Linking Anonymous Transactions: The Consistent View Attack

Andreas Pashalidis* and Bernd Meyer

Siemens AG, Corporate Technology, Otto-Hahn-Ring 6, 81739 Munich, Germany {andreas.pashalidis, bernd.meyer}@siemens.com

Abstract. In this paper we study a particular attack that may be launched by cooperating organisations in order to link the transactions and the pseudonyms of the users of an anonymous credential system. The results of our analysis are both positive and negative. The good (resp. bad) news, from a privacy protection (resp. evidence gathering) viewpoint, is that the attack may be computationally intensive. In particular, it requires solving a problem that is polynomial time equivalent to ALLSAT. The bad (resp. good) news is that a typical instance of this problem may be efficiently solvable.

1 Introduction

Anonymous credential or 'pseudonym' systems enable a user to interact with organisations using distinct pseudonyms that hide their relation to each other and to the user's identity. In particular, the user can obtain a credential (a statement of a designated type that attests to one or more of his attributes) using one of his pseudonyms from one organisation and then 'show' it to another organisation using a different pseudonym, such that transactions of issuing and showing credentials do not reveal the identity of the user. Pseudonym systems must prevent users from showing credentials that have not been issued (i. e. they must guarantee 'credential unforgeability'), and prevent users from pooling their credentials (for example, to collectively obtain a new credential that each user individually would not be able to). This latter property is usually referred to as 'credential non-transferability' (see, for example, [2, 11]). Note that a number of pseudonym systems have been proposed in the literature (e. g. [2–6, 9, 12, 15]).

As a result of the property of credential non-transferability, it is possible for cooperating organisations to link user transactions based on the type of the credential. If, for example, only one credential is ever issued with a particular set of attributes, i. e. type, then clearly all credential showings containing this set of attributes can be linked to each other because they must have been initiated by a single user.

^{*} This author is also affiliated with the Information Security Group at Royal Holloway, University of London.

In this paper, we extend the above simplistic observation to the setting with arbitrarily many users, pseudonyms, and credential types. We show that linking transactions and pseudonyms in this setting requires solving an NP-complete problem. Moreover, we show that linking pseudonyms and transactions in *all* permissible ways is polynomial time equivalent to ALLSAT, i. e. the problem of enumerating all boolean assignments that satisfy a given boolean formula [7]. We stress that linking transactions and pseudonyms in this way does not require breaking any cryptographic properties of the underlying system.

The rest of the paper is organised as follows. The next section briefly overviews related work. Section 3 describes our attack and Section 4 provides an analysis of its complexity. Finally, Section 5 concludes.

2 Related work

In [8], Kesdogan, Agrawal, and Penz present the 'disclosure' attack which may be launched against a MIX network, and which bears certain similarities to our 'consistent view' attack. In particular, in both attacks, the adversary collects information during the normal operation of the system and then, based on this information, tries to defeat a certain user privacy property that the system is meant to protect. However, the MIX network model that is used in [8] cannot be directly applied to the setting of anonymous credential systems due to the inherent differences of the two types of system. A few notable differences between the disclosure attack and the consistent view attack that arise as a result of this incompatibility are the following.

- The disclosure attack is a traffic analysis attack where senders and recipients are typically identified by the use of network layer identifiers (e.g. IP addresses). In the consistent view attack, by contrast, users are identified based on application layer identifiers, i.e. pseudonyms.
- The disclosure attack is on the anonymity of users, while the consistent view attack is on the unlinkability of the pseudonyms and transactions of users (see, for example, [14] for a treatment of these two types of attack).
- The adversary of the disclosure attack is an external player. In the consistent view attack it is an internal player, i. e. a set of cooperating organisations.
- In the disclosure attack, the adversary is required to read all the messages that enter and leave the MIX network. The consistent view attack, by contrast, is based on data that is typically found in the audit records of the system, i.e. it does not require the adversary to acquire data from the network.

3 The attack

Our attack is based on the following assumptions. We assume that the system's lifetime is limited, i. e. that only a finite number of events occur. We also assume that no two users have the same pseudonym, and that the pseudonym system has

the 'credential non-transferability' property (or, equivalently, that the system has the 'credential unforgeability' property and that users do not share their credentials).

In order to describe the attack, we require some notation. We denote by P and T the finite sets of pseudonyms and credential types, respectively, that are used in the system. Due to our assumption that no two users have the same pseudonym, there exists a partition $Q_1, \ldots, Q_k \subseteq P$ that divides the set of pseudonyms P into as many equivalence classes (i. e. disjoint subsets) as there are users in the system, such that, for all $1 \le j \le k$, the class Q_j contains only the pseudonyms of the user j. We write $p \equiv q$ if $p, q \in P$ belong to the same class with respect to a partition of P.

The adversary \mathcal{A} in our setting is a subset of cooperating organisations and proceeds in two consecutive phases, namely the 'learning phase' and the 'linking phase'. We now describe these phases.

Learning phase: As users obtain and show credentials during the lifetime of the system, \mathcal{A} creates and maintains records, as follows. For each issuing and showing event that occurs, \mathcal{A} records the pseudonym that was used, the type of the credential, and the type of the event (i. e. issuing or showing). We call the resulting collection of records, the 'history file' \mathcal{H} . Formally, \mathcal{H} is a finite list. The entries of \mathcal{H} are of the form $(p,t,event) \in P \times T \times \{issue,show\}$. Note that, as a result of the 'credential unforgeability' property of the underlying pseudonym system, for every (p,t,show)-entry in \mathcal{H} there exists at least one preceding (q,t,issue)-entry. Also note that \mathcal{A} does not get to know the identities of the users; \mathcal{A} only sees their pseudonyms¹.

Linking phase: \mathcal{A} examines the recorded history file \mathcal{H} and divides the pseudonyms into equivalence classes. The result of this phase is a set of partitions of P that all satisfy the constraints implied by the events of issuing and showing in the history file \mathcal{H} . In the optimal case, \mathcal{A} is able to uniquely link the pseudonyms and the transactions that occur in \mathcal{H} .

We require some more notation, as follows. A partition of P is called \mathcal{H} -consistent if and only if, for each $(p,t,show) \in \mathcal{H}$, there exists a preceding $(q,t,issue) \in \mathcal{H}$ such that $p \equiv q$ with respect to this partition. In this case, \mathcal{H} is also said to admit the partition. The \mathcal{H} -consistent partitions represent all the information, in an information-theoretical sense, that \mathcal{A} can extract from the history file \mathcal{H} .

The partition of P that contains a single element equal to P is always \mathcal{H} -consistent and is called the trivial partition. Note that, since we assume that the pseudonym system has the 'credential non-transferability' property, there exists an \mathcal{H} -consistent partition that divides P into exactly those equivalence classes that correspond to the users in the attacked system.

¹ In the sequel we assume that all pseudonyms in P appear in \mathcal{H} , i.e. we ignore the case where pseudonyms have been established in the system but have not been used to obtain or show a credential.

We now define the LINKING problem, which \mathcal{A} must solve in the linking phase.

Definition 1. (LINKING) On input \mathcal{H} , obtained from the learning phase, output descriptions of all non-trivial \mathcal{H} -consistent partitions of P.

If an adversary that solves LINKING outputs only a single partition, then it has unambiguously linked all pseudonyms and transactions in the system. If an adversary that solves LINKING does not output any partitions, then the only \mathcal{H} -consistent partition is the trivial one. This represents the scenario where all pseudonyms belong to a single user.

Remark 1: In practical terms, the learning phase is a timing attack where the cooperating organisations maintain clocks that are sufficiently synchronised to enable them to unambiguously establish the order of events that occur in the system and keep records. Then they proceed to the linking phase. Although, as we show below, this may be a resource intensive task, cooperating organisations are in an advantageous position as they can, by definition, pool their resources.

Remark 2: Using the attack, organisations can link different pseudonyms of a single user. In theory, the identity of this user remains unknown. In practice, however, the anonymity of this user may be affected since, if one pseudonym p can be associated with the user's real identity, then all pseudonyms that are linked to p can be associated with that identity, too.

Remark 3: While some pseudonym systems permit users to show a credential an arbitrary number of times (e.g. [2]), others impose an upper limit on the number of times that a credential may be shown. In [5], for example, a credential may be shown only once without loss of unlinkability. Moreover, certain anonymous credential systems enable users to 'selectively disclose' a subset of the attributes that are encoded into a credential (e.g. [1]). It follows from the construction of our reduction that our attack applies to all above types of anonymous credential system.

4 The complexity of the linking phase

In this section we show that LINKING is polynomial time equivalent to ALLSAT, i.e. the problem of enumerating all satisfying truth assignments of a given boolean formula. First, we prove that the problem of deciding whether or not a given history file admits a non-trivial partition, is NP-complete. We call this problem the 'decision version' of LINKING.

Definition 2. (DV – LINKING) Given a history file \mathcal{H} , obtained from the learning phase, decide whether or not a non-trivial \mathcal{H} -consistent partition exists.

Theorem 1. DV – LINKING is NP-complete.

In order to prove this theorem, we provide a polynomial time reduction of CIRCUIT SATISFIABILITY, as defined in [13, p. 328], to DV – LINKING. We assume that the input and the output of the reduction is encoded in a

'reasonable' way, i.e. that the length of the encoding of a boolean circuit is polynomially bounded in the number of its inputs, gates, and interconnecting wires, and that the length of the encoding of the history file is polynomially bounded in the number of pseudonyms, credential types, and events in the list. Furthermore, we assume that boolean circuits are acyclic and consist only of gates of type NAND where each gate has two inputs. The above assumptions are made without loss of generality (see, for example, [10, 13]).

Given a description of a boolean circuit, our reduction generates a history file where each (p, t, show)-entry is preceded by a (q, t, issue)-entry. In particular, given the description of a circuit C with n inputs, a single output, and m interconnecting wires (excluding the inputs and the output), the reduction generates a history file \mathcal{H}_C in which the set of pseudonyms that appear is

$$P = \{in_1, \dots, in_n, w_1, \dots, w_m, out, true, false\},$$
(1)

where in_1, \ldots, in_n correspond to the inputs of C, w_1, \ldots, w_m to its interconnecting wires, and *out* to its output. The pseudonyms true and false are auxiliary pseudonyms that are used for the representation of boolean values.

The history file \mathcal{H}_C is constructed in a way that guarantees that, if it admits a non-trivial \mathcal{H}_C -consistent partition, this partition will have exactly two elements $Q_1, Q_2 \subset P$ where $true \in Q_1$, $false \in Q_2$, and that, therefore, for all $p \in P$, either $p \equiv true$ or $p \equiv false$. Furthermore, for all such partitions, setting the inputs of C that correspond to pseudonyms in Q_1 to 'true' and the remaining inputs (i. e. those that correspond to pseudonyms in Q_2) to 'false' yields a satisfying truth assignment for C.

The history file \mathcal{H}_C consists of two parts, namely the 'setup' part and the 'main' part. The setup part is constructed using the setup algorithm which is shown in Figure 1. Note that this algorithm generates 3(n+m)+2 entries in \mathcal{H}_C , and that the amount of different types appearing in these entries is n+m+1.

Lemma 1. The entries in the setup part of \mathcal{H}_C ensure that any non-trivial \mathcal{H}_C -consistent partition has exactly two elements $Q_1, Q_2 \subset P$ with true, out $\in Q_1$ and false $\in Q_2$.

Proof. It follows from the entries that are added to \mathcal{H}_C in Step 4 of the setup algorithm that, for all $p \in P - \{out\}$, either $p \equiv true$ or $p \equiv false$. Consider an \mathcal{H}_C -consistent partition $Q_1, \ldots, Q_k \subseteq P$. If the partition is such that $true \equiv false$, then $p \equiv true \equiv false$ for all $p \in P$. Thus, the partition is the trivial partition $Q_1 = P$. If the partition contains two sets $Q_1, Q_2 \subset P$ with $true \in Q_1$ and $false \in Q_2$, then, since either $p \equiv true$ or $p \equiv false$ for all $p \in P - \{out\}$, it follows that $Q_1 \cup Q_2 = P$. Moreover, the entries added to \mathcal{H}_C in Step 6 imply that $out \equiv true$. The result follows.

The main part of \mathcal{H}_C encodes the gates in C. We first describe an algorithm that encodes a single NAND-gate G into \mathcal{H}_C , and leave the encoding of the entire circuit for later.

As determined by the setup algorithm, each gate G is associated with three pseudonyms. Let $a, b \in P$ be the pseudonyms that correspond to the two inputs

Setup algorithm (input: a description of a boolean circuit C):

- 1. Generate the set of pseudonyms P according to Equation 1 and uniquely assign each pseudonym (except true and false) to either an input, an interconnecting wire, or the output of C, as described above.
- 2. Generate the set of types $T = \{t_1, \ldots, t_{4m+n+4}\}.$
- 3. Start with an empty list \mathcal{H}_C and set the global counter $i \leftarrow 0$.
- 4. For each $p \in P \{true, false, out\}$ do the following.
 - (a) Increase i by one.
 - (b) Append the three entries $(true, t_i, issue)$, $(false, t_i, issue)$, and $(p, t_i, show)$ in this order to \mathcal{H}_C .
- 5. Increase i by one.
- 6. Append the entries $(out, t_i, issue)$ and $(true, t_i, show)$ in this order to \mathcal{H}_C .

Figure 1: Generation of the setup part of \mathcal{H}_C

of G and $c \in P$ be the pseudonym that corresponds to its output. The NAND-gate algorithm, shown in Figure 2, adds entries for the encoding of G to the main part of \mathcal{H}_C . It is assumed that the algorithm runs after the setup algorithm has completed.

NAND-gate algorithm (input: pseudonyms a, b, c that are associated with a gate G, a history file \mathcal{H}_C , a counter value i):

- 1. Increase i by one.
- 2. Append the three entries $(a, t_i, issue), (c, t_i, issue), (true, t_i, show)$ in this order to \mathcal{H}_C .
- 3. Increase i by one.
- 4. Append the three entries $(b, t_i, issue)$, $(c, t_i, issue)$, $(true, t_i, show)$ in this order to \mathcal{H}_C .
- 5. Increase i by one.
- 6. Append the four entries $(a, t_i, issue)$, $(b, t_i, issue)$, $(c, t_i, issue)$ $(false, t_i, show)$ in this order to \mathcal{H}_C .

Figure 2: Generation of the encoding of a NAND-gate

Note that the NAND-gate algorithm generates 10 entries in which three different types appear.

Lemma 2. The entries generated by the NAND-gate algorithm, together with the entries generated by the setup algorithm, encode gate G into \mathcal{H}_C , i. e. they ensure that, for all non-trivial \mathcal{H}_C -consistent partitions, it holds that $c \equiv f$ also if and only if $a \equiv b \equiv t$ rue.

Proof. By Lemma 1 it follows that, for any given non-trivial \mathcal{H}_C -consistent partition of P and for all $p \in \{a, b, c\}$, either $p \equiv true$ or $p \equiv false$. Furthermore, the entries that are generated by the NAND-gate algorithm enforce that each of the sets $\{a, c\}$ and $\{b, c\}$ contains at least one element that is equivalent to true, and that at least one element in $\{a, b, c\}$ is equivalent to false.

We now show that all non-trivial \mathcal{H}_C -consistent partitions of P are such that the pseudonyms a, b and c are equivalent with either true or false in a way that is consistent with the boolean behaviour of G. Consider any such partition. If $a \equiv b \equiv true$, then the last four entries imply that $c \equiv false$. Also, no contradiction arises from the first six entries. Since $c \equiv false$, the behaviour of G is correctly encoded in this case. In all other cases (i.e. if $a \equiv b \equiv false$ or if $a \not\equiv b$), the first six entries imply that $c \equiv true$. Also, no contradiction arises from the last four entries. Since $c \equiv true$, the behaviour of G is correctly encoded in these cases as well.

The encoding of the entire circuit C into \mathcal{H}_C amounts to calling the NAND-gate algorithm for each gate G in C in turn, and setting the pseudonyms a, b and c to those that correspond to G, as this correspondence was determined by the setup algorithm. Note the total number of gates in the circuit is m+1: one for each interconnecting wire, plus one output gate. This results in 10(m+1) entries in the main part of \mathcal{H}_C , where 3(m+1) different types appear. Thus, the total amount of entries in \mathcal{H}_C is 13m+3n+12, and the total amount of different types that appear is 4m+n+4.

We can now prove Theorem 1.

Proof. There exists an obvious polynomial time algorithm that, given a partition of P and a history file \mathcal{H} , checks whether or not the partition is consistent with all events in \mathcal{H} . Thus, $DV - LINKING \in NP$.

By Lemma 1, the setup part of \mathcal{H}_C makes sure that all non-trivial \mathcal{H}_{C} -consistent partitions of P are such that all pseudonyms that are associated with the inputs and the interconnecting wires of C are either equivalent to true or false. Furthermore, the pseudonym that corresponds to the output of C is equivalent to true. By Lemma 2, it can be proven using induction on the number of gates in C that the main part of \mathcal{H}_C guarantees that all non-trivial \mathcal{H}_{C} -consistent partitions are such that the pseudonyms are either equivalent to true or false in a way that is consistent with the boolean behaviour of the circuit. Thus, DV - LINKING is NP-complete.

It follows from the construction of our reduction that the partition of the pseudonyms in_1, \ldots, in_n , which correspond to the inputs of the circuit, uniquely determines the partition of all other pseudonyms in the system, in accordance with the boolean behaviour of the circuit. Therefore, there exists a one-to-one correspondence between the non-trivial \mathcal{H}_C -compliant partitions and the satisfying truth assignments for the circuit.

By a standard complexity-theoretical argument, the algorithm for the verification of the non-trivial \mathcal{H} -consistency of a given partition can be efficiently

transformed into a family of boolean circuits of polynomial size. We thus arrive at the following corollary.

Corollary 1. LINKING is polynomial time equivalent to ALLSAT.

5 Conclusion

In this paper we studied the 'consistent view' attack that may be launched by cooperating organisations in order to link the transactions and the pseudonyms of the users of an anonymous credential system. The attack is based on the information in a 'history file' that describes the events that take place during the lifetime of the system. We showed that extracting all the information from such a history file is polynomial time equivalent to solving ALLSAT, by providing a polynomial time reduction. This, however, is a statement about the worst-case complexity of the attack. Our reduction produces history files that are unlikely to be similar in structure to the history files of real-world pseudonym systems; in a typical real-world scenario, extracting all the information from the history file is therefore likely to be significantly more efficient than polynomial equivalence to ALLSAT might suggest. Moreover, since LINKING can be formulated as an instance of ALLSAT, the adversary can use state-of-the-art SAT solvers [7, 16].

Unfortunately, making more precise statements about the complexity and the efficiency of our attack in a real-world scenario requires making assumptions about the behaviour of the users in the system. It is conceivable that there may exist certain user strategies that lead to history files that, with at least non-negligible probability, make the 'consistent view' attack computationally expensive. However, this is rather unlikely, as the existence of a generic strategy could be used as the basis to prove that $P \neq NP$.

Studying the efficiency of the attack under a reasonable user behaviour model is a subject for further research.

Acknowledgements

The authors would like to thank Michael Braun, Erwin Heß, Anton Kargl, Chris J. Mitchell, Torsten Schütze, and Susanne Wetzel for their help and encouragement. The authors would also like to thank the anonymous reviewers for their stimulating comments.

References

- S. Brands. Rethinking Public Key Infrastructures and Digital Certificates Building in Privacy. The MIT Press, Cambridge, Massachusetts, 2000.
- 2. J. Camenisch and A. Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In B. Pfitzmann, editor, Advances in Cryptology EUROCRYPT 2001, International Conference on the

- Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceedings, volume 2045 of Lecture Notes in Computer Science, pages 93–118. Springer Verlag, Berlin, 2001.
- 3. J. Camenisch and A. Lysyanskaya. Signature schemes and anonymous credentials from bilinear maps. In M. Franklin, editor, *Proceedings of the 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19 CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 56–72. Springer Verlag, Berlin, 2004.
- 4. D. Chaum. Showing credentials without identification: Transferring signatures between unconditionally unlinkable pseudonyms. In J. Seberry and J. Pieprzyk, editors, *Advances in Cryptology AUSCRYPT 90*, volume 453 of *Lecture Notes in Computer Science*, pages 246–264. Springer Verlag, Berlin, 1990.
- L. Chen. Access with pseudonyms. In E. Dawson and J. D. Golic, editors, Cryptography: Policy and Algorithms, International Conference, Brisbane, Queensland, Australia, July 3-5, 1995, Proceedings, number 1029 in Lecture Notes in in Computer Science, pages 232–243. Springer Verlag, Berlin, 1995.
- 6. I. Damgård. Payment systems and credential mechanisms with provable security against abuse by individuals. In S. Goldwasser, editor, Advances in Cryptology CRYPTO '88: Proceedings, number 403 in Lecture Notes in Computer Science, pages 328–335. Springer Verlag, Berlin, 1990.
- 7. H. Jin and F. Somenzi. Prime clauses for fast enumeration of satisfying assignments to boolean circuits. In *DAC '05: Proceedings of the 42nd Annual Conference on Design Automation*, pages 750–753, New York, NY, USA, 2005. ACM Press.
- 8. D. Kesdogan, D. Agrawal, and S. Penz. Limits of anonymity in open environments. In F. Petitcolas, editor, *Proceedings of the 5th International Information Hiding Workshop, IH 2002*, volume 2578 of *Lecture Notes in Computer Science*, pages 53–69. Springer Verlag, Berlin, 2003.
- 9. A. Lysyanskaya, R. L. Rivest, A. Sahai, and S. Wolf. Pseudonym systems. In H. M. Heys and C. M. Adams, editors, Selected Areas in Cryptography, 6th Annual International Workshop, SAC'99, Kingston, Ontario, Canada, August 9-10, 1999, Proceedings, volume 1758 of Lecture Notes in Computer Science, pages 184–199. Springer Verlag, Berlin, 2000.
- 10. M. Mano. Digital Design. Prentice Hall, third edition, 2001.
- 11. A. Pashalidis and C. Mitchell. A security model for anonymous credential systems. In S. J. Y. Deswarte, F. Cuppens and L. Wang, editors, *Information Security Management, Education and Privacy, Proceedings of the 3rd Working Conference on Privacy and Anonymity in Networked and Distributed Systems (I-NetSec'04)*, pages 183–199. Kluwer Academic Publishers, August 2004.
- 12. G. Persiano and I. Visconti. An efficient and usable multi-show non-transferable anonymous credential system. In A. Juels, editor, *Proceedings of the Eighth International Financial Cryptography Conference (FC '04)*, volume 3110 of *Lecture Notes in Computer Science*, pages 196–211. Springer Verlag, Berlin, 2004.
- M. Sipser. Introduction to the Theory of Computation. PWS Publishing Company, 1997.
- S. Steinbrecher and S. Köpsell. Modelling unlinkability. In R. Dingledine, editor, Privacy Enhancing Technologies, Third International Workshop, PET 2003, Dresden, Germany, March 26-28, 2003, Revised Papers, volume 2760 of Lecture Notes in Computer Science, pages 32-47. Springer Verlag, Berlin, 2003.
- 15. E. R. Verheul. Self-blindable credential certificates from the Weil pairing. In C. Boyd, editor, ASIACRYPT '01: Proceedings of the 7th International Conference

- on the Theory and Application of Cryptology and Information Security, volume 2248 of Lecture Notes in Computer Science, pages 533–551. Springer Verlag, Berlin, 2001.
- 16. L. Zhang and S. Malik. The quest for efficient boolean satisfiability solvers. In A. Voronkov, editor, Automated Deduction CADE-18, 18th International Conference on Automated Deduction, Copenhagen, Denmark, July 27-30, 2002, Proceedings, volume 2392 of Lecture Notes in Computer Science, pages 295–313. Springer Verlag, Berlin, 2002.