, 1, show)) \qquad \text{and} \qquad \mathcal{H}_2 := ((p_3, 2, issue)).$$

Algorithm 1 would return $cnf := (p_1 = p_2)$ and $P_{assign} := \{p_1, p_2\}$, which implies $P_{free} := \{p_3\}$. Thus the kernel consist of exactly one kernel-partition $\mathcal{S} := \{\{p_1, p_2\}\}$. The solutions resulting from the combination of the kernel-solution $\mathcal{S}$ with P_{free} are $\mathcal{S}_1 := \{\{p_1, p_2, p_3\}\}$ and $\mathcal{S}_2 := \{\{p_1, p_2\}, \{p_3\}\}$, which are also all solutions of the history.

Note that the CNF obtained by algorithm 1 is a formal description of the solution space, where the solutions are not always obvious to see. But we can use this CNF to generate particular solutions in two ways. The first way is to transform the CNF into a disjunctive normal form (DNF). In this case each of the disjunctive clauses of the DNF is a solution of the Idemix game.

The second option is to use the CNF as a verifier of a hypothesis (i.e. a partition of pseudonyms). This means that we compute a partition and then check wether the variable

assignment associated with the partition satisfies the CNF. If the CNF is satisfied by the partition, then the partition is a solution, otherwise it is not a solution. A partition of the pseudonym set P can be computed by the sterling formula of second kind.

If there is only a single, and thus unique, possibility to partition the pseudonyms in k sets, that also satisfies the CNF, then this solution equals the true situation. If the attacker has found this unique solution, and can ensure that it is the only one, he has identified the true links between the pseudonyms and wins the game.

3.3 NP-completeness of Idemix Game

The computation of the solution space of the Idemix game can be proven to be NP-complete. We will present an efficient reduction from minimal vertex coloring to the Idemix game to show that this game is at least as hard as vertex coloring. It is well known, that the minimal vertex coloring problem is NP-complete [1], thus Idemix game is at least NP-complete. In order to prove that the game is not harder than NP-complete, we will show that it belongs to the polynomial time verifyable problems, which are equivalent to the class of NP problems.

3.3.1 Reduction of Vertex Coloring to Idemix Game

The minimal vertex coloring problem is the problem of finding the least number of colors to color the vertexes of an undirected simple graph, so that adjacent vertexes are assigned distinct colors. Let $G = (V, E)$ be an undirected graph, where $V = \{p_1, \ldots, p_n\}$ is the set of vertexes and $E \subseteq V \times V$ is the set of edges. For simplicity we assume w.l.o.g. that there is no vertex in G, which is not adjacent to another vertex, as those vertexes can be colored arbitrary.

For each edge $e_i \in E$, where $e_i = \{p_{i_1}, p_{i_2}\}$, we construct the following observation of the Idemix game:

$$(p_{i_1}, issue, i), (p_{i_2}, issue, i) \quad \Longleftrightarrow \quad [p_{i_1}] \neq [p_{i_2}] \tag{1}$$

Let $\mathcal{H}$ be the history consisting of all observations constructed by equation (1), then each solution of the Idemix game on $\mathcal{H}$ is a partition of V, so that no adjacent vertexes belong to the same segment. If we assign a distinct color to each segment of a solution $\mathcal{S}$, then $\mathcal{S}$ becomes a vertex coloring of G and particularly the solution with the least number of segments is a minimal vertex coloring of G.

3.3.2 Polynomial Time Verifiability

It was shown in section 3.2.2 that we can represent the solution space of the Idemix game as an FOL term consisting of equalities and inequalities of pseudonym classes, where the pseudonym classes can be interpreted as variables. A partition of the pseudonyms can be considered as a variable assignment of the pseudonym classes. Thereby each segment of the partition is associated with a distinct segment number and all pseudonym classes equal to the same segment are to be assigned to the segment number. If the logical value of the FOL term after the described variable assignment is *true*, then the partition is a solution of the Idemix game, otherwise it is no solution. This process of verification can obviously be done in polynomial time.
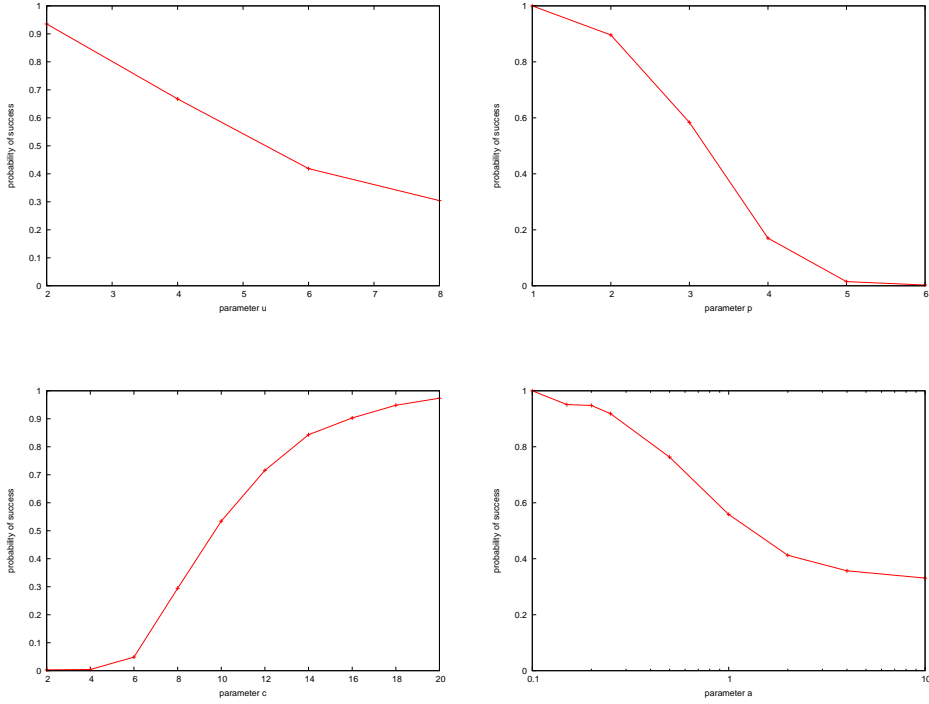
Figure 1: Experimental Results: Likelyhood of successful attacks.

4 Experimental Results

In order to estimate realistic values for winning the games as described in the previous sections, we have written a simulation of the system in order to create sets of observations and test the feasibility of attacking the proposed scheme.

We took a fixed set of users, with a varying number of pseudonyms, and generated a fixed amount of observations that an attacker would make in this setting. The users started each with an empty set of credentials. At each time a random user was uniformly chosen out of the set of users. For this user the relative amount of credentials he has issued $r = \frac{\text{no. of user's credentials}}{\text{all credential types}}$ was determined. The probability that the user considered to take an *show*-action was equally to $p_{show} = r^\alpha$ for a fixed value α, and the probability for an *issue*-action $p_{issue} = 1 - p_{show}$. All actions were done with a randomly chosen pseudonym out of the set of a user's pseudonyms.

With the parameter α we could vary the users' behavior: values $\alpha < 1$ simulated users that acquired only a small set of credential types in the first few rounds and then continued to show them repeatedly. Values for $\alpha > 1$ simulated users that acquired credentials before showing them.

For our experimental results we used as a default setup $u = 5$ users, $p = 3$ pseudonyms per user, $c = 10$ credential types, $o = 1000$ observations, and $\alpha = 1$. To show the impact of the single variables we kept all but one of these five parameters fix and varied the fifth along an interval. The results are depicted in the figure 1. Note that we left out the plot for o due to

10

space constraints and the fact, that for a given setting, the probability of a successful attack converges against a fixed value between 0 and 1 for $\lim o \to \infty$.

All plots show the probability of a successful attack, i.e. the probability that an attacker can uniquely link all the pseudonyms. As can be clearly seen in the first two plots, the likelihood of a successful attack reduces drastically with the number of users and the number of pseudonyms per user. However, the third plot shows that the probability of success is raised by a multiplicity of credentials.

The fourth plot shows the impact of different user behavior α: if the users fetch a small amount of credential types and then continue with show-actions, it is clear that they are more likely to be linked. The other way around, we see that it is positive for a user's privacy to first accumulate as much credentials as possible, before actually using them.

5 Discussion and Conclusion

In this work we have evaluated the issue-show information leakage problem in an information theoretic setting. We presented an algorithm to describe this information in FOL. Our description approach is, with slight modifications, also applicable to the consistent view problem [12], but we did not show it in this paper due to space limitations. It could also be shown, that the Idemix game is also NP-complete by reducing the vertex coloring problem on it. This reduction shows that Idemix game is related to graph theory and we expect to derive more assertions about the structures covered in this game by applying results from the graph theory.

We have investigated a disclosure attack on the most prominent anonymous credential system, Idemix [2]: if the adversary knows the number of participating users k, then the Idemix system is insecure[5]. We have identified precise conditions, that the adversary can use to check his success. Moreover, we have done some experiments to show the information flow in such a setting. The experiments show, that under certain circumstances, Idemix provides no protection, but also that the users can influence the degree of anonymity by their behavior.

Information theoretic settings as introduced by Claude Shannon (see for the unicity distance approach [14] and [8]) and used here in our work are interesting if someone is interested in what is maximally possible. Moreover, it discloses interesting discrete structures inherent to the system. In general, in our work we have identified that the anonymous credentials are insecure, if the identified conditions are valid. Fortunately, anonymous credentials are only information theoretically insecure. Does this mean that the anonymous credentials are as secure as the used cryptography? This is an open problem and future work.

Moreover, in the future, we will extend our models by the insight from our theoretical work and evaluate it for more realistic settings. In the future we will not only extend the evaluation setting of our work but also the adversary model from passive to active.

References

[1] R. L. Brooks. On Colouring the Nodes of a Network. volume 37 of *Cambridge Philos. Soc.*, pages 194 – 197, 1941.

---

[5]Note that the system will fail even without the knowledge of k with a smaller probability. Proof is missing due to space limitations.

[2] J. Camenisch and A. Lysyanskaya. An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation. In *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques (EUROCRYPT '01)*, pages 93 – 118, London, UK, 2001. Springer-Verlag.

[3] D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Commun. ACM*, 28(10):1030–1044, 1985.

[4] D. Chaum and J.-H. Evertse. A secure and privacy-protecting protocol for transmitting personal information between organizations. In *Proceedings on Advances in cryptology – CRYPTO '86*, pages 118–167, London, UK, 1987. Springer-Verlag.

[5] L. Chen. Access with pseudonyms. In *Proceedings of the International Conference on Cryptography: Policy and Algorithms*, pages 232–243, London, UK, 1995. Springer-Verlag.

[6] S. Clauss, D. Kesdogan, and T. Kölsch. Privacy enhancing identity management: Protection against re-identification and profiling. In *Digital Identity Management*, pages 84–93. ACM, 2005.

[7] I. B. Damgard. Payment systems and credential mechanisms with provable security against abuse by individuals. In *CRYPTO '88: Proceedings on Advances in cryptology*, pages 328–335, New York, NY, USA, 1990. Springer-Verlag New York, Inc.

[8] M. E. Hellman. An extension of the shannon theory approach to cryptography. *IEEE Trans. on Info. Theory*, IT-23:289–294, 1977.

[9] D. Kesdogan, D. Agrawal, and S. Penz. Limits of Anonymity in Open Environments. In F. Petitcolas, editor, *Information Hiding: 5th International Workshop (IH2002)*, volume 2578 of *LNCS*, pages 53 – 69. Springer-Verlag, October 2002.

[10] D. Kesdogan, D. Agrawal, V. Pham, and D. Rauterbach. Fundamental Limits on the Anonymity Provided by the Mix Technique. *IEEE Symposium on Security and Privacy*, May 2006.

[11] A. Lysyanskaya, R. L. Rivest, A. Sahai, and S. Wolf. Pseudonym systems. In *SAC '99: Proceedings of the 6th Annual International Workshop on Selected Areas in Cryptography*, pages 184–199, London, UK, 2000. Springer-Verlag.

[12] A. Pashalidis and B. Meyer. Linking Anonymous Transactions: The Consistent View Attack. In G. Danezis and P. Golle, editors, *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006)*, Cambridge, UK, June 2006. Springer.

[13] A. Pfitzmann and M. Köhntopp. Anonymity, unobservability, and pseudonymity - a proposal for terminology. In *Workshop on Design Issues in Anonymity and Unobservability*, volume 2009 of *Lecture Notes in Computer Science*, pages 1–9. Springer, 2000.

[14] C. Shannon. Communication theory of secrecy systems. *Bell Sys. Tech. J.*, 28:656–715, 1949.

[15] S. Steinbrecher and S. Köpsell. Modelling unlinkability. In *Privacy Enhancing Technologies*, volume 2760 of *Lecture Notes in Computer Science*, pages 32–47. Springer, 2003.