# Secure Group Key Establishment Revisited

Jens-Matthias Bohli<sup>1</sup>, María Isabel González Vasco<sup>2</sup>, and Rainer Steinwandt<sup>3</sup>

<sup>1</sup> Institut für Algorithmen und Kognitive Systeme, Universität Karlsruhe (TH), 76128 Karlsruhe, Germany;

bohli@ira.uka.de

<sup>2</sup> Departamento de Matemática Aplicada, Universidad Rey Juan Carlos, c/ Tulipán, s/n, 28933 Madrid, Spain;

mariaisabel.vasco@urjc.es

<sup>3</sup> Dept. of Mathematical Sciences, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA;

rsteinwa@fau.edu

Abstract. We examine the popular proof models for group key establishment of Bresson et al. [BCPQ01,BCP01] and point out missing security properties addressing malicious protocol participants. We show that established group key establishment schemes from CRYPTO 2003 and ASIACRYPT 2004 do not fully meet these new requirements. Next to giving a formal definition of these extended security properties, we prove a variant of the explored proposal from ASIACRYPT 2004 secure in this stricter sense. Our proof builds on the Computational Diffie Hellman (CDH) assumption and the random oracle model.

Keywords: Group Key Establishment, Provable Security, Malicious Insiders

# 1 Introduction

Group key establishments allow  $n \geq 2$  principals to agree upon a common secret key. It turns out that the design of such schemes faces some qualitatively new challenges that in this form do not arise in the two-party case. An excellent introduction and survey of the subject is given by Boyd and Mathuria [BM04]. To allow a rigorous security analysis, a framework for modelling group key establishments has been developed [BCP01,BCPQ01,KY03] going back on [BR93,BR95,BCK98,CK01,Sho99,BPR00]. A recent overview of those indistinguishability-based models is given in [CBH05].

An issue that compared to the two-party case becomes much more relevant is the protocol behavior in the presence of malicious participants: For groups with  $n \gg 2$  participants, assuming all participants to strictly follow the protocol specification can be a a rather strong assumption. Thus the question of security guarantees in the presence of malicious insiders naturally arises. Unfortunately, so far for the quite popular security models [BCP01] and [BCPQ01] extensions to malicious participants have hardly been explored and analyzed. In this type of model, security analysis typically restricts to the case of honest participants (see, e. g., [BN03,KY03,CBHM05]).

Our contribution. We examine two group key establishment protocols put forward in [KY03,KLL04] and point out insider attacks that are not covered by the model [BCPQ01] in which they are proven secure. We give a formal definition of some security notions motivated by these attacks. We put forward a notion of session integrity that can be seen as a correctness guarantee in the presence of malicious insiders. Further on, we suggest a way to formalize an agreement property in the style of [BCPQ01]. In a sense, these new notions round off the model [BCPQ01] and cover most known attacks that can be carried out in the presence of malicious insiders. In the model we consider, a long term key is associated with a principal  $U_i$ , i.e., it is identical for all protocol instances  $\Pi_i^{s_i}$  run by  $U_i$ . Once  $U_i$ 's long term key gets compromised through a corruption, we consider  $U_i$  as dishonest and do not try to establish security guarantees for individual protocol instances of  $U_i$ . In particular, we do not discuss key-compromise impersonation attacks where knowledge of a principal  $U_i$ 's long term secret is exploited to impersonate principals  $U_j \neq U_i$  towards  $U_i$ .

To give a flavor of how to design secure protocols in the new strict sense that we suggest, in Section 4.1 a "hidden" security feature in a proposal from [KY03] is proven. Finally, in Section 4.2 we present a modification of [KLL04] and prove it secure in our model—using the Computational Diffie Hellman (CDH) assumption as well as the random oracle model.

Related work. While we are not aware of a formal treatment of malicious insiders in the frameworks of [BCP01,BCPQ01], the issue of malicious insiders in group key establishment protocols has been addressed by several authors already. In particular, Saeednia and Safavi-Naini put forward several security classes for group key establishments [SSN98], and their class D.2 imposes that it is infeasible for any coalition of malicious insiders to break the authenticity of the conference key without no insider detecting the fraud. Also Cheng et al.'s list of attacks in [CVC04] mentions insider attacks, and the work of Tzeng [Tze00], for instance, shows that it is feasible to derive protocols with well-specified security guarantees even if a subset of the protocol participants acts maliciously.

Frameworks guaranteeing universal composability provide another approach to model key establishment protocols [Ste02] where insider attacks or failures are already considered [CS04,KS05a]. Katz and Shin [KS05a] in particular point out that their definition of insider impersonation attacks is stronger than the numerous varieties of insider attacks considered in [SSN98,CVC04] and present a protocol compiler to obtain protocols that are secure in this model. On the other hand, no efficient two-round protocol as [KLL04] is available in the UC framework and the protocol we give in Section 4.2 cannot be obtained by the protocol compiler of Katz and Shin. The formulation of key agreement in this setting is also unclear [HMQS03]. The definition of agreement in [KS05a] differs from the one we give below, and in particular does not quantify the influence maliciously acting protocol participants have on the session key. Another significant difference between the approach in [KS05a] and the one below is the role of the session identifier. Unlike [KS05a] we do not assume a session identifier to be available from a protocol-external context, but take it as a goal of the group key

establishment to come up with a session identifier that can serve as non-secret identifier for the established key.

# 2 Security Model and Security Goals

As indicated already, our basic security model is the proof model [BCPQ01] in the way it is used by Katz and Yung in [KY03]. Before we motivate and describe our extensions we give a short summary of the model:

Participants. We model the (potential) protocol participants as a finite set  $\mathcal{U}$  of fixed size with each  $U_i$  being a probabilistic polynomial time (ppt) Turing machine. Each protocol participant  $U_i \in \mathcal{P} \subseteq \mathcal{U}$  may execute a polynomial number of protocol instances in parallel. We will refer to instance  $s_i$  of principal  $U_i$  as instance  $\Pi_i^{s_i}$  ( $i \in \mathbb{N}$ ). Each such instance may be taken for a process executed by  $U_i$  and has assigned seven variables  $\mathsf{state}_i^{s_i}$ ,  $\mathsf{sid}_i^{s_i}$ ,  $\mathsf{pid}_i^{s_i}$ ,  $\mathsf{sk}_i^{s_i}$ ,  $\mathsf{term}_i^{s_i}$ ,  $\mathsf{used}_i^{s_i}$  and  $\mathsf{acc}_i^{s_i}$ :

 $\mathsf{used}_i^{s_i}$  indicates whether this instance is or has been used for a protocol run. The  $\mathsf{used}_i^{s_i}$  flag can only be set through a protocol message received by the instance due to a call to the Send-oracle (see below);

 $\mathsf{state}_i^{s_i}$  keeps the state information during the protocol execution;

 $term_i^{s_i}$  shows if the execution has terminated;

 $\mathsf{sid}_i^{s_i}$  denotes a (non-secret) session identifier that can serve as identifier for the session key  $\mathsf{sk}_i^{s_i}$ ;

 $\operatorname{pid}_{i}^{s_{i}}$  stores the set of identities of those principals that  $\Pi_{i}^{s_{i}}$  aims at establishing a key with—including  $U_{i}$  himself;

 $\mathsf{acc}_i^{s_i}$  indicates if the protocol instance was successful, i.e., the principal accepted the session key;

 $\mathsf{sk}_i^{s_i}$  stores the session key once it is accepted by the instance  $\Pi_i^{s_i}$ . Before acceptance, it stores a distinguished NULL value.

For more details on the usage of the variables see [BPR00]. We suppose that an instance  $\Pi_i^{s_i}$  must accept the session key constructed at the end of the corresponding protocol instance if no deviation from the protocol specification occurs.

Communication network. We assume arbitrary point-to-point connections among the principals to be available. As connections are potentially under adversarial control (cf. the adversarial model below) the network is non-private and fully asynchronous.

Adversarial model. The adversary  $\mathcal{A}$  is modeled as a ppt Turing machine and considered to be active:  $\mathcal{A}$  has full control of the communication network and may delay, eavesdrop, suppress, alter and insert messages at will. To make the adversary's capabilities explicit, the subsequently listed oracles are used that can be executed by  $\mathcal{A}$ .

Send $(U_i, s_i, M)$  This sends the message M to the instance  $\Pi_i^{s_i}$ . If  $\Pi_i^{s_i}$  sends a message in the protocol right after receiving M, then the Send-oracle returns this message to the adversary. If the oracle is called with an unused instance  $\Pi_i^{s_i}$  and  $M = \{U_1, \ldots, U_n\}$  with  $U_i \in \{U_1, \ldots, U_n\}$ , then  $\Pi_i^{s_i}$  spid $_i^{s_i}$ -value is initialized to M, the used  $_i^{s_i}$ -flag is set and  $\Pi_i^{s_i}$  processes the first step of the protocol. This means that in this session,  $U_i$  aims at establishing a common key with the principals specified in M.

Reveal $(U_i, s_i)$  yields the session key  $\mathsf{sk}_i^{s_i}$  provided that it is defined , i.e., if  $\mathsf{acc}_i^{s_i} = \mathsf{true}$  and  $\mathsf{sk}_i^{s_i} \neq \mathsf{NULL}$ . Otherwise the distinguished NULL-value is returned.

Corrupt $(U_i)$  reveals the long term secret key  $SK_i$  of  $U_i$  to the adversary. Given a concrete protocol run, involving instances  $\Pi_i^{s_i}$  of principals  $U_1, \ldots, U_n$ , we say that principal  $U_{i_0} \in \{U_1, \ldots, U_n\}$  is *honest* if and only if no query of the form  $Corrupt(U_{i_0})$  has ever been made by the adversary.

Test $(U_i, s_i)$  Only one query of this form is allowed for the adversary  $\mathcal{A}$ . Provided that  $\mathsf{sk}_i^{s_i}$  is defined, (i. e.  $\mathsf{acc}_i^{s_i} = \mathsf{true}$  and  $\mathsf{sk}_i^{s_i} \neq \mathsf{NULL}$ ),  $\mathcal{A}$  can execute this oracle query at any time when being activated. Then with probability 1/2 the session key  $\mathsf{sk}_i^{s_i}$  and with probability 1/2 a uniformly chosen random session key is returned.

As the session identifier  $\mathsf{sid}_i^{s_i}$  is intended to serve as public identifier for the session key  $\mathsf{sk}_i^{s_i}$ , we also grant the adversary access to any session identifiers of her choice.

Initialization. Before the actual key establishment protocol is executed for the first time, an initialization phase takes place where for each principal  $U_i \in \mathcal{P}$  a public key/secret key pair  $(SK_i, PK_i)$  is generated<sup>1</sup>,  $SK_i$  is revealed to  $U_i$  only, and  $PK_i$  is given to all principals and the adversary.

Correctness. This property basically expresses that the protocol will establish a good key without adversarial interference and allows us to exclude "useless" protocols. We take a group key establishment protocol for *correct* if in the absence of attacks a common key along with a common identifier is established:

**Definition 1.** A group key establishment protocol  $\mathcal{P}$  is called correct if upon honest delivery of all messages and no Corrupt-queries being made, a single execution of the protocol for establishing a key among  $U_1, \ldots, U_n$  involves n instances  $\Pi_1^{s_1}, \ldots, \Pi_n^{s_n}$  and ensures that with overwhelming probability all instances:

- $\ accept, \ i. \ e., \ \mathsf{acc}_1^{s_1} = \dots = \mathsf{acc}_n^{s_n} = \mathsf{true}.$
- obtain a common session identifier  $\operatorname{sid}_1^{s_1} = \cdots = \operatorname{sid}_n^{s_n}$  which is globally unique.

<sup>&</sup>lt;sup>1</sup> For the sake of simplicity we assume these key pairs to be generated by a trusted party, i.e., we do not consider malicious parties who try to generate incorrect key pairs. Also, we do not consider scenarios where only low-entropy secrets, like passwords, are available for authentication.

- have accepted the same session key  $\mathsf{sk}_1^{s_1} = \cdots = \mathsf{sk}_n^{s_n} \neq \mathsf{NULL}$  associated with the common session identifier  $\operatorname{sid}_1^{s_1}$ .

  - know their partners  $\operatorname{pid}_1^{s_1} = \operatorname{pid}_2^{s_2} = \cdots = \operatorname{pid}_n^{s_n} = \{U_1, \dots, U_n\}$ .

Partnering. For detailing the security definition, we will have to specify under which conditions a Test-query may be executed. To do so we fix the following notion of partnering.

Freshness. A Test-query should only be allowed to those oracles holding a key that is not for trivial reasons known to the adversary. An instance  $\Pi_i^{s_i}$  is called *fresh* if none of the following two conditions hold:

- For some  $U_j \in \mathsf{pid}_i^{s_i}$  a  $\mathsf{Corrupt}(U_j)$  query was executed before a query of the form  $\mathsf{Send}(U_k, s_k, *)$  has taken place where  $U_k \in \mathsf{pid}_i^{s_i}$ .
- The adversary  $\mathcal{A}$  queried Reveal $(U_j, s_j)$  with  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  being partnered.

The idea here is that revealing a session key from an oracle  $\Pi_i^{s_i}$  trivially yields the session key of all oracles partnered with  $\Pi_i^{s_i}$ , and hence this kind of "attack" will be excluded in the security definition. While the second condition seems pretty natural, imposing the first condition might look too restrictive. It is adopted from [BGVS05] and aims at precluding (insider) attacks where, once a subset of the principals has computed the session key, this subset is corrupted and the last outgoing messages are altered and correctly signed so that some honest protocol participants end up with a different session identifier but identical session key. Then the honest participants are not partnered with the corrupted ones, and breaking the security of the protocol becomes trivial. In the next section, we will discuss such a scenario for a protocol of Katz and Yung.

Security. The security definition of [BCPQ01] can be summarized as follows. As a function of the security parameter k we define the advantage  $Adv_{\mathcal{A}}(k)$  of a ppt adversary A in attacking protocol P as

$$\mathsf{Adv}_A := |2 \cdot \mathsf{Succ} - 1|$$

where Succ is the probability that the adversary queries Test on a fresh instance  $\Pi_i^{s_i}$  and guesses correctly the bit b used by the Test oracle.

**Definition 3.** We call the group key establishment protocol P secure if for any ppt adversary A the function  $Adv_A = Adv_A(k)$  is negligible.

#### 3 **Extended Security Properties**

Unfortunately, established protocols that are proven secure in the above model can be vulnerable to annoyingly simple attacks if one considers a slightly broader scenario. In this section we explore the protocol of Katz and Yung [KY03] that goes back to Burmester and Desmedt [BD95] and a very efficient protocol from Kim, Lee and Lee [KLL04]. We present new attacks on these protocols, but we stress that these attacks were not considered in the security model where they are proven secure: Hence our discussion does not invalidate the security proofs given by the authors. Nevertheless we think such vulnerabilities are relevant and should indeed be prevented.

#### 3.1 Attacks on a Proposal of Katz and Yung

At CRYPTO 2003, Katz and Yung put forward a three round group key agreement [KY03] building on the protocol of [BD95]. In an initialization phase a finite cyclic group  $\mathbb G$  of prime order q and a generator g of  $\mathbb G$  is chosen such that the Decisional Diffie Hellman (DDH) assumption holds. We summarize the fundamentals of the protocol for establishing a key among  $\{U_1,\ldots,U_n\}$ , where indices are to be taken in a cycle. A detailed overview of the exchanged messages is given in Figure 1. Arbitrary point to point connections among participants are available, and a broadcast is understood as simultaneous point to point delivery of messages to all intended recipients. The participants exchange nonces in the first round to get a unique session. In the following, the participants broadcast  $z_i = g^{r_i}$  and compute a Diffie-Hellman key with each of their neighbors. In the third round, the participants compute the quotient of the keys shared with their two neighbors  $X_i = (z_{i+1}/z_{i-1})^{r_i}$  and broadcast this value. It is now possible for all participants to compute the key  $\mathsf{sk}_i^{s_i} = (z_{i-1})^{nr_i} \cdot X_i^{n-1} \cdot X_{i+1}^{n-2} \cdots X_{i+n-2}$ .

Using a model close to the one outlined in Section 2, in [KY03] this protocol is shown to be secure. At this, it is assumed that the signature scheme used is not only secure against existential forgeries under adaptive chosen message attacks, but with overwhelming probability also prevents an attacker from producing a different signature for an already signed message.

Violating the integrity of a session. Let us assume the adversarial goal is now to prevent a certain session unnoticeably from succeeding, forcing some involved principals to obliviously compute a different session key with the same session identifier. (In Definition 4 we will formalize resilience against this type of attack as integrity.) Say n > 3 and  $\operatorname{ord}(g)$  are coprime, then the adversary  $\mathcal A$  can mount the following (insider) attack:

- 1.  $\mathcal{A}$  corrupts  $U_1$  and  $U_3$  (henceforth blocking any communication from and to these parties).
- 2. In the the 3rd protocol round,  $\mathcal{A}$  computes  $X_1, X_3$  as specified, but then sets  $\widetilde{X}_1 := X_3, \widetilde{X}_3 := X_1$ . Now  $\mathcal{A}$  signs  $(U_1 || 2 || \widetilde{X}_1 || t)$  with  $U_1$ 's signing key, signs  $(U_3 || 2 || \widetilde{X}_3 || t)$ 's with  $U_3$ 's signing key and then broadcasts  $(U_1 || 2 || \widetilde{X}_1 || \sigma_1^{\text{II}})$  and  $(U_3 || 2 || \widetilde{X}_3 || \sigma_3^{\text{II}})$ . In other words,  $\mathcal{A}$  swaps the  $X_i$ -contributions of  $U_1$  and  $U_3$ .

Now all protocol participants compute the same session identifier, all of them receive the same messages, but with overwhelming probability the (honest) participants  $U_2$  and  $U_4$  will have derived different session keys: With the notation

#### Round 1:

**Broadcast** Each  $U_i$  chooses a random nonce  $t_i \in \{0,1\}^k$  and broadcasts  $(U_i||0||t_i)$ .

**Computation** Each  $U_i$  waits until messages  $(U_j||0||t_j)$  for all  $U_j$  arrived and sets  $t := t_1||\dots||t_n|$ .

#### Round 2:

Computation Each  $U_i$  chooses a random  $r_i \in \mathbb{Z}_q$  and computes  $z_i = g^{r_i}$  and a signature  $\sigma_i^{I}$  of  $(1||z_i||\mathsf{pid}_i^{s_i}||t)$ .

**Broadcast** Each  $U_i$  broadcasts  $(U_i||1||z_i||\sigma_i^{\mathrm{I}})$ .

**Check** Each  $U_i$  waits for all incoming messages  $(U_j||1||z_j||\sigma_j^{\rm I})$  and checks all signatures  $\sigma_j^{\rm I}$ .

#### Round 3:

**Computation** Each  $U_i$  computes  $X_i = (z_{i+1}/z_{i-1})^{r_i}$  and a signature  $\sigma_i^{II}$  of  $(2\|X_i\|\operatorname{pid}_i^{s_i}\|t)$ .

**Broadcast** Each  $U_i$  broadcasts  $(U_i||2||X_i||\sigma_i^{II})$ .

**Check** Each  $U_i$  waits for all incoming messages  $(U_j||2||X_j||\sigma_j^{\text{II}})$  and checks all signatures  $\sigma_i^{\text{II}}$ .

**Key computation:** Each participant  $U_i$  computes the session key  $\mathsf{sk}_i^{s_i} = (z_{i-1})^{nr_i} \cdot X_i^{n-1} \cdot X_{i+1}^{n-2} \cdots X_{i+n-2}$ . The session identifier  $\mathsf{sid}_i^{s_i}$  is the concatenation of all messages that were sent and received.

Fig. 1. A group key establishment protocol from CRYPTO 2003 [KY03].

from Figure 1 a simple computation shows that the quotient of  $U_2$ 's and  $U_4$ 's session keys is  $X_3^n \cdot (z_2/z_4)^{r_3n} = 1_{\mathbb{G}}$  without and  $X_1^n \cdot (z_2/z_4)^{r_3n}$  with  $U_1$  and  $U_3$  swapping their  $X_i$ -contributions in the 3rd protocol round. Thus, in the latter case the keys derived by  $U_2$  and  $U_4$  coincide with negligible probability only. This actually violates the *correctness* of [KY03] where matching keys are required for all instances that accepted in the same session. As this is a non-trivial attack we restricted the correctness explicitly to honest executions and will introduce a broadened requirement—named *integrity*—in Section 3.3.

Moreover it is now easy to see that in the above scheme not every participant contributes to the session key. In fact, the key can be completely determined by an adversary corrupting two neighboring principals  $U_i$ ,  $U_{i+1}$ . In Section 4.1 we prove that corrupting only one principal does not suffice for successfully attacking this scheme in a similar fashion.

# 3.2 Attacks on a Proposal of Kim, Lee and Lee

At ASIACRYPT 2004, Kim, Lee and Lee presented an efficient authenticated group key agreement protocol [KLL04], which is claimed to take precautions against "illegal members or system faults". No formal definition or security proof for this is provided, however, and below we will see that the protocol does not meet strong security guarantees as one malicious participant is sufficient to violate integrity and to mount an impersonation attack.

Figure 2 outlines Kim, Lee and Lee's proposal for establishing a key among  $\{U_1, \ldots, U_n\}$ , where again indices are to be taken in a cycle. Similarly as in the

proposal of Katz and Yung, during an initialization phase a cyclic group G of prime order q along with a generator q is chosen such that the CDH assumption holds; the hash function  $H(\cdot)$  is modelled as random oracle and again broadcast is understood as simultaneous point to point delivery of messages. The protocol begins with the participants broadcasting  $y_i = g^{x_i}$ , again to establish Diffie-Hellman keys  $t_i^L, t_i^R$  with their two neighbored participants. In the second round the participants broadcast the XOR sum  $T_i = t_i^L \oplus t_i^R$  of their two keys to allow all participants to compute all shared keys. Moreover they broadcast a nonce  $k_i$ as contribution to the session key, though one participant broadcasts his nonce encrypted  $k_n \oplus t_n^R$ . Now all participants can compute the nonces and the session key  $\mathsf{sk}_i^{s_i} = H(k_1 \| \dots \| k_n \| 0).$ 

```
Round 1:
```

**Computation** Each  $U_i$  chooses  $k_i \in \{0,1\}^k$ ,  $x_i \in \mathbb{Z}_q^*$  and computes  $y_i = g^{x_i}$ , only  $U_n$  computes additionally  $H(k_n||0)$ . Each  $U_i$  except  $U_n$  sets  $M_i^I = y_i$ and  $U_n$  sets  $M_n^{\rm I} = H(k_n || 0) || y_n$ . Each  $U_i$  computes a signature  $\sigma_i^{\rm I}$  on  $M_i^1 \| \mathsf{pid}_i^{s_i} \| 0.$ 

Broadcast Each  $U_i$  broadcasts  $(M_i^{\rm I} \| \sigma_i^{\rm I})$ . Check Each  $U_i$  checks all signatures  $\sigma_j^{\rm I}$  of incoming messages  $(M_j^{\rm I} \| \sigma_j^{\rm I})$ .

#### Round 2:

Computation Each  $U_i$  computes  $t_i^L = H(y_{i-1}^{x_i} || \operatorname{pid}_i^{s_i} || 0), t_i^R$ However, the participants  $U_i$  computes a signature  $\sigma_i^{\mathrm{II}} = H(y_{i+1}^{\mathrm{II}} \| \mathrm{pid}_i^{s_i} \| 0)$  and  $T_i = t_i^{\mathrm{II}} \oplus t_i^{\mathrm{R}}$ , only  $U_n$  computes additionally  $k_n \oplus t_n^{\mathrm{R}}$ . The participants  $U_1, \ldots, U_{n-1}$  set  $M_i^{\mathrm{II}} = k_i \| T_i, U_n$  sets  $M_n^{\mathrm{II}} = k_n \oplus t_n^{\mathrm{R}} \| T_n$  and each  $U_i$  computes a signature  $\sigma_i^{\mathrm{II}}$  of  $M_i^{\mathrm{II}} \| \mathrm{pid}_i^{s_i} \| 0$ .

**Broadcast** Each  $U_i$  broadcasts  $(M_i^{\text{II}} || \sigma_i^{\text{II}})$ .

**Check** Firstly, each  $U_i$  checks all signatures  $\sigma_j^{\text{II}}$  of incoming messages. Then each  $U_i$  checks if  $T_1 \oplus \cdots \oplus T_n = 0$ , computes  $t_n^R$  to obtain  $k_n$  from  $U_n$ 's message and checks the commitment  $H(k_n||0)$  for  $k_n$ .

**Key computation:** Each participant  $U_i$  computes the session key  $\mathsf{sk}_i^{s_i} =$  $H(k_1\|\ldots\|k_n\|0).$ 

Fig. 2. A group key establishment protocol from ASIACRYPT 2004 [KLL04].

Attacks on the security and integrity. In [KLL04] it is not specified how to generate the session identifier  $sid_i^{s_i}$ , and it turns out that the standard method of concatenating all messages an oracle sent and received is not enough to prove it secure: For n > 3, an adversary A could proceed as follows to provoke a situation where  $U_1$  and  $U_3$  end up with different session identifiers (hence not being partnered) but identical session key  $sk_1^{s_1} = sk_3^{s_3}$ :

- 1.  $\mathcal{A}$  executes a complete protocol run and eavesdrops the message  $(M_1^{\mathrm{I}} \| \sigma_1^{\mathrm{I}})$ broadcast by  $U_1$  in Round 1.
- 2.  $\mathcal{A}$  initiates another protocol execution, but in Round 1 replaces the message sent from  $U_1$  to  $U_3$  with the old  $(M_1^1 || \sigma_i^1)$ -value eavesdropped in the previous protocol run.

Because of  $U_3$  not being a neighbor of  $U_1$ , this message substitution does not affect the computation of the session key, but with overwhelming probability  $U_3$ ends up with a session identifier different from the session identifier computed by  $U_1$  and the respective oracles of  $U_1$  and  $U_3$  will not be partnered. Therefore, the key of  $U_1$  can be revealed but  $U_3$  remains fresh. Also the attack from Section 3.1 aiming at a different session key under the same session identifier applies here in an analogous way.

To avoid this kind of "trivial" problems, subsequently we assume the session identifier  $sid_i^{s_i}$  to be derived as

$$\operatorname{sid}_{i}^{s_{i}} = H(k_{1} \| \dots \| k_{n-1} \| H(k_{n} || 0)),$$

so that identical session identifiers with overwhelming probability correspond with identical session keys. However, this session identifier still allows a single protocol run ending up with different session identifiers. This can be provoked by simply having a malicious participant  $U_1$  in Round 2 sending different  $k_1$ -values to the other protocol participants (instead of broadcasting one  $k_1$ -value).

An impersonation attack Independent from the choice of the session identifier and potentially more severe is the following impersonation attack. For n > 2participants an adversary A can impersonate participants as follows:

- 1. First, she gets herself a protocol transcript of a successful key establishment among principals  $U_1, \ldots, U_n$ . Next,  $\mathcal{A}$  reveals  $U_1$ 's long term secret by querying  $Corrupt(U_1)$ .
- 2. Now  $\mathcal{A}$  initializes unused oracles of  $U_3, \ldots, U_n$  with  $\operatorname{pid}_j^{s_j} = \{U_1, \ldots, U_n\}$ . 3. In Round 1 she replays the message that  $U_2$  sent in the previously eavesdropped key establishment and participates honestly for  $U_1$ .
- 4. In Round 2,  $\mathcal{A}$  again replays  $U_2$ 's message from the eavesdropped protocol run. On behalf of  $U_1$  the adversary computes

$$T_1 := T_2 \oplus \cdots \oplus T_n$$

and broadcasts the signed message  $(M_1^{\text{II}} \| \sigma_1^{\text{II}})$  with  $M_1^{\text{II}} = k_1 \| T_1$ .

Now all participants can compute the session key and will accept it as common secret key among  $U_1, \ldots, U_n$  although the honest principal  $U_2$  never took part in the session.

#### **Definition of Extended Security Goals** 3.3

The model in [BR93] goes further in its definition of security than the model in [BCPQ01]: Building on the notion of a matching session, a protocol is called secure if besides the usual negligible advantage in guessing the session key it also holds, that a matching session results in the participants accepting the same key. In a group key establishment protocol it is more appropriate to identify matching sessions via a session identifier. So in analogy to the two-party case, a matching session identifier should result in matching session keys. In the case of  $n \gg 2$  protocol participants, there is significant potential for insider attacks, however, and malicious insiders indeed matter in group key establishment protocols: Even if several principals are dishonest, there can still remain numerous honest protocol participants.

Granted, in the presence of malicious participants the adversary always learns the session key, for honest participants the situation can still differ. For some applications it could even be more relevant to prevent the case in which honest principals share mismatching keys than the case where the correct key is shared with unintended principals. For instance, if the keys serve as access control passwords for shared data, then the above attacks result in situations in which principals assume others to have access rights which they may actually not have. We therefore propose the following notions to extend the security of group key establishments.

Session integrity. Motivated by the security definition of [BR93] and [CK01] we introduce an integrity property also for group key establishments to prevent sessions to mix up under adversarial influence. This property will basically be an extension of the correctness to active adversaries and malicious insiders.

We have seen an attack in Section 3.1 on the protocol of Katz and Yung where a malicious participant could run a protocol where two honest participants ended up with the same session identifier but different session keys. Such a situation possibly invalidates vital assumptions on the application level without the participants having a chance to detect it.

Recall also that the notion of correctness of a key establishment prevents partners from accepting when they have different  $\operatorname{pid}_i^{s_i}$ -values—i.e. they aim at establishing the key with different sets of users. We want to keep this property also in the presence of malicious insiders. For instance, assume a group key establishment, where a malicious  $U_1$  could convince  $U_2$  to have partners  $\{U_1, U_2, U_3\}$  and  $U_3$  to have partners  $\{U_1, U_2, U_3, U_4\}$  when indeed only  $U_1, U_2$  and  $U_3$  have a common key. Then the honest principals  $U_2$  and  $U_3$  will not agree whether the subsequent application is confidential with respect to  $U_4$  or not. To avoid such situations, in our definition of integrity we impose that a matching session identifier should also result in a matching partner identifier.

**Definition 4.** We say a correct group key establishment protocol fulfills integrity if with overwhelming probability all instances of honest principals that have accepted with the same session identifier  $\operatorname{sid}_j^{s_j}$  hold identical session keys  $\operatorname{sk}_j^{s_j}$  and associate this key with the same principals  $\operatorname{pid}_j^{s_j}$ .

Strong entity authentication. Entity authentication is a relevant issue for key establishment even excluding the possibility of corrupted participants. It is considered in the model [BR93] and in the models for password-based key establishment following [BPR00]. Again malicious insiders are significantly stronger in violating this property, as seen in the attack scenarios in Section 3.2. An approach to define entity authentication formally was made in [JG04]. For our

security model dealing with group key establishment we rephrase this definition as follows.

**Definition 5.** Strong entity authentication to an oracle  $\Pi_i^{s_i}$  is provided if both  $\mathsf{acc}_i^{s_i} = \mathsf{true}$  and for all honest  $U_j \in \mathsf{pid}_i^{s_i}$  with overwhelming probability there exists an oracle  $\Pi_j^{s_j}$  with  $\mathsf{sid}_j^{s_j} = \mathsf{sid}_i^{s_i}$  and  $U_i \in \mathsf{pid}_j^{s_j}$ .

For the two-party case, the above definition is close to the mutual authentication requirements in [JG04], but instead of imposing exchanged messages seen by instances  $\Pi_i^{s_i}$  and  $\Pi_i^{s_j}$  to be equal, we require equality of the session identifiers.

Key agreement. Člearly, key freshness can never be guaranteed in the presence of malicious participants if some incomplete subset of principals is able to predetermine the key. However, if the key establishment is contributory, that is, if all parties must be involved in the construction of the key, we can at least provide some freshness guarantees. This kind of contributory key establishment protocols is usually referred to as key agreement protocols; however, some caution has to be taken here, as different notions of key agreement exist and not all of them suit our purposes. Before defining the type of key agreement we have in mind, it is worth to motivate this type of requirement:

- A protocol participant could be a program embedded in an environment that
  prevents protocol-external communication. In such a situation controlling
  (parts) of the session key may still be feasible, even if communicating a
  learned value is not.
- Even partial control over the session key can be useful, if on the application level only parts of the established session key are actually used for a specific purpose, say symmetric encryption. Another part of the established session key could be used for a purpose that is less relevant for the adversary.

The notion of key agreement we use is motivated by the discussion in [MWW98] and imposes a quantitative restriction on the influence principals have on the derived session key. To express this security requirement, we split the adversarial action into two parts  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ . This separation is only for the ease of explanation, and we allow  $\mathcal{A}_1$  and  $\mathcal{A}_2$  to freely exchange state information a. In a precomputation phase,  $\mathcal{A}_1$  tries to identify a favorable subset  $\kappa$  of the key space  $\mathcal{K}$  and a protocol participant  $U_i$  that seems likely to accept a key from  $\kappa$ . For instance,  $\kappa$  could be the set of all potential session keys with the most significant half of the bits being 0. The algorithm  $\mathcal{A}_2$  controls the malicious participants during the actual protocol run and tries to establish a session key that is contained in  $\kappa$ .

**Definition 6.** Let  $t \in \{1, ..., |\mathcal{P}|\}$ , P a key establishment protocol, and for a fixed pair of ppt algorithms  $(\mathcal{A}_1, \mathcal{A}_2)$  consider the following game:

- 1. The initialization phase of P establishing the long term keys is executed.
- 2. Having access to the public keys, the Send and Reveal-oracle and being allowed up to t-1 calls to the Corrupt oracle,  $\mathcal{A}_1$  outputs a quadruple  $(i, s_i, \chi_{\kappa}, a)$  with state information a and such that
  - $-U_i$  is honest with used<sub>i</sub><sup>s<sub>i</sub></sup> = false;

- $-\chi_{\kappa}$  is a boolean-valued ppt algorithm with  $\kappa := \{sk \in \mathcal{K} : \chi_{\kappa}(sk) = \mathsf{true}\}\$  such that  $|\kappa|/|\mathcal{K}|$  is negligible in the security parameter.
- 3. Upon input of the state information a,  $A_2$  tries to make  $\Pi_i^{s_i}$  accept a session key  $sk_i^{s_i} \in \kappa$ ; for this,  $A_2$  has access to the Send and Reveal-oracle, but may call the Corrupt-oracle only with an argument  $\neq U_i$  and as long as the total number of Corrupt-queries of  $A_1$  and  $A_2$  is  $\leq t-1$ .

If there is no such pair  $(A_1, A_2)$  with  $A_2$  succeeding with non-negligible probability, then we refer to P as being t-contributory. Moreover, by a key agreement, we mean a  $|\mathcal{P}|$ -contributory key establishment.

Summarizing, we take a group key establishment protocol for secure if it is correct, a proper subset of dishonest principals cannot predetermine the key, and it provides the "usual" confidentiality guarantees, integrity, and strong entity authentication:

**Definition 7.** We say a group key establishment protocol is secure against t malicious participants if it is a correct (t+1)-contributory protocol in the sense of Definition 1 and Definition 6, secure in the sense of Definition 3, and assuming at most t principals are dishonest, it offers integrity in the sense of Definition 4 and provides strong entity authentication to all participating instances in the sense of Definition 5. A group key establishment secure against  $|\mathcal{P}|-1$  malicious participants is referred to as a secure group key agreement.

# 4 Secure Authenticated Group Key Agreement

# 4.1 Looking back to Katz and Yung

To illustrate our extended model we show that the protocol of Katz and Yung is partially secure in this sense. The generation of the session identifier has to be modified, though. We moreover assume that all participants check for  $\prod_i X_i = 1$  before accepting the key.

**Proposition 1.** Suppose that in the protocol of Katz and Yung described in Figure 1 all participants check for  $\prod_i X_i = 1$  before accepting the key. Then, with session identifier  $\operatorname{sid}_i^{s_i} = \operatorname{pid}_i^{s_i} || t$  (in this point diverging from [KY03]), we obtain a key establishment protocol secure against one malicious participant.

*Proof.* For correctness and security according to Definition 3 the proof of [KY03] applies. In the sequel, we assume only one participant in the key establishment is allowed to act maliciously.

Integrity. Let us suppose an adversary  $\mathcal{A}$  aims at violating integrity as defined in Definition 4, however, she is only able to make a corrupt call to, say, principal  $U_i$ . The adversarial goal is to make two honest principals that accept a fixed session  $\operatorname{sid}_i^{s_j}$  have either different session keys or hold an incorrect  $\operatorname{pid}_i^{s_j}$  value.

Though, once the session is fixed its  $sid_j^{s_j}$  contains  $pid_j^{s_j}$ , shared thus to all honest principals.

Let us see why she cannot violate integrity either, by forcing honest principals that have accepted to share different keys. Here are the concrete messages  $\mathcal{A}$  can alter:

- (i) the messages in the first round, especially since they are not authenticated. Anyway sending an invalid message at this stage will result in blocking of a particular connection, and not have influence on accepting principals.
- (ii) the value  $z_i$  that  $U_i$  broadcasts in the second round. The latter is actually only used by  $U_{i-1}$  and  $U_{i+1}$ . Obviously, the protocol is still correct if  $U_i$  sends different values  $z_i$  and  $z'_i$  to its neighbors, sending the same one is rather to save an exponentiation.
- (iii) the value  $X_i$  that  $U_i$  broadcasts in the third round. However, the message is implicitly fixed by the values  $X_j$  of the honest participants as  $X_i = (\prod_{j \neq i} X_j)^{-1}$ .

Entity authentication. The concatenation of the nonces t computed in the first round is fresh as long as one honest oracle is involved. Since all participants compute the signature over the message  $(1||z_i||\mathsf{pid}_i^{s_i}||t)$ , it is assured that at the end of the second round all honest instances have knowledge of the session identifier if it is chosen as  $\mathsf{sid}_i^{s_i} = \mathsf{pid}_i^{s_i}||t$ , and in particular they hold the same  $\mathsf{pid}_i^{s_i}$ .

2-Contributory. Note that the adversary cannot influence the key by a dedicated choice of one principal's "random" choices in the first two rounds. Obviously, the random nonce in the first round does not influence the key. In the second round the adversary chooses values for a Diffie-Hellman key exchange. Assumed it is not allowed to choose the exponent  $r_i = 0$  the probability that the resulting key is included in the negligible fraction of the key space specified by the adversary is negligible; this is also true, even allowing exponent 0, for n > 3.

### 4.2 A Secure Two-Round Protocol

As shown in Section 3.2, the proposal of Kim, Lee and Lee in Figure 2 does not offer the discussed security guarantees. In Figure 3 we present—with the notation from Section 3.2—a variant of the protocol that again consists of two rounds, but in the presence of malicious participants offers the security guarantees from Definition 7. We changed the protocol so that all participants  $U_i$  except  $U_n$  send their contribution  $k_i$  to the session key already in the first round. Thus the session key is fixed by the messages of the first round. This allows the participants in the second round to send a confirmation of the key material, namely  $H(\operatorname{pid}_i^{s_i}||k_1||\dots||k_{n-1}||H(k_n))$ , to certify that all of them will compute the same session key. Therewith, the attacks from Section 3.2 are effectively defeated.

The protocol allows  $U_n$  a rushing attack, waiting in the first round until  $k_1, \ldots, k_{n-1}$  are known and then choosing  $k_n$  depending on these. This attack is

```
Round 1:

Computation Each U_i chooses k_i \in \{0,1\}^k, x_i \in \mathbb{Z}_q^* and computes y_i = g^{x_i}, U_n computes additionally H(k_n). Each U_i except U_n sets M_i^{\mathrm{I}} = k_i \| y_i and U_n sets M_n^{\mathrm{I}} = H(k_n) \| y_n. Each U_i computes a signature \sigma_i^{\mathrm{I}} of M_i^{\mathrm{I}}.

Broadcast Each U_i broadcasts (M_i^{\mathrm{I}} \| \sigma_i^{\mathrm{I}}).

Check Each U_i checks all signatures \sigma_j^{\mathrm{I}} of incoming messages (M_j^{\mathrm{I}} \| \sigma_j^{\mathrm{I}}).

Round 2:

Computation Each U_i computes t_i^L = H(y_{i-1}^{x_i}), t_i^R = H(y_{i+1}^{x_i}), T_i = t_i^L \oplus t_i^R and \mathrm{sid}_i^{s_i} = H(\mathrm{pid}_i^{s_i} \| k_1 \| \dots \| k_{n-1} \| H(k_n)), only U_n computes additionally k_n \oplus t_n^R. The participants U_1, \dots, U_{n-1} set M_i^{\mathrm{II}} = \mathrm{sid}_i^{s_i} \| T_i, U_n sets M_n^{\mathrm{II}} = k_n \oplus t_n^R \| \mathrm{sid}_n^{s_n} \| T_n and each U_i computes a signature \sigma_i^{\mathrm{II}} of M_i^{\mathrm{II}}.

Broadcast Each U_i broadcasts (M_i^{\mathrm{II}} \| \sigma_i^{\mathrm{II}}).

Check Firstly, each U_i checks all signatures \sigma_j^{\mathrm{II}} of incoming messages. Then each U_i checks if T_1 \oplus \cdots \oplus T_n = 0 and \mathrm{sid}_i^{s_i} = \mathrm{sid}_j^{s_j} (j = 1, \dots, n). Moreover, each U_i (i < n) checks the commitment H(k_n) for k_n.

Key computation: Each participant U_i computes the session key \mathrm{sk}_i^{s_i} = H(\mathrm{pid}_i^{s_i} \| k_1 \| \dots \| k_n).
```

Fig. 3. A secure group key agreement protocol.

also possible in the protocol of Kim et al., however in the second round and only for the participants except  $U_n$ . One may argue that stronger security requirements on the key *agreement* property of the protocol should be imposed, so that the adversary cannot predetermine any bit of the session key (cf. [MWW98]). Transforming the above protocol accordingly is possible for the price of slightly increasing the computational effort of the involved parties: Instead of broadcasting the  $k_i$ -values in Round 1, in the first round only commitments  $H(k_i)$  are sent and the  $k_i$  values are transmitted in the second round. However, the present protocol fulfills the requirements according to Definition 6.

**Proposition 2.** Suppose that the CDH assumption holds for  $(\mathbb{G}, g)$ ,  $H(\cdot)$  is a random oracle and the underlying signature scheme is existentially unforgeable under adaptive chosen message attacks. Then the protocol in Figure 3 is a secure group key agreement in the sense of Definition 7.

*Proof.* Let  $q_s$  and  $q_{ro}$  be polynomial bounds for the number of the adversary's queries to the Send respectively the random oracle. We begin by defining three events that will occur in several places throughout the proof, and we give bounds for the probability of these events that are negligible in k.

Forge is the event that the adversary succeeds in forging an authenticated message  $M_U \| \sigma_U$  for participant U without having queried  $\mathsf{Corrupt}(U)$  and where  $M_U$  was not output by any of U's instances. An adversary  $\mathcal A$  that can reach Forge can be used for forging a signature for a given public key: This key is assigned to one of the n principals and  $\mathcal A$  succeeds in the intended forgery with probability  $\geq \frac{1}{n} \cdot P(\mathsf{Forge})$ . Thus, using  $\mathcal A$  as black box we can derive an

attacker defeating the existential unforgeability of the underlying signature scheme S with probability

$$\begin{array}{c} \mathsf{Adv}_S^{\mathrm{cma}} \geq \frac{1}{n} \cdot P(\mathsf{Forge}) \\ \iff P(\mathsf{Forge}) \leq \ n \cdot \mathsf{Adv}_S^{\mathrm{cma}}. \end{array}$$

Here  $\mathsf{Adv}_S^{\mathsf{cma}}$  denotes the advantage of the adversary in violating the existential unforgeability under adaptive chosen message attack of the signature scheme, which is negligible by assumption. Thus, the event Forge occurs with negligible probability only.

Collision is the event that the random oracle produces a collision. A Send query causes at most 3 random oracle calls. Thus, the total number of random oracle queries is bounded by  $3q_s+q_{ro}$  and the probability that a collision of the random oracle occurs is

$$P(\mathsf{Collision}) \le \frac{(3q_s + q_{ro})^2}{2^k},$$

which is negligible in k.

Repeat is the event that an uncorrupted participant chooses a nonce  $k_i$  that was previously used by an oracle of some principal. There are at most  $q_s$  used oracles that may have chosen a nonce  $k_i$  and thus Repeat happens with a probability

$$P(\mathsf{Repeat}) \leq \frac{q_s^2}{2^k},$$

again negligible in k.

Security. To prove the security according to Definition 3, we consider a sequence of games. In these games we let the adversary  $\mathcal{A}$  interact with a simulator, that in Game 0 offers the original protocol environment to  $\mathcal{A}$ , and subsequently we change the simulator's behavior in several small steps without affecting  $\mathcal{A}$ 's success probability significantly. Keeping track of the changes between subsequent games, in the last game we will be able to derive the desired negligible upper bound on  $\mathsf{Adv}_{\mathcal{A}}$ .

**Game 0:** In this game the protocol participants' instances are faithfully simulated for the adversary, i.e., the adversary's situation is the same as in the real model.

$$\mathsf{Adv}^{\mathrm{Game}\ 0}_{\mathcal{A}} = \mathsf{Adv}_{\mathcal{A}}.$$

Game 1: This game is aborted if one of the events Forge, Collision or Repeat occurs. Otherwise the game is identical with Game 0 and the adversary cannot detect the difference. Thus, for adversary  $\mathcal{A}$ 's advantage we have

$$|\mathsf{Adv}_A^{\mathsf{Game}\ 1} - \mathsf{Adv}_A^{\mathsf{Game}\ 0}| \le P(\mathsf{Forge}) + P(\mathsf{Collision}) + P(\mathsf{Repeat}).$$

**Game 2:** This game differs from Game 1 in the simulator's response in Round 2. If the simulator has to output the message of an instance  $\Pi_i^{s_i}$  and none of the

neighbors  $U_{i-1}$  or  $U_{i+1}$  is corrupted, then the simulator chooses random values from  $\{0,1\}^k$  for  $t_i^L = t_{i-1}^R$  and  $t_i^R = t_{i+1}^L$  instead of querying the random oracle. To keep consistent, the same values have to be used in the neighbored instances subsequently. By the random oracle assumption, the adversary can only detect the difference by querying the random oracle for  $y_{i-1}^{x_i} = y_i^{x_{i-1}}$ . The adversary cannot know  $x_i$  or  $x_{i-1}$  because  $U_i$  and  $U_{i-1}$  are uncorrupted and the messages cannot have been forged by the adversary through the modification in Game 1.

An adversary  $\mathcal{A}$  that distinguishes Game 1 and Game 2 can be used as black box to solve a CDH instance. Two instances  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  are selected by randomly choosing two different users  $U_i, U_j \in \mathcal{P}$  and two numbers  $s_i, s_j \in \{1, \ldots, q_s\}$ . A given CDH instance  $(g^a, g^b)$  is then assigned to  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  such that these instances will use  $g^a$  respectively  $g^b$  as their message for the first round. Then a random index  $z \in \{1, \ldots, q_{ro}\}$  is chosen and the adversary's z-th query to the random oracle is taken for the answer to the CDH challenge. The answer to the CDH challenge is correct if  $\mathcal{A}$  distinguished the games and  $s_i, s_j$  and z were determined correctly. So we have

$$|\mathsf{Adv}_{\mathcal{A}}^{\mathrm{Game}\ 2} - \mathsf{Adv}_{\mathcal{A}}^{\mathrm{Game}\ 1}| \leq \mathsf{Succ}_{(\mathbb{G},g)}^{\mathrm{CDH}} \cdot q_{ro} \cdot q_{s}^{2},$$

where  $\mathsf{Succ}_{(\mathbb{G},g)}$  is an—under the CDH assumption existing—negligible upper bound for the success probability of the above algorithm to solve CDH.

**Game 3:** In this game the simulator changes the computation of the session key. Having received all messages of Round 2 for an instance  $\Pi_i^{s_i}$ , the simulator checks if all  $U_j \in \mathsf{pid}_i^{s_i}$  are uncorrupted. If so, then the simulator chooses a session key  $\mathsf{sk}_i^{s_i} \in \{0,1\}^k$  at random instead of querying the random oracle. For consistency the simulator will later assign the same key to all partnered instances.

The only way for the adversary to detect the difference is by querying the random oracle for  $H(\operatorname{pid}_i^{s_i}||k_1||\dots||k_n)$ . However, about  $k_n$  only  $H(k_n)$  is known to the adversary. Thus, the adversary can only guess a random value for  $k_n$  and query the random oracle at most  $q_{ro}$  times. This results in:

$$|\mathsf{Adv}^{\mathrm{Game}\ 3}_{\mathcal{A}} - \mathsf{Adv}^{\mathrm{Game}\ 2}_{\mathcal{A}}| \leq rac{q_{ro}}{2^k}.$$

None of the partners of the adversary's Test-instance are allowed to be corrupted or revealed (see Definition 3). Thereby, those instances were affected in Game 3 and use a random value as session key. Therefore, the adversary has only a probability of  $\frac{1}{2}$  for guessing the bit of Test, yielding  $Adv_{\mathcal{A}}^{\text{Game 3}} = 0$ . Putting the probabilities together we recognize the adversary's advantage in the real model as negligible:

$$\mathsf{Adv}_{\mathcal{A}} \leq P(\mathsf{Forge}) + P(\mathsf{Collision}) + P(\mathsf{Repeat}) + \mathsf{Succ}^{\mathsf{CDH}}_{(\mathbb{G},g)} \cdot q_{ro} \cdot q_s^2 + \frac{q_{ro}}{9^k} \cdot q_{ro}^2 + \frac{q_$$

Integrity. Let NoIntegrity be the event that some instance violates the condition imposed in Definition 4. To determine the probability of NoIntegrity let  $U_i$  and  $U_j$  be any two honest principals whose instances  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  accept ( $\operatorname{acc}_i^s = \operatorname{true}$ ) with a matching session identifier  $\operatorname{sid} := \operatorname{sid}_i^{s_i} = \operatorname{sid}_j^{s_j}$ . The session identifier  $\operatorname{sid}$  is unique if uncorrupted principals contributed fresh nonces  $k_i$  (unless Repeat) and the random oracle is collision-free (unless Collision). Moreover the messages of uncorrupted principals cannot be forged (unless Forge) by the adversary. Thus  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  must have received each other's message  $\operatorname{sid}_i^{s_i} \|T_i\|\sigma_i^{\mathrm{II}}$  respectively  $\operatorname{sid}_j^{s_j} \|T_j\|\sigma_j^{\mathrm{II}}$ , where necessarily  $\operatorname{sid} := \operatorname{sid}_i^{s_i} = \operatorname{sid}_j^{s_j}$  matched due to the check phase.

The construction of sid assures that  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  hold the same pid :=  $\operatorname{pid}_i^{s_i} = \operatorname{pid}_j^{s_j}$  (obtained in the respective instance's initialization) and know the same values  $k_1, \ldots, k_{n-1}$  and  $H(k_n)$ . Again by collision-freeness of the random oracle,  $\Pi_i^{s_i}$  and  $\Pi_j^{s_j}$  have received the same  $k_n$  and therewith compute the same session key  $\operatorname{sk}_i^{s_i} = \operatorname{sk}_j^{s_j}$ . Thus, putting things together we obtain the desired negligible upper bound

$$P(\mathsf{NoIntegrity}) \leq P(\mathsf{Collision}) + P(\mathsf{Repeat}) + P(\mathsf{Forge}).$$

Entity authentication. Let EntAuthFail be the event that strong entity authentication fails. We consider entity authentication in Game 1. Let  $U_i$  be any principal with an instance  $\Pi_i^{s_i}$  that has accepted. It is easy to see that entity authentication is provided to  $\Pi_i^{s_i}$ : Since  $\Pi_i^{s_i}$  has accepted, in Round 2 it received messages including the session identifier sid from all principals  $U \in \operatorname{pid}_i^{s_i}$  (unless Forge). As above, in absence of Collision, Repeat and Forge, the session identifier is unique and the message cannot be replayed from a past session. Thus every honest partner holds the same session identifier sid and for the reasons stated above also the partner identifiers  $\operatorname{pid}_i^{s_i}$  and  $\operatorname{pid}_j^{s_j}$  match. Therewith entity authentication is violated with a probability

$$P(\mathsf{EntAuthFail}) \leq P(\mathsf{Collision}) + P(\mathsf{Repeat}) + P(\mathsf{Forge}).$$

Key Agreement. The values relevant for deriving the session key are only the values  $k_i$  that participant  $U_i$  chooses in the first round. An honest participant chooses a fresh value with probability 1 - P(Repeat). Thus a corrupted participant  $U_n$ , who can know the inputs of  $U_1, \ldots, U_{n-1}$  can only choose between a polynomial set of keys, bounded by the number of random oracle queries  $q_{ro}$ .

Finally, correctness of the protocol in Figure 3 is straightforward, and hence the proposition follows.

5 Conclusion

Building on established models for analyzing group key establishment protocols, the tools suggested in this paper offer a possibility to explore security properties of group key establishment protocols in the presence of malicious participants. The introduced framework in particular allows to show that a protocol proposed by Katz and Yung in [KY03] offers security guarantees against a single malicious participant "for free", whereas a proposal of Kim, Lee and Lee [KLL04] fails to do so. However, as shown in the last section, security against malicious participants is achievable in two rounds: without sacrificing efficiency, the discussed proposal of [KLL04] can be modified to offer rather strong security guarantees even in the presence of malicious participants.

# Acknowledgment

We are indebted to Dominique Unruh and the anonymous reviewers for their insightful comments and remarks.

### References

- [BCK98] Mihir Bellare, Ran Canetti, and Hugo Krawczyk. A Modular Approach to the Design and Analysis of Authentication and Key Exchange Protocols. In *Proceedings of STOC 98*, pages 419–428. ACM, 1998.
- [BCP01] Emmanuel Bresson, Olivier Chevassut, and David Pointcheval. Provably Authenticated Group Diffie-Hellman Key Exchange - The Dynamic Case. In Colin Boyd, editor, Advances in Cryptology - ASIACRYPT 2001, volume 2248 of Lecture Notes in Computer Science, pages 290–309. Springer, 2001.
- [BCPQ01] Emmanuel Bresson, Olivier Chevassut, David Pointcheval, and Jean-Jacques Quisquater. Provably Authenticated Group Diffie-Hellman Key Exchange. In Pierangela Samarati, editor, *Proceedings of the 8th ACM Conference on Computer and Communications Security (CCS-8)*, pages 255–264. ACM, 2001.
- [BD95] Mike Burmester and Yvo Desmedt. A Secure and Efficient Conference Key Distribution System. In Alfredo De Santis, editor, Advances in Cryptology — EUROCRYPT'94, volume 950 of Lecture Notes in Computer Science, pages 275–286. Springer, 1995.
- [BGVS05] Jens-Matthias Bohli, María Isabel González Vasco, and Rainer Steinwandt. Burmester-Desmedt Tree-Based Key Transport Revisited: Provable Security. Cryptology ePrint Archive: Report 2005/360, 2005. At the time of writing available electronically at http://eprint.iacr.org/2005/360.
- [BM04] Colin Boyd and Anish Mathuria. Protocols for Authentication and Key Establishment. Springer, 2004.
- [BN03] Colin Boyd and Juan Manuel González Nieto. Round-optimal Contributory Conference Key Agreement. In Yvo Desmedt, editor, *Proceedings of PKC 2003*, volume 2567 of *Lecture Notes in Computer Science*, pages 161–174. Springer, 2003.
- [BPR00] Mihir Bellare, David Pointcheval, and Phillip Rogaway. Authenticated Key Exchange Secure Against Dictionary Attacks. In Bart Preneel, editor, Advances in Cryptology — EUROCRYPT'00, volume 1807 of Lecture Notes in Computer Science, pages 139–155. Springer, 2000.

- [BR93] Mihir Bellare and Phillip Rogaway. Entity authentication and key distribution. In Douglas R. Stinson, editor, Advances in Cryptology—CRYPTO '93, volume 773 of Lecture Notes in Computer Science, pages 232–249. Springer, 1993
- [BR95] Mihir Bellare and Phillip Rogaway. Provably secure session key distribution—the three party case. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing, STOC'95*, pages 57–66. ACM Press, 1995.
- [CBH05] Kim-Kwang Raymond Choo, Colin Boyd, and Yvonne Hitchcock. Examining Indistinguishability-Based Proof Models for Key Establishment Protocols. In Bimal Roy, editor, Advances in Cryptology ASIACRYPT 2005, volume 3788 of Lecture Notes in Computer Science, pages 585–604. Springer, 2005.
- [CBHM05] Kim-Kwang Raymond Choo, Colin Boyd, Yvonne Hitchcock, and Greg Maitland. On Session Identifiers in Provably Secure Protocols: The Bellare-Rogaway Three-Party Key Distribution Protocol Revisited. In Carlo Blundo and Stelvio Cimato, editors, Fourth Conference on Security in Communication Networks - SCN 2004 Proceedings, volume 3352 of Lecture Notes in Computer Science, pages 351–366. Springer, 2005.
- [CK01] Ran Canetti and Hugo Krawczyk. Analysis of Key-Exchange Protocols and Their Use for Building Secure Channels. In Birgit Pfitzmann, editor, Advances in Cryptology – EUROCRYPT 2001, volume 2045 of Lecture Notes in Computer Science, pages 453–474. Springer, 2001.
- [CS04] Christian Cachin and Reto Strobl. Asynchronous Group Key Exchange with Failures. In *Proceedings of the 23rd ACM Symposium on Principles of Distributed Computing (PODC 2004)*, pages 357–366. ACM Press, 2004.
- [CVC04] Zhaohui Cheng, Luminita Vasiu, and Richard Comley. Pairing-Based One-Round Tripartite Key Agreement Protocols. Cryptology ePrint Archive: Report 2004/079, 2004. At the time of writing available electronically at http://eprint.iacr.org/2004/079.
- [HMQS03] Dennis Hofheinz, Jörn Müller-Quade, and Rainer Steinwandt. Initiator-Resilient Universally Composable Key Exchange. In Einar Snekkenes and Dieter Gollmann, editors, Computer Security, Proceedings of ESORICS 2003, volume 2808 of Lecture Notes in Computer Science, pages 61–84. Springer, 2003.
- [JG04] Shaoquan Jiang and Guang Gong. Password Based Key Exchange with Mutual Authentication. In Helena Handschuh and M. Anwar Hasan, editors, Selected Areas in Cryptography: 11th International Workshop, SAC 2004, volume 3357 of Lecture Notes in Computer Science, pages 267–279. Springer, 2004.
- [KLL04] Hyun-Jeong Kim, Su-Mi Lee, and Dong Hoon Lee. Constant-Round Authenticated Group Key Exchange for Dynamic Groups. In Pil Joong Lee, editor, Advances in Cryptology ASIACRYPT'04, volume 3329 of Lecture Notes in Computer Science, pages 245–259. Springer, 2004.
- [KS05a] Jonathan Katz and Ji Sun Shin. Modeling Insider Attacks on Group Key-Exchange Protocols. Cryptology ePrint Archive: Report 2005/163, 2005. At the time of writing available electronically at http://eprint.iacr.org/ 2006/163. Full version of [KS05b].
- [KS05b] Jonathan Katz and Ji Sun Shin. Modeling Insider Attacks on Group Key-Exchange Protocols. In 12th ACM Conference on Computer and Communications Security, pages 180–189. ACM Press, 2005.

- [KY03] Jonathan Katz and Moti Yung. Scalable Protocols for Authenticated Group Key Exchange. In Dan Boneh, editor, Advances in Cryptology — CRYPTO'03, volume 2729 of Lecture Notes in Computer Science, pages 110–125. Springer, 2003.
- [MWW98] Chris J. Mitchell, Mike Ward, and Piers Wilson. Key control in key agreement protocols. IEE Electronics Letters, 34(10):980–981, 1998.
- [Sho99] Victor Shoup. On Formal Models for Secure Key Exchange. Cryptology ePrint Archive: Report 1999/012, 1999. At the time of writing available electronically at http://eprint.iacr.org/1999/012.
- [SSN98] Shahrokh Saeednia and Rei Safavi-Naini. Efficient Identity-Based Conference Key Distribution Protocols. In Colin Boyd and Ed Dawson, editors, Information Security and Privacy, Third Australasian Conference, ACISP'98, volume 1438 of Lecture Notes in Computer Science, pages 320–331. Springer, 1998.
- [Ste02] Michael Steiner. Secure Group Key Agreement. PhD thesis, Universität des Saarlandes, 2002. At the time of writing available at http://www.semper.org/sirene/publ/Stei\_02.thesis-final.pdf.
- [Tze00] Wen-Guey Tzeng. A Practical and Secure Fault-Tolerant Conference-Key Agreement Protocol. In Hideki Imai and Yuliang Zheng, editors, Third International Workshop on Practice and Theory in Public Key Cryptosystems, PKC 2000, volume 1751 of Lecture Notes in Computer Science, pages 1–13. Springer, 2000.