Unsatisfiability of Comparison-Based Non-Malleability for Commitments

(Short Paper)

Denis Firsov[1,2], Sven Laur[3], and Ekaterina Zhuchko[2,3]

[1] Guardtime, Tallinn, Estonia
denis.firsov@guardtime.com
[2] Tallinn University of Technology, Tallinn, Estonia
ekzhuc@ttu.ee
[3] Tartu University, Tartu, Estonia
swen@math.ut.ee

Abstract. There are two distinct formulations of non-malleability of commitments found in the literature: the comparison-based definition and the simulation-based definition. In this paper, we prove that the comparison-based definition is unstatisfiable by any realistic commitment scheme. Our proof is fully formalized in the EasyCrypt theorem prover.

Keywords: cryptography · commitments · comparison-based · non-malleability · formal methods · EasyCrypt.

1 Introduction

A commitment scheme is one of the fundamental primitives in cryptography. Intuitively, we can think of a commitment as a locked box containing a message. Only the sender who produced the commitment knows the secret opening key which can unlock the box and reveal the message. The sender can send this box to a receiver and then at a later stage give him the opening key to unlock it.

The most fundamental security properties of commitments are hiding and binding. We say that a commitment is hiding if an adversary is unable to see the message without the opening key (the box which contains the message should not be transparent). We say that a commitment is binding if, once the sender committed to a message and sent the commitment to the receiver, the sender cannot open the commitment to a different message (the box should not have any secret backdoors or double bottoms). But these properties do not prevent all of the attacks and most notably the "man-in-the-middle" attacks.

The *non-malleability* property aims to protect commitments against man-in-the-middle attacks. In such an attack, we have Mallory who is an active adversary between two parties: Alice and Bob. Let's assume that Alice sends a commitment c of a message m to Bob. However, all of their communication goes through the man-in-the-middle adversary Mallory who can modify the commitment or simply not deliver it. The goal of Mallory is to generate a commitment c' (based only on the commitment c) to another message m' which is non-trivially related to the

original message m.[4] Later, for a successful attack, Mallory must generate an opening d' for commitment c' when it sees the Alice's opening d for commitment c.

A classical motivating example where non-malleability would be needed is that of a blind auction. Consider an auction where participants bid for an item by publishing commitments to their bids. At the end, bidders open their commitments and the highest bid wins. If the commitment scheme is malleable, an adversary could participate in the auction by posting for each of the other bids a commitment to a bid that is only one dollar higher. In this case, the adversary would have an unfair advantage. Moreover, the adversary has no need to learn the exact amounts that other bidders have placed. The goal of non-malleability definitions is to prevent these types of attacks.

There have been several attempts to formally define non-malleability of commitments. Most notably, Crescenzo et al. presented a simulation-based definition [1]. The main idea of their definition is to compare the success probability of an adversary and its simulator. The adversary sees a commitment c of a message m and must produce a commitment c' of a message m' which must be non-trivially related to m. At the same time, the simulator must also produce a message similarly related to m, but without seeing any of the derivatives of m (e.g., commitment on m). If the difference between success probabilities is negligible then the commitment scheme is considered simulation-based non-malleable.

Later, Laur et al. introduced a new formulation of non-malleability which is now known as *comparison-based definition* [5,7,6,3]. The goal of this definition is to phrase non-malleability without referring to a simulator. This was motivated by the fact that definitions formulated in terms of simulators are more complicated to falsify by presenting a specially programmed adversary.

The original intention of this paper was to analyze the comparison based non-malleability of commitments introduced by Laur et al. [5] and prove that it implies hiding and binding of the commitment scheme. However, after we started our formal analysis and specified the definition precisely in EasyCrypt, we were able to conjecture and then prove that the definition is unsatisfiable by any realistic commitment schemes, but is satisfiable by a completely non-binding "constant"-commitment scheme. Taking both discoveries into account we claim that the comparison-based non-malleability as defined by Laur et. al. [5] is unfit for any practical and theoretical purposes.

Our result is formalized in the EasyCrypt theorem prover and the proof-scripts can be found in the supplementary material [2]. The results presented in this paper follow our formal EasyCrypt development. However, for the purpose of readability we present them in the standard mathematical notation.

2 Comparison-Based Non-Malleability

Definition 1 (Commitment Scheme). *A commitment scheme is a triple of efficient algorithms* (*Gen*, *Commit*, *Verify*) *where:*

[4] An example of a non-trivial relation could be that the message m' is the same as m except all occurrences of *"PAY TO: Alice"* are replaced with *"PAY TO: Mallory"*.

- *Gen: is a distribution of public keys (also known as public parameters) of a commitment scheme.*
- *Commit*(pk, m)*: is a distribution of commitment-opening pairs which is parameterized by a public key* pk *and a message* m*.*
- *Verify*(pk, m, c, d)*: is a deterministic function which verifies the commitment* c *on the message* m *with respect to the opening key* d*.*

The commitment scheme is ***functional*** *iff all commitment-opening pairs produced by Commit*(pk, m) *verify on* m*:*

$$\forall\ c\ d\ pk\ m,\ pk \in Gen \wedge (c,d) \in Commit(pk,m) \implies Verify(pk,m,c,d) = 1.$$

(Here, $x \in D$ *denotes that* x *is in support of distribution* D*.)*

Let us give a formal definition of comparison-based non-malleability introduced by Laur et al. [5].

Definition 2 (Laur et al.). *A commitment scheme* $C = (Gen, Commit, Verify)$ *is* ***comparison-based non-malleable*** *iff for any efficient adversary* A*, the advantage* $AdvC(C, A)$ *is negligible, where*

$$AdvC(C,A) := |\mathsf{Pr}\,[r \leftarrow GN_0(C,A).main() : r = 1] - \mathsf{Pr}\,[r \leftarrow GN_1(C,A).main() : r = 1]\,|.$$

```
1: module GN_0(C, A)
2:     proc main() = {
3:         pk ←$ Gen
4:         M ← A.init(pk)
5:         m ←$ M
6:         (c, d) ←$ Commit(pk, m)
7:         (c', R) ← A.commit(c)
8:         (d', m') ← A.decommit(d)
9:         v ← Verify(pk, m', c', d')
10:        return v ∧ R(m, m') ∧ c ≠ c'
11:    }
12: end
```

```
1: module GN_1(C, A)
2:     proc main() = {
3:         pk ←$ Gen
4:         M ← A.init(pk)
5:         m ←$ M; n ←$ M
6:         (c, d) ←$ Commit(pk, m)
7:         (c', R) ← A.commit(c)
8:         (d', m') ← A.decommit(d)
9:         v ← Verify(pk, m', c', d')
10:        return v ∧ R(n, m') ∧ c ≠ c'
11:    }
12: end
```

(For simplicity of presentation, in Def. 2 the adversary computes a single commitment c' while in the original definition of Laur et al. the adversary was allowed to return n commitments and $n+1$-place relation R. In our EasyCrypt formalization, we work with the original definition, but in the paper we show the simplified version since this detail is irrelevant for the main unsatifiability result.)

Both games are parameterized by a commitment scheme C and an adversary A. In the game GN_0, adversary A is given the public key pk and is asked to compute a message distribution $\mathcal{M}$. A message m is then sampled from $\mathcal{M}$ and a commitment-opening pair (c, d) is computed with respect to m. Next, adversary A is given the commitment c and asked to produce a commitment c'

and a relation R. After that, A is given the opening d and asked to produce an opening-message pair (d', m). The adversary wins the game if the pair (c', d') is valid with respect to m', the relation R is satisfied by a pair (m, m') and A's commitment c' is different from c. The only difference in the game GN_1 is that a second message n is sampled from the message distribution (independently from m). The commitment-opening pair is still computed with respect to the message m, but the winning condition of GN_1 considers whether $R(n, m')$ holds (line 10).

The adversary's overall advantage is defined in terms of its ability to distinguish between games GN_0 and GN_1. In other words, A has to win one game and lose the other in order to increase the advantage. This means that to be successful, the adversary has to find the exact relation R which will hold given the pair (m, m') and will not hold given the pair (n, m'), or vice versa.

2.1 (Un)satisfiability of the Comparison-Based Definition

In this section, we show that Def. 2 is not satisfiable by any realistic[5] commitment scheme. More specifically, we construct a single adversary which can break the comparison-based non-malleability of any realistic commitment scheme with unacceptably high probability. Moreover, we also define a paradoxical and completely non-binding "constant"-commitment scheme which satisfies Def. 2. Taking both discoveries into consideration it must be sufficient to claim that the comparison-based non-malleability as defined by Laur et. al. [5] is unfit for any practical and theoretical purposes.

Theorem 1. *For any functional commitment scheme* $C = (\mathit{Gen}, \mathit{Commit}, \mathit{Verify})$ *the adversary* A *(see Figure 1 for the definition) has the following comparison-based non-malleability advantage:*

$$\mathit{AdvC}(C, A) = \frac{1}{4} - \frac{1}{4} \cdot \mathsf{Pr}\left[\begin{array}{c} pk \xleftarrow{\$} \mathit{Gen};\ (c, d) \xleftarrow{\$} \mathit{Commit}(pk, 0); \\ (c', d') \xleftarrow{\$} \mathit{Commit}(pk, 0) : c = c' \end{array}\right].$$

(If commitments generated by Commit are sufficiently random then $\mathit{AdvC}(C, A)$ *is not negligible.)*

Proof. The adversary A is defined as follows (see Figure 1): in the initialization phase, the adversary returns a uniform distribution of booleans. During the commit phase A receives the commitment c and generates a commitment-opening pair (c', d') on $m' = 0$. Moreover, the relation $R(m, m')$ is also fixed and will only hold true if $m = 0$ and $m' = 0$. During the "decommit" phase, A receives opening d and checks if it opens c with message $m = 0$. If so, A returns $(d', 0)$ as the opening-message pair. If the verification fails, the adversary intentionally loses the game (denoted by $\perp$).

[5] We assume that in realistic schemes commitment values contain a sufficient amount of randomness.

Figure 1 Adversary for Comparison-Based Non-Malleability

```
1: module A
2:     var pk, c, c', d'
3:     proc init(pk) = {
4:         A.pk ← pk                          ▷ save the public key in the global variable
5:         return {0,1}                       ▷ {0,1} is a uniform distribution of bits
6:     }
7:     proc commit(c) = {
8:         A.c ← c                            ▷ the commitment is stored for the next phase
9:         (c', d') ←$ Commit(pk, 0)
10:        R ← λm0 m1.m0 = 0 ∧ m1 = 0
11:        return (R, c')
12:    }
13:    proc decommit(d) = {
14:        if Verify(pk, 0, c, d) then
15:            return (d', 0)
16:        end if
17:        return ⊥                           ▷ denotes a pair which always fails the verification
18:    }
19: end
```

In order to calculate the adversary's advantage we argue as follows:

$$
\begin{aligned}
&\mathsf{Pr}\left[r \leftarrow GN_0(C,A).main() : r = 1\right] - \mathsf{Pr}\left[r \leftarrow GN_1(C,A).main() : r = 1\right] \\
&\overset{(1)}{=} \left(\mathsf{Pr}\left[GN_0(A).main() : m = 0\right] - \mathsf{Pr}\left[GN_0(A).main() : m = 0, c = c'\right]\right) \\
&\quad - \left(\mathsf{Pr}\left[GN_1(A).main() : m = 0, n = 0\right] - \mathsf{Pr}\left[GN_1(A).main() : m = 0, n = 0, c = c'\right]\right) \\
&\overset{(2)}{=} \frac{1}{2} - \mathsf{Pr}\left[GN_0(A).main() : m = 0, c = c'\right] \\
&\quad - \frac{1}{4} + \frac{1}{2} \cdot \mathsf{Pr}\left[GN_0(A).main() : m = 0, c = c'\right] \\
&\overset{(3)}{=} \frac{1}{4} - \frac{1}{2} \cdot \mathsf{Pr}\left[\begin{array}{c} pk \xleftarrow{\$} Gen;\ \mathcal{M} \leftarrow A.init(pk);\ m \xleftarrow{\$} \mathcal{M}; \\ (c,d) \xleftarrow{\$} Commit(pk,m);\ (c',d') \xleftarrow{\$} Commit(pk,0) : \\ m = 0, c = c' \end{array}\right] \\
&\overset{(4)}{=} \frac{1}{4} - \frac{1}{4} \cdot \mathsf{Pr}\left[\begin{array}{c} pk \xleftarrow{\$} Gen;\ (c,d) \xleftarrow{\$} Commit(pk,0); \\ (c',d') \xleftarrow{\$} Commit(pk,0) : c = c' \end{array}\right].
\end{aligned}
$$

In step (1), we observe that for any functional scheme the commitment verification (i.e., $Verify(pk, 0, c', d') = 1$) is guaranteed to succeed. Also, we rewrite the winning probability in terms of an event complement to the $c \neq c'$ condition. In step (2), we can restate all the probabilities in relation to GN_0 by observing that n is independent from m and making explicit the probability of sampling $n = 0$ as a coefficient. In step (3), we compute the probabilities and inline the game GN_0. In step (4), we observe that the remaining probabiltiy expression is non-zero only when $m = 0$, so we can simplify the game further.

Observe that the following probability can be safely assumed to be negligible for any realistic commitment scheme which produces sufficiently random

commitments:

$$\mathsf{Pr}\left[\begin{array}{l} pk \leftarrow Gen;\ (c,d) \xleftarrow{\$} Commit(pk,m); \\ \quad (c',d') \xleftarrow{\$} Commit(pk,m) : c = c' \end{array}\right].$$

The reason why the adversary A is able to have a non-negligible advantage is because it could "intentionally lose" in the decommit phase. Once it receives the opening d, it can easily verify the content of the given commitment c and if verification fails, intentionally lose the game. Finally, we find it interesting that this analysis shows that the comparison-based definition cannot be instantiated with any realistic commitment scheme, but could be proved for some paradoxical schemes. Indeed, we can define the following "constant"-commitment scheme:

$$\begin{aligned} &Gen := \{*\} \\ &Commit(pk,m) := (*,m) \\ &Verify(pk,m,c,d) := \textbf{\textit{if}}\ m = d\ \textbf{\textit{then}}\ 1\ \textbf{\textit{else}}\ 0 \end{aligned}$$

The public key and commitments are elements of a singleton set (denoted by $*$) and the opening of a commitment is a message itself. This commitment-scheme is functional, perfectly hiding, and completely non-binding. Moreover, since the winning condition $c \neq c'$ of GN_0 and GN_1 is never satisfied then we conclude that the above "constant"-coimmitment scheme is perfectly non-malleable according to Def. 2. This must be understood as another reason to abandon that definition of non-malleability.

3 Conclusions

The problem of inadequate definitions in cryptography is not new [4]. The errors in definitions may take many years to be discovered and the impact of these errors can range from a minimal nuisance to an actual threat that can be realised as an attack in the real world.

In our investigation, we were surprised to find the definition of comparison-based non-malleability unsatisfiable. The paper [5] radiates confidence of the authors that their definition is not only satisfiable, but that some constructions provide unreasonably high level of security. Moreover, the paper is well-cited with more than 200 citations to that date. However, according to our best knowledge, we are the first to spot the mistake. We attribute our discovery of unsatisfiability to the fact that our investigation was carried out in the formal setting of the EasyCrypt theorem prover. Although the idea behind the proof is fairly simple, the formal derivation took considerable effort (the formalization is 600 loc).

Finally, this work stresses the need to provide higher assurance to the cryptographic security proofs. We believe that formal methods provide a solution which ensures rigor necessary for the mission critical systems.

In the future, we plan to investigate alternative definitions of comparison-based non-malleability.

References

1. Giovanni Di Crescenzo, Jonathan Katz, Rafail Ostrovsky, and Adam Smith, *Efficient and non-interactive non-malleable commitment*, Cryptology ePrint Archive, Report 2001/032, 2001, `https://ia.cr/2001/032`.
2. Denis Firsov, Ekaterina Zhuchko, and Sven Laur, *Formal analysis of non-malleability for commitments in EasyCrypt*, `https://github.com/dfirsov/comparison-based-non-malleabiltiy-unsat`, 2022.
3. Sameh Khalfaoui, Jean Leneutre, Arthur Villard, Jingxuan Ma, and Pascal Urien, *Security analysis of out-of-band device pairing protocols: A survey*, Wireless Communications and Mobile Computing **2021** (2021), 1–30.
4. Neal Koblitz and Alfred Menezes, *Critical perspectives on provable security: Fifteen years of "another look" papers*, Advances in Mathematics of Communications **13** (2019), 517–558.
5. Sven Laur and Kaisa Nyberg, *Efficient mutual data authentication using manually authenticated strings*, International Conference on Cryptology and Network Security, Springer, 2006, pp. 90–107.
6. Ming Li and et al., *Secure ad-hoc trust initialization and key management in wireless body area networks*, ACM TRANS. SENSOR NETW (2012).
7. Shahab Mirzadeh, Haitham Cruickshank, and Rahim Tafazolli, *Secure device pairing: A survey*, IEEE Communications Surveys Tutorials **16** (2014), no. 1, 17–40.