Perfect Forward Security of SPAKE2

Michel Abdalla\textsuperscript{1,2} and Manuel Barbosa\textsuperscript{3}

\textsuperscript{1} DIENS, École normale supérieure, CNRS, PSL University, Paris, France
michel.abdalla@ens.fr
\textsuperscript{2} INRIA, Paris, France
\textsuperscript{3} FCUP and INESC TEC, Porto, Portugal
mbb@fc.up.pt

\textbf{Abstract.} SPAKE2 is a balanced password-authenticated key exchange (PAKE) protocol, proposed by Abdalla and Pointcheval at CTRSA 2005. Due to its simplicity and efficiency, SPAKE2 is one of the balanced PAKE candidates currently under consideration for standardization by the CFRG, together with SPEKE, CPace, and J-PAKE. In this paper, we show that SPAKE2 achieves perfect forward security in the random-oracle model under the Gap Diffie-Hellman assumption. Unlike prior results, which either did not consider forward security or only proved a weak form of it, our results guarantee the security of the derived keys even for sessions that were created with the active involvement of the attacker, as long as the parties involved in the protocol are not corrupted when these sessions take place. Finally, our proofs also demonstrate that SPAKE2 is flexible with respect to the generation of its global parameters $M$ and $N$. This includes the cases where $M$ is a uniform group element and $M = N$ or the case where $M$ and $N$ are chosen as the output of a random oracle.

1 Introduction

Password-authenticated key exchange (PAKE) allows users to establish session keys among themselves with the help of a short secret, known as a password, which can be drawn from a small set of possible values. Passwords can be represented in shorter human-readable formats and distributed/stored using a wider range of mechanisms, which are important requirements in many applications. One important use-case is when secrets must be memorized by humans.

Since the seminal work by Bellovin and Merritt \cite{BM92}, several PAKE protocols have appeared in the literature, achieving different levels of security (such as indistinguishability-based security or universal composability) under a variety of assumptions.
Recently, the Crypto Forum Research Group (CFRG)\textsuperscript{4}, which is an IRTF (Internet Research Task Force) research group focused on the discussion and review of uses of cryptographic mechanisms, started a PAKE selection process with the goal of providing recommendations for password-based authenticated key establishment in IETF protocols. Originally, 4 candidates were under consideration by the CFRG in the balanced PAKE category: SPEKE [Jab97, Mac01, HS14], SPAKE2 [AP05, LK19], J-PAKE [HR10, ABM15], and CPace [HL18, HL19]. Currently, only SPAKE2 and CPace are still under consideration in the balanced PAKE category.

Security analyses for SPAKE2. In the original security analysis [AP05], Abdalla and Pointcheval proved that SPAKE2 achieves implicit authentication in the indistinguishability-based model by Bellare, Pointcheval, and Rogaway [BPR00]. Their security analysis, however, did not take corruption queries into account, which are needed for proving forward security. To address this shortcoming, Becerra, Ostrev, and Skrobot [BOS18] recently provided a proof of weak forward security for SPAKE. Unlike perfect forward security, weak forward security only guarantees the security of sessions created without an active intervention by the attacker as long as the involved parties remain uncorrupted during the execution of these sessions [KOY03, Kra05].

In addition to proving weak forward security in [BOS18], the authors also provide a proof of perfect forward security for a variant of SPAKE2 which includes explicit authentication and in which one of the flows is not encrypted with the password. The latter result, however, seems to be a particular case of the scheme proven secure by Abdalla et al. in [ABC+06].

Our contributions. In this work, we provide further security analyses for SPAKE2 [AP05, LK19]. More precisely, we demonstrate that SPAKE2 achieves perfect forward security even without key confirmation. Our proof of security is based on the Gap Diffie-Hellman assumption [OP01] in the random-oracle model [BR93]. Note that this does not contradict the impossibility result about the perfect forward security of 2-message key-exchange protocols in the HMQV [Kra05] paper since the SPAKE2 protocol assumes the pre-existence of secure shared state between parties which is the password itself. We also include an analysis of the protocol when key confirmation is added, which yields a better bound.

Finally, our proofs also demonstrate that SPAKE2 is flexible with respect to the generation of its global parameters $M$ and $N$. This includes the cases where $M$ is a uniform group element and $M = N$ or the case where $M$ and $N$ are chosen as the output of a random oracle.

Difference to previous version. The proof of security of SPAKE2 without key confirmation has been updated to refine the analysis of a bad event that was not fully covered by the initial version. To deal with this, we added an extra game hop, for which we give two alternative justifications. In the first one, which yields a tight reduction to GapSqDH, we assume an algebraic adversary. In the second one, we get a looser reduction to GCDH via the splitting lemma of [PS00, AP05]. We also added the analysis of SPAKE2 with key confirmation.

2 Preliminaries

Notation. We write $x \leftarrow y$ for the action of assigning the value $y$ to the variable $x$. We write $x_1, \ldots, x_n \leftarrow X$ for sampling $x_1, \ldots, x_n$ from a finite set $X$ uniformly at random. If $A$ is a probabilistic algorithm we write $y_1, \ldots, y_n \leftarrow A(x_1, \ldots, x_n)$ for the action of running $A$ on inputs $x_1, \ldots, x_n$ with independently chosen coins, and assigning the result to $y_1, \ldots, y_n$. \texttt{PPT} as usual abbreviates probabilistic polynomial-time. We use notation $T[x]$ to denote access to a dictionary/table $T$ at index $x$, and $\{}$ to denote the empty table. We abuse notation and use $T$ to also represent the set of assigned indices in table $T$.

Games. We use the code-based game-playing language [BR04]. Each game has an \texttt{INITIALIZE} and a \texttt{FINALIZE} procedure. It also has specifications of procedures to respond to an adversary’s various

\textsuperscript{4} https://irtf.org/cfrg
queries. A game is run with an adversary $A$ as follows. First Initialize runs and its outputs are passed to $A$. Then $A$ runs and its oracle queries are answered by the procedures of the game. When $A$ terminates, its output is passed to Finalize which returns the outcome of the game.

3 SPAKE2

Fig. 1 shows a SPAKE2 protocol execution between a user $U$ and a server $S$.

\begin{center}
\begin{tabular}{ll}
\textbf{Client} & \textbf{Server} \\
$pw \in \mathcal{P}, M, N \in G$ & \\
accept $\leftarrow$ F & accept $\leftarrow$ F \\
$x \leftarrow Z_q$ & $y \leftarrow Z_q$ \\
$X \leftarrow g^x$ & $Y \leftarrow g^y$ \\
$X^* \leftarrow X \cdot M^{pw}$ & $U, X^* \rightarrow S, Y^*$ \\
$Y \leftarrow Y^* / N^{pw}$ & $X \leftarrow X^* / M^{pw}$ \\
$K \leftarrow H(U, S, X^*, Y^*, pw, Y^*)$ & $Y^* \leftarrow Y \cdot N^{pw}$ \\
accept $\leftarrow$ T & accept $\leftarrow$ T \\
\end{tabular}
\end{center}

Fig. 1. The SPAKE2 protocol [AP05].

4 Security Model

Our proof of security uses the indistinguishability-based model by Bellare, Pointcheval, and Rogaway [BPR00] and its extension to multiple Test queries proposed by Abdalla, Fouque, and Pointcheval (AFP) [AFP05, AFP06]. In the latter, all Test queries are answered with the same challenge bit $b$.

The security game already instantiated with SPAKE2 can be seen as $G_1$ in Fig. 3. The name spaces for servers $S$ and users $U$ are assumed to be disjoint; oracles reject queries inconsistent with these name spaces.

In the following, we describe more precisely the state of a party instance, the notion of partnering, and the freshness condition used in the proof.

**Definition 4.1 (Instance state).** The state of an instance $i$ at party $P$, denoted $\pi^i_P$, is a tuple of the form $(e, tr, K, ac)$ where

- $e$ is the secret exponent ($x$ or $y$) chosen by the party in that instance
- $tr$ is a session trace of the form $(U, S, X^*, Y^*)$
- $K$ is the accepted session key
- $ac$ is a boolean flag that indicates whether the instance accepted ($ac = T$) or is expecting a response ($ac = F$)

We will use $\pi^i_P.e$, $\pi^i_P.tr$, $\pi^i_P.K$, $\pi^i_P.ac$ to denote the individual components of the state.
Definition 4.2 (Partnering). A server instance $\pi^S_i$ and a user instance $\pi^U_j$ are partnered iff

$$\pi^S_i.\text{ac} = T \land \pi^U_j.\text{ac} = T \land \pi^S_i.\text{tr} = \pi^U_j.\text{tr}$$

Two user instances are never partnered. Two server instances are never partnered.

This is a definition in the style of matching-conversations. Note that partnering together with correctness implies that the associated keys are the same.

Definition 4.3 (Freshness). Instance $i$ at party $P$ is fresh, written $\text{Fresh}(P,i)$ if and only if all the following conditions hold:

1. the instance accepted;
2. the instance was not queried to $\text{Test}$ or $\text{Reveal}$ before;
3. At least one of the following conditions holds:
   (a) The instance accepted during a query to $\text{Execute}$;
   (b) There exists more than one partner instance;
   (c) No partner instance exists and $\text{Corrupt}(U,S)$ was not called prior to acceptance;
   (d) A unique fresh partner instance exists.

5 Assumptions

The perfect forward security of SPAKE2 is based on the difficulty of solving the Computational Diffie-Hellman (CDH) problem by attackers that have access to a Decisional Diffie-Hellman (DDH) oracle. This is usually known as the Gap Diffie-Hellman problem (GCDH) [OP01].

The proof is carried out in the Random-Oracle (RO) model and, to prove perfect forward secret, we need to restrict our attention to algebraic adversaries [FKL18]. Note that the GCDH problem has been shown to be equivalent to the standard discrete logarithm problem (DL) in the Algebraic and Generic Group models [FKL18, Los19] so, our proof implies reductions to the DL problem in these idealized models of computation.

Our proof is structured to highlight which problem an attacker would need to solve to break the protocol. On the one hand, we give tight reductions to well-known computational assumptions. On the other hand, it follows sequence of hops that clarifies how the assumptions map to the generation of SPAKE2 global parameters and runtime operation. We show that, as expected, passive behavior by the attacker implies that it needs to solve the standard Gap CDH problem to break the established session. Intuitively attackers have access to a DDH oracle by comparing random oracle outputs to the keys revealed by the game. Active attackers that may try to take advantage of leaked passwords must compute the CDH of the global parameters $M$ and $N$ or at least compute the squared version of this problem $\text{CDH}(M,M)$, with the help of a DDH oracle. We show that the protocol is flexible with respect to the generation procedure, provided that the previous problem is hard to solve. This includes, in particular, the case where $M$ is a uniform group element and $M = N$.

To this end, we formalize a family of assumptions parametrized by a distribution $\mathcal{D}$ that outputs a pair of elements in $\mathbb{Z}_q$, where $q$ is a large prime. The assumptions are stated with respect to a group $G$ of order $q$ and generator $g$.

---

5 There is a subtlety here which further justifies our presentation with gap assumptions. Our proof in the algebraic model uses the gap oracles on values queried by attackers to a random oracle. As random oracles are an abstraction of local computation, it is perhaps unreasonable to ask that the algebraic adversary provides us with representations of those values. If this is not required from the adversary, then our proof does not necessarily imply a reduction to DL for algebraic attackers, and we must rely on the gap assumptions. For this reason, in our proof we only make explicit use of the algebraic nature of the adversary on messages that it explicitly delivers.
Definition 5.1 (Gap Computational Diffie-Hellman (G-GCDH)). The G-GCDH assumption states that, for any ppt adversary $A$, the following probability is small.

$$\text{Adv}_{A}^{\text{G-GCDH}} := \Pr \left[ Z = g^{xy} : Z \leftarrow A^{\text{GDH}}(g^{x}, g^{y}); x \leftarrow U_{\mathbb{Z}_{q}}; \ y \leftarrow \mathcal{D}(x) \right]$$

The standard GCDH assumption is a particular case when $\mathcal{D}$ is the uniform distribution over $\mathbb{Z}_{q}$. We denote this by $U$-GCDH. The Gap Squared Diffie-Hellman (GSqDH) assumption is another particular case, when we restrict the previous case to $\mathcal{D}$ outputting $x = y$. For simplicity, in our analysis, we will restrict our attention to global parameter sampling operations where we exclude the possibility that $x = 0$ or $y = 0$. This means that the bound for our proof includes an extra statistical test when we remove this restriction (this will be of the form $(q_s + q_e)/q$ for standard GCDH). We will also rely on the weaker Gap Discrete Logarithm assumption.

Definition 5.2 (Gap Discrete Logarithm (GDL)). The GDL assumption states that, for any ppt adversary $A$, the following probability is small.

$$\text{Adv}_{A}^{\text{GDL}} := \Pr \left[ x' = x : x' \leftarrow A^{\text{DDH}}(g^{x}); x \leftarrow U_{\mathbb{Z}_{q}} \right]$$

Finally, for the version of our result where we do not rely on the assumption that the adversary is algebraic, we require the following assumption introduced in [AP05], which we extend to the gap version.

Definition 5.3 (Gap Discrete Logarithm Password-based Chosen-basis Computational Diffie-Hellman assumption (GPCCDH)). The G-GPCCDH assumption states that, for any ppt adversary $A$, the following probability is small.

$$\text{Adv}_{A}^{\text{G-GPCCDH}} := \Pr \left[ C = \text{CDH}(A_{1}/B_{1}^{\text{pw}}, A_{2}/B_{2}^{\text{pw}}) \mid b_{1} \leftarrow U_{\mathbb{Z}_{q}}; b_{2} \leftarrow \mathcal{D}(b_{1}) \right. \text{\quad} A_{1} \leftarrow g^{b_{1}}; B_{1} \leftarrow g^{b_{2}} \text{\quad} (A_{2}, \text{st}) \leftarrow A_{2}^{\text{GPCCDH}}(g^{b_{1}})(B_{1}, B_{2}, A_{1}) \text{\quad} \text{pw} \leftarrow \mathcal{P} \text{\quad} C \leftarrow A_{2}^{\text{GPCCDH}}(\text{pw}, \text{st}) \right]$$

It was shown in [AP05] that the assumption above without the DDH oracle is implied by the standard CDH problem. The proof extends naturally to gap versions of the problems, so we have the following.

Lemma 5.4 ([AP05], Lemma 3.4). For all GPCCDH adversaries $A$ there exists an adversary $B$ such that:

$$\frac{2}{|\mathcal{P}|} \geq \text{Adv}_{A}^{\text{G-GPCCDH}} \geq \frac{1}{|\mathcal{P}|} + \epsilon \Rightarrow \text{Adv}_{B}^{\text{G-GCDH}} \geq \frac{|\mathcal{P}|}{128} \cdot \epsilon^{3}.$$

Here we present the lemma for the version where $1/|\mathcal{P}|$ is large when compared to $\epsilon$, i.e., we are dealing with passwords with significantly lower entropy than the security parameter.

In the proof of SPAKE2 with key confirmation, we will assume that the Key Derivation Function (KDF) is a Pseudorandom-Function (PRF). This assumption is implied by the standard notions of security for KDFs [Kra10].

Definition 5.5 (Pseudorandom Function). A Key Derivation Function KDF with key space $\mathcal{K}$, domain $\mathcal{D}$ and range $\mathcal{R}$ is PRF-secure if, for any ppt adversary $A$, the following advantage term is small.

$$\text{Adv}_{A}^{\text{PRF}} := \Pr \left[ b = 1 : b \leftarrow A^{\text{KDF}}(\mathcal{K}, \cdot)(K \leftarrow \mathcal{K}) \right] - \Pr \left[ b = 1 : b \leftarrow A^{F}(\cdot); F \leftarrow U_{\mathcal{F}} \right]$$

Here $\mathcal{F}$ is the set of all functions with domain $\mathcal{D}$ and range $\mathcal{R}$. 

In our proof, we actually only require a PRF that is pseudorandom for one input per key, and we will denote this weaker requirement by 1PRF.

Finally, in the proof of SPAKE2 with key confirmation, we require the standard unforgeability property for a one-time MAC.

**Definition 5.6 (One-time MAC security).** A MAC scheme with key space $K$ and domain $D$ is one-time unforgeable if, for any ppt adversary $A$, the following advantage term is small.

$$\text{Adv}^{\text{UF}}_A := \Pr\left[t = \text{MAC}(K, m) : (m, t) \leftarrow A(); K \leftarrow K\right]\right]$$

6 Perfect Forward Security Proof

**Theorem 6.1.** SPAKE2 is tightly PFS-secure in the random-oracle model under the Gap Squared Diffie-Hellman assumption. More precisely, for every algebraic attacker $A$ against SPAKE2, there exist attackers $B_1$ and $B_3$ against the Gap Diffie-Hellman problem, attacker $B_2$ against the Gap Discrete Logarithm problem, and attacker $B_4$ against the Gap Squared Diffie-Hellman problem such that

$$\text{Adv}^{\text{spake2}}_A(\cdot) \leq \frac{2q_s}{|\mathcal{P}|} + \text{Adv}^{\text{GCDH}}_{B_1}(\cdot) + \text{Adv}^{\text{GDL}}_{B_2}(\cdot) + \text{Adv}^{D^*-\text{GCDH}}_{B_3}(\cdot) + \text{Adv}^{\text{GSqDH}}_{B_4}(\cdot) + \frac{(q_s + q_e)^2}{2q}$$

Here there is an extra $\frac{2q_s}{|\mathcal{P}|}$ term that is a consequence of our game hopping strategy, which has been chosen for clarity by separating the weak forward secrecy property from the strong one that relies on algebraic adversaries. Using a different sequence of hops, one can remove the extra term by aggregating the two bad events that give rise to these terms into a single one.

**Theorem 6.2.** SPAKE2 is PFS-secure in the random-oracle model under the Gap Computational Diffie-Hellman assumption. More precisely, for every attacker $A$ against SPAKE2, there exist attackers $B_1$ and $B_3$ against the Gap Diffie-Hellman problem, attacker $B_2$ against the Gap Discrete Logarithm problem, and attacker $B_4$ against the Gap GPCCDH Diffie-Hellman problem such that

$$\text{Adv}^{\text{spake2}}_A(\cdot) \leq \frac{q_s}{|\mathcal{P}|} + \text{Adv}^{\text{GCDH}}_{B_1}(\cdot) + \text{Adv}^{\text{GDL}}_{B_2}(\cdot) + \text{Adv}^{D^*-\text{GCDH}}_{B_3}(\cdot) + 2q_s \cdot \text{Adv}^{\text{GPCCDH}}_{B_4}(\cdot) + \frac{(q_s + q_e)^2}{2q}$$

Note that in the above bound there exists an extra $2q_s/|\mathcal{P}|$ within the GPCCDH advantage, since we only have guarantees based on GCDH once that advantage term grows above this limit, so this bound is worse than the one in Theorem 6.1.

We give a joint proof for both theorems, since they only differ in the proof of the last game hop. The code of all games and adversaries is given in Appendix A.

**Proof.** We prove the security of SPAKE2 using a sequence of games shown in Figs. 3 to 5. In Table 1 we give a summary of the meaning of bad events in each game.

**GAME G_1.** This is the initial security game, so we have:

$$\text{Adv}^{\text{spake2}}_A(\cdot) = |\Pr[G_1 \Rightarrow T] - 1/2|$$

**GAME G_2.** From game $G_1$ to game $G_2$ we introduce a bad flag $\text{bad}_2$. The $\text{bad}_2$ flag is set whenever

- a session is about to be accepted on the server side that collides on $(U, S, X^*, Y^*)$ with any previously accepted session;
- a session is about to be accepted on the user side that collides on $(U, S, X^*, Y^*)$ with any previously accepted user session.
If $\text{bad}_2$ is set, then the oracle call is ignored and the session is not accepted.

Note that this implies that all accepted sessions on the server (resp. client) side are unique and that no session can have more than one partner. The two games are identical until $\text{bad}$ occurs, and this event can be bound by a statistical term. More precisely, for an attacker placing at most $q_e$ queries to $\text{EXECUTE}$ and $q_s$ queries to the $\text{SEND}$ oracles, we have for large $q$ denoting the order of the group:

$$|\Pr[G_1 \Rightarrow T] - \Pr[G_2 \Rightarrow T]| \leq \frac{(q_e + q_s)^2}{2q}$$

Here, $q_e + q_s$ denotes the maximum number of accepted sessions (we count passively partnered sessions as one here) and this is a birthday bound computed pessimistically for an attacker that makes accepted sessions collide by fixing the user, server and one of the transmitted group elements in all accepted sessions.

GAME $G_3$. In this game we take advantage of the fact that accepted sessions have unique traces (modulo partnering) to make the freshness condition explicit in the game (we extend the state of sessions to keep track of freshness with a new field $fr$). We also remove the code that refers to $\text{bad}_2$. Since nothing changes, we have

$$\Pr[G_3 \Rightarrow T] = \Pr[G_2 \Rightarrow T]$$

GAME $G_4$. In this game we make the keys of sessions accepted as a result of calls to $\text{EXECUTE}$ independent from the random oracle accessible to the attacker. These sessions obtain keys from a new random oracle $T_e$ that is indexed by the session trace $(U, S, X^*, Y^*)$. Note that, due to the guarantee of session uniqueness on the server and client sides, there is no room for ambiguity. We set a bad flag $\text{bad}_4$ if ever the answers given to the attacker could be inconsistent with the previous game: either because a new accepted session in $\text{EXECUTE}$ collides with a previous call to $H$, or because a new call to $H$ collides with a previously accepted session in $\text{EXECUTE}$. The games are identical until $\text{bad}$ occurs, so we have:

$$|\Pr[G_3 \Rightarrow T] - \Pr[G_4 \Rightarrow T]| \leq \Pr[G_4 \Rightarrow \text{bad}_4]$$

The probability of $\text{bad}_4$ occurring in $G_4$ can be tightly reduced to the standard Gap CDH problem, which means that:

$$\Pr[G_4 \Rightarrow \text{bad}_4] \leq \text{Adv}^{GCDH}_{U}(\cdot)$$

The reduction to Gap CDH is given in Lemma 6.4.

GAME $G_5$. In game $G_5$ we again rearrange code in preparation for the next hop. We remove the code that deals with bad events, and we change $\text{EXECUTE}$ so that nothing depends on passwords. We can do this because the random oracle that generates these keys only depend on the trace, which can be sampled independently of passwords. We have

$$\Pr[G_5 \Rightarrow T] = \Pr[G_4 \Rightarrow T]$$

GAME $G_6$. In this game we change the way in which we generate keys for $\text{SEND}$ queries. We treat corrupt sessions, for which the attacker may trivially compute the key, as before. The remaining keys are derived using an independent random oracle $T_s$. We set a bad flag $\text{bad}_6$ if ever there could be an inconsistency with the main random oracle. As before, this means that:

$$|\Pr[G_5 \Rightarrow T] - \Pr[G_6 \Rightarrow T]| \leq \Pr[G_6 \Rightarrow \text{bad}_6]$$

We also observe that, since all fresh sessions that can be tested now have keys derived using random oracles independent from $H$, accepted keys of fresh sessions are independently distributed from anything else in the game. This means that the attacker has no information on bit $b$ and has therefore 0 advantage in this game. From the previous equations we can derive that:

$$\text{Adv}^{\text{SPAKE2}}_{\mathcal{A}}(\cdot) \leq \frac{(q_s + q_e)^2}{2q} + \text{Adv}^{GCDH}_{U}(\cdot) + \Pr[G_6 \Rightarrow \text{bad}_6]$$
To bound the probability of $\text{bad}_6$ we need to continue our sequence of hops. 

**GAME $G_7$.** We modify the way in which $X^*$ and $Y^*$ are generated, removing the component that depends on the password. The distribution of the $X^*$ and $Y^*$ values does not change with respect to the previous game, but we need to adapt the way in which the Diffie-Hellman tuples are computed, to make sure we obtain the same values as in the previous game. After we do this change, passwords associated with fresh sessions are still (and only) used to check for the bad event $\text{bad}_6$. We add a flag $\text{bad}_7$ to account for the unlikely case that a password occurring in the game (either queried by the adversary or sampled by the challenger) collides with the secret exponents used in $X^*$ and $Y^*$ sampled by the challenger. If this happens, we reset $\text{bad}_6$ at the end of the game. The probability of this event occurring can be easily reduced to the GDL problem, as shown in Lemma 6.5. We therefore have:

$$\Pr[G_6 \Rightarrow \text{bad}_6] \leq \Pr[G_7 \Rightarrow \text{bad}_6] + \text{Adv}^{\text{GDL}}_{G_2}(\cdot)$$

**GAME $G_8$.** We now change the checking of $\text{bad}_6$ to allow delaying all password samplings as late as possible in the game: upon corruption or at the end of the game. This means we stop verifying password equality when checking for $\text{bad}_6$ in the SEND oracles (there is no impact in the random oracle simulation because the attacker provides the password explicitly there). Without additional changes we might not be able to tightly relate the probability of $\text{bad}_6$ in this new game to its probability in the previous game, since we have relaxed the check that detects it. We therefore need to add additional checks that allow us to continue the proof. There are two important cases to distinguish:

- the attacker causes $\text{bad}_6$ for a password that has not been generated yet (this includes all occurrences in the SEND oracles and some in the random oracle simulation).
- the attacker causes $\text{bad}_6$ by querying the random oracle on the correct password (intuitively it does not need to guess the password in this case);

For the second case we set a bad flag $\text{bad}_8^2$ and no longer set $\text{bad}_6$.

For the other cases we add the potentially problematic random oracle entries to a list $T_{\text{bad}}$, so that we can delay the analysis of potential inconsistencies until the end of the game. At the end of the game we split the checking of $\text{bad}_6$ into two cases. We set a flag $\text{bad}_8$ if in the $T_{\text{bad}}$ log of problematic calls there are two entries consistent with a single entry in $T_s$, i.e., with different passwords; in this case we no longer set $\text{bad}_6$. Finally, we check for $\text{bad}_6$ simply by going through the list of problematic entries and checking whether there is a password match.

The combined actions of the three bad flags in this game account for all possible cases in which $\text{bad}_6$ can occur in game $G_7$, so we can write:

$$\Pr[G_6 \Rightarrow \text{bad}_7] \leq \Pr[G_8 \Rightarrow \text{bad}_6] + \Pr[G_8 \Rightarrow \text{bad}_6^4] + \Pr[G_8 \Rightarrow \text{bad}_6^2] + \Pr[G_8 \Rightarrow \text{bad}_6^2]$$

The probability of $\text{bad}_8$ occurring in $G_8$ can be tightly reduced to Gap CDH, where we use here the generalized version for any parameter distribution $D$ where the CDH problem is hard to compute with the help of a DDH oracle; the reduction is given in Lemma 6.6.

We complete the analysis of this game by further observing that the probability of $\text{bad}_6$ happening is easy to bound. Indeed, the size of the log in which $\text{bad}_6$ is checked has size at most $q_s$ (since there can be no duplicates for the same trace) and all entries are added to this set before the corresponding password is sampled. We therefore have

$$\Pr[G_8 \Rightarrow \text{bad}_6] \leq \frac{q_s}{|P|}$$

Putting these results together, we obtain:

$$\Pr[G_7 \Rightarrow \text{bad}_6] \leq \Pr[G_8 \Rightarrow \text{bad}_6] + \Pr[G_8 \Rightarrow \text{bad}_6^4] + \Pr[G_8 \Rightarrow \text{bad}_6^2] \leq \frac{q_s}{|P|} + \text{Adv}^{D\text{-GCDH}}_{G_2}(\cdot) + \Pr[G_8 \Rightarrow \text{bad}_6^2]$$

\footnote{Note that this does not weaken our result, as one can take $D$ to be the uniform distribution. On the other hand, it makes it clear that different parameter generation procedures can be used and our proof still applies.}
To complete the proof we need to bound the probability of $\text{bad}_8^2$ occurring in $G_8$, and so we continue our sequence of hops.

**GAME $G_9$.** For clarity of presentation, in game $G_9$ we perform a cleanup, removing all code unrelated to $\text{bad}_8^2$. We also add the possibility that the attacker passes an additional parameter $\text{alg}$ when it delivers a message. This additional parameter is ignored for now, so clearly

$$\Pr[G_8 \Rightarrow \text{bad}_8^2] = \Pr[G_9 \Rightarrow \text{bad}_8^2]$$

To complete the proof for Theorem 6.2 we give in Lemma 6.8 a reduction of the above probability to GPCCDH, which yields the desired bound:

$$\text{Adv}_{\mathcal{A}}^{\text{PAKE2}}(\cdot) \leq \frac{q_s}{|P|} + \text{Adv}^{G^\text{CDH}}_{B_1}(\cdot) + \text{Adv}^{G^\text{DL}}_{B_2}(\cdot) + \text{Adv}^{\text{GPCCDH}}_{B_3}(\cdot) + 2q_s \cdot \text{Adv}^{G^\text{PCDH}}_{B_4}(\cdot) + \frac{(q_s + q_e)^2}{2q}$$

We proceed with a final game hop to complete the proof for Theorem 6.1.

**GAME $G_{10}$.** This is our final game. Here we restrict our attention to algebraic adversaries. This means that, when performing an active attack, the adversary provides an additional input $\text{alg}$ to the $\text{SEND}$ oracles that gives a representation of the group element it is producing in terms of the other group elements it observed in the game. These elements include $g$, $M$, and all the honestly generated $X^* = M^2$ and $Y^* = N^y$. We note that this means that any group element produced by $\mathcal{A}$ can be rewritten as a representation of the form $\text{alg} = [(g^s), (M^e), (N^c)]$.

We now add a bad flag $\text{bad}_{10}$ that excludes some password guessing events related to $\text{bad}_8^2$. When these happen, we no longer set $\text{bad}_8^2$, whereas this would have been set in the previous game. This means that:

$$\Pr[G_9 \Rightarrow \text{bad}_8^2] \leq \Pr[G_{10} \Rightarrow \text{bad}_8^2] + \Pr[G_{10} \Rightarrow \text{bad}_{10}]$$

Event $\text{bad}_{10}$ is detected whenever a password is sampled at the moment of corruption, and it hits an already fixed representation of a group element produced by the attacker. The offending values are $b + \lambda(c - \text{pw})$ for group elements delivered at $\text{SENDTERM}$ and $b - \text{pw} + \lambda c$ for group elements delivered at $\text{SENDRESP}$. If this event is detected, then we never set $\text{bad}_8^2$, so in this game the analysis of $\text{bad}_8^2$ can proceed under the assumption that $\text{bad}_{10}$ did not occur. The probability of this event is easy to bound as

$$\Pr[G_{10} \Rightarrow \text{bad}_{10}] \leq \frac{q_s}{|P|}$$

The remaining cases where we set $\text{bad}_8^2$ can be reduced to the Gap Squared Diffie-Hellman problem (GSqDH), as shown in Lemma 6.7. This completes the proof and yields the bound.

$$\text{Adv}_{\mathcal{A}}^{\text{PAKE2}}(\cdot) \leq \frac{2q_s}{|P|} + \text{Adv}^{G^\text{CDH}}_{B_1}(\cdot) + \text{Adv}^{G^\text{DL}}_{B_2}(\cdot) + \text{Adv}^{\text{GPCCDH}}_{B_3}(\cdot) + \text{Adv}^{G^\text{PCDH}}_{B_4}(\cdot) + \frac{(q_s + q_e)^2}{2q}$$

A different sequence of games permits removing a $q_s/|P|$ term, yielding the expected bound for PAKE.

**Corollary 6.3.** SPAKE2 is weak PFS-secure in the random-oracle model under the Gap Computational Diffie-Hellman assumption. More precisely, for every attacker $\mathcal{A}$ against SPAKE2, there exist attackers $B_1$ and $B_3$ against the Gap Diffie-Hellman problem and attacker $B_2$ against the Gap Discrete Logarithm problem such that

$$\text{Adv}_{\mathcal{A}}^{\text{PAKE2}}(\cdot) \leq \frac{q_s}{|P|} + \text{Adv}^{G^\text{CDH}}_{B_1}(\cdot) + \text{Adv}^{G^\text{DL}}_{B_2}(\cdot) + \text{Adv}^{\text{GPCCDH}}_{B_3}(\cdot) + \frac{(q_s + q_e)^2}{2q}$$

**Proof.** In the weak forward secrecy model, corruption of password $\text{pw}_{1/3}$ makes all fresh session where the adversary was active immediately unfresh and, furthermore, corruption is not allowed if such a
session was already tested. It is easy to see that a simple adaptation of the previous proof can be done for this setting, where the probability of event \( \text{bad}_2 \) occurring is 0 and the advantage of the adversary is still 0; for such cases where the event would occur we simply return \( T_s[\text{tr}] \) as the answer to the problematic \( H \) query.

We now prove the auxiliary lemmas supporting the proofs of Theorems 6.1 and 6.2.

**Lemma 6.4.** For every attacker \( \mathcal{A} \), there exists an attacker \( \mathcal{B}_1 \), such that

\[
\Pr[\mathcal{G}_4 \Rightarrow \text{bad}_4] \leq \text{Adv}^{G\text{-GCDH}}_{\mathcal{B}_1}()
\]

*Proof.* Let us consider the GCDH attacker \( \mathcal{B}_1 \) in Fig. 6. It gets a generalized challenge of the form \((X_1, \ldots, X_{q_c}, Y_1, \ldots, Y_{q_c})\) and finds \( \text{CDH}(X_i, Y_j) \), for some \( 1 \leq i, j \leq q_c \). This problem is tightly equivalent to \( U\text{-GCDH} \) by random self-reducibility. The attacker embeds \( X_i \) and \( Y_j \) in traces for protocol executions that result from calls to \( \text{EXECUTE} \). For all calls to \( \text{SEND} \) oracles, the attacker runs everything as in \( \mathcal{G}_4 \). The DDH oracle is used to check for the bad event, in which case a solution to the GCDH problem was found. It uses the DDH oracle whenever the rules of the game require it, and it can also use it to check if \( \text{bad}_4 \) occurred, in which case it can solve the Gap CDH challenge.\(^7\)

**Lemma 6.5.** For every attacker \( \mathcal{A} \), there exists an attacker \( \mathcal{B}_2 \), such that

\[
\Pr[\mathcal{G}_7 \Rightarrow \text{bad}_7] \leq \text{Adv}^{\text{DL}}_{\mathcal{B}_2}()
\]

*Proof.* Let us consider a GDL attacker \( \mathcal{B}_2 \) in Fig. 6. It gets a challenge of the form \( A \) and finds \( \text{DL}(A) \). The attacker uses \( A \) in \( \text{SEND} \) queries to compute \( X^* = A \cdot g^a \) and \( Y^* = A \cdot g^\Delta \). Note that the distribution of these values is correct. The checking of CDH tuples, which is needed for correctly maintaining random-oracle consistency, is carried out using the DDH oracle.

If \( \text{bad}_7 \) occurs, our reduction recovers the discrete logarithm of \( A \) offset by a known quantity. Indeed, suppose the attacker guessed \( \text{pw} = x \) or \( \text{pw} = y \) in game \( \mathcal{G}_7 \), which means we have \( X^* = g^{\text{px}} \) or \( Y^* = g^{\text{py}} \). In our reduction we program both \( X^* \) and \( Y^* \) as \( g^{a+\Delta} \), where \( a \) is the unknown discrete logarithm, and we know all of \( m, \text{pw} \) and \( \Delta \). This means that, if the guess was successful, we can compute the discrete logarithm as \( a = m\text{pw} - \Delta \) or \( a = n\text{pw} - \Delta \).

**Lemma 6.6.** For every attacker \( \mathcal{A} \), there exists an attacker \( \mathcal{B}_3 \), such that

\[
\Pr[\mathcal{G}_7 \Rightarrow \text{bad}_3] \leq \text{Adv}^{\mathcal{D}^*\text{-GCDH}}_{\mathcal{B}_3}()
\]

*Proof.* Let us consider a \( \mathcal{D}^*\text{-GCDH} \) attacker \( \mathcal{B}_3 \) in Fig. 6. It gets a challenge of the form \( (M, N) \) and finds \( \text{CDH}(M, N) \). The attacker uses \( (M, N) \) as the variables of the same name in game \( \mathcal{G}_8 \). The checking of CDH tuples, which is needed for correctly maintaining random-oracle consistency and detecting bad events, is carried out using the DDH oracle.

Once the game terminates, the attacker checks for the occurrence of \( \text{bad}_3 \) and if offending entries exist, it recovers \((X^*, Y^*, \text{pw}, Z)\) and \((X^*, Y^*, \text{pw}', Z')\) such that \( \text{pw} \neq \text{pw}' \) and the following holds for known \( x \) or known \( y \), and known \( \text{pw} \) and \( \text{pw}' \):

\[
\text{CDH}(g^{mx}/M^{\text{pw}}, g^{ny}/N^{\text{pw}}) = Z \land \text{CDH}(g^{mx}/M^{\text{pw}'}, g^{ny}/N^{\text{pw}'}) = Z'
\]

Letting \( Z = g^z \), \( Z' = g^{z'} \), we can rewrite these equations as

\[
\begin{align*}
mx - \text{pw} \cdot ny - \text{pw} &= z \\
mx - \text{pw}' \cdot ny - \text{pw}' &= z'
\end{align*}
\]

\(^7\) Note that it actually suffices to have a restricted DDH oracle in which one of the inputs is fixed as in the Strong Diffie-Hellman (SDH) problem \([ABR01]\), so a tight reduction to SDH is also possible. Furthermore, storing all possible values that cause the bad event and returning one at random gives a reduction to CDH that loses a linear factor in the maximum size of the random-oracle table.
Let us assume that the attacker knows $x$ (the other case is symmetric). We scale the equations and rearrange the formula as follows:

\[
\begin{align*}
  m(x - pw)(x - pw') \cdot n(y - pw) &= z \cdot (x - pw') \\
  m(x - pw')(x - pw) \cdot n(y - pw') &= z' \cdot (x - pw)
\end{align*}
\]

Subtracting the two equations, we get:

\[
z(x - pw') - z'(x - pw) = mn(x - pw)(x - pw')(pw' - pw)
\]

And we can derive a formula for the desired CDH:

\[
g^{mn} = \left(\frac{Z^{x-pw'}}{Z^{pw'-x}}\right)^{\frac{1}{(x-pw)(x-pw')(pw' - pw)}}
\]

This formula can be computed by the attacker, provided that $x \neq pw$ and $x \neq pw'$, which we know to be the case as otherwise the bad event would not have occurred due to the action of bad7.

\[\Box\]

**Lemma 6.7.** For every algebraic attacker $A$, there exists an attacker $B_4$, such that

\[
Pr[G_10 \Rightarrow \text{bad}_3] \leq \text{Adv}_{G_{\text{SQDH}}}^{\text{B}_4}()
\]

**Proof.** The reduction is given in Fig. 7. In the unlikely event that $M = 1$ there is nothing to do, as the problem is trivial to solve. We can also assume that the conditions checked to activate bad7 all failed, as otherwise bad3 would have been reset. This implies, in particular, that there are no entries in $T_s$ such that the secret exponent $x$ or $y$ collides with the password $pw_{US}$.

Given $M$, the reduction sets $N = M^\lambda$, for random non-zero $\lambda$, which yields correctly distributed global parameters. Similarly to what happened in the proof of Lemma 6.6, we use the DDH oracle to ensure the simulation is perfect, so it remains to show that our reduction can compute $\text{CDH}(M, M)$ whenever bad3 occurs. There are two cases to consider, depending on whether we are dealing with a session accepted by a user or by a server.

**Sessions accepted in SendTerm.** We know that the following equation holds.

\[
Z = \text{CDH}\left(\frac{M^x}{M^{pw}}, \frac{g^a M^b N^c}{N^{pw}}\right)
\]

This can be rewritten as:

\[
Z = \text{CDH}(M^{x-pw}, g^a M^b N^c) \leftrightarrow \frac{Z}{M^{a(x-pw)}} = \text{CDH}(M^{x-pw}, M^b N^{c-pw})
\]

Replacing $N = M^\lambda$ and rearranging the exponents we get:

\[
\frac{Z}{M^{a(x-pw)}} = \text{CDH}(M, M)^{(x-pw)(b + \lambda(c-pw))}
\]

We can therefore recover the desired SqDH provided that $(x - pw)(b + \lambda(c - pw)) \neq 0$. So see that this can never happen, note $b + \lambda(c - pw) \neq 0$, as otherwise $\text{bad}_{10}$ would have happened and this entry would not be in the list, and we also excluded $x = pw$ above. This means that we can compute the SqDH.

**Sessions accepted in SendResp.** In this case we know that the following equation holds.

\[
Z = \text{CDH}\left(\frac{g^a M^b N^c}{M^{pw}}, \frac{N^y}{N^{pw}}\right)
\]

This can be rewritten as:

\[
Z = \text{CDH}(g^a M^{b-pw} N^c, N^{y-pw})
\]
Replacing $N = M^\lambda$ and rearranging the exponents we get

$$Z = \text{CDH}(g^a M^{b-\text{pw}+\lambda c}, M^{\lambda(y-\text{pw})}) \iff \frac{Z}{M^{a\lambda(y-\text{pw})}} = \text{CDH}(M, M^{(b-\text{pw}+\lambda c)(y-\text{pw})})$$

We can therefore recover the desired SqDH provided that $(b-\text{pw}+\lambda c)(y-\text{pw}) \neq 0$. So see that this can never happen, note that we know that $y - \text{pw} \neq 0$ and it must be the case that $b - \text{pw} + \lambda c \neq 0$, as otherwise $\text{bad}^3$ would have happened and this entry would not be in the list. This means that the SqDH can be computed.

**Lemma 6.8.** For every attacker $A$, there exists an attacker $B_5$, such that

$$\Pr[G \Rightarrow \text{bad}^2] \leq 2q_s \cdot \text{Adv}_{B_5}^{\text{GPCCDH}}()$$

**Proof.** The reduction is given in Fig. 7. Our adversary $B_5$ guesses a query in the range $[0..q_s-1]$ and chooses a role uniformly at random to guess whether the challenge should be injected on the responder side or on the initiator side. The behavior conditioned on these guesses is shown in the figure as two different versions of $B_5$, one for initiator side and another for responder side. Similarly to what happened in the proof of Lemma 6.6, we use the DDH oracle to ensure the simulation is perfect, and the correct result is produced as long as the guesses are correct. On the initiator side, the reduction is slightly more intricate as one needs to embed $A_1$ in all possible initiator sessions that might lead to the bad event. The bound on the theorem follows from the fact that the reduction only works if our attackers guesses are correct.

**7 Security with Key Confirmation**

The result we obtained in the previous section provides guarantees even if the derived secret keys are used before a key confirmation step. However, in most practical settings a key confirmation step exists and, in this case, a much better security bound can be proved. This is what we do in this section. We show the modified protocol in Fig. 2. Note that a one-time secure MAC suffices.

**Theorem 7.1.** SPAKE2 with key confirmation is tightly PFS-secure in the random-oracle model under the GCDH assumption. More precisely, for every attacker $A$ against SPAKE2, there exist attackers $B_6$ and $B_5$ against the Gap Diffie-Hellman problem, attacker $B_2$ against the Gap Discrete Logarithm problem, attacker $B_7$ against the pseudorandomness of KDF and attacker $B_8$ against the unforgeability of the MAC scheme such that

$$\text{Adv}_{A}^{\text{SPAKE2}}() \leq \frac{q_s}{|P|} + \text{Adv}_{B_6}^{\text{GCDH}}() + \text{Adv}_{B_7}^{\text{GDL}}() + \text{Adv}_{B_8}^{\text{DPF-CDH}}() + q_s \left( \text{Adv}_{B_2}^{\text{PRF}}() + \text{Adv}_{B_2}^{\text{UF}}() \right) + \frac{(q_s + q_e)^2}{2q}$$

**Proof.** Again we use a sequence of hops, but we explain here only the differences to the proofs in the previous section. We give the games in Figs. 8 to 10 in Appendix B.

The hops up to game $G_4$ are identical to those for the proofs in the previous section. In game $G_4$, as before, we exclude the possibility of the adversary breaking a session where it behaved passively but, in addition to execute queries, we handle also the sequences of send queries where the adversary establishes sessions with traces $(U, S, X^*, Y^*)$ where the messages were generated by the game. We give a modified reduction to deal with this case in Lemma 7.3 and note that here the gap oracle is not only used to detect the correct CDH answer, but to maintain consistency of all other sessions where the attacker is active.

We now carry on the the discussion of the following hops in detail from game $G_6$.

**GAME G_6.** This game introduces an independent random oracle for keys derived for (possibly aborted) fresh sessions. Flag $\text{bad}_6$ corresponds to the attacker observing an inconsistency in the hash queries for $H$. As in the previous proof, we have

$$\text{Adv}_{A}^{\text{SPAKE2}}() \leq \frac{(q_s + q_e)^2}{2q} + \text{Adv}_{B_6}^{\text{GCDH}}() + \text{Adv}_{B_7}^{\text{GDL}}() + \Pr[G_7 \Rightarrow \text{bad}_6],$$
<table>
<thead>
<tr>
<th>$G_2$</th>
<th>bad$_2$</th>
<th>Accepted trace on responder side collides with any previously accepted trace or accepted trace on initiator side collides with trace of any previously accepted initiator session. Bounded by a statistical term.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_4$</td>
<td>bad$_4$</td>
<td>Random oracle query consistent with the trace accepted as a result of a query to \texttt{Execute}. Such an event would allow attacker to detect an inconsistency between the outputs of \texttt{Test} or \texttt{Reveal} and the random oracle. Bounded by a tight reduction to GCDH.</td>
</tr>
<tr>
<td>$G_6$</td>
<td>bad$_6$</td>
<td>Random oracle query consistent with the trace accepted as a result of a query to one of the \texttt{Send} oracles and declared fresh. Such an event would allow attacker to detect an inconsistency between the outputs of \texttt{Test} or \texttt{Reveal} and the random oracle. Adversary’s view is independent of $b$ in this game, but bad$_6$ is hard to bound.</td>
</tr>
<tr>
<td>$G_7$</td>
<td>bad$_7$</td>
<td>The meaning is the same as in the previous game, but it now is only triggered in cases where the other bad event in this game does not occur. Either there is a call to the random oracle with $pw = x$ or $pw = y$, where $x$ or $y$ is one of the random values sampled in the game. Or a password was sampled by the game that also collides with one of these values. Bounded by a tight reduction to GDL.</td>
</tr>
<tr>
<td>$G_8$</td>
<td>bad$_8$</td>
<td>The meaning is the same as in the previous games, but it now is only triggered in cases where the other bad event in this game does not occur. At this point this event can only happen for random oracle queries placed before the associated password was sampled in the game. Furthermore, we know there can be at most $q_s$ entries, which results in a $q/</td>
</tr>
<tr>
<td>$G_9$</td>
<td>bad$_9$</td>
<td>Same meaning as in the previous game and same probability of occurrence. We fork here the proofs of Theorems 6.1 and 6.2. For the latter we immediately bound the event with a tight reduction to GPCCDH, which yields a loose reduction to GCDH via the splitting lemma of [PS00, AP05]. For Theorem 6.2 we delay bounding this event until $G_{10}$</td>
</tr>
<tr>
<td>$G_{10}$</td>
<td>bad$_{10}$</td>
<td>Same meaning as in the previous game, but now only set when the other bad flag in this game is not set. Bound using a reduction to GSGDH assuming an algebraic adversary. The adversary committed to a representation of a group element in a \texttt{Send} query that predicts a password sampled later in the game. The probability of this event occurring can be easily bound as $q_s/</td>
</tr>
</tbody>
</table>

| Table 1. Meaning of the bad events in the different games. |
Client

\( pw \in P, M, N \in G \)

\begin{align*}
\text{accept} & \leftarrow F \\
x & \leftarrow Z_q \\
X & \leftarrow g^x \\
X^* & \leftarrow X \cdot M^{pw} \\
Y & \leftarrow Y^* / N^{pw} \\
(K_a, K_e) & \leftarrow H(U, S, X^*, Y^*, pw, X^{'}) \\
(K_{U,S}, K_{S,U}) & \leftarrow \text{KDF}(K_a, \text{const}) \\
\text{tag}_{S,U} & \leftarrow \text{MAC}(K_{S,U}, \text{sid}) \\
\text{tag}_{U,S} & \leftarrow \text{MAC}(K_{U,S}, \text{sid}) \\
\text{accept} & \leftarrow T
\end{align*}

Server

\begin{align*}
\text{accept} & \leftarrow F \\
y & \leftarrow Z_q \\
Y & \leftarrow g^y \\
X^* & \leftarrow X^* / M^{pw} \\
Y^* & \leftarrow Y \cdot N^{pw} \\
(K_a, K_e) & \leftarrow H(U, S, X^*, Y^*, pw, X^{'}) \\
(K_{U,S}, K_{S,U}) & \leftarrow \text{KDF}(K_a, \text{const}) \\
\text{tag}_{S,U} & \leftarrow \text{MAC}(K_{S,U}, \text{sid}) \\
\text{tag}_{U,S} & \leftarrow \text{MAC}(K_{U,S}, \text{sid}) \\
\text{accept} & \leftarrow T
\end{align*}

Fig. 2. The SPAKE2 protocol [AP05] with key confirmation. Here, \( \text{const} \) is a fixed point in the domain of the KDF, \( \text{sid} = (U, S, X^{'}, Y^{'}) \) and \( K_e \) is the accepted session key.
where the term corresponding to $B_6$ highlights the fact that the reduction that deals with passive sessions is the one given in Lemma 7.3. Furthermore, as before, the adversary’s view in this game is independent of the secret bit $b$ and it only remains to bound the probability of $\text{bad}_6$ occurring. Let us take a closer look at game $G_6$ to confirm that this is the case.

First of all, let us confirm that all the session keys corresponding to fresh accepted sessions have been moved to an independent random oracle, whereas all revealable (accepted) non-fresh keys have not. Notice that random oracle queries are added to $T_s$ rather than $T$ in two occasions:

- When a $\text{Send}$ query starts a responder session for an uncorrupted password; this session may or may not turn out to be accepted and, if it is, it may or may not be fresh.
- When a $\text{Send}$ query is about to conclude a fresh initiator session (such an entry has not yet been previously added to $T_s$ since we never do this for sessions where the attacker is passive); this session may or may not be accepted due to MAC verification.

For the latter type of entries, if the MAC verification passes, then the session will be accepted as fresh, so we must ensure that such an entry is never removed from $T_s$ in this case. Removal could only occur in the CORRUPT oracle, but here the check performed guarantees that entries are only removed from $T_s$ when no accepted fresh session with that trace exists.

For the first type of entry, we must ensure that the entry is removed from $T_s$ whenever it might be accepted as unfresh. This is ensured by the corruption oracle: it removes any such entries upon corruption and reprograms the global random oracle with the corresponding key material when corruption occurs before any session with that trace is accepted.

Note further that the corrupt oracle also cleans up any entries in $T_s$ corresponding to traces of potentially fresh sessions that the attacker failed to conclude before corruption (e.g., by delivering an inconsistent MAC). This means that, after corruption of $pw_{US}$, the only entries that remain in $T_s$ corresponding to $(U,S)$ interactions must correspond to at least one fresh accepted session.

Now let us confirm that any inconsistencies detected by the adversary in its queries to the global random oracle (because it has observed in the game the output of the independent random oracle or some usage of it) are correctly signalled by activating the $\text{bad}_6$ flag.

The attacker may observe the action of the independent random oracle via the following oracle queries:

- $\text{Reveal}$ or $\text{Test}$ queries to accepted fresh sessions.
- MAC values computed using keys derived from entries in $T_s$ rather than the global random oracle.

For all entries added and never removed from $T_s$, the $\text{bad}_6$ flag is activated if a matching entry is ever queried to $H$ in the game. This covers the first class of observations. MAC checks and computations for such sessions are handled similarly.

The possibility remains, however, that the attacker might observe a MAC value related to an entry in $T_s$ that is later moved to $T$. If the query to $H$ that causes the inconsistency comes while the offending entry is in $T_s$, then $\text{bad}_6$ is activated as before. Otherwise, note that entries moved from $T_s$ to $T$ retain the key material, so subsequent queries to $H$ will actually be consistent.

**GAME $G_7$.** The intuition of this hop is identical to that for game $G_7$ in the proofs of the previous section: we change the way we generate messages in $\text{Send}$ queries and set a bad flag if, due to some unlikely event, one of the random exponents generated by the game to do this occurs as a password in the game. This event can be bounded by a reduction to GDL, as shown in Lemma 6.5, so we have:

$$\mathsf{Adv}_{\mathcal{A}}^{\text{SPAKE2}}(\cdot) \leq \frac{(q_s + q_e)^2}{2q} + \mathsf{Adv}_{\mathcal{B}_1}^{\text{U-GCDH}}(\cdot) + \mathsf{Adv}_{\mathcal{B}_2}^{\text{GDL}}(\cdot) + \Pr[G_7 \Rightarrow \text{bad}_6]$$

**GAME $G_8$.** In this game we reject any tags submitted by the attacker to a $\text{Send}$ oracle that would result in a new fresh session being accepted, when it has no information on the corresponding authentication key. More in detail:
If a fresh session is about to be accepted on the initiator side, and there does not exist a matching session on the responder side, then the attacker has never observed any computation based on the corresponding $T_s$ entry.

If a fresh session is about to be accepted on the responder side, and there does not exist a matching session on the initiator side, then the attacker may have observed a MAC computed with $K_{S,U}$, but it has no information about $K_{U,S}$.

Note that, by rejecting these sessions, the attacker can now only complete fresh sessions by behaving passively.\footnote{Forging MACs that result the same trace being accepted as in passive behaviour is not considered an active attack. This maps to the standard existential unforgeability notion.} We add a bad flag $\texttt{bad}_6$ to deal with cases where the tag would have been accepted in the previous game and, when this flag is activated, the potentially fresh session is forgotten and removed from the independent random oracle. This also means that, in this game, $\texttt{bad}_6$ can now only occur after corruption for sessions where the attacker behaved passively, which is why we can further restrict the setting of the bad fact to non-corrupted passwords. To see that this is indeed the case, observe that $\texttt{bad}_6$ happens while entries are in $T_s$, but in this game we remove from $T_s$ traces corresponding to active behaviour before the corresponding password is corrupted (as otherwise these sessions would not lead to fresh sessions anyway).

The probability of $\texttt{bad}_6$ occurring can be bounded using the PRF property of KDF and the standard (one-time) unforgeability of the MAC scheme. A formal proof is given in Lemma 7.4. Therefore we have

$$\Pr[\mathcal{G}_7 \Rightarrow \texttt{bad}_6] \leq \Pr[\mathcal{G}_8 \Rightarrow \texttt{bad}_6] + q_s \cdot \left( \text{Adv}_{G}^{\text{PRF}}() + \text{Adv}_{B}^{\text{UF}}() \right)$$

GAME $\mathcal{G}_9$. In this game we introduce changes similar to those we introduced in $\mathcal{G}_8$ of the proofs of Theorems 6.1 and 6.2 in moving all password-related checks to the end of the game. However, unlike in the previous section, we can now move all occurrences of $\texttt{bad}_6$ to the end, since we handled the problematic queries placed after corruption in the previous game hops. We move checks to the end by introducing a $T_{\text{bad}}$ list, and a final bad event $\texttt{bad}_9$ for executions where the attacker is able to construct two different consistent entries in this list for the same trace and different passwords.

Analysis of $\texttt{bad}_9$ and $\texttt{bad}_6$ in game $\mathcal{G}_9$ are identical to the analyses of $\texttt{bad}_6$ and $\texttt{bad}_8$ in game $\mathcal{G}_8$ of the proofs in the previous section: $\texttt{bad}_6$ is the probability of passwords guessing, whereas $\texttt{bad}_10$ can be reduced to GCDH as proved in Lemma 6.6. We therefore have

$$\Pr[\mathcal{G}_8 \Rightarrow \texttt{bad}_6] \leq \Pr[\mathcal{G}_9 \Rightarrow \texttt{bad}_6] + \Pr[\mathcal{G}_9 \Rightarrow \texttt{bad}_9] \leq \frac{q_e |P|}{|T|} + \text{Adv}_{GCDH}^{\text{GCDH}}()$$

This concludes the proof.\footnote{Forging MACs that result the same trace being accepted as in passive behaviour is not considered an active attack. This maps to the standard existential unforgeability notion.}

The above proof also establishes that SPAKE2 with key confirmation provides explicit authentication defined formally as:

- if a client $U$ accepts a session with server $S$, then $S$ has a pending unique session where it has derived the same key and session identifier ($U, S, X^*, Y^*$);
- if a server $S$ accepts a session with client $U$, then $U$ has already accepted a unique session with the same key and session identifier ($U, S, X^*, Y^*$).

We state this as the following corollary, which shows that impersonation can only be achieved with negligible probability beyond guessing the password.

**Corollary 7.2.** SPAKE2 with key confirmation provides explicit authentication.

**Proof.** The proof follows the same sequence of hops as that of Theorem 7.1. We can observe that in the final game a session is accepted only if the adversary behaved passively, which means that explicit authentication defined formally as:

- if a client $U$ accepts a session with server $S$, then $S$ has a pending unique session where it has derived the same key and session identifier ($U, S, X^*, Y^*$);
- if a server $S$ accepts a session with client $U$, then $U$ has already accepted a unique session with the same key and session identifier ($U, S, X^*, Y^*$).
authentication is guaranteed with the same bound. More precisely, the attacker can break explicit authentication with probability at most
\[ \frac{q_s}{|P|} + \text{Adv}_{B_{6}}^\mathcal{U}\text{-GCDH}(\cdot) + \text{Adv}_{B_{2}}^\mathcal{GDL}(\cdot) + \text{Adv}_{B_{3}}^{\mathcal{P}^\star}\text{-GCDH}(\cdot) + q_s \cdot \left( \text{Adv}_{B_{7}}^{1\text{PRF}}(\cdot) + \text{Adv}_{B_{8}}^{1\text{UF}}(\cdot) \right) + \frac{(q_s + q_e)^2}{2q} \]

\[ \square \]

**Remark.** We note that the proof strategy we use here for dealing with sessions where the attacker behaves passively actually shows that SPAKE2 is secure in a stronger model where the adversary does not have access to the EXECUTE oracle and can instead adaptively decide how to orchestrate a bounded number of \( q_e \) passive attacks concurrently with a bounded number of \( q_s \) active attacks. For this, the use of the gap oracle is essential in the proof step that handles passive attacks. The strategy we used in the previous section does not use this strategy for clarity, but the same observations apply.

**Lemma 7.3.** For every attacker \( \mathcal{A} \), there exists an attacker \( \mathcal{B}_6 \), such that
\[ \Pr[H_4 \Rightarrow \text{bad}_4] \leq \text{Adv}_{\mathcal{B}_6}^{\mathcal{U}\text{-GCDH}}(\cdot) \]

**Proof.** Let us consider the GCDH attacker \( \mathcal{B}_6 \) in Fig. 11. It gets a generalized challenge of the form \( (X_1, \ldots, X_{q_e+q_s}, Y_1, \ldots, Y_{q_e+q_s}) \) and finds CDH(\( X_i, Y_j \)), for some \( 1 \leq i, j \leq q_e \). This problem is tightly equivalent to \( \mathcal{U}\text{-GCDH} \) by random self-reducibility. The attacker embeds \( X_{k_s} \) and \( Y_{k_s} \) in all game-generated \( X^* \) and \( Y^* \), which means that it cannot compute the CDH for any session. However, it can use the DDH oracle to check when \( H \) queries for a given group element \( Z \) match traces where CDH(\( X, Y \) = \( Z \). Additionally, the DDH oracle is used to check for the bad event, in which case a solution to the GCDH problem was found. \[ \square \]

**Lemma 7.4.** For every attacker \( \mathcal{A} \), there exist attackers \( \mathcal{B}_7 \) and \( \mathcal{B}_8 \), such that
\[ \Pr[H_8 \Rightarrow \text{bad}_8] \leq q_s \cdot \left( \text{Adv}_{\mathcal{B}_7}^{1\text{PRF}}(\cdot) + \text{Adv}_{\mathcal{B}_8}^{1\text{UF}}(\cdot) \right) \]

**Proof.** Consider the modified version of game \( \mathcal{G}_4 \) we present in Figure 11 (game \( \mathcal{G}_8^c \)) and let us consider for now the behaviour of this game when \( c = F \). In this case, the only difference between the two games is that game \( \mathcal{G}_8^F \) only sets \( \text{bad}_8 \) in one special \( \ell \)-th query to \( \text{SEND} \), while the original version of \( \mathcal{G}_8 \) might set the flag in many \( \text{SEND} \) queries. However, this special query is chosen uniformly at random and is outside of the adversary’s view, so it must be the case that, conditioning on the probability that the randomly chosen \( \ell \) correctly guesses the query that causes the first occurrence of \( \text{bad}_8 \) in game \( \mathcal{G}_8 \), the probability of \( \text{bad}_8 \) occurring in the two games is the same. This implies,
\[ \Pr[H_8 \Rightarrow \text{bad}_8] = \Pr[H_8^F \Rightarrow \text{bad}_8 | \ell \text{ is correct}] \leq q_s \Pr[H_8^F \Rightarrow \text{bad}_8] \]

We can now focus on game \( \mathcal{G}_8^c \). Observe that parameter \( c = T \) replaces the keys produced by the KDF with random ones for the \( \ell \)-th query. This means that it is easy to construct an attacker \( \mathcal{B} \) such that
\[ \text{Adv}_{\mathcal{B}_7}^{1\text{PRF}}(\cdot) = |\Pr[H_8^c \Rightarrow \text{bad}_8] - \Pr[H_8^T \Rightarrow \text{bad}_8]| \]

We omit the details of this reduction. At this point, it suffices to bound the probability of \( \text{bad}_8 \) in game \( \mathcal{G}_8^c \). We do this via a direct reduction to the one-time unforgeability property of the MAC scheme. This is shown in Figure 11 as attacker \( \mathcal{B}_8 \). Again the reduction is simple: for the critical \( \ell \)-th query, the attacker fixes the authentication key that causes \( \text{bad}_8 \) implicitly as the external MAC key for which it is trying to forge. When \( \text{bad}_8 \) occurs, it returns the offending trace and authenticator as a forgery. \[ \square \]
8 Implications

Nothing-up-your-sleeve global parameters. Our security proof applies without change to any distribution of the global parameters \((M, N)\) under which computing the CDH between the two with the help of a DDH oracle is assumed to be hard. This means that simply sampling \((M, N)\) uniformly at random is a good choice when the GCDH assumption holds, and so is choosing \(M = H(K)\) and \(N = H(K')\) when \(H\) is modeled as a random oracle. Here the choice of \(K \neq K'\) means one is relying on the standard GCDH assumption, whereas \(K = K'\) leads to \(M = N\), which means relying on the squared GCDH problem.

Removing the global parameters. Our proof extends to a variant of SPAKE2 where \((M, N)\) are not generated in a global setup. Instead one can use \((M, N) = H(U, S)\) and \(N = H(S, U)\), where different values of these parameters are precomputed for each user-server pair. The online efficiency of the protocol is not affected. Modeling \(H\) as a random oracle, this means that each \((U, S)\) pair now poses an independent GCDH challenge to the active attacker. Our proof can be easily modified to cover this case with the same bound by observing that bad events that involve multiple sessions are always defined between sessions established for the same user-server \((U, S)\).

Reusing passwords. Our model assumes that user passwords are sampled independently for each pair \((U, S)\). The simple case of password reuse, where passwords are either repeats or they are sampled independently can be easily addressed by extending the corrupt oracle to exclude additional trivial attacks: it should declare as corrupt all \((U', S')\) pairs that use the same password. The same proof applies and the bound is not affected in this case, as the entropy of non-repeat passwords is assumed to be unaffected.

The treatment of more complex distributions follows in the same lines as described in [KOY09, Section 2.2.1].

Acknowledgments. We would like to thank David Pointcheval for helpful discussions regarding the proof of Lemma 6.6 and Doreen Riepel for discussions that helped us clarify the proofs of Lemmas 6.5 and 6.6. This work was supported in part by the ERC Project aSCEND (H2020 639554) and by the French ANR ALAMBIC Project (ANR-16-CE39-0006). Manuel Barbosa was supported by grant SFRH/BSAB/143018/2018 awarded by FCT, Portugal.

References


Perfect Forward Security of SPAKE2


A Games and adversaries for the proof of Theorems 6.1 and 6.2
\begin{align*}
&\text{proc Initialize(} T, K \text{)} \quad \text{G5} \\
&\quad \text{if } m = \text{msg}_2 \text{ and } \text{msg}_2 = 0 \text{ then } \text{Return} (0) \\
&\quad \text{else if } \text{msg}_2 = 1 \text{ then } \text{Return} (1) \\
&\quad \text{else } \text{Return} (0) \\
&\text{sendResp(} T, \text{msg}_3 \text{)} \\
&\quad \text{if } \text{msg}_3 = 0 \text{ then } \text{Return} (0) \\
&\quad \text{else if } \text{msg}_3 = 1 \text{ then } \text{Return} (1) \\
&\quad \text{else } \text{Return} (0) \\
&\text{sendResp(} T, \text{msg}_4 \text{)} \\
&\quad \text{if } \text{msg}_4 = 0 \text{ then } \text{Return} (0) \\
&\quad \text{else if } \text{msg}_4 = 1 \text{ then } \text{Return} (1) \\
&\quad \text{else } \text{Return} (0) \\
&\text{Finalize(} T, K \text{)} \\
&\quad \text{Return} (0) \\
\end{align*}

\text{Fig. 4. Security proof for SPAKE2. Games 5 to 7.}
SendTerm:

\[ \text{If } \pi_j \neq \pi_i \text{ Return } \perp \]
\[ \text{Else } \]
\[ X \leftarrow \{ (x, U, S, \pi_j, \perp), \ast, F, F \} \]
\[ \text{Return } (X, \ast^*) \]

ProveSendInit\((u, i, S)\): pmod: \( \ast \leftarrow \{ \} \); \( T \leftarrow \{ \} \); \( T_0 \leftarrow \{ \} \); \( T_{\pi_i} \leftarrow \{ \} \text{; } \text{Return } \perp \)

ProveSend\((u, i, S)\): \( \ast \leftarrow \{ \} \); \( T \leftarrow \{ \} \); \( T_0 \leftarrow \{ \} \); \( T_{\pi_i} \leftarrow \{ \} \text{; } \text{Return } \perp \)

ProveSend\((u, i, S)\): \( \ast \leftarrow \{ \} \); \( T \leftarrow \{ \} \); \( T_0 \leftarrow \{ \} \); \( T_{\pi_i} \leftarrow \{ \} \text{; } \text{Return } \perp \)

ProveSend\((u, i, S)\): \( \ast \leftarrow \{ \} \); \( T \leftarrow \{ \} \); \( T_0 \leftarrow \{ \} \); \( T_{\pi_i} \leftarrow \{ \} \text{; } \text{Return } \perp \)

Perfect Forward Security of SPAKE2 23
 ADVISORY B₁

Procedure Initialize(A)

Procedure SendTerm(P, i)

Procedure SendResp(P, i)

Procedure SendK(P, i)

Procedure Finalize(k')

ADVISORY B₂

Procedure Initialize(4)

Procedure SendTerm(4)

Procedure SendResp(4)

Procedure SendK(4)

Procedure Finalize(4)

ADVISORY B₃

Procedure Initialize(M, N)

Procedure SendTerm(U, S)

Procedure SendResp(U, S, X*)

Procedure SendK(U, S, X*)

Procedure Finalize(U, S, X*)

Fig. 6. Security proof for SPAKE2. GCDH and GDL attackers.
B Games and adversary for the proof of Theorem 7.1
Fig. 7. Security proof for SPAKE2. Algebraic reduction to GCDH and alternative reduction to GPCCDH.
Fig. 8. Security proof for SPAKE2 with confirmation. Games 1 to 4.
Fig. 9. Security proof for SPAKE2 with confirmation. Games 5 to 7.
\[
\begin{align*}
\text{SendInit}(\pi_i) & \quad \text{return } b = \text{true} \\
\text{Return}(b) & \quad \text{return } b = \text{true} \\
\text{SendTermInit}(\pi_i) & \quad \text{return } b = \text{true} \\
\text{Return}(b) & \quad \text{return } b = \text{true} \\
\text{SendTermInit}(\pi_i) & \quad \text{return } b = \text{true} \\
\text{Return}(b) & \quad \text{return } b = \text{true} \\
\end{align*}
\]

Fig. 10. Security proof for SPAKE2 with confirmation. Games 8 to 9.