Efficient Mutual Data Authentication Using Manually Authenticated Strings

Sven Laur[2] and Kaisa Nyberg[1,2]

[1] Nokia Research Center, Finland kaisa.nyberg@nokia.com
[2] Helsinki University of Technology, Finland {slaur,knyberg}@tcs.hut.fi

Abstract. Solutions for an easy and secure setup of a wireless connection between two devices are urgently needed for WLAN, Wireless USB, Bluetooth and similar standards for short range wireless communication. All such key exchange protocols employ data authentication as an unavoidable subtask. As a solution, we propose an asymptotically optimal protocol family for data authentication that uses short manually authenticated out-of-band messages. Compared to previous articles by Vaudenay and Pasini the results of this paper are more general and based on weaker security assumptions. In addition to providing security proofs for our protocols, we focus also on implementation details and propose practically secure and efficient sub-primitives for applications.

1 Introduction

In this paper we consider the problem of setting up a shared secret key in an ad hoc manner, that is, in isolation from any key management facility and without pre-shared secrets. Consider two parties Alice and Bob who want to establish a shared secret key over an insecure network without any prior authenticated information. If adversaries are passive, that is, no malicious messages are sent to the network and all messages are delivered unaltered, then exchanging public keys for Diffie-Hellman or similar public key based key exchange protocol is sufficient. However, an active adversary, Charlie, can launch an man-in-the-middle attack. Namely, Charlie can replace a desired secure channel from Alice to Bob by a pair of secure channels, one from Alice to Charlie and one from Charlie to Bob. The attack is transparent to legitimate users without prior authenticated information. Thus secure key exchange is impossible without authenticated channels. The main question is how much information, *authentic out-of-band messages (OOB messages)*, must be sent over the authenticated channel to achieve reasonable security level. We are aiming at an application, where keys are exchanged between various electronic devices and authentic communication is done by an ordinary user who either enters messages into devices or compares output displays. The latter severely limits a plausible size of OOB messages: one could consider 4–6 decimal digits as optimal and 16 hexadecimal characters as an absolute limit. Other possible OOB channels include various visual or audible signals like blinking lights, images, phone calls etc.

Most urgently such a solution is needed for WLAN: the current use of pre-shared keys degrades both practical usability and security. The home users should have a clear and manageable procedure to set up a secure wireless network so that it is easy to add and remove devices from the network. Hence, the WiFi Alliance is working on a better solution. Recently, manual data authentication using short authenticated strings received practical applications in ad hoc key agreement. Phil Zimmermann released a software called Zfone and an Internet draft to offer security to Voice over IP [ZJC06]. A similar protocol (See Protocol 3) was adopted by USB-IF for Wireless USB devices [WUS06] and manual data authentication is going to be incorporated into Bluetooth [BT06].

A formal security model for such protocols consists of three bidirectional asynchronous channels, where messages can be arbitrarily delayed. In-band communication is routed from Alice to Bob via an active adversary Charlie, who can drop, modify or insert messages. The out-of-band channel between Alice and Bob is authentic but has low bandwidth and Charlie can arbitrarily delay[3] OOB messages. The model captures nicely all threats in wireless environment, as malicious adversary with a proper equipment can indeed change the network topology and thus reroute, drop, insert and modify messages. However, security is not the only objective. User-friendliness, low resource consumption and simple setup assumptions are equally important. There should be no public key infrastructure, as it is almost impossible to guarantee authenticity and availability of public keys to the humongous number of electronic devices. Also, protocols should use only symmetric primitives if possible.

All currently known user-aided key exchange and data authentication protocols can be divided into two different groups: protocols with authenticated but public OOB messages [Hoe05,CCH06,Vau05,LAN05,PV06a,PV06b,NSS06] and protocols with confidential passwords. Password-protected key exchange, see [BM92,KOY01] and Mana III in [GMN04], is needed when a user wants to establish a secure connection between devices that have input only, for example, devices with keyboards but no display. The main application for the manual data authentication is also a cryptographically secure but still user-friendly ad hoc key agreement between two or more network devices.

Our contribution. In this paper, we clarify and extend our preliminary results [LAN05]. In particular, we show that the previously presented manual cross authentication protocols [LAN05,PV06b] are indeed instantiations of the same protocol family that uses a commitment scheme to temporarily hide a secret key needed for data authentication. Compared to the results by Pasini and Vaudenay [PV06b], our security proofs (Sec. 4) are more modular and assumptions on used primitives are weaker and geared towards practice. We explicitly consider implementation details, that is, how to choose practical primitives (Sec. 5). Given a data authentication protocol it can be combined with the Diffie-Hellman key agreement in a secure way by taking the Diffie-Hellman key, or the pair of the public keys, as the data to be authenticated. But the designers of the practical protocols from [ZJC06,WUS06] have taken a different approach by using the Diffie-Hellman key shares as the source of randomness. In Sec. 3, we extend our proof of security also for such a case. In App. A, we consider security in any computational context and show that, under reasonable assumptions, security is not abruptly degraded if several protocols are executed in parallel. As an important theoretical result, we show that all asymptotically optimal (unilateral) manual data authentication protocols have a

[3] For example, the adversary can distract the user who compares the output of two devices.

certain structure (App. B, Theorem 5) and that there are no asymptotically optimal two round protocols for data authentication (Corollary 1).

2 Cryptographic preliminaries

Our results are formalised in exact security framework where security reductions are precise and thus reveal quantitative differences between various assumptions. Security goals are formalised through games between a challenger and a t-time adversary [4] A who tries to violate some design property. Advantage of A is a non-trivial success probability $\mathrm{Adv}^{\mathsf{sec}}(A)$ in the game sec. The description of sec is omitted when the shape of $\mathrm{Adv}^{\mathsf{sec}}(A)$ reveals the complete structure of sec. We consider asymptotic complexity only w.r.t. the time-bound t. Let $g(t) = \mathcal{O}(f(t))$ if $\limsup_{t\to\infty} g(t)/f(t) < \infty$. If the working time of adversary can be unbounded, we talk about *statistical security.*

Let $x \leftarrow \mathcal{X}$ denote independent random draws from a set $\mathcal{X}$ and $y \leftarrow A(x_1, \dots, x_n)$ denote assignment according to a randomised algorithm A with inputs $x_1, \dots, x_n$.

Keyed hash functions. A keyed hash function $h : \mathcal{M} \times \mathcal{K} \to \mathcal{T}$ has two arguments: the first argument corresponds to a data and the second to a key. In our context an applicable tag space $\mathcal{T}$ is relatively small (may contain as few as 10^4 elements) and we need information theoretic properties. A hash function h is ε_u*-almost universal*, if for any two inputs $x_0 \neq x_1$, $\Pr[k \leftarrow \mathcal{K} : h(x_0, k) = h(x_1, k)] \leq \varepsilon_u$ and ε_u*-almost XOR universal* if for any $x_0 \neq x_1$ and y, $\Pr[k \leftarrow \mathcal{K} : h(x_0, k) \oplus h(x_1, k) = y] \leq \varepsilon_u$.

We need a special notion of almost regular functions when the key is divided into two sub-keys, i.e., $h : \mathcal{M} \times \mathcal{K}_a \times \mathcal{K}_b \to \mathcal{T}$. A hash function h is $(\varepsilon_a, \varepsilon_b)$*-almost regular* w.r.t. the sub-keys if for each data $x \in \mathcal{M}$, tag $y \in \mathcal{T}$ and sub-keys $\widehat{k}_a \in \mathcal{K}$, $\widehat{k}_b \in \mathcal{K}_b$, we have $\Pr[k_a \leftarrow \mathcal{K}_a : h(x, k_a, \widehat{k}_b) = y] \leq \varepsilon_a$ and $\Pr[k_b \leftarrow \mathcal{K}_b : h(x, \widehat{k}_a, k_b) = y] \leq \varepsilon_b$. In particular, $(\varepsilon_a, \varepsilon_b)$-almost regularity implies that the inequalities hold even if y is drawn from a distribution that is independent from k_a and k_b. Finally, a hash function h is ε_u*-almost universal w.r.t. the sub-key* k_a if for any two data $x_0 \neq x_1$ and $k_b, \widehat{k}_b \in \mathcal{K}_b$, we have $\Pr[k_a \leftarrow \mathcal{K} : h(x_0, k_a, k_b) = h(x_1, k_a, \widehat{k}_b)] \leq \varepsilon_u$. We say that h is *strongly* ε_u*-almost universal w.r.t. the sub-key* k_a if for any $(x_0, k_b) \neq (x_1, \widehat{k}_b)$, we have $\Pr[k_a \leftarrow \mathcal{K} : h(x_0, k_a, k_b) = h(x_1, k_a, \widehat{k}_b)] \leq \varepsilon_u$. Note that $\varepsilon_u, \varepsilon_a, \varepsilon_b \geq 1/|\mathcal{T}|$ and the word 'almost' is skipped in the definitions if the latter equality holds.

Commitment schemes. A commitment scheme $\mathcal{C}om$ is specified by a triple of algorithms $(\mathsf{Gen}, \mathsf{Com}, \mathsf{Open})$. A setup algorithm Gen generates public parameters pk of the commitment scheme. The commitment function $\mathsf{Com}_{\mathsf{pk}} : \mathcal{M} \times \mathcal{R} \to \mathcal{C} \times \mathcal{D}$ transforms data $m \in \mathcal{M}$ into a commitment string c of fixed length and a decommitment value d. Usually $d = (m, r)$, where $r \in \mathcal{R}$ is the used randomness. Finally, correctly formed commitments can be opened, i.e., $\mathsf{Open}_{\mathsf{pk}}(c, d) = m$ for all $(c, d) = \mathsf{Com}_{\mathsf{pk}}(m, r)$. Incorrect decommitment values yield to a special abort value $\perp$. We often use a shorthand $\mathsf{Com}_{\mathsf{pk}}(m)$ to denote $\mathsf{Com}_{\mathsf{pk}}(m, r)$ with $r \leftarrow \mathcal{R}$. Basic properties of commitment

[4] We explicitly assume that adversarial code is executed on a universal Turing or RAM machine.

schemes are defined by hiding and binding games. A commitment scheme is (t, ε_1)-hiding if any t-time adversary A achieves advantage

$$\mathrm{Adv}^{\mathsf{hid}}_{\mathcal{C}om}(A) = 2 \cdot \left| \Pr\begin{bmatrix} \mathsf{pk} \leftarrow \mathsf{Gen}, s \leftarrow \{0,1\}, (x_0, x_1, \sigma) \leftarrow A(\mathsf{pk}) \\ (c_s, d_s) \leftarrow \mathsf{Com}_{\mathsf{pk}}(x_s) : \; A(\sigma, c_s) = s \end{bmatrix} - \frac{1}{2} \right| \leq \varepsilon_1 \ .$$

A commitment scheme is (t, ε_2)-binding if any t-time adversary A achieves advantage

$$\mathrm{Adv}^{\mathsf{bind}}_{\mathcal{C}om}(A) = \Pr\begin{bmatrix} \mathsf{pk} \leftarrow \mathsf{Gen}, (c, d_0, d_1) \leftarrow A(\mathsf{pk}) : \\ \perp \neq \mathsf{Open}_{\mathsf{pk}}(c, d_0) \neq \mathsf{Open}_{\mathsf{pk}}(c, d_1) \neq \perp \end{bmatrix} \leq \varepsilon_2 \ .$$

Non-malleable commitment schemes. Many notions of non-malleable commitments have been proposed in cryptographic literature [CIO98,FF00,DG03] starting from the seminal article [DDN91] by Dolev, Dwork and Naor. All these definitions try to capture requirements that are necessary to defeat man-in-the-middle attacks. We adopt the modernised version of [CIO98]—non-malleability w.r.t. opening. The definition is slightly weaker than the definition given in [DG03], as we assume that committed messages are independent from public parameters pk. Such choice allows to define non-malleability without a simulator similarly to the framework of non-malleable encryption [BS99].

Intuitively, a commitment scheme is non-malleable, if given a valid commitment c, it is infeasible to generate related commitments $c_1, \ldots, c_n$ that can be successfully opened after seeing a decommitment value d. Formally, an adversary is a quadruple $A = (A_1, A_2, A_3, A_4)$ of efficient algorithms where (A_1, A_2, A_3) represents an active part of the adversary that creates and afterwards tries to open related commitments and A_4 represents a distinguisher (sometimes referred as a target relation). The adversary succeeds if A_4 can distinguish between two environments World_0 (real world) and World_1 (environment where all adversaries are harmless). In both environments, Challenger computes $\mathsf{pk} \leftarrow \mathsf{Gen}$ and then interacts with adversary A:

1. Challenger sends pk to A_1 that outputs a description of an efficient message sampler MGen and a state σ_1. Then Challenger draws two independent samples $x_0 \leftarrow \mathsf{MGen}$, $x_1 \leftarrow \mathsf{MGen}$ and computes a challenge commitment $(c, d) \leftarrow \mathsf{Com}_{\mathsf{pk}}(x_0)$.
2. Challenger sends c, σ_1 to A_2 that computes a state σ_2 and a commitment vector $(c_1, \ldots, c_n)$ with arbitrary length. If some $c_i = c$ then Challenger stops A with $\perp$.
3. Challenger sends d, σ_2 to A_3 that must produce a *valid* decommitment vector $(d_1, \ldots, d_n)$. More precisely, Challenger computes $y_i = \mathsf{Open}_{\mathsf{pk}}(c_i, d_i)$. If some $y_i = \perp$ then Challenger stops A with $\perp$.[5]
4. In World_0 Challenger invokes $A_4(x_0, y_1, \ldots, y_n, \sigma_2)$ with the correct sample x_0 whereas in World_1 Challenger invokes $A_4(x_1, y_1, \ldots, y_n, \sigma_2)$ with the sample x_1.

[5] The latter restriction is necessary, as otherwise A_3 can send n bits of information to A_4 by refusing to open some commitments. The same problem has been addressed [DG03] by requiring that behaviour of A_4 should not change if y_i is replaced with $\perp$. The latter is correct but somewhat cumbersome, as static program analysis of A_4 is undecidable in theory. Also, in a real life protocol a honest party always halts when $y_i = \perp$ as in our model.

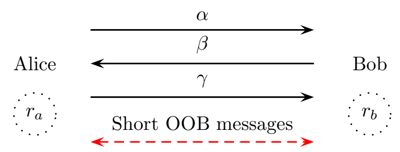

Fig. 1. Three round manual authentication protocol

The working time of A is the total time taken to run $A_1, \ldots, A_4$ and MGen. A commitment scheme is (t, ε)-non-malleable iff for any t-time adversary A the advantage of distinguishing the two worlds is

$$\mathrm{Adv}^{\mathsf{nm}}_{\mathit{Com}}(A) = |\Pr[A_4 = 0|\mathsf{World}_0] - \Pr[A_4 = 0|\mathsf{World}_1]| \leq \varepsilon \ .$$

The definition given above is natural in the concrete security framework, as it is conceptually clear and easy to apply. Also, the equivalence result between simulation and comparison based definition of non-malleable encryption [BS99] can be directly generalised.[6] Moreover, the definition of non-malleable encryption is stronger and therefore non-malleable encryption schemes (including CCA2 secure encryption schemes) can be used as non-malleable commitments provided that the public parameters pk are generated by the trusted party. See Sec. 5 for more detailed discussion.

3 Manual data authentication and key exchange protocols

Formal security model. Consider a three round *manual cross-authentication* protocol for data (depicted in Fig. 1) where messages α, β, γ are routed via an active adversary Charlie who can drop, delay, modify and insert messages. A low bandwidth out-of-band channel between Alice and Bob is bidirectional and authentic, but Charlie can arbitrarily delay OOB messages. As communication is asynchronous, Charlie can arbitrarily reorder in-band messages, e.g., Bob can receive $\widehat{\alpha}$ from Charlie even before Alice has sent α. Throughout the article, the hatted messages are received from Charlie and subscripts a and b denote values Alice and Bob compute locally. In particular, r_a, r_b denote random coins and m_a, m_b input data of Alice and Bob. The desired common output is (m_a, m_b) if both parties reach accepting state.

We assume that Alice and Bob send two OOB messages $\mathsf{oob}_{a\rightarrow b}$ and $\mathsf{oob}_{b\rightarrow a}$ in a fixed order during the protocol. Additionally, we require that the OOB messages have been specified in such a way that either both Alice and Bob accept the output or neither of them does. Often, the second OOB message just indicates whether the sender reached the accepted state. Charlie succeeds in deception if at the end of the protocol Alice and Bob reach the accepting state but $(m_a, \widehat{m}_b) \neq (\widehat{m}_a, m_b)$. A protocol is *correct* if Alice and Bob always reach the accepting state when Charlie does not intervene.

[6] Substitutions in the definitions and proofs of [BS99] are straightforward, except that there is no decommitment oracle and an isolated sub-adversary A_3 has to compute decommitment values.

Let A be an adversarial algorithm used by Charlie. Then the advantage of A against data authentication protocol is defined as

$$\mathrm{Adv}^{\mathsf{forge}}(A) = \max_{m_a,m_b} \Pr\left[\text{Alice and Bob accept } (m_a, \widehat{m}_b) \neq (\widehat{m}_a, m_b)\right]$$

where probability is taken over random coins of A and the honest participants. An authentication protocol is (t, ε)-secure if for any t-time adversary A, $\mathrm{Adv}^{\mathsf{forge}}(A) \leq \varepsilon$. We use the same adversarial model for user-aided key exchange protocols. Here, the number of exchanged messages might be larger than three and the desired common output consists of a fresh key key and a unique session identifier sid. A key exchange protocol is (ε, t)-immune against active attacks if, for any t-time adversary A,

$$\mathrm{Adv}^{\mathsf{forge}}(A) = \Pr\left[\text{Alice and Bob accept } (\mathsf{sid}, \mathsf{key}_a) \neq (\mathsf{sid}, \mathsf{key}_b)\right] \leq \varepsilon \enspace .$$

A (ε, t)-immune key exchange protocol is secure when it can resist passive attacks: ε quantifies the maximal security drop against active compared to passive attacks.

Clearly, the protocol outcome is determined by the first OOB message. Moreover, for the ℓ-bit message there exists an efficient deception strategy that with $\mathrm{Adv}^{\mathsf{forge}}(A) = 2^{-\ell}$. A protocol family is *asymptotically optimal* if it is possible to choose sub-primitives so that security level reaches asymptotically $2^{-\ell}$, see App. B for further discussion.

Other authors [Vau05,PV06a,PV06b] have used more complex framework [BR93] to define security. Such approach is needed only if consecutive runs of authentication protocols are not statistically independent, i.e., protocols use long-lived authentication keys. In our case, all protocol runs are statistically independent, i.e., given m_a and m_b or sid, a potential adversary can always perfectly simulate all protocol messages. Therefore, our protocols are secure in any computational context, see App. A.

New protocol families. Our new construction for cross-authentication protocols covers all currently known asymptotically optimal protocol families: a construction given by Pasini and Vaudenay [PV06b] and our earlier results [LAN05]. The protocol is depicted in Fig. 2. We explicitly assume that all public parameters are generated correctly by a trusted authority, i.e., we assume the common reference string model. Such assumption is not farfetched, as almost all communication standards provide some public parameters, e.g., descriptions of hash functions.

Protocol

1. Alice computes $(c, d) \leftarrow \mathsf{Com}_{\mathsf{pk}}(k_a)$ for random $k_a \leftarrow \mathcal{K}_a$ and sends (m_a, c) to Bob.
2. Bob chooses random $k_b \leftarrow \mathcal{K}_b$ and sends (m_b, k_b) to Alice.
3. Alice sends d to Bob, who computes $k_a \leftarrow \mathsf{Open}_{\mathsf{pk}}(c, d)$ and halts if $k_a = \perp$.
   Both parties compute a test value $\mathsf{oob} = h(m_a||m_b, k_a, k_b)$ from the received messages.
4. Both parties accept (m_a, m_b) iff the local ℓ-bit test values oob_a and oob_b coincide.

Specification: h is a keyed hash function with sub-keys k_a, k_b where $\mathcal{K}_a$ is a message space of commitment scheme. The hash function h and the public parameters pk of the commitment scheme are fixed and distributed by a trusted authority.

Fig. 2. Three round cross-authentication protocol Mana IV with ℓ-bit OOB messages.

Protocol

1. Alice computes $(c, d) \leftarrow \mathsf{Com}_{\mathsf{pk}}(k_a)$ for $k_a = g^a, a \leftarrow \mathbb{Z}_q$ and sends (id_a, c) to Bob.
2. Bob computes $k_b = g^b$ for random $b \leftarrow \mathbb{Z}_q$ and sends (id_b, k_b) to Alice.
3. Alice sends d to Bob, who computes $k_a \leftarrow \mathsf{Open}_{\mathsf{pk}}(c, d)$ and halts if $k_a = \perp$.
   Both parties compute $\mathsf{sid} = (\mathsf{id}_a, \mathsf{id}_b)$ and $\mathsf{oob} = h(\mathsf{sid}, k_a, k_b)$ from the received messages.
4. Both parties accept $\mathsf{key} = (g^a)^b = (g^b)^a$ iff the ℓ-bit test values oob_a and oob_b coincide.

Specification: h is a keyed hash function with sub-keys $k_a, k_b \in G$ where $G = \langle g \rangle$ is a q element Decisional Diffie-Hellman group; G is a message space of commitment scheme. Public parameters pk and G are fixed and distributed by a trusted authority. Device identifiers id_a and id_b must be unique in time, for example, a device address followed by a session counter.

Fig. 3. Manually authenticated Diffie-Hellman protocol MA–DH with ℓ-bit OOB messages

The Bluetooth authentication mechanisms are undergoing the standardisation phase and the current proposal for the standard [BT06] includes an instantiation of Mana IV (NUMERIC COMPARISON) among other methods. Our security analysis provides the necessary theoretical validation (A more detailed discussion is given in Sec. 5).

One can use the Mana IV protocol to authenticate the transcript of the classical Diffie-Hellman key exchange and thus prevent active attacks. Another reasonable alternative, proposed by Zimmermann and Wireless-USB standard group, is to fuse both protocols into a single one (See Fig. 3). Such solution reduces the number of random bits and computational complexity. Both are scarce resources in small electronic devices. The MA–DH protocol does not directly correspond to these protocols, as it uses commitments to hide g^a whereas these practical protocols use a cryptographic hash function $\mathcal{H}$ instead and set $c = \mathcal{H}(g^a)$. As a result our security proofs do not directly apply for protocols [ZJC06,WUS06]. Still the results give a some insight and provide suggestions how to achieve provable security (See Sec. 5).

Related work. The protocols by Pasini and Vaudenay [PV06b, Fig. 2 and 4] do not directly follow the structure of Mana IV, since in their first message $\alpha = c$ where $(c, d) = \mathsf{Com}_{\mathsf{pk}}(m_a||r_a)$ and $r_a \leftarrow \mathcal{K}_a$. In our security model, we can always assume that Charlie knows m_a, as m_a can be hardwired into the adversarial code. Therefore, if we send $\alpha = (m_a, c)$, the security level does not drop and sending m_a under the commitment becomes unnecessary. As the authenticated data m_a can be many kilobytes long, it also increases the message space for the commitment scheme. The latter can significantly decrease efficiency, as all currently known provably secure non-malleable commitment schemes are based on asymmetric cryptography.

A modified scheme with $(c, d) \leftarrow \mathsf{Com}_{\mathsf{pk}}(r_a)$ and $\alpha = (m_a, c)$ is a specific instance of Mana IV. We also note that in the security proofs of [Vau05,PV06b] it is assumed that the commitment is either a *simulation sound trapdoor commitment scheme* or a hiding one, even if adversary is allowed to query values for non-challenge commitments $c \neq c_s$. Both of these notions imply non-malleability [MY04], hence our assumptions are weaker. Moreover, in Sec. 4, we show that non-malleability of $\mathcal{C}om$ is also necessary, in a certain sense, to the security of the protocol. Finally, a secure fusion of [PV06b] and Diffie-Hellman key exchange similarly to MA–DH becomes problematic (See Sec. 5).

4 Security analysis of Mana IV and MA–DH protocols

The structure of Mana IV and MA–DH protocols forces adversary, Charlie, to fix data $\widehat{m}_a$ and $\widehat{m}_b$ before the complete hash key (k_a, k_b) becomes public. Hence, Charlie must either directly attack the hash function h or some property of commitment scheme to get extra knowledge about the hash key. A good message authentication code h provides security against simple substitution attacks and basic properties of commitment scheme along with non-malleability safeguard against more complex attacks.

Theorem 1 (Statistically binding commitments). *For any t, there exists $\tau = t + \mathcal{O}(1)$ such that if $\mathcal{C}om$ is (τ, ε_1)-hiding, ε_2-binding and (τ, ε_3)-non-malleable and h is $(\varepsilon_a, \varepsilon_b)$-almost regular and ε_u-almost universal w.r.t. the sub-key k_a then the Mana IV protocol is $(2\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3 + \max\{\varepsilon_a, \varepsilon_b, \varepsilon_u\}, t)$-secure. If additionally h is also strongly ε_u-almost universal w.r.t. the sub-key k_a, then the MA–DH protocol is $(2\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3 + \max\{\varepsilon_a, \varepsilon_b, \varepsilon_u\}, t)$-immune against active attacks.*

Theorem 2 (Computationally binding commitments). *For any t, there exists $\tau = 2t + \mathcal{O}(1)$ such that if $\mathcal{C}om$ is (τ, ε_1)-hiding, (τ, ε_2)-binding and (τ, ε_3)-non-malleable and h is $(\varepsilon_a, \varepsilon_b)$-almost regular and ε_u-almost universal w.r.t. the sub-key k_a then the Mana IV protocol is $(2\varepsilon_1 + \varepsilon_2 + \sqrt{\varepsilon_2} + \varepsilon_3 + \max\{\varepsilon_a, \varepsilon_b, \varepsilon_u\}, t)$-secure. If additionally h is also strongly ε_u-almost universal w.r.t. the sub-key k_a, then the MA–DH protocol is $(2\varepsilon_1 + \varepsilon_2 + \sqrt{\varepsilon_2} + \varepsilon_3 + \max\{\varepsilon_a, \varepsilon_b, \varepsilon_u\}, t)$-immune against active attacks.*

Proof. For clarity, the proof is split into Lemmata 1–5, as all (including passive) attacks can be divided into four disjoint classes. Combining the corresponding upper bounds on success probabilities proves the claims. □

Theorems 1 and 2 have several noteworthy implications. First, the Mana IV and MA–DH protocols are indeed asymptotically optimal, see Def. 1 in App. B, as one can choose h such that $\max\{\varepsilon_a, \varepsilon_b, \varepsilon_u\} = 2^{-\ell}$ and under standard complexity assumptions there exist commitment schemes where $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are negligible w.r.t. the security parameter if allowed working time τ is polynomial. Secondly, statistically binding commitments give better security guarantee than computationally binding ones: ε_2 vs. $\sqrt{\varepsilon_2}$. The latter is caused by the "non-trivial" reduction technique in Lemma 5. Thirdly, the slight difference in security objectives of Mana IV and MA–DH protocol manifests itself as an extra requirement to h. This is quite natural: if $m_a = \widehat{m}_a, m_b = \widehat{m}_b$ but $(k_a, \widehat{k}_b) \neq (\widehat{k}_a, k_b)$, we get a correct output for Mana IV but incorrect output for MA–DH, as $\mathsf{sid}_a = \mathsf{sid}_b$ but $\mathsf{key}_a \neq \mathsf{key}_b$. Finally, if Decisional Diffie-Hellman assumption holds for G, then MA–DH is approximately $2^{-\ell}$ secure key exchange protocol.

We give a formal security proof of Mana IV and MA–DH by constructing black box reductions corresponding to four different attack types. These reductions have the following structure: given an adversary A that is good in deception, we construct an adversary A^* that breaks some property of the commitment scheme. Generic construction of A^* is depicted on Fig. 4: in order to win a security game A^* simulates the original protocol and communicates with Challenger. As the communication is asynchronous, A can reorder protocol messages α, β, γ. Let $\mathsf{msg}_1 \prec \mathsf{msg}_2$ denote that msg_1 was output on a communication channel before msg_2. As Alice and Bob are honest, temporal restrictions $\alpha \prec \widehat{\beta} \prec \gamma$ and $\widehat{\alpha} \prec \beta \prec \widehat{\gamma}$ hold for all executions.

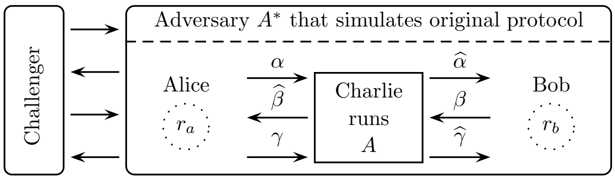

Fig. 4. Generic reduction scheme

An execution path is almost normal (denoted as norm) if the second round is completed before A starts the third round, i.e., $\alpha, \widehat{\alpha}, \beta, \widehat{\beta} \prec \gamma, \widehat{\gamma}$. Otherwise, one of the mutually exclusive events $\gamma \prec \beta$ or $\widehat{\gamma} \prec \widehat{\beta}$ must occur. For brevity, let $\mathsf{d\text{-}forge}$ denote that Alice and Bob accept $(m_a, \widehat{m}_b) \neq (\widehat{m}_a, m_b)$ in the Mana IV protocol and $\mathsf{k\text{-}forge}$ denote that Alice and Bob accept $(\mathsf{id}_a, \widehat{\mathsf{id}}_b, \mathsf{key}_a) \neq (\widehat{\mathsf{id}}_a, \mathsf{id}_b, \mathsf{key}_b)$ in the MA–DH protocol. Note that all probabilities in Lemmata 1–5 are taken over random coins of Gen, A and Alice and Bob and for a fixed input data (m_a, m_b) or identifiers $(\mathsf{id}_a, \mathsf{id}_b)$. As all proofs are quite straightforward but tedious, only the proof of Lemma 1 covers all details. All other proofs are more compact: some elementary steps are left to the reader.

Attacks based on almost normal execution paths. In the simplest attack, Charlie attacks directly h by altering only m_a, m_b, k_b and possibly γ. Charlie's aim here is to cleverly choose $\widehat{k}_b$ so that $\mathsf{oob}_a = \mathsf{oob}_b$. An attack where $k_b \neq \widehat{k}_b$ but other messages are unaltered can be successful against MA–DH but not against Mana IV. Strong ε_u-universality w.r.t the sub-key k_a provides appropriate protection against such attacks.

Lemma 1. *For any t, there exists $\tau = t + \mathcal{O}(1)$ such that if $\mathcal{C}om$ is (τ, ε_1)-hiding and (τ, ε_2)-binding and h is ε_u-almost universal w.r.t. the sub-key k_a, then for any t-time adversary A and input data (m_a, m_b)*

$$\Pr\left[\mathsf{d\text{-}forge} \wedge \mathsf{norm} \wedge c = \widehat{c}\right] \leq \varepsilon_u \cdot \Pr\left[\mathsf{norm} \wedge c = \widehat{c}\right] + \varepsilon_1 + \varepsilon_2 \ . \tag{1}$$

If additionally h is strongly ε_u-almost universal w.r.t. the sub-key k_a, then for any pair of identifiers $(\mathsf{id}_a, \mathsf{id}_b)$

$$\Pr\left[\mathsf{k\text{-}forge} \wedge \mathsf{norm} \wedge c = \widehat{c}\right] \leq \varepsilon_u \cdot \Pr\left[\mathsf{norm} \wedge c = \widehat{c}\right] + \varepsilon_1 + \varepsilon_2 \ . \tag{2}$$

Proof. ANALYSIS OF MANA IV. Assume a t-time algorithm A violates (1). Then $\Pr\left[\mathsf{d\text{-}forge} \wedge \mathsf{norm} \wedge c = \widehat{c} \wedge k_a = \widehat{k}_a\right] \geq \Pr\left[\mathsf{d\text{-}forge} \wedge \mathsf{norm} \wedge c = \widehat{c}\right] - \varepsilon_2$, or otherwise Alice and A together can open the commitment c to two different values $k_a \neq \widehat{k}_a$ with probability more than ε_2. The latter contradicts (τ, ε_2)-binding for $\tau = t + \mathcal{O}(1)$.

Next, we construct A^* that wins the hiding game, i.e., given pk outputs (x_0, x_1, σ) and afterwards given a commitment c_s for $s \leftarrow \{0, 1\}$, can correctly guess the bit s. The adversary A^* acts in the following way:

1. Given pk, chooses $k_a, k_a^* \leftarrow \mathcal{K}_a$ as (x_0, x_1) and sends $(k_a, k_a^*, \mathsf{pk})$ to Challenger.

2. When Challenger replies c_s for $(c_s, d_s) = \mathsf{Com}_{\mathsf{pk}}(x_s)$, A^* simulates a faithful execution of Mana IV with $\alpha = (m_a, c_s)$ until A queries γ. A^* stops the simulation and halts with $\perp$, if there is a protocol failure, $\neg\mathsf{norm}$ or $c \neq \widehat{c}$.
3. If $h(m_a||\widehat{m}_b, k_a, \widehat{k}_b) = h(\widehat{m}_a||m_b, k_a, k_b)$ and $(m_a, \widehat{m}_b) \neq (\widehat{m}_a, m_b)$ outputs a guess $s = 0$, else outputs a guess $s = 1$.

Now, consider when the simulation diverges from the real run of Mana IV with the same randomness r_a and r_b. If $s = 0$ then $(c_0, d_0) = \mathsf{Com}_{\mathsf{pk}}(k_a)$ and Step 3 does not reflect the protocol outcome in three disjoint cases: (a) abnormal execution or $c \neq \widehat{c}$, (b) $\widehat{\gamma}$ is not a valid decommitment (d-forge does not happen) and (c) $k_a \neq \widehat{k}_a$. Therefore, we get $\Pr[A^* = 0|s = 0] \geq \Pr[\mathsf{d\text{-}forge} \wedge \mathsf{norm} \wedge c = \widehat{c} \wedge k_a = \widehat{k}_a]$. For $s = 1$, we get $\Pr[A^* \neq \perp|s = 1] = \Pr[\mathsf{norm} \wedge c = \widehat{c}]$, as simulation is perfect until A queries γ. Since c_1 and k_a are statistically independent, all values computed by A are independent from k_a and thus $\Pr[A^* = 0|s = 1, A^* \neq \perp] \leq \varepsilon_u$. We arrive at a contradiction, as these bounds imply $\mathrm{Adv}^{\mathsf{hid}}(A^*) = |\Pr[A^* = 0|s = 0] - \Pr[A^* = 0|s = 1]| > \varepsilon_1$ and A^* runs in time $t + \mathcal{O}(1)$.

ANALYSIS OF MA–DH. Lets update only the forgery test in the last step of A^*:

3. If $h(\mathsf{id}_a||\widehat{\mathsf{id}}_b, k_a, \widehat{k}_b) = h(\widehat{\mathsf{id}}_a||\mathsf{id}_b, k_a, k_b)$ and $(\mathsf{id}_a, \widehat{\mathsf{id}}_b, \widehat{k}_b) \neq (\widehat{\mathsf{id}}_a, \mathsf{id}_b, k_b)$ output a guess $s = 0$, else output a guess $s = 1$.

Similarly to Mana IV, $\Pr[A^* = 0|s = 0] \geq \Pr[\mathsf{k\text{-}forge} \wedge \mathsf{norm} \wedge c = \widehat{c} \wedge k_a = \widehat{k}_a]$ and $\Pr[A^* \neq \perp|s = 1] = \Pr[\mathsf{norm} \wedge c = \widehat{c}]$, since the remaining code of A^* is identical. The new forgery test forces a restriction $(x_0, k_b) \neq (x_1, \widehat{k}_b)$ instead of $x_0 \neq x_1$ and we need strongly ε_u-universal h to bound $\Pr[A^* = 0|s = 1, A^* \neq \perp] \leq \varepsilon_u$. $\square$

Note 1. Strong ε_u-universality is necessary for the security of the MA–DH protocol, see Sec. 5 for a concrete counter example.

Another alternative is a direct attack against non-malleability where A tries to create "cleverly" related sub-keys k_a and $\widehat{k}_a$ to bypass the security check.

Lemma 2. *For any t, there exists $\tau = t + \mathcal{O}(1)$ such that if $\mathcal{C}om$ is (τ, ε_3)-non-malleable and h is $(\varepsilon_a, \varepsilon_b)$-almost regular, then for any t-time adversary A and inputs (m_a, m_b) or session identifier $(\mathsf{id}_a, \mathsf{id}_b)$*

$$\Pr[\mathsf{d\text{-}forge} \wedge \mathsf{norm} \wedge c \neq \widehat{c}] \leq \varepsilon_a \cdot \Pr[\mathsf{norm} \wedge c \neq \widehat{c}] + \varepsilon_3 \ , \tag{3}$$

$$\Pr[\mathsf{k\text{-}forge} \wedge \mathsf{norm} \wedge c \neq \widehat{c}] \leq \varepsilon_a \cdot \Pr[\mathsf{norm} \wedge c \neq \widehat{c}] + \varepsilon_3 \ . \tag{4}$$

Proof. Let A be a t-time algorithm that violates (3). Then we can build an adversary $A^* = (A_1^*, A_2^*, A_3^*, A_4^*)$ that can break non-malleability of the commitment scheme:

1. Given pk, A_1^* outputs a uniform sampler over $\mathcal{K}_a$ and a state $\sigma_1 = (\mathsf{pk}, m_a, m_b)$. Challenger computes $x_0, x_1 \leftarrow \mathcal{K}_a$ and $(c, d) \leftarrow \mathsf{Com}_{\mathsf{pk}}(x_0)$.
2. Given c, σ_1, A_2^* simulates the protocol with $k_b \leftarrow \mathcal{K}_b$ and stops before A demands γ. A^* stops the simulation and halts with $\perp$, if there is a protocol failure, $\neg\mathsf{norm}$ or $c = \widehat{c}$. Otherwise, A_2^* outputs a commitment $\widehat{c}$ and σ_2 containing enough information to resume the simulation including $(m_a, \widehat{m}_a, m_b, \widehat{m}_b, k_b, \widehat{k}_b)$.

3. Given d, σ_2, A_3^* resumes the simulation and outputs $\widehat{d}$ as a decommitment value.
4. If A_3^* was successful in opening $\widehat{c}$ then $A_4^*(x_s, y, \sigma_2)$ sets $k_a \leftarrow x_s$ and $\widehat{k}_a \leftarrow y$ and computes $\mathsf{oob}_a = h(m_a||\widehat{m}_b, k_a, \widehat{k}_b)$ and $\mathsf{oob}_b = h(\widehat{m}_a||m_b, \widehat{k}_a, k_b)$. If $\mathsf{oob}_a = \mathsf{oob}_b$ but $(m_a, \widehat{m}_b) \neq (\widehat{m}_a, m_b)$, then A_4^* outputs a guess $s = 0$, else outputs 1.

Again, consider where the simulation can diverge from the real execution of Mana IV. In both worlds, we can have a discrepancy if execution is abnormal or $c = \widehat{c}$. In World_0, Step 4 provides a perfect simulation whereas in World_1 k_a is independent of all variables computed by A. Therefore, using same methodology as before

$$\Pr\left[A_4^* = 0|\mathsf{World}_0\right] = \Pr\left[\mathsf{d\text{-}forge} \wedge \mathsf{norm} \wedge c \neq \widehat{c}\right] \ ,$$
$$\Pr\left[A_4^* = 0|\mathsf{World}_1\right] \leq \varepsilon_a \cdot \Pr\left[\mathsf{norm} \wedge c \neq \widehat{c}\right] ,$$

as h is $(\varepsilon_a, \varepsilon_b)$-almost regular. A contradiction as $\mathrm{Adv}^{\mathsf{nm}}(A^*) > \varepsilon_3$. For the MA–DH protocol, we have to refine the forgery test in Step 4 similarly to the proof of Lemma 1, but otherwise the analysis is exactly the same. □

Note 2. Obviously, non-malleability w.r.t. every target relation is not necessary. In particular, if h is fixed then it is necessary and sufficient that $\mathcal{C}om$ is secure for all adversaries having the same structure as in Lemma 2. The latter requirement is weaker than complete non-malleability, however, one has to reconsider the condition if h is substituted with a different function and technically such condition is not easier to prove.

Attacks based on abnormal execution paths. The remaining two attack patterns are easy to analyse, since they are direct attacks against binding and hiding properties. If $\widehat{\gamma} \prec \widehat{\beta}$ then successful A can predict k_a given only c and thus win the hiding game.

Lemma 3. *For any t there exists $\tau = t + \mathcal{O}(1)$ such that if $\mathcal{C}om$ is (τ, ε_1)-hiding, h is $(\varepsilon_a, \varepsilon_b)$-almost regular. Then for any t-time adversary A and input (m_a, m_b) or session identifier* $(\mathsf{id}_a, \mathsf{id}_b)$

$$\Pr\left[\mathsf{d\text{-}forge} \wedge \widehat{\gamma} \prec \widehat{\beta}\right] \leq \varepsilon_1 + \varepsilon_a \cdot \Pr\left[\widehat{\gamma} \prec \widehat{\beta}\right] \ , \tag{5}$$
$$\Pr\left[\mathsf{k\text{-}forge} \wedge \widehat{\gamma} \prec \widehat{\beta}\right] \leq \varepsilon_1 + \varepsilon_a \cdot \Pr\left[\widehat{\gamma} \prec \widehat{\beta}\right] \ . \tag{6}$$

Proof. Let A be a t-time adversary that violates (5). If $\widehat{\gamma} \prec \widehat{\beta}$, the Bob's control value oob_b is fixed before A receives γ. Consider A^* that plays the hiding game:

1. Given pk, chooses $k_a, k_a^* \leftarrow \mathcal{K}_a$ as (x_0, x_1) and sends $(k_a, k_a^*, \mathsf{pk})$ to Challenger.
2. When Challenger replies c_s for $(c_s, d_s) = \mathsf{Com}_{\mathsf{pk}}(x_s)$, A^* simulates a faithful execution of Mana IV with $\alpha = (m_a, c_s)$ until A outputs $\widehat{\beta}$. A^* stops the simulation and halts with $\perp$, if there is a protocol failure, $\widehat{\gamma} \not\prec \widehat{\beta}$ or $\mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d}) = \perp$.
3. Next A^* computes $\widehat{k}_a = \mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d})$, $\mathsf{oob}_a = h(m_a||\widehat{m}_b, k_a, \widehat{k}_b)$ and $\mathsf{oob}_b = h(\widehat{m}_a||m_b, \widehat{k}_a, k_b)$. If $\mathsf{oob}_a = \mathsf{oob}_b$ and $(m_a, \widehat{m}_b) \neq (\widehat{m}_a, m_b)$ outputs 0, else 1.

Again, consider where the simulation can diverge from the real protocol. If $s = 0$ then only $\widehat{\gamma} \not\prec \widehat{\beta}$ can cause the difference. For $s = 1$, simulation is perfect until γ is queried and thus $\Pr[A^* \neq \perp | s = 1] = \Pr[\widehat{\gamma} \prec \widehat{\beta}]$. As k_a is independent from oob_b, $\widehat{m}_b$ and $\widehat{k}_b$, then $\Pr[A^* = 0 | s = 1, A^* \neq \perp] \leq \varepsilon_a$ follows from $(\varepsilon_a, \varepsilon_b)$-almost regularity. A contradiction, as $\mathrm{Adv}^{\mathsf{hid}}(A^*) > \Pr[\mathsf{d\text{-}forge} \wedge \widehat{\gamma} \prec \widehat{\beta}] - \varepsilon_a \cdot \Pr[\widehat{\gamma} \prec \widehat{\beta}] > \varepsilon_1$. Same algorithm with a redefined forgery check is suitable for the MA–DH protocol. □

To win the remaining case $\gamma \prec \beta$, adversary A must double open $\widehat{c}$ to succeed. For statistically binding commitments, the reduction is simple. Analysis of computational binding commitments is more complex.

Lemma 4. *If $\mathcal{C}om$ is statistically ε_2-binding and h is $(\varepsilon_a, \varepsilon_b)$-almost regular, then for any adversary A and input (m_a, m_b) or session identifier $(\mathsf{id}_a, \mathsf{id}_b)$*

$$\Pr[\mathsf{d\text{-}forge} \wedge \gamma \prec \beta] \leq \varepsilon_2 + \varepsilon_b \cdot \Pr[\gamma \prec \beta] \quad , \tag{7}$$

$$\Pr[\mathsf{k\text{-}forge} \wedge \gamma \prec \beta] \leq \varepsilon_2 + \varepsilon_b \cdot \Pr[\gamma \prec \beta] \quad . \tag{8}$$

Proof. For each commitment $\widehat{c}$, fix a canonical $\widehat{k}_a$ such that $\widehat{k}_a = \mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d}_0)$ for some $\widehat{d}_0$. If $\gamma \prec \beta$ then oob_a is fixed before k_b. Now, the probability that different k_b values lead to different valid openings $\widehat{k}'_a \neq \widehat{k}_a$ is at most ε_2. Otherwise, one can find valid double openings $\mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d}_0) \neq \mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d}_1)$ just by enumerating all possible protocol runs. Now $\Pr[k_b \leftarrow \mathcal{K} : \mathsf{oob}_a = h(\widehat{m}_a || m_b, \widehat{k}_a, k_b)] \leq \varepsilon_b$, as k_b is independent form $\widehat{k}_a$ and oob_a and thus both claims follow. □

Lemma 5. *For any t there exists $\tau = 2t + \mathcal{O}(1)$ such that if $\mathcal{C}om$ is (τ, ε_2)-binding and h is $(\varepsilon_a, \varepsilon_b)$-almost regular, then for any t-time adversary A and inputs (m_a, m_b)*

$$\Pr[\mathsf{d\text{-}forge} \wedge \gamma \prec \beta] \leq \varepsilon_b \cdot \Pr[\gamma \prec \beta] + \sqrt{\varepsilon_2} \quad , \tag{9}$$

$$\Pr[\mathsf{k\text{-}forge} \wedge \gamma \prec \beta] \leq \varepsilon_b \cdot \Pr[\gamma \prec \beta] + \sqrt{\varepsilon_2} \quad . \tag{10}$$

Proof. Let A be a t-time adversary that violates (9). Consider A^* that

1. Simulates protocol run until A queries β and stores $\widehat{c}$. Halts if $\gamma \not\prec \beta$.
2. Provides $k_b^0, k_b^1 \leftarrow \mathcal{K}_b$ and outputs $\widehat{c}$ with the corresponding replies $\widehat{d}_0$ and $\widehat{d}_1$.

For a fixed pk and $\widehat{c}$, let $\varepsilon_{\mathsf{pk},\widehat{c}} = \Pr[\mathsf{d\text{-}forge} | \gamma \prec \beta, \mathsf{pk}, \widehat{c}]$ denote the forgery probability w.r.t. a single challenge k_b at Step 2 and

$$\delta_{\mathsf{pk},\widehat{c}} = \Pr[\perp \neq \mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d}_0) \neq \mathsf{Open}_{\mathsf{pk}}(\widehat{c}, \widehat{d}_1) \neq \perp | \gamma \prec \beta, \mathsf{pk}, \widehat{c}]$$

the success probability at Step 2. Then $\delta_{\mathsf{pk},\widehat{c}} \geq \varepsilon_{\mathsf{pk},\widehat{c}}(\varepsilon_{\mathsf{pk},\widehat{c}} - \varepsilon_b)$, since h is $(\varepsilon_a, \varepsilon_b)$-almost regular and oob_b^0 is fixed before k_b^1. Using a special case of Jensen's inequality, $\mathrm{E}(X^2) \geq \mathrm{E}(X)^2$ for any distribution of X, we get

$$\begin{aligned} \Pr[\mathsf{success} | \gamma \prec \beta] &= \sum_{\mathsf{pk},\widehat{c}} \Pr[\mathsf{pk} = \mathsf{Gen}, \widehat{c} | \gamma \prec \beta] \, (\varepsilon_{\mathsf{pk},\widehat{c}}^2 - \varepsilon_b \varepsilon_{\mathsf{pk},\widehat{c}}) \\ &\geq \Pr[\mathsf{d\text{-}forge} | \gamma \prec \beta]^2 - \varepsilon_b \Pr[\mathsf{d\text{-}forge} | \gamma \prec \beta] \quad . \end{aligned}$$

As $\Pr[\mathsf{d\text{-}forge} | \gamma \prec \beta] > \varepsilon_b$, we get $\Pr[\mathsf{success} | \gamma \prec \beta] \geq (\Pr[\mathsf{d\text{-}forge} | \gamma \prec \beta] - \varepsilon_b)^2$. Now from $\Pr[\gamma \prec \beta] \geq \Pr[\gamma \prec \beta]^2$, we obtain a contradiction

$$\mathrm{Adv}^{\mathsf{bind}}(A^*) \geq \Pr[\gamma \prec \beta]^2 \, (\Pr[\mathsf{d\text{-}forge} | \gamma \prec \beta] - \varepsilon_b)^2 > \varepsilon_2 \quad .$$

The same proof is valid also for the MA–DH protocol. □

Note 3. There are several alternatives to Lemma 5 that offer various tradeoffs between time τ and ε_2 depending how many times A is probed with different values of k_b. As A may totally break the Mana IV protocol on ε_2 fraction public parameters pk and do nothing for other values of pk, we cannot get a better bound than $\Pr[\mathsf{d\text{-}forge} \wedge \gamma \prec \beta] \leq \varepsilon_b \cdot \Pr[\gamma \prec \beta] + \varepsilon_2$ with black-box reductions. In our earlier work [LAN05], we used knowledge extraction techniques to obtain more complex reductions.

Note 4. Compared to proofs in [Vau05,PV06b] Lemma 5 seems to be inefficient and cumbersome. However, Vaudenay et al uses a different notion of binding—de facto they postulate Lemma 5 for a certain h as a security requirement. In asymptotic sense these notions are equivalent (there are polynomial reduction between them), but the exact security framework reveals that their condition is quantitatively much stronger.

In practical applications, commitments are constructed from cryptographic hash functions like SHA-1 and classical binding is more appropriate notion, since it leads directly to collision resistance. Secondly, Vaudenay's approach does not generalise for more complex constructions of h.

5 Practical Implementation Details

Security constraints. Mana IV and MA–DH protocols are secure in any computational context if (a) random values are never reused, (b) protocol outputs are never used before reaching the accepting state, (c) there are no multiple protocol instances between the *same* device pair at any time. Then a single protocol instance has same security guarantees as in Theorems 1 and 2. See App. A for a formal proof and discussion.

Hash functions. To instantiate Mana IV and MA–DH protocols, we need hash functions $h : \mathcal{M} \times \mathcal{K}_a \times \mathcal{K}_b \to \{0,1\}^\ell$ that are $(\varepsilon_a, \varepsilon_b)$-almost regular and (strongly) ε_u-almost universal w.r.t. the sub-key k_a. In our preliminary work [LAN05], we proposed a construction $h(m, k_a, k_b) = h_0(m, f(k_a, k_b))$ where h_0 is a ε_u-almost universal and ε_a-regular and $f : \mathcal{K}_a \times \mathcal{K}_b \to \{0,1\}^m$ is regular w.r.t. sub-keys k_a, k_b and for any $k_b \neq \widehat{k}_b$ the distribution of pairs $(f(k_a, k_b), f(k_a, \widehat{k}_b))$ for $k_a \leftarrow \mathcal{K}_a$ is statistically δ-close to uniform distribution. Then it is straightforward to verify that h is $(\varepsilon_a, \varepsilon_a)$-almost regular and $\max\{\varepsilon_a + \delta, \varepsilon_u\}$-almost universal, since for $k_b \neq \widehat{k}_b$ keys $f(k_a, k_b)$ are $f(k_a, \widehat{k}_b)$ almost independent.

As a concrete example let $f : \{0,1\}^{2m} \times \{0,1\}^{2m} \to \{0,1\}^m$ be defined as follows: $f(x_0||x_1, y) = x_0 y \oplus x_1$ in $\mathsf{GF}(2^m)$ if $x_0 \neq 0$ and $f(0^m||x_1, y) = x_1 \oplus y$ otherwise. Clearly, f is regular w.r.t. the sub-keys and $f(x_0, x_1, y_1) \oplus f(x_0, x_1, y_2) = x_0(y_1 \oplus y_2)$ covers $\mathsf{GF}(2^m) \setminus \{0\}$ when $y_1 \neq y_2$ and $x_0 \neq 0$. Hence, f is $(\varepsilon_a, \varepsilon_a)$-almost regular and $\max\{2^{-m+1} + \varepsilon_a, \varepsilon_u\}$-secure. Note that for proper choice of m, $2^{-m+1} \ll 2^{-\ell}$.

Pasini et al., [PV06b] proposed a construction $h(m, k_a, k_b) = h_0(m, k_a) \oplus k_b$ where h_0 is ε_u-almost XOR universal and ε_a-almost regular w.r.t. k_a. The latter is $(\varepsilon_a, 2^{-\ell})$-almost regular and strongly ε_u-almost universal. But such construction cannot be used in the MA–DH protocol, as k_b is typically at least 200 bits long. If we compress k_b in some manner, i.e., compute $h(m_a||m_b, k_a, k_b) = h_0(m_a||m_b, k_a) \oplus h_1(m_b, k_b)$ then the resulting hash function is only ε_u-almost universal. A malicious adversary can choose

$(m_b, k_b) \neq (m_b, \widehat{k}_b)$ such that $h_1(m_b, k_b) = h_1(m_b, \widehat{k}_b)$. Since ℓ is small in practical protocols, such pair can be found in real time and Charlie can indeed break the MA–DH protocol by choosing $\widehat{k}_b = g^c$ for $c \leftarrow \mathbb{Z}_q$ in this way. As a result Charlie and Alice share a common key. If Bob is a wireless router, then Charlie has successfully completed the attack, as he can transfer Alice's communication to Bob using secure channel between himself and Bob. Hence, the proper choice of h is extremely important.

For practical purposes $\mathcal{M} = \{0,1\}^{512}$ is sufficiently big, as one can always use a collision resistant hash functions to compress longer messages. And for such parameters many efficient ε_u-almost (XOR) universal and perfect hash functions are known with $\varepsilon_u \leq 2^{-\ell+1}$ (See [Sti91,BJKS93,NGR05] for some concrete examples).

Some practical proposals [BT06, p. 13, 21] propose use cryptographic hash functions to construct h. The latter is a plausible though heuristic choice, as long as statistical tests do not reveal a significant deviation from desired parameters $\varepsilon_a, \varepsilon_b, \varepsilon_u$. Otherwise, the potential adversary can discover and exploit these weaknesses.

Non-malleable commitment schemes. The simplest construction of a non-malleable commitment scheme is based on a CCA2 secure encryption scheme. Let $\mathsf{Enc}_{\mathsf{pk}} : \mathcal{M} \times \mathcal{R} \rightarrow \mathcal{C}$ be a deterministic encryption rule where $r \in \mathcal{R}$ denotes randomness used to encrypt a message. Define $(c, d) \leftarrow \mathsf{Com}_{\mathsf{pk}}(x, r)$ as $c = \mathsf{Enc}_{\mathsf{pk}}(x, r)$ and $d = (x, r)$ and $\mathsf{Open}_{\mathsf{pk}}(c, d) = m$ if $\mathsf{Enc}_{\mathsf{pk}}(x, r) = c$ and $\perp$ otherwise. Then the corresponding commitment scheme is non-malleable provided that pk is generated by a trusted party. We suggest Cramer-Shoup or Desmedt-Kurosawa encryption schemes [CS98,KD04], as the public key is a random tuple of group elements and can be easily generated without the secret key. RSA-OAEP is also CCA2 secure in a random oracle model [FOPS01]. Nevertheless, the key pk must be securely distributed, since *a priori* non-malleability w.r.t. pk does not guarantee non-malleability w.r.t. related keys pk_1 and pk_2.

All these constructions are *too inefficient* for small electronic devices and they offer too high levels of security. Recall that $\ell \lesssim 14$ and thus a commitment scheme should be roughly $(2^{80}, 2^{-20})$-non-malleable. Secure distribution of pk is another problem. In principle, it can be managed as there is only single public key, but may still not be well accepted for industrial applications. There are commitment schemes that are non-malleable without commonly shared pk, but these are very inefficient in practice.

In reality, a cryptographic hash functions like SHA-1 are used instead of commitments, as such constructions are hundred times faster and there are no setup assumptions. Let $\mathcal{H}$ be a collision resistant hash function. Then the hash commitment is computed as $(c, d) \leftarrow \mathsf{Com}(x, r)$ with $c = \mathcal{H}(x||r)$ and $d = (x, r)$ or, as in HMAC, $c = \mathcal{H}(r \oplus \mathsf{opad}||\mathcal{H}(r \oplus \mathsf{ipad}||x))$ with $d = r$ (See [BT06, p. 13] as an example). Both constructions are *a priori* not hiding. We would like to have a provably secure construction. In theory, we could use one-wayness of $\mathcal{H}$ and define commitment with hard-core bits but this leads to large commitments. Instead, we use Bellare-Rogaway random oracle design principle to heuristically argue that a hash commitment based on the OAEP padding is a better alternative. Recall that the OAEP padding is $c = \mathcal{H}(s, t), s = (x||0^{k_0}) \oplus g(r), t = r \oplus f(s)$. The corresponding commitment c along with $d = r$ is provably hiding and binding if g is pseudorandom, f is random oracle, and $\mathcal{H}$ is collision resistant. A priori SHA-1 and SHA-512 are not known to be non-malleable, as it has never been a design goal. On the other hand, the security proof of OAEP [FOPS01] shows CCA2 security (non-malleability) provided that $\mathcal{H}$ is a partial-domain one-way permutation. More specifically, it should be infeasible to find s given $h(s,t),\ s \leftarrow \mathcal{M}_1, t \leftarrow \mathcal{M}_2$. The partial one-wayness follows for $r, t \in \{0,1\}^{80}$ if we assume that $\mathcal{H}$ is at least $(2^{160}, 2^{-20})$-collision resistant as we can enumerate all possible t values to get a collision. The other assumption that h is a permutation is important in the proof. Therefore, we can only *conjecture* that the proof can be generalised and the OAEP padding provides a non-malleable commitment scheme.

Hence, an important theoretical task is to provide efficient but provably hiding and non-malleable but efficient padding construction for hash commitments. Also, one could reprove Lemma 1 and Lemma 3 without assuming hiding from $\mathcal{C}om$, as in both proofs we do not need hiding of k_a but just Charlie's inability to control Alice's $\mathtt{oob}_a$. Practical implementations [ZJC06,WUS06] of the MA–DH protocol use $c = \mathcal{H}(g^a)$ and such a relaxed security proof would bridge the gap between theory and practice.

Acknowledgements. We would like to thank N. Asokan for joint work on the initial solution and for many useful discussions and comments, and Emilia Käsper for helpful suggestions. The first author was partially supported by the Finnish Academy of Sciences and Estonian Doctoral School in Information and Communication Technologies.

References

[BJKS93] J. Bierbrauer, T. Johansson, G. Kabatianskii, and B. Smeets. On families of hash functions via geometric codes and concatenation. In *Proc. of CRYPTO '93*, LNCS 773. Springer, 1993.

[BM92] S. Bellovin and M. Merrit. Encrypted Key Exchange: Password-Based Protocols Secure Against Dictionary Attacks. In *Proc. of the IEEE Symposium on Security and Privacy*, pages 72–84, 1992.

[BR93] Mihir Bellare and Phillip Rogaway. Entity Authentication and Key Distribution. In *Proc. of CRYPTO '93*, LNCS 773, pages 232–249. Springer, 1993.

[BS99] Mihir Bellare and Amit Sahai. Non-malleable Encryption: Equivalence between Two Notions, and an Indistinguishability-Based Characterization. In *Proc. of CRYPTO '99*, LNCS 1666, pages 519–536. Springer, 1999.

[BT06] Bluetooth Special Interest Group. Simple Pairing Whitepaper (Revision V10r00). `http://www.bluetooth.com/Bluetooth/Apply/Technology/Research/Simple_Pairing.htm`, 2006.

[CCH06] M. Cagalj, S. Capkun, and J.-P. Hubaux. Key agreement in peer-to-peer wireless networks. *Proc. of the IEEE*, 94(2):467–478, Feb 2006.

[CIO98] Giovanni Di Crescenzo, Yuval Ishai, and Rafail Ostrovsky. Non-interactive and non-malleable commitment. In *STOC '98*, pages 141–150, 1998.

[CS98] Ronald Cramer and Victor Shoup. A Practical Public Key Cryptosystem Provably Secure Against Adaptive Chosen Ciphertext Attack. In *Proc. of CRYPTO '98*, LNCS 1462, pages 13–25. Springer, 1998.

[DDN91] Danny Dolev, Cynthia Dwork, and Moni Naor. Non-malleable cryptography. In *STOC '91*, pages 542–552, New York, NY, USA, 1991. ACM Press.

[DG03] Ivan Damgård and Jens Groth. Non-interactive and reusable non-malleable commitment schemes. In *STOC 2003*, pages 426–437, 2003.

[FF00] Marc Fischlin and Roger Fischlin. Efficient Non-malleable Commitment Schemes. In *Proc. of CRYPTO 2000*, LNCS 1880, pages 413–431. Springer, 2000.

[FOPS01] Eiichiro Fujisaki, Tatsuaki Okamoto, David Pointcheval, and Jacques Stern. RSA-OAEP Is Secure under the RSA Assumption. In *Proc. of CRYPTO 2001*, LNCS 2139, pages 260–274, 2001.

[GMN04] Christian Gehrmann, Chris J. Mitchell, and Kaisa Nyberg. Manual authentication for wireless devices. *RSA Cryptobytes*, 7(1):29–37, January 2004.

[Hoe05] Jaap-Henk Hoepman. Ephemeral Pairing on Anonymous Networks. In *Proc. of SPC 2005*, LNCS 3450, pages 101–116. Springer, 2005.

[KD04] Kaoru Kurosawa and Yvo Desmedt. A New Paradigm of Hybrid Encryption Scheme. In *Proc. of CRYPTO 2004*, LNCS 3152, pages 426–442. Springer, 2004.

[KOY01] Jonathan Katz, Rafail Ostrovsky, and Moti Yung. Efficient Password-Authenticated Key Exchange Using Human-Memorable Passwords . In *Proc. of EUROCRYPT 2001*, LNCS 2045, pages 475–494. Springer, 2001.

[LAN05] Sven Laur, N. Asokan, and Kaisa Nyberg. Efficient Mutual Data Authentication Using Manually Authenticated Strings: Preleiminary Version. Cryptology ePrint Archive, Report 2005/424, 2005. `http://eprint.iacr.org/`.

[LN06] Sven Laur and Kaisa Nyberg. Efficient Mutual Data Authentication Using Manually Authenticated Strings: Extended Version. Cryptology ePrint Archive, Report 2005/424, 2006. `http://eprint.iacr.org/`.

[MY04] Philip D. MacKenzie and Ke Yang. On Simulation-Sound Trapdoor Commitments. In *Proc. of EUROCRYPT*, LNCS 3027, pages 382–400. Springer, 2004.

[NGR05] Kaisa Nyberg, Henri Gilbert, and Matt Robshaw. Galois MAC with forgery probability close to ideal. General Public Comments on NIST Cryptopage, 2005.

[NSS06] Moni Naor, Gil Segev, and Adam Smith. Tight Bounds for Unconditional Authentication Protocols in the Manual Channel and Shared Key Models. In *Proc. of CRYPTO 2006*, LNCS 4117, pages 214–231 Springer, 2006.

[PV06a] Sylvain Pasini and Serge Vaudenay. An Optimal Non-interactive Message Authentication Protocol. In *Proc. of CT-RSA 2006*, LNCS 3860, pages 280–294. Springer, 2006.

[PV06b] Sylvain Pasini and Serge Vaudenay. SAS-Based Authenticated Key Agreement. In *PKC 2006*, LNCS 3958, pages 395–409. Springer, 2006.

[Sti91] D. R. Stinson. Universal Hashing and Authentication Codes. In *Proc. of CRYPTO '91*, LNCS 576, pages 74–85. Springer, 1991.

[Vau05] Serge Vaudenay. Secure Communications over Insecure Channels Based on Short Authenticated Strings. In *Proc. of CRYPTO 2005*, LNCS 3621, pages 309–326. Springer, 2005.

[WUS06] Association Models Supplement to the Certified Wireless Universal Serial Bus Specification, 2006. `http://www.usb.org/developers/wusb/`.

[ZJC06] Philip Zimmermann, Alan Johnston, and Jon Callas. ZRTP: Extensions to RTP for Diffie-Hellman Key Agreement for SRTP draft-zimmermann-avt-zrtp-01, March 2006.

A Security in arbitrary computational context

Assume that Mana IV and MA–DH protocols are run so that the security constraints presented in Sec. 5 are fullfilled. Then a protocol instance is uniquely determined by the time, origin and destination of the OOB message and a potential adversary cannot interleave OOB messages. This restriction can be trivially fullfilled—there is no need to exchange more than one key at a time and multiple messages can be sent together.

Consider ideal implementations of cross-authentication and Diffie-Hellman key exchange protocols. In ideal world, given m_a and m_b adversary can either deliver them to Alice and Bob or drop messages. Similarly, given g^a, g^b and sid, adversary can either do a passive attack against the key exchange protocol or interrupt it. Now consider a security game sec that defines security of a complex protocol. Next, theorem shows that the security drop compared to the ideal implementation is at most ε.

Theorem 3. *Let t_p be the total computational time needed to complete a complex protocol Π. For any t-time adversary A such that $\mathrm{Adv}^{\mathsf{sec}}_{\mathsf{real}}(A) = \delta$ in the real protocol, there exists a $(t + t_p)$-time adversary A^* that achieves $\mathrm{Adv}^{\mathsf{sec}}_{\mathsf{ideal}}(A^*) \geq \delta - \varepsilon$, if used Mana IV or MA–DH protocol is at least $(t + t_p + \mathcal{O}(1), \varepsilon)$-secure.*

Proof (Sketch). Since the source and destination of OOB messages together with the time uniquely reveal the corresponding Mana IV or MA–DH instance, it is straightforward to verify that honest Alice and Bob accept $m_a||\widehat{m}_b \neq \widehat{m}_a||m_b$ or $(\mathsf{sid}_a, \mathsf{key}_a) \neq (\mathsf{sid}_b, \mathsf{key}_b)$ with probability at most ε. Otherwise, we can simulate the surrounding computational context and break a stand-alone Mana IV or MA–DH protocol instance with larger probability than ε. Of course, the preceding history that determines (m_a, m_b) or sid should be fixed in such attack. As all previously used random coins can be hardwired, the corresponding attack still takes $t + t_p + \mathcal{O}(1)$ steps.

Now consider an adversary A^* that tries to win the game sec in the ideal world. It simulates a real protocol run to the adversary A. Essentially, A^* provides A a direct access to the ideal world except for the absent Mana IV or MA–DH protocol instance. Given (m_a, m_b) or $(\mathsf{id}_a, g^a, \mathsf{id}_b, g^b)$, A^* simulates the corresponding protocol to A. If A succeeds in deception, then A^* halts. Otherwise it simulates the end result of the protocol in the ideal world, i.e., delivers all messages unaltered or drops them. Note that when A^* does not halt then there is no discrepancy between the ideal and real protocol run. Since $\Pr[A^* \text{ halts}] \leq \varepsilon$ due to the first part of the proof, the result follows. □

Note 5. If many protocol instances can be run in parallel between the same device pair, then there are no security guarantees. When more than 2^ℓ protocols run in parallel, then Charlie can certainly swap OOB messages so that at least one attacked protocol reaches accepting state. Of course, such attack is not practical.

B Theoretical limitations

In the following, we show that there are no asymptotically optimal two round manual message authentication protocols. In other words, two round protocols are inherently less secure. However, the exact quantification of such security drop is out of our scope.

Here it is advantageous to consider unilateral authentication protocols since unilateral authentication is a special case of cross authentication. In a manual unilateral authentication protocol Sender wants to transfer a authentic message m to Receiver. The restrictions to the protocols are same: protocol must be correct, the order of all messages is fixed ahead and the first OOB message oob determines the protocol outcome. Let ℓ the maximal length of oob. We explicitly assume that $\ell \leq 30$, since for sufficiently large ℓ (say 160 bits) one can use collision resistant hash functions to protect authenticity, e.g., send $\mathsf{oob} = h(m)$. We also assume that the length of inputs m is larger than ℓ or otherwise we can send m as the first OOB message. Note that a simple collision attack where Charlie interacts honestly but uses $2^{\ell+1}$ pre-tabulated values for input and randomness gives a formal proof to the "folklore" bound.

Theorem 4. *Let π be a correct unilateral authentication protocol with fixed message order and let ℓ be the maximal length of the first out-of-band message. Then there exists a strategy A such that $\mathrm{Adv}^{\mathsf{forge}}_{\pi}(A) \geq 2^{-\ell}$.*

Proof. The proof is omitted due to the lack of space. The complete proof is given in the extended version of the current article [LN06]. □

Such strategy is feasible to follow in real-time for $\ell \leq 30$, as necessary pre-tabulated values can be hardwired into the code of A and then the computational resources needed to construct the actual program code are irrelevant.

Next we show that no two round protocols can achieve security bounds arbitrarily close to $2^{-\ell}$. A practical protocol must be secure against the attacks that take super-linear time w.r.t. the honest protocol run or otherwise the security margin is too small.

Definition 1. *We say that a protocol family $\{\pi_k\}$ with a fixed message ordering is asymptotically optimal when the maximal advantage ε_k with respect to the time-bound t_k approaches $\varepsilon_k \to 2^{-\ell}$ and t_k is at least super-linear in the protocol complexity.*

In principle, unilateral authentication protocols can have arbitrary structure. If we assume asymptotic optimality from the protocol family, then we can show that for large enough k, $\mathsf{oob}(m, r_r, r_s)$ is almost uniform w.r.t. to Receiver's randomness r_r and Sender's randomness r_s, and with high probability only a single value $\mathsf{oob}(m, r_r, r_s)$ leads to acceptance. Formally, we need a concept of uniform convergence to state these properties. A parametrised sequence $x_k(a)$ converges uniformly $x_k(a) \rightrightarrows x$ with respect to the parameter $a \in \{0,1\}^*$, if $\lim\limits_k \sup\limits_a x_k(a) = \lim\limits_k \inf\limits_a x_k(a) = x$.

Theorem 5. *Let $\{\pi_k\}$ be an asymptotically optimal and correct protocol family, let probability $\Pr[\cdot|\pi_k]$ be taken over honest runs of π_k and let $\mathsf{two\text{-}oob}$ denote the event that more than one value of oob lead to acceptance. Then next statements are true:*

(a) $\Pr[\mathsf{oob}|m, r_s, \pi_k] \rightrightarrows 2^{-\ell}$ *w.r.t. the parameters m, r_s,*
(b) $\Pr[\mathsf{two\text{-}oob}|m, \pi_k] \rightrightarrows 0$ *w.r.t. the parameter m,*
(c) $\Pr[\mathsf{oob}|r_r, m, \pi_k] \rightrightarrows 2^{-\ell}$ *w.r.t. the parameters r_r and m.*

Proof. Follows from an extended analysis of the collision attack, see [LN06]. □

Corollary 1. *There are no asymptotically optimal and correct two round protocol families with a fixed message order for unilateral authentication.*

Proof. Omitted, see the extended version of [LN06]. □

The stated result is rather weak, since it does not quantify how close to the optimal bound the deception probability of two round protocols can go. More accurate analysis is still an open question.