Security Analysis of Signal's PQXDH Handshake

Version 2.0[*]

Rune Fiedler[1] Felix Günther[2]

[1] Cryptoplexity, Technische Universität Darmstadt, Germany
[2] IBM Research Europe – Zurich, Switzerland
rune.fiedler@cryptoplexity.de, mail@felixguenther.info

Abstract. Signal recently deployed a new handshake protocol named PQXDH to protect against "harvest now, decrypt later" attacks of a future quantum computer. To this end, PQXDH adds a post-quantum KEM to the Diffie–Hellman combinations of the prior X3DH handshake.

In this work, we give a reductionist security analysis of Signal's PQXDH handshake in a game-based security model that captures the targeted "maximum-exposure" security against both classical and quantum adversaries, allowing fine-grained compromise of user's long-term, semi-static, and ephemeral key material. We augment prior such models to capture not only the added KEM component but also the signing of public keys, which prior analyses did not capture but which adds an additional flavor of post-quantum security in PQXDH. We then establish fully parameterized, concrete security bounds for the classical and post-quantum session key security of PQXDH, and discuss how design choices in PQXDH make a KEM binding property necessary and how a lack of domain separation reduces the achievable security.

Our discussion of KEM binding and domain separation complements the concurrent tool-based analysis of PQXDH by Bhargavan, Jacomme, Kiefer, and Schmidt (USENIX Security 2024), which pointed out a potential re-encapsulation attack if the KEM shared secret does not bind the public key. In contrast to the tool-based analysis, we analyze all protocol modes of PQXDH and its "maximum-exposure" security. We further show that both Kyber (used in PQXDH) and the NIST standard ML-KEM (expected to replace Kyber) satisfy a novel binding notion we introduce and rely on for our PQXDH analysis, which may be of independent interest.

1 Introduction

Billions of people today use messaging apps such as Facebook Messenger, Google Messages, Skype, Signal, or WhatsApp, in which the Signal end-to-end encryption protocol [Sig] secures the communication. The Signal protocol consists of two main components: The initial handshake protocol, which allows two parties to derive a shared session key while authenticating each other, and the Double Ratchet protocol, which allows renewing the session key in an ongoing session to achieve forward secrecy and post-compromise security. Until 2023, Signal used X3DH [MP16b] as initial handshake protocol. In 2023, Signal released PQXDH [KS23], which modifies X3DH to add protection against quantum adversaries; in particular against "harvest now, decrypt later" attacks where an adversary records communication today to break its confidentiality when a cryptographically-relevant quantum computer becomes available in the future.

In X3DH, each user has Diffie–Hellman (DH) keys of different lifetimes (long-term, semi-static, and ephemeral). Combining several DH keys of two users Alice and Bob in the initial key agreement ensures

[*] We provide an overview of major changes between versions in Appendix A.

mutual authentication and, as long as one of the combinations of keys remains uncompromised, that the derived session key is secure. To protect against "harvest now, decrypt later" quantum adversaries, PQXDH [KS23] adds a post-quantum key encapsulation mechanism (KEM) to X3DH while keeping the existing, well-understood [CCD+17] handshake structure unmodified. Furthermore, X3DH signs the involved semi-static DH key and PQXDH additionally the KEM key under the the user's long-term key.

In a nutshell, the protocol message flow is as follows (where the KEM appears only in PQXDH): Bob sends his long-term, semi-static, and ephemeral DH shares, as well as an ephemeral KEM public key for a full handshake; a reduced handshake omits the ephemeral keys and has a semi-static KEM public key instead. Alice generates an ephemeral DH key pair, encapsulates against Bob's KEM key, and sends her ephemeral DH key and KEM ciphertext to Bob. Both parties derive the session key from DH shared secrets from three or four DH key combinations (long-term/semi-static, ephemeral/long-term, ephemeral/semi-static, and, if present, ephemeral/ephemeral) and the KEM shared secret.

Security Analyses of X3DH and PQXDH

The development of PQXDH was accompanied by formal, tool-based verification conducted by Bhargavan, Jacomme, Kiefer, and Schmidt (BJKS) [BJKS23, BJK23], subsequently (and concurrently to this work) expanded and published in [BJKS24]. Their analysis provided improvements and security assurance for PQXDH (leading to revisions of the protocol description [KS23, KS24]), using both the symbolic-analysis tool ProVerif [BC] and the computational-analysis tool CryptoVerif [Bla]. Security of Signal's original X3DH handshake (as well as its ratcheting protocol) was established by Cohn-Gordon, Cremers, Dowling, Garratt, and Stebila (CCDGS) [CCD+17] through a reductionist security analysis in what we refer to as a "maximum-exposure"[1], game-based security model. Vatandas, Gennaro, Ithurburn, and Krawczyk [VGIK20] analyzed the deniability of X3DH, and Fiedler and Janson [FJ24] the deniability of PQXDH.

The core difference between the BJKS analysis [BJKS24] of PQXDH and the CCDGS analysis [CCD+17] for X3DH is that while the tool-based analysis of BJKS provide machine-checked assurance, their ProVerif results remain on the symbolic, protocol-logic level assuming perfect cryptography and their CryptoVerif model does not account for all "maximum-exposure" attack vectors that Signal aims to protect against. Concretely, for proof complexity reasons, the CryptoVerif model of BJKS focuses on a reduced set of simpler properties and, for example, only captures the compromise of long-term user keys (but not of semi-static and ephemeral keys), does not model the full security contributed by certain key combinations, and separately considers attacks only against the classical DH components and only against the post-quantum KEM component. Such fine-grained compromise, however, is a main reason for combining multiple secrets in Signal's X3DH and PQXDH handshakes, and combined classic/post-quantum hybrid guarantees a main reason behind the PQXDH design specifically. Furthermore, BJKS only model the semi-static KEM keys in PQXDH but not the ephemeral ones, sidestepping a domain separation issue in the protocol that allows signed ephemeral and semi-static KEM public keys to be confused. Hence, there remains a notable gap between the security guarantees formally confirmed so far and those aimed at by PQXDH.

Our Contributions

This work completes the analysis picture for PQXDH by providing a reductionist security analysis of PQXDH (Revision 3 [KS24], throughout this paper) in a "maximum-exposure", game-based security model following that for X3DH by CCDGS [CCD+17].

Augmented game-based security model. To this end, we first augment the coded game-based security model of Brendel, Fiedler, Günther, Janson, and Stebila (BFGJS) [BFG+22, BFG+21] to capture the

[1] The adversary may compromise nearly all secrets as long as trivial wins are excluded.

added KEM components in PQXDH. In particular, we model that the KEM public key and the semi-static DH shares are signed in PQXDH (instead of assuming them to be authentically distributed like in prior game-based models [CCD$^+$17, BFG$^+$22, BFG$^+$21]). This requires careful adaptation of related corruption modeling, which adds both model and proof complexity but leads to a security model which more closely[2] captures the real-world deployment. Similarly, we specifically capture the key confusion of ephemeral and semi-static KEM public keys possibly in PQXDH and when we can still expect security, enabling us to capture the ephemeral KEM public keys omitted in the model of BJKS [BJKS24].

Concrete security bound for PQXDH. We then analyze the PQXDH handshake with respect to its security against both classical and quantum adversaries. We establish classic security based on the Gap Diffie–Hellman (GapDH) assumption holding in the DH group, the KEM providing one-way security under chosen-ciphertext attacks (OW-CCA), a certain form of binding property (which we expand on below), and low key-collision probability and correctness errors of the KEM, the signature scheme being existentially unforgeable (EUF-CMA), and the key derivation function behaving like a random oracle. In particular, we confirm that combining DH and KEM keys yields a combined, *hybrid* security bound, compared to the separate analyses in BJKS [BJKS24]. Against quantum adversaries, we establish security in the standard model, based on the KEM providing indistinguishability under chosen-ciphertext attacks (IND-CCA) and the same binding and correctness properties as before, and the key derivation function being a secure PRF when keyed by the KEM shared secret. Our resulting security theorems for PQXDH are fully parameterized, giving concrete security bounds for each component.

Design discussion and KEM binding. We conclude with a discussion of the PQXDH design, with insights that complement those by BJKS [BJKS24]. For example, our model and proofs pinpoint the stronger security achievable if PQXDH were to properly domain-separate ephemeral and semi-static KEM public keys. Furthermore, we argue how the key derivation approach taken by Signal's X3DH and PQXDH leads to requiring a certain *binding* property from the KEM in our analysis, and how this requirement could be easily avoided.

For background, BJKS in their analysis discovered the following potential "KEM re-encapsulation attack", a class of attacks introduced by Cremers, Dax, and Medinger (CDM) [CDM24], on PQXDH: An adversary can learn a KEM shared secret by compromising the involved KEM secret key, and then re-encapsulating this shared secret under another, uncompromised KEM public key. As a consequence, decapsulations under two distinct public keys yield the same shared secret, and two sessions with different KEM public keys compute the same session key, violating the protocol's security goals.

The initial BJKS analysis [BJKS23, BJK23] excluded this attack by modeling that the KEM public key is added to the associated data when AEAD-encrypting messages, a suggestion that the PQXDH description however only follows optionally: it argues that Kyber [SAB$^+$], used in the Signal implementation, already prevents the re-encapsulation attack [KS24, Sections 3.3 and 4.12]. Our analysis covers Signal's approach, providing fine-grained insight into the properties required of the KEM to ensure security of the initial key agreement, *without* relying on the AEAD. Concretely, we require that the derived shared secret *binds* the involved public key and ciphertext, in a setting where the adversary knows the involved keys *and the randomness* to generate them. Capturing binding under revealed randomness is necessary to treat the "maximum-exposure" security of PQXDH. We prove that both Kyber [SAB$^+$] (currently deployed in PQXDH) and the new NIST standard ML-KEM [Nat24] (provisioned to replace Kyber in PQXDH) satisfy this binding property.

We view these KEM binding results as being of independent interest: They complement a range of binding notions for KEMs defined in concurrent work by CDM [CDM24], yet their notion that would fit

[2] Security modeling still always requires trade-offs in abstraction, see our discussion in Section 6.

```
$\mathcal{G}^{\mathsf{prfsec}}_{\mathsf{PRF}}(\mathcal{A})$:
$K \leftarrow\!\!\$ \{0,1\}^{l_k}$
$g \leftarrow\!\!\$ \{\text{functions } f\colon \{0,1\}^{l_i} \to \{0,1\}^{l_o}\}$
$b \leftarrow\!\!\$ \{0,1\}$
$b' \leftarrow\!\!\$ \mathcal{A}^{\mathrm{PRF}}()$
return $[\![b' = b]\!]$

$\mathrm{PRF}(x)$:
if $b = 0$
    return $\mathsf{PRF}(K, x)$
else
    return $g(x)$
```

Figure 1: Security game for PRF security for a function PRF defined in Definition 2.1.

the PQXDH setting is *not* satisfied by $\mathsf{ML\text{-}KEM}$ [Sch24]. Furthermore, BJKS expanded their initial analysis in [BJKS24], published after this work, to also include a CryptoVerif analysis assuming a KEM binding property and proving it for Kyber. We discuss and relate the binding notion we introduce and require in our PQXDH analysis to the framework of CDM and to the notion used by BJKS in more detail in Section 3. Binding properties also arise in other key-combiner settings, e.g., in multi-recipient KEMs [AHK$^+$23] and X-Wing [BCD$^+$24], where the shared secret shall bind the ciphertext.

Related work

Several fully post-quantum replacements for $\mathsf{X3DH}$ have been considered in the literature: Hashimoto, Katsumata, Kwiatkowski, and Prest [HKKP22] propose SC-DAKE, a generic construction based on KEMs and a ring signature. BFGJS [BFG$^+$22] propose a similar construction called SPQR based on KEMs and a designated verifier signature. Dobson and Galbraith [DG22] lift $\mathsf{X3DH}$ to supersingular isogenies with their construction called SI-X3DH, which is however broken by the SIDH attack [CD23, MMP$^+$23, Rob23]. Collins, Huguenin-Dumittan, Nguyen, Rolin, and Vaudenay [CHDN$^+$24] propose K-Waay, based on adapting split KEMs [BFG$^+$20]. In this work, we focus on Signal's deployed PQXDH protocol.

2 Preliminaries

2.1 Notation

We refer to the ith element of a list or vector l with $l[i]$. For $n \in \mathbb{N}$, $[n]$ denotes $\{1, \ldots, n\}$. We write a deterministic or randomized algorithm A with input x and output y as $y \leftarrow A(x)$ resp. $y \leftarrow\!\!\$ A(x)$ (or also $A(x) \to y$ resp. $A(x) \$\!\!\to y$), for the latter with explicit randomness r we write $y \leftarrow B(x; r)$. We sample an element e uniformly at random from a set S with $e \leftarrow\!\!\$ S$. We write $\Pr[\mathcal{G}(\mathcal{A})]$ for the probability that a game $\mathcal{G}$ involving an adversary $\mathcal{A}$ outputs true.

2.2 Pseudorandom Functions

We recall the definition of pseudorandom functions.

Definition 2.1. *Let* $\mathsf{PRF}\colon \{0,1\}^{l_k} \times \{0,1\}^{l_i} \to \{0,1\}^{l_o}$ *be an efficient keyed function with key length* l_k*, input length* l_i*, and output length* l_o*. We say that* PRF *is* $(t, \epsilon, q_{\mathsf{PRF}})$–PRF-secure*, if for any adversary* $\mathcal{A}$ *against* $\mathcal{G}^{\mathsf{prfsec}}_{\mathsf{PRF}}(\mathcal{A})$ *defined in Figure 1 with running time at most* t *and making at most* q_{PRF} *queries to the* PRF *oracle, we have that*

$$\mathsf{Adv}^{\mathsf{prfsec}}_{\mathsf{PRF}}(\mathcal{A}) := \left| \Pr\left[\mathcal{G}^{\mathsf{prfsec}}_{\mathsf{PRF}}(\mathcal{A}) = 1\right] - \frac{1}{2} \right| \leq \epsilon.$$

```
$\underline{\mathcal{G}^{\mathsf{EUF\text{-}CMA}}_{\mathsf{SIG}}(\mathcal{A})}$:
    $(pk, sk) \leftarrow_{\$} \mathsf{SIG.KGen}()$
    $Q \leftarrow \emptyset$
    $(m^*, \sigma^*) \leftarrow_{\$} \mathcal{A}^{\textsc{Sign}}(pk)$
    return $[\![\mathsf{Vf}(pk, m^*, \sigma^* \wedge m^* \notin Q]\!]$

$\underline{\textsc{Sign}(m)}$:
    $Q \leftarrow Q \cup \{m\}$
    return $\mathsf{SIG.Sig}(sk, m)$
```

Figure 2: Security game for EUF-CMA for a signature scheme $\mathsf{SIG} = (\mathsf{KGen}, \mathsf{Sig}, \mathsf{Vf})$ defined in Definition 2.3.

2.3 Signatures

Definition 2.2 (Signature scheme)**.** *A signature scheme* $\mathsf{SIG} = (\mathsf{KGen}, \mathsf{Sig}, \mathsf{Vf})$ *with associated message space* $\mathcal{M}$ *is a triple of* PPT *algorithms:*

- $\mathsf{KGen}() \xrightarrow{\$} (pk, sk)$*: The probabilistic key generation outputs a public-key/secret-key pair* (pk, sk)*.*
- $\mathsf{Sig}(sk, m) \xrightarrow{\$} \sigma$*: The probabilistic signing takes as input a secret key sk and a message m and outputs a signature* σ*.*
- $\mathsf{Vf}(pk, m, \sigma) \rightarrow d$*: The deterministic verification takes as input a public key pk, a message m, and a signature* σ *and outputs a decision* $d \in \{0, 1\}$*. If* $d = 1$ *we say the signature is valid.*

A signature scheme is correct, if for all $(pk, sk) \leftarrow_{\$} \mathsf{KGen}()$ *and for all* $m \in \mathcal{M}$ *it holds that* $\Pr[\mathsf{Vf}(pk, m, \mathsf{Sig}(sk, m)) = 1]$.

Definition 2.3 (EUF-CMA Security of Signatures)**.** *Let* $\mathsf{SIG} = (\mathsf{KGen}, \mathsf{Sig}, \mathsf{Vf})$ *be a signature scheme. We say that* SIG *is* $(t, \epsilon, q_{\mathrm{Sig}})$–EUF-CMA*-secure, if for any adversary* $\mathcal{A}$ *against Game* $\mathcal{G}^{\mathsf{EUF\text{-}CMA}}_{\mathsf{SIG}}(\mathcal{A})$ *defined in Figure 2 with running time at most t and making at most* q_{Sig} *queries to the* Sign *oracle, we have that*

$$\mathsf{Adv}^{\mathsf{EUF\text{-}CMA}}_{\mathsf{SIG}}(\mathcal{A}) := \Pr\left[\mathcal{G}^{\mathsf{EUF\text{-}CMA}}_{\mathsf{SIG}}(\mathcal{A})\right] \leq \epsilon.$$

2.4 Hash Functions

We follow the "human ignorance" approach of Rogaway [Rog06] to capture practical, unkeyed hash functions, giving explicit constructions for adversaries against, e.g., the collision resistance of a hash function in our results.

Definition 2.4 (Preimage resistance of hash functions)**.** *Let* $\mathsf{H}\colon \{0,1\}^* \rightarrow \{0,1\}^n$ *be a (hash) function. We define the advantage of an adversary* $\mathcal{A}$ *against the* preimage resistance *of* H*, distinguishing whether challenges are sampled from (a subset of) the domain* $\mathcal{M} \subseteq \{0,1\}^*$ *(*PR-d*) or from the range* $\{0,1\}^n$ *(*PR-r*), as*

$$\mathsf{Adv}^{\mathsf{PR\text{-}d}}_{\mathsf{H},\mathcal{M}}(\mathcal{A}) := \Pr\left[d = \mathsf{H}(m') \mid m \leftarrow_{\$} \mathcal{M}, d \leftarrow \mathsf{H}(m), m' \leftarrow_{\$} \mathcal{A}(d)\right], \quad \textit{resp.}$$
$$\mathsf{Adv}^{\mathsf{PR\text{-}r}}_{\mathsf{H}}(\mathcal{A}) := \Pr\left[d = \mathsf{H}(m') \mid d \leftarrow_{\$} \{0,1\}^n, m' \leftarrow_{\$} \mathcal{A}(d)\right].$$

Definition 2.5 (Collision resistance of hash functions)**.** *Let* $\mathsf{H}\colon \{0,1\}^* \rightarrow \{0,1\}^n$ *be a (hash) function. We define the advantage of an adversary* $\mathcal{A}$ *against the* collision resistance *(*CR*) of* H *as*

$$\mathsf{Adv}^{\mathsf{CR}}_{\mathsf{H}}(\mathcal{A}) := \Pr\left[\mathsf{H}(m) = \mathsf{H}(m') \wedge m \neq m' \mid (m, m') \leftarrow_{\$} \mathcal{A}()\right].$$

Our analysis further uses that finding two input values that collide in the truncated (256-bit) output is hard. Prior work discusses the related (weak) near collision resistance [MvV97, PS14, JKN21], partial collision resistance [WY05, LT11, YCW14], and truncated collision resistance [JK18]. We model this as collision resistance of a hash function H truncated to the first 256 bits, which we denote by H_{256}.

```
$\underline{\mathcal{G}^{\mathsf{GDH}}_{(\mathbb{G},g,q)}(\mathcal{A})\text{:}}$
    $a, b \leftarrow\!\!\$\ \mathbb{Z}_q$
    $g^z \leftarrow\!\!\$\ \mathcal{A}^{\mathrm{DDH}}((\mathbb{G}, g, q), a, b)$
    **return** $[\![g^z = g^{ab}]\!]$

$\underline{\mathrm{DDH}(g^x, g^y, g^z)\text{:}}$
    **return** $[\![\mathsf{CDH}(g^x, g^y) = g^z]\!]$
```

Figure 3: Security game for the GapDH problem defined in Definition 2.7.

SHA3. Bertoni, Daemen, Peters, and Van Assche [BDPV08] show that SHA3 is indifferentiable from a random oracle when the underlying permutation is modeled as a random permutation. Under this assumption, finding a collision or preimage in SHA3 or $\mathsf{SHA3}_{256}$ is as hard as finding it in a random oracle yielding digests of the same length.

2.5 Diffie–Hellman Key Exchange

Definition 2.6 (Diffie–Hellman Key Exchange)**.** *A* Diffie–Hellman Key Exchange (DH) scheme *is a tuple of algorithms* $(\mathsf{KGen}, \mathsf{DH})$*, defined as follows:*

- $\mathsf{KGen}() \overset{\$}{\rightarrow} (pk, sk)$*: This probabilistic algorithm returns a key pair* (pk, sk)
- $\mathsf{DH}(pk_A, sk_B) \rightarrow \mathsf{DH}_{AB}$*: On input a public key* pk_A *and a secret key* sk_B*, this deterministic algorithm returns the shared DH secret* DH_{AB}*.*

We say that a DH key exchange DH *is* correct *if, for every* $(pk_A, sk_A), (pk_B, sk_B) \leftarrow\!\!\$\ \mathsf{KGen}()$*, it holds that*

$$\Pr[\mathsf{DH}(pk_A, sk_B) = \mathsf{DH}(pk_B, sk_A)] = 1.$$

For notational convenience we allow the arguments to be in arbitrary order, i.e., $\mathsf{DH}(pk_A, sk_B) = \mathsf{DH}(sk_A, pk_B)$. We write $\mathsf{CDH}(pk_A, pk_B)$ to refer to the DH shared secret DH_{AB} when we do not care which secret key is used.

We rely on the GapDH problem [OP01, CCD$^+$17], i.e., that the computational DH problem is hard given a decisional DH oracle.

Definition 2.7 (GapDH problem)**.** *Let* $(\mathbb{G}, g, q)$ *be a group of order* q *with generator* g*. We say that the* $(t, \epsilon_{\mathsf{GDH}}, q_{\mathrm{DDH}})$-GapDH *problem holds in* $(\mathbb{G}, g, q)$ *if for any adversary* $\mathcal{A}$ *with running time at most* t *making at most* q_{DDH} *queries to its* DDH *oracle, we have that*

$$\mathsf{Adv}^{\mathsf{GDH}}_{(\mathbb{G},g,q)}(\mathcal{A}) := \Pr\Big[\mathcal{G}^{\mathsf{GDH}}_{(\mathbb{G},g,q)}(\mathcal{A})\Big] \leq \epsilon_{\mathsf{GDH}},$$

where $\mathcal{G}^{\mathsf{GDH}}_{(\mathbb{G},g,q)}(\mathcal{A})$ *is defined in Figure 3.*

3 Key Encapsulation Mechanisms and Binding Properties

We first recap the basic syntax, correctness, and security of key encapsulation mechanisms (KEMs), before discussing the novel KEM binding properties.

Definition 3.1 (Key Encapsulation Mechanisms)**.** *A* key encapsulation mechanism $\mathsf{KEM} = (\mathsf{KGen}, \mathsf{Enc}, \mathsf{Dec})$ *consists of the following three algorithms:*

- $\mathsf{KGen}() \overset{\$}{\rightarrow} (pk, sk)$*: The probabilistic* key generation *with randomness space* $\mathcal{R}_{\mathsf{KEM.KGen}}$ *outputs a public-key/secret-key pair with* $(pk, sk) \in \mathcal{PK} \times \mathcal{SK}$*.*

```
$\mathcal{G}_{\mathsf{KEM}}^{\mathsf{IND\text{-}CCA}}(\mathcal{A})$:
$b \leftarrow_\$ \{0,1\}$
$(pk, sk) \leftarrow_\$ \mathsf{KEM.KGen}()$
$(ct^*, ss_0^*) \leftarrow_\$ \mathsf{KEM.Enc}(pk)$
$ss_1^* \leftarrow_\$ \mathcal{SS}$
$b' \leftarrow_\$ \mathcal{A}^{\textsc{Decaps}}(pk, ct^*, ss_b^*)$
return $[\![b' = b]\!]$
```

```
$\mathcal{G}_{\mathsf{KEM}}^{\mathsf{OW\text{-}CCA}}(\mathcal{A})$:
$(pk, sk) \leftarrow_\$ \mathsf{KGen}()$
$(ct^*, ss^*) \leftarrow_\$ \mathsf{Enc}(pk)$
$ss' \leftarrow_\$ \mathcal{A}^{\textsc{Decaps}}(pk, ct^*)$
return $[\![ss' = ss^*]\!]$
```

```
$\textsc{Decaps}(ct)$:
if $ct = ct^*$
    return $\bot$
else
    return $\mathsf{Dec}(sk, ct)$
```

Figure 4: Security games for IND-CCA security (left) and OW-CCA security (middle) of KEM = (KGen, Enc, Dec) defined in Definition 3.3 and Definition 3.4, respectively.

- $\mathsf{Enc}(pk) \xrightarrow{\$} (ct, ss)$: *The probabilistic* encapsulation *algorithm with randomness space* $\mathcal{R}_{\mathsf{KEM.Enc}}$ *takes as input a public key* $pk \in \mathcal{PK}$ *and outputs a ciphertext* $ct \in \mathcal{C}$ *and the therein encapsulated shared secret* $ss \in \mathcal{SS}$*.*
- $\mathsf{Dec}(sk, ct) \to ss'$: *The deterministic* decapsulation *algorithm takes as input a ciphertext* $ct \in \mathcal{C}$ *and secret key* sk *and outputs* $ss' \in \mathcal{SS} \cup \{\bot\}$*, where* $\bot$ *indicates an error.*

We say that a KEM KEM = (KGen, Enc, Dec) *is* δ-correct *if, for every key pair* $(pk, sk) \leftarrow_\$ \mathsf{KGen}()$*, and every encapsulation* $(ct, ss) \leftarrow_\$ \mathsf{Enc}(pk)$*, we have*

$$\Pr\left[ss' \neq ss \mid ss' \leftarrow \mathsf{Dec}(sk, ct)\right] \leq \delta.$$

In our analysis, we sometimes need to rule out that honestly generated KEM public keys collide. We capture that probability in the following. For Kyber [SAB+] and ML-KEM [Nat24], such public-key collisions boil down to random 256-bit seeds colliding under SHA3-512.

Definition 3.2 (KEM public-key collision probability)**.** *We define the* public-key collision probability *of a KEM via a function* $\gamma_{\mathsf{coll}} \colon \mathbb{N} \to [0,1]$*, letting* $\gamma_{\mathsf{coll}}(n)$ *denote the probability that two among* n *honestly generated public keys collide:*

$$\gamma_{\mathsf{coll}}(n) := \Pr\Big[pk_i = pk_j \wedge i \neq j \mid (pk_i, sk_i) \leftarrow_\$ \mathsf{KGen}() \text{ for } i \in [1,n]\Big].$$

In our analysis, we require *indistinguishability under chosen-ciphertext attacks* (IND-CCA) against quantum adversaries and *one-way security under chosen-ciphertext attacks* (OW-CCA) against classical adversaries of the KEM scheme, where the former notion implies the latter.

Definition 3.3 (IND-CCA Security of KEMs)**.** *Let* KEM = (KGen, Enc, Dec) *be a KEM. We say that* KEM *is* $(t, \epsilon, q_{\mathrm{Dec}})$–IND-CCA*-secure, if for any adversary* $\mathcal{A}$ *against Game* $\mathcal{G}_{\mathsf{KEM}}^{\mathsf{IND\text{-}CCA}}(\mathcal{A})$ *defined in Figure 4 with running time at most* t *and making at most* q_{Dec} *queries to the* Decaps *oracle, we have that*

$$\mathsf{Adv}_{\mathsf{KEM}}^{\mathsf{IND\text{-}CCA}}(\mathcal{A}) := \left|\Pr\Big[\mathcal{G}_{\mathsf{KEM}}^{\mathsf{IND\text{-}CCA}}(\mathcal{A})\Big] - \frac{1}{2}\right| \leq \epsilon.$$

Definition 3.4 (OW-CCA Security of KEMs)**.** *Let* KEM = (KGen, Enc, Dec) *be a KEM. We say that* KEM *is* $(t, \epsilon, q_{\mathrm{Dec}})$–OW-CCA*-secure, if for any adversary* $\mathcal{A}$ *against Game* $\mathcal{G}_{\mathsf{KEM}}^{\mathsf{OW\text{-}CCA}}(\mathcal{A})$ *defined in Figure 4 with running time at most* t *and making at most* q_{Dec} *queries to the* Decaps *oracle, we have that*

$$\mathsf{Adv}_{\mathsf{KEM}}^{\mathsf{OW\text{-}CCA}}(\mathcal{A}) := \Pr\Big[\mathcal{G}_{\mathsf{KEM}}^{\mathsf{OW\text{-}CCA}}(\mathcal{A})\Big] \leq \epsilon.$$

3.1 Binding Properties

Due to the way PQXDH includes (only) the KEM shared secret in its key derivation (but not the KEM public key or ciphertext), our security analysis relies on a novel binding property of the KEM scheme. In a nutshell, we ask that it is hard to find two distinct ciphertexts or (honestly generated) public keys so that the corresponding decapsulations output the same shared secrets, even when given full control over the ciphertexts and knowing corresponding secret keys and the *randomness* used to generate these keys. Revealing also the key generation randomness to the adversary is necessary to treat the "maximum-exposure" security of PQXDH, in which an adversary can reveal a session's random coins (cf. Section 4). The PQXDH setting, in which an adversary interacts with many KEM keys, also motivates our notion being *multi-user*.

Aligning with language of Cremers, Dax, and Medinger (CDM) [CDM24] from their concurrent work systematizing *binding* notions for KEMs, our notion asks that the KEM shared secret binds both public key and ciphertext used to produce it, under leakage of key generation randomness (and hence secret keys), which is a novel variant we denote as LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$. In general, their notion X-BIND-P-Q captures that the components in set $P \in \{\{ss\},\{ct\},\{ss,ct\}\}$ bind the components in set $Q \in \{\{pk\},\{ss\},\{ct\}\}$, where the KEM keys are chosen by the adversary ($X = \mathit{MAL}$), honestly generated and leaked to the adversary ($X = \mathit{LEAK}$), or honestly generated without leakage of secrets ($X = \mathit{HON}$). We reproduce their LEAK and MAL notions for ss simultaneously binding pk and ct in our syntax[3] on the right-hand side of Figure 5.

In the following, we formalize our new LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ binding property and relate it to the notions in the CDM framework [CDM24]. In particular, we show in Section 3.3 below that both Kyber and the NIST standard ML-KEM satisfy the LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ property, while the notion from [CDM24] that implies ours, MAL-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$, is not satisfied by ML-KEM [Sch24]. While Signal's current implementation of PQXDH uses Kyber [SAB+], it is already prepared[4] to transition to the new NIST standard ML-KEM [Nat24].

Definition 3.5 (LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$)**.** *Let* $\mathsf{KEM} = (\mathsf{KGen},\mathsf{Enc},\mathsf{Dec})$ *be a KEM and* $n \geq 2$. *We say that* KEM *is* (t,ϵ,n)-LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$-secure*, if for any adversary* $\mathcal{A}$ *against Game* $\mathcal{G}_{\mathsf{KEM},n}^{\mathit{LEAK}^{+r}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{A})$ *defined in Figure 5 with running time at most* t*, we have that*

$$\mathsf{Adv}_{\mathsf{KEM},n}^{\mathit{LEAK}^{+r}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{A}) := \Pr\Big[\mathcal{G}_{\mathsf{KEM},n}^{\mathit{LEAK}^{+r}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{A})\Big] \leq \epsilon.$$

Figure 6 visualizes the relations between LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ and the corresponding notions from [CDM24]: The MAL variant is strictly stronger than the LEAK^{+r} variant, which in turn is strictly stronger than the LEAK variant. Notably, ML-KEM separates the MAL and our LEAK^{+r} variants; we show the other implications and separations below, incl. relating our multi-user notion to the 2-user case.

Relation to SH-CR [BJKS24]. The tool-based analysis in BJKS [BJKS24] introduces a similar KEM property to the binding notions we discuss here; dubbed "semi-honest collision resistance" (SH-CR): There, one party decapsulates a malicious ciphertext for an honestly generated but compromised key pair, and another party encapsulates against an adversarially-chosen public key. The adversary wins if the shared secrets from decapsulation and encapsulation match and either the ciphertexts or the public keys differ.

[3] Note that in the syntax of [CDM24] the KEMs shared secret ss is simply called a "key", denoted k. Accordingly, in their notation, $P \in \{\{\mathsf{k}\},\{ct\},\{\mathsf{k},ct\}\}$ and $Q \in \{\{pk\},\{\mathsf{k}\},\{ct\}\}$. We will stick to our syntax and hence, e.g., write MAL-BIND-SS-CT corresponding to their notion MAL-BIND-K-CT.

[4] https://github.com/signalapp/libsignal/commit/0670f0d

```
$\mathcal{G}^{LEAK^{+r}\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\}}_{\mathsf{KEM},n}(\mathcal{A})$:
for $k \in [n]$
    $r_k \leftarrow\$ \mathcal{R}_{\mathsf{KEM.KGen}}$
    $(pk_k, sk_k) \leftarrow \mathsf{KEM.KGen}(; r_k)$
$(i, ct_i, j, ct_j) \leftarrow\$ \mathcal{A}((pk_k, sk_k, r_k)_{k\in[n]})$
    // $\mathcal{A}$ gets all key pairs and KGen randomness
$ss_i \leftarrow \mathsf{KEM.Dec}(sk_i, ct_i)$
$ss_j \leftarrow \mathsf{KEM.Dec}(sk_j, ct_j)$
if $ss_i = \perp \vee ss_j = \perp$: return false
return $[\![ss_i = ss_j \wedge (ct_i \neq ct_j \vee pk_i \neq pk_j)]\!]$
```

```
$\mathcal{G}^{X\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\}}_{\mathsf{KEM}}(\mathcal{A})$:
if $X = LEAK$:
    $(pk_0, sk_0) \leftarrow\$ \mathsf{KEM.KGen}()$
    $(pk_1, sk_1) \leftarrow\$ \mathsf{KEM.KGen}()$
    $(ct_0, ct_1) \leftarrow\$ \mathcal{A}(pk_0, sk_0, pk_1, sk_1)$
if $X = MAL$:
    $(pk_0, sk_0, pk_1, sk_1, ct_0, ct_1) \leftarrow\$ \mathcal{A}()$
$ss_0 \leftarrow \mathsf{KEM.Dec}(sk_0, ct_0)$
$ss_1 \leftarrow \mathsf{KEM.Dec}(sk_1, ct_1)$
if $ss_0 = \perp \vee ss_1 = \perp$: return false
return $[\![ss_0 = ss_1 \wedge (ct_0 \neq ct_1 \vee pk_0 \neq pk_1)]\!]$
```

Figure 5: Security games for our (multi-user) $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ notion (left) and the $LEAK$-BIND-SS-$\{CT, PK\}$ and MAL-BIND-SS-$\{CT, PK\}$ notion from [CDM24] (right). In all games, the adversary's goal is to produce colliding shared secrets ss under distinct ciphertexts ct or public keys pk.

$$\begin{tikzcd}
MAL\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\} \arrow[r, bend left, "\text{Thm. 3.6}"] & LEAK^{+r}\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\} \text{ for n users} \arrow[l, "/" marking, "\text{Thm. 3.10 (ML-KEM)}"] \arrow[d, bend left, "\text{Thm. 3.8}"] \\
LEAK\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\} \arrow[r, "/" marking, "\text{Thm. 3.9}"'] & LEAK^{+r}\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\} \text{ for 2 users} \arrow[l, bend right, "\text{Thm. 3.7}"'] \arrow[u, dashed, bend left, "\text{Thm. 3.8}"]
\end{tikzcd}$$

Figure 6: Relations between our $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ binding notion and the corresponding MAL, $LEAK$ notions from [CDM24] for KEMs. Solid arrows indicate implications, dashed ones loose implications, and crossed-out ones separations. Annotations indicate the respective theorems.

Intuitively, this notion is incomparable to ours, since it allows malicious key pairs but does not give the key generation randomness to the adversary. More formally, we show that a KEM can fulfill SH-CR but not $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$, and vice-versa:

- $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ but not SH-CR: Given a $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$–secure scheme, modify key generation to add a constant 0 bit to the secret key and decapsulation returns an extra bitstring appended to the secret key as shared secret if the extra bit in the secret key is set to 1. This change does not affect $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ since the special case is never triggered for honestly generated key pairs, but it allows breaking SH-CR.
- SH-CR but not $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$: given a SH-CR–secure scheme, modify key generation to add a trapdoor in the secret key (as we will do for the separation of $LEAK$-BIND-SS-$\{CT, PK\}$ and $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ in Theorem 3.9), for which decapsulation returns a special shared secret. This does not help the SH-CR adversary, who does not learn the key generation randomness, but it allows breaking $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ as in Theorem 3.9.

3.2 Relations of Binding Properties

First, we show the implication from MAL to $LEAK^{+r}$ (Theorem 3.6) and from $LEAK^{+r}$ to $LEAK$ (Theorem 3.7). Second, we show that our notion is equivalent for n and 2 users, up to a factor of n^2

(Theorem 3.8). Third, we give a separation between LEAK and LEAK^{+r} (Theorem 3.9). Lastly, we show that ML-KEM and Kyber are LEAK^{+r} (Theorem 3.10 and Theorem 3.11), while ML-KEM is known to not be MAL [Sch24], giving us a separation for LEAK^{+r} and MAL.

Note that for the following theorems we assume KEMs with a public key space $|\mathcal{PK}| \geq 2$.

Theorem 3.6 (MAL-BIND-SS-$\{\mathit{CT},\mathit{PK}\} \implies \mathit{LEAK}^{+r}$-BIND-$\mathit{SS}$-$\{\mathit{CT},\mathit{PK}\}$)**.** *If a KEM* KEM *is* (t',ϵ')*-*MAL-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure, then it is also* (t,ϵ,n)*-*LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure with* $t \approx t' + n$ *and* $\epsilon \leq \epsilon'$*.*

Proof. An adversary $\mathcal{A}$ against LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ must produce ciphertexts ct_i, ct_j such that $ss_i = ss_j$ and one (or both) of $pk_i \neq pk_j$ and $ct_i \neq ct_j$ hold.

We let the adversary $\mathcal{B}$ against MAL-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ truthfully simulate $\mathcal{G}_{\mathsf{KEM},n}^{\mathit{LEAK}^{+r}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{A})$ for $\mathcal{A}$, i.e., honestly computing keys pk_k for $k \in [n]$, obtaining i, ct_i, j, ct_j from $\mathcal{A}$. Then, $\mathcal{B}$ outputs $(pk_i, sk_i, pk_j, sk_j, ct_i, ct_j)$. If $\mathcal{A}$ wins, we have that $ss_i = ss_j$ while $pk_i \neq pk_j$ or $ct_i \neq ct_j$, so $\mathcal{B}$ also wins in the MAL-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ game, i.e., $\epsilon \leq \epsilon'$. The adversary $\mathcal{B}$ runs KGen n times and $\mathcal{A}$ once, resulting in $t \approx t' + n$. □

Theorem 3.7 (LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\} \implies \mathit{LEAK}$-BIND-$\mathit{SS}$-$\{\mathit{CT},\mathit{PK}\}$)**.** *If a KEM* KEM *is* (t',ϵ',n)*-*LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure, then it is also* (t,ϵ)*-*LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure with* $t \approx t'$ *and* $\epsilon \leq \epsilon'$ *for any* n*.*

Proof. The adversary $\mathcal{B}$ against LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ runs the adversary $\mathcal{A}$ against LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ on the first two key pairs (excluding key generation randomness) of its own input, thereby truthfully simulating the LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ game for $\mathcal{A}$. When $\mathcal{A}$ terminates, $\mathcal{B}$ returns the output of $\mathcal{A}$. If $\mathcal{A}$ wins, then so does $\mathcal{B}$, i.e., $\epsilon \leq \epsilon'$. The adversary $\mathcal{B}$ runs $\mathcal{A}$ once, resulting in $t \approx t'$. □

Theorem 3.8 (LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$: n users $\iff$ 2 users)**.** *Let* $n \geq 2$*. If a KEM* KEM *is* (t,ϵ,n)*-*LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure, then it is also* $(t,\epsilon,2)$*-*LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure, and if a KEM* KEM *is* $(t',\epsilon',2)$*-*LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure, then it is also* $(t, n^2 \cdot \epsilon', n)$*-*$\mathit{LEAK}^{+r}$-BIND-$\mathit{SS}$-$\{\mathit{CT},\mathit{PK}\}$*–secure for* $t \approx t' + n$*.*

Proof. The "$\Longrightarrow$" direction straightforwardly holds by letting the adversary against the n-user variant ignore the remaining $n-2$ keys.

The "$\Longleftarrow$" direction holds via a guessing argument: Let the adversary $\mathcal{B}$ against the 2-user variant guess two distinct key indices $i, j \in [n]$, embed the two keys obtained in its game in these positions, and sample the remaining $n-2$ keys itself. If the adversary $\mathcal{A}$ against the n-user variant uses i (and possibly j) for its attack, $\mathcal{B}$ is successful if $\mathcal{A}$ is. The adversary $\mathcal{B}$ runs $\mathcal{A}$, guesses two indices, and runs KGen $n-2$ times, resulting in $t \approx t' + n$. The chance of $\mathcal{B}$ guessing correctly is $\frac{1}{n^2}$, establishing the claim. □

Theorem 3.9 (LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\} \not\Longrightarrow \mathit{LEAK}^{+r}$-BIND-$\mathit{SS}$-$\{\mathit{CT},\mathit{PK}\}$)**.** *Assuming a* LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure KEM* KEM′ *and a preimage-resistant (*PR-d*) hash function* H *(when sampling from* $\{0,1\}^{256}$*), there exists a KEM* KEM *which is* LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure but not* LEAK^{+r}-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$*–secure. Concretely, for any* LEAK-BIND-SS-$\{\mathit{CT},\mathit{PK}\}$ *adversary* $\mathcal{A}$ *against* KEM′ *we construct efficient adversaries* $\mathcal{B}_1$*,* $\mathcal{B}_2$*,* $\mathcal{B}_3$ *such that*

$$\mathsf{Adv}_{\mathsf{KEM},2}^{\mathit{LEAK}^{+r}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{B}_1) = 1 \text{ and}$$
$$\mathsf{Adv}_{\mathsf{KEM}}^{\mathit{LEAK}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{A}) \leq \mathsf{Adv}_{\mathsf{H}}^{\mathsf{PR}\text{-}\mathsf{d}}(\mathcal{B}_2) + \mathsf{Adv}_{\mathsf{KEM}'}^{\mathit{LEAK}\text{-}\mathsf{BIND}\text{-}\mathit{SS}\text{-}\{\mathit{CT},\mathit{PK}\}}(\mathcal{B}_3).$$

```
KGen():
    $z \leftarrow_\$ \{0,1\}^{256}$
    $(pk, \tilde{sk}) \leftarrow_\$ \mathsf{PKE.KGen}()$
    $sk \leftarrow (\tilde{sk}, pk, \mathsf{H}(pk), z)$
    $\textbf{return } (pk, sk)$

Enc(pk):
    $m \leftarrow_\$ \{0,1\}^{256}$
    $\boxed{m \leftarrow \mathsf{H}(m)}$
    $(ss', r) \leftarrow \mathsf{G}(m \| \mathsf{H}(pk))$
    $ct \leftarrow \mathsf{PKE.Enc}(pk, m; r)$
    $\boxed{ss \leftarrow \mathsf{J}(ss' \| \mathsf{H}(ct))}$
    $\dbox{ss \leftarrow ss'}$
    $\textbf{return } (ct, ss)$

Dec(sk, ct):
    $(\tilde{sk}, pk, h, z) \leftarrow sk$
    $m' \leftarrow \mathsf{PKE.Dec}(\tilde{sk}, ct)$
    $(ss', r') \leftarrow \mathsf{G}(m' \| h)$
    $ct' \leftarrow \mathsf{PKE.Enc}(pk, m'; r')$
    $\textbf{if } ct = ct'$:
        $\boxed{ss \leftarrow \mathsf{J}(ss' \| \mathsf{H}(ct))}$
        $\dbox{ss \leftarrow ss'}$
    $\textbf{else}$
        $\boxed{ss \leftarrow \mathsf{J}(z \| \mathsf{H}(ct))}$
        $\dbox{ss \leftarrow \mathsf{J}(z \| ct)}$
    $\textbf{return } ss$
```

Figure 7: Algorithmic description of ML-KEM and Kyber using functions H, G, and J and a public key encryption scheme PKE. $\boxed{\text{Solid boxes}}$ are exclusive to Kyber, $\dbox{\text{dashed boxes}}$ are exclusive to ML-KEM.

Proof. Let $\mathsf{KEM}' = (\mathsf{KGen}', \mathsf{Enc}', \mathsf{Dec}')$ with shared secret space $\mathcal{SS}'$ and randomness space $\mathcal{R}_{\mathsf{KGen}'}$ for key generation be LEAK-BIND-SS-$\{\mathit{CT}, \mathit{PK}\}$–secure, and consider the following modification KEM with shared secret space $\mathcal{SS}' \cup \{ss^*\}$ where ss^* denotes a distinguished shared secret with $ss^* \notin \mathcal{SS}'$ and randomness space $\{0,1\}^{256} \times \mathcal{R}_{\mathsf{KGen}'}$ for key generation.

```
KGen(;r):
    $(x, r_{KG}) \leftarrow r$
    $y \leftarrow \mathsf{H}(x)$
    $(pk', sk') \leftarrow \mathsf{KGen}'(; r_{KG})$
    $\textbf{return } (pk', (sk', y))$

Enc(pk):
    $\textbf{return } (\mathsf{Enc}'(pk), \perp)$

Dec(sk, pk, ct):
    $(sk', y) \leftarrow sk$
    $(ct', z) \leftarrow ct$
    $\textbf{if } y = \mathsf{H}(z)$
        $\textbf{return } ss^*$
    $\textbf{return } \mathsf{Dec}'(sk', pk, ct')$
```

KEM is not LEAK^{+r}-BIND-SS-$\{\mathit{CT}, \mathit{PK}\}$-secure: An adversary $\mathcal{B}_1$ gets $(pk_1, sk_1, r_1, pk_2, sk_2, r_2)$ as input, parses the randomness as $(x_1, r_{KG,1}) \leftarrow r_1$ and $(x_2, r_{KG,2}) \leftarrow r_2$, and outputs the two ciphertexts $(0, x_1), (1, x_2)$ (omitting the indices for the two users). Both decapsulations trigger the special case. Hence, the shared secrets collide, while the ciphertexts differ (at least) in their first component and $\mathcal{B}_1$ always wins.

KEM is LEAK-BIND-SS-$\{\mathit{CT}, \mathit{PK}\}$-secure: If $\mathcal{A}$ triggers the special condition in the decapsulation, then it has found a preimage under H (for randomly sampled 256-bit preimages), breaking preimage resistance. Otherwise, the adversary cannot take advantage of the extra components in the secret key and ciphertext. Hence, breaking LEAK-BIND-SS-$\{\mathit{CT}, \mathit{PK}\}$ of KEM implies breaking LEAK-BIND-SS-$\{\mathit{CT}, \mathit{PK}\}$ of KEM', which we have excluded by assumption. □

3.3 Kyber and ML-KEM

Figure 7 gives an algorithmic description of Kyber [SAB+] and the new NIST standard ML-KEM [Nat24]. On a high level, the scheme uses a public key encryption scheme PKE in an FO transform [FO99], as well as three functions H, G, and J (where J is called KDF in Kyber), instantiated as SHA3-256, SHA3-512, and SHAKE256 with 256 bits output, respectively. Our interest being in the binding properties, we focus here on the internals of decapsulation. The decapsulation algorithm first decrypts the PKE ciphertext ct to m'.

Then, it hashes m' and h (the hashed public key, stored in the secret key) under G onto ss', r'. It uses r' as randomness for re-encrypting m'. If the re-encrypted ciphertext matches ct, Kyber outputs $\mathsf{J}(ss' \| \mathsf{H}(ct))$ as shared secret while ML-KEM outputs ss' directly. Otherwise, both implicitly reject by computing the shared secret as $\mathsf{J}(z \| \mathsf{H}(ct))$ for Kyber resp. $\mathsf{J}(z \| ct)$ for ML-KEM, where z is a secret random 256-bit string, which is part of the secret key.

In the following, we show that ML-KEM and Kyber satisfy the $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ binding property, assuming collision resistance of all involved hash functions, (only for ML-KEM) random-oracle properties of H and J, and (only for Kyber) preimage resistance of G_{256} for randomly sampled images.

Theorem 3.10 (ML-KEM is $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$–secure)**.** *Let $\mathcal{A}$ be an adversary against the (t, ϵ, n)–$LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ security of ML-KEM and assume H, J behave like independent random oracles. Then we can construct adversaries $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ (given in the proof) with running time $\approx t$ such that*

$$\epsilon \leq \frac{n^2}{2^{256}} + 2 \cdot \mathsf{Adv}_{\mathsf{J}}^{\mathsf{CR}}(\mathcal{B}_1) + \frac{q_{\mathrm{RO}}^2}{2^{256}} + \mathsf{Adv}_{\mathsf{H}}^{\mathsf{CR}}(\mathcal{B}_2) + 2 \cdot \mathsf{Adv}_{\mathsf{G}_{256}}^{\mathsf{CR}}(\mathcal{B}_3),$$

where q_{RO} is the number of random oracle queries made by $\mathcal{A}$.

When treating G, J as *quantum-accessible* (independent) random oracles, following Zhandry [Zha15] we get a query complexity of $\Theta(2^{256/3})$ for finding a collision between the two instead of the $\frac{q_{\mathrm{RO}}^2}{2^{256}}$ term in the theorem bound.

Proof. We analyze the probability of $\mathcal{A}$ winning over a series of game hops.

Game 0. We start with the original binding game $\mathcal{G}_{\text{ML-KEM},n}^{LEAK^{+r}\text{-BIND-}SS\text{-}\{CT,PK\}}(\mathcal{A})$:

$$\mathsf{Adv}_{\text{ML-KEM},n}^{LEAK^{+r}\text{-BIND-}SS\text{-}\{CT,PK\}}(\mathcal{A}) = \mathsf{Adv}_{\text{ML-KEM}}^{\mathcal{G}_0}(\mathcal{A}).$$

We distinguish several cases for the adversary to win, depending on how the adversary fulfills the winning condition and if the decapsulation accepts (line 15 in Figure 7 evaluates to true) or implicitly rejects (i.e., decapsulation branches into line 17). We denote line 15 being taken with A_x and line 17 with R_x, where $x \in \{i, j\}$ are the key indices in the binding game. We index any intermediate values from decapsulation of ct_i and ct_j with i and j, respectively. We assume a successful adversary, i.e., $ss_i = ss_j$. Note that R_x and A_x are mutually exclusive. The cases are:

A. Both decapsulations reject and the ciphertexts are for two distinct public keys, i.e., $R_i \wedge R_j \wedge pk_i \neq pk_j$.

B. Both decapsulations reject and the ciphertexts are distinct, i.e., $R_i \wedge R_j \wedge ct_i \neq ct_j$.

C. One decapsulation rejects, i.e., wlog. $R_i \wedge A_j$.

D. Both decapsulations accept and the ciphertexts are for two distinct public keys, i.e., $A_i \wedge A_j \wedge pk_i \neq pk_j$.

E. Both decapsulations accept and the ciphertexts are distinct, i.e., $A_i \wedge A_j \wedge pk_i = pk_j \wedge ct_i \neq ct_j$.

We treat these cases as events in $\mathcal{G}_0$ and indicate the occurrence of event X by $\mathcal{G}_0[X]$. By the union bound we get:

$$\mathsf{Adv}_{\text{ML-KEM}}^{\mathcal{G}_0}(\mathcal{A}) \leq \sum_{X \in \{\text{A,B,C,D,E}\}} \mathsf{Adv}_{\text{ML-KEM}}^{\mathcal{G}_0[X]}(\mathcal{A})$$

Case A (Both decapsulations reject, distinct public keys: $R_i \wedge R_j \wedge pk_i \neq pk_j$).

Here, we bound the probability of the adversary producing ciphertexts that both get implicitly rejected during decapsulation and decapsulate to the same shared secret under distinct public keys.

Game A.0. This is the game conditioned on $R_i \wedge R_j \wedge pk_i \neq pk_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[R_i \wedge R_j \wedge pk_i \neq pk_j]}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}).$$

Game A.1 (Colliding z values). We let $\mathcal{G}_{\mathrm{A.0}}$ return false if there is a collision among the z values in the n secret keys. Since the z values are 256-bit strings sampled uniformly at random, by the birthday bound we get:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathcal{G}_{\mathrm{A.1}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) + \frac{n^2}{2^{256}}.$$

Game A.2 (Collision under J). Now, we have established $z_i \| ct_i \neq z_j \| ct_j$ (since $z_i \neq z_j$), yet $\mathsf{J}(z_i \| ct_i) = \mathsf{J}(z_j \| ct_j)$. We build a reduction $\mathcal{B}_1$ that outputs $z_i \| ct_i, z_j \| ct_j$ as a collision under J:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.1}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{J}}(\mathcal{B}_1).$$

Case B (Both decapsulations reject, distinct ciphertexts: $R_i \wedge R_j \wedge ct_i \neq ct_j$).

Here, we bound the probability of the adversary producing two distinct ciphertexts that both get implicitly rejected during decapsulation and decapsulate to the same shared secret.

Game B.0. This is the game conditioned on $R_i \wedge R_j \wedge ct_i \neq ct_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[R_i \wedge R_j \wedge ct_i \neq ct_j]}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}).$$

Since $ct_i \neq ct_j$, we also have $z_i \| ct_i \neq z_j \| ct_j$, yet $\mathsf{J}(z_i \| ct_i) = \mathsf{J}(z_j \| ct_j)$. As in the prior case, the reduction $\mathcal{B}_1$ that outputs $z_i \| ct_i, z_j \| ct_j$ produces a collision under J:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{J}}(\mathcal{B}_1).$$

Case C (One decapsulation rejects: $R_i \wedge A_j$).

Here, we bound the probability of the adversary producing exactly one ciphertext that gets implicitly rejected during decapsulation while both ciphertexts decapsulate to the same shared secret.

Game C.0. This is the game conditioned on $R_i \wedge A_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{C.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[R_i \wedge A_j]}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}).$$

For $\mathcal{A}$ to win, it needs to create a collision $\mathsf{J}(z_i \| ct_i) = ss_i = ss_j = \mathsf{G}_{256}(m'_j \| h_j)$ between J and G. Assuming both J and G behave like (independent) random oracles, the probability of finding such a collision with q_{RO} many random oracle queries is upper bounded by:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{C.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) \leq \frac{q_{\mathrm{RO}}^2}{2^{256}}.$$

Case D (Both decapsulations accept, distinct public keys: $A_i \wedge A_j \wedge pk_i \neq pk_j$).

Here, we bound the probability of the adversary producing ciphertexts that both get accepted during decapsulation and decapsulate to the same shared secret under distinct public keys.

Game D.0. This is the game conditioned on $A_i \wedge A_j \wedge pk_i \neq pk_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[A_i \wedge A_j \wedge pk_i \neq pk_j]}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}).$$

Since we know that $pk_i \neq pk_j$ and $\mathsf{G}_{256}(m_i \| \mathsf{H}(pk_i)) = ss_i = ss_j = \mathsf{G}_{256}(m_j \| \mathsf{H}(pk_j))$, it must be that $\mathcal{A}$ provides us with a collision either in H or in G_{256}. Once more, we can distinguish two cases:

1. The first case is $m_i \| \mathsf{H}(pk_i) = m_j \| \mathsf{H}(pk_j)$. Then we can build a reduction $\mathcal{B}_2$ that outputs pk_i, pk_j as collision under H.
2. Otherwise, $m_i \| \mathsf{H}(pk_i) \neq m_j \| \mathsf{H}(pk_j)$, and we can build reduction $\mathcal{B}_3$ that outputs $m_i \| \mathsf{H}(pk_i), m_j \| \mathsf{H}(pk_j)$ as collision under G_{256}.

Jointly,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{H}}(\mathcal{B}_2) + \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{G}_{256}}(\mathcal{B}_3).$$

Case E (Both decapsulations accept, distinct ciphertexts: $A_i \wedge A_j \wedge pk_i = pk_j \wedge ct_i \neq ct_j$).

Here, we bound the probability of the adversary producing ciphertexts that both get accepted during decapsulation and decapsulate to the same shared secret under the same public key.

Game E.0. This is the game conditioned on $A_i \wedge A_j \wedge pk_i = pk_j \wedge ct_i \neq ct_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{E.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[A_i \wedge A_j \wedge pk_i = pk_j \wedge ct_i \neq ct_j]}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}).$$

We know that $pk_i = pk_j$, and hence $h_i = \mathsf{H}(pk_i) = \mathsf{H}(pk_j) = h_j$ in $\mathsf{G}_{256}(m_i \| h_i) = ss_i = ss_j = \mathsf{G}_{256}(m_j \| h_j)$. Note that we must have $m_i \neq m_j$: Otherwise, the re-encryption step in both decapsulations would result in identical ciphertexts $ct_i = ct_j$, which contradicts the condition of event E.

As in the prior case, the reduction $\mathcal{B}_3$ that outputs $m_i \| H(pk_i), m_j \| H(pk_j)$ produces a collision under G_{256}:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{E.0}}}_{\mathsf{ML\text{-}KEM}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{G}_{256}}(\mathcal{B}_3).$$

Collecting the bounds yields the claim. □

The Kyber proof does not need to consider collisions *across* hash functions, allowing us to avoid relying on the random oracle. Instead, we additionally assume G_{256} to be preimage-resistant for randomly sampled images.

Theorem 3.11 (Kyber is *LEAK*$^{+r}$-BIND-*SS*-$\{CT, PK\}$–secure)**.** *Let $\mathcal{A}$ be an adversary against the (t, ϵ, n)–LEAK^{+r}-BIND-SS-$\{CT, PK\}$ security of* Kyber*. Then we can construct adversaries $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$, $\mathcal{B}_4$, $\mathcal{B}_5$ (given in the proof) with running time $\approx t$ such that*

$$\epsilon \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{H}}(\mathcal{B}_1) + \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{H}}(\mathcal{B}_5) + \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{G}_{256}}(\mathcal{B}_4) + n \cdot \mathsf{Adv}^{\mathsf{PR\text{-}r}}_{\mathsf{G}_{256}}(\mathcal{B}_3) + 4 \cdot \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{J}}(\mathcal{B}_2) + \frac{n^2}{2^{256}}.$$

Proof. We analyze the probability of $\mathcal{A}$ winning over a series of game hops.

Game 0. We start with the original binding game $\mathcal{G}_{\mathsf{Kyber},n}^{LEAK^{+r}\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\}}(\mathcal{A})$:

$$\mathsf{Adv}_{\mathsf{Kyber},n}^{LEAK^{+r}\text{-}\mathsf{BIND}\text{-}SS\text{-}\{CT,PK\}}(\mathcal{A}) = \mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_0}(\mathcal{A}).$$

We distinguish several cases for the adversary to win, depending on how the adversary fulfills the winning condition and if the decapsulation accepts (line 15 in Figure 7 evaluates to true) or implicitly rejects (i.e., decapsulation branches into line 17). We denote line 15 being taken with A_x and line 17 with R_x, where $x \in \{i, j\}$ are the key indices in the binding game. We index any intermediate values from decapsulation of ct_i and ct_j with i and j, respectively. We assume a successful adversary, i.e., $ss_i = ss_j$. Note that R_x and A_x are mutually exclusive. The cases are:

A. Both ciphertexts are distinct, i.e., $ct_i \neq ct_j$.

B. Both decapsulations reject and the ciphertexts are for two distinct public keys, i.e., $R_i \wedge R_j \wedge pk_i \neq pk_j$.

C. One decapsulation rejects and the ciphertexts are for two distinct public keys, i.e., wlog. $R_i \wedge A_j \wedge pk_i \neq pk_j$.

D. Both decapsulations accept and the ciphertexts are for two distinct public keys, i.e., $A_i \wedge A_j \wedge pk_i \neq pk_j$.

We treat these cases as events in $\mathcal{G}_0$ and indicate the occurrence of event X by $\mathcal{G}_0[X]$. By the union bound we get:

$$\mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_0}(\mathcal{A}) \leq \sum_{X \in \{\text{A},\text{B},\text{C},\text{D}\}} \mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_0[X]}(\mathcal{A})$$

Case A (distinct ciphertexts: $ct_i \neq ct_j$).

Here, we bound the probability of the adversary producing two distinct ciphertexts that decapsulate to the same shared secret.

Game A.0. This is the game conditioned on $ct_i \neq ct_j$ being satisfied.

$$\mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_{\text{A.0}}}(\mathcal{A}) = \mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_0[ct_i \neq ct_j]}(\mathcal{A}).$$

We know that $\mathsf{J}(x_i \| \mathsf{H}(ct_i)) = ss_i = ss_j = \mathsf{J}(x_j \| \mathsf{H}(ct_j))$ for x_i being either ss_i' (in case of an accepting decapsulation) or z_i (in case of an implicitly rejecting decapsulation), and x_j likewise. Since we know that $ct_i \neq ct_j$, we can now either build a reduction $\mathcal{B}_1$ that outputs ct_i, ct_j as collision under H, or a reduction $\mathcal{B}_2$ that outputs $x_i \| \mathsf{H}(ct_i), x_j \| \mathsf{H}(ct_j)$ as collision under J:

$$\mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_{\text{A.0}}}(\mathcal{A}) \leq \mathsf{Adv}_{\mathsf{H}}^{\mathsf{CR}}(\mathcal{B}_1) + \mathsf{Adv}_{\mathsf{J}}^{\mathsf{CR}}(\mathcal{B}_2).$$

Case B (Both decapsulations reject, distinct public keys: $R_i \wedge R_j \wedge pk_i \neq pk_j$).

Here, we bound the probability of the adversary producing ciphertexts that both get implicitly rejected during decapsulation and decapsulate to the same shared secret under distinct public keys.

Game B.0. This is the game conditioned on $R_i \wedge R_j \wedge pk_i \neq pk_j$ being satisfied.

$$\mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_{\text{B.0}}}(\mathcal{A}) = \mathsf{Adv}_{\mathsf{Kyber}}^{\mathcal{G}_0[R_i \wedge R_j \wedge pk_i \neq pk_j]}(\mathcal{A}).$$

Game B.1 (Colliding z values). We let $\mathcal{G}_{\text{B.0}}$ return false if there is a collision among the z values in the n secret keys. Since the z values are 256-bit strings sampled uniformly at random, by the birthday bound we get:

$$\mathsf{Adv}^{\mathcal{G}_{\text{B.0}}}_{\mathsf{Kyber}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathcal{G}_{\text{B.1}}}_{\mathsf{Kyber}}(\mathcal{A}) + \frac{n^2}{2^{256}}.$$

Game B.2 (Collision under J). Now, we have established $z_i \| \mathsf{H}(ct_i) \neq z_j \| \mathsf{H}(ct_j)$ (since $z_i \neq z_j$), yet $\mathsf{J}(z_i \| \mathsf{H}(ct_i)) = \mathsf{J}(z_j \| \mathsf{H}(ct_j))$. The reduction $\mathcal{B}_2$ that outputs $z_i \| \mathsf{H}(ct_i), z_j \| \mathsf{H}(ct_j)$[5] produces a collision under J:

$$\mathsf{Adv}^{\mathcal{G}_{\text{B.1}}}_{\mathsf{Kyber}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{J}}(\mathcal{B}_2).$$

Case C (One decapsulation rejects, distinct public keys: $R_i \wedge A_j \wedge pk_i \neq pk_j$).

Here, we bound the probability of the adversary producing exactly one ciphertext that gets implicitly rejected during decapsulation, while both ciphertexts decapsulate to the same shared secret under distinct public keys.

Game C.0. This is the game conditioned on $R_i \wedge A_j \wedge pk_i \neq pk_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\text{C.0}}}_{\mathsf{Kyber}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[R_i \wedge A_j \wedge pk_i \neq pk_j]}_{\mathsf{Kyber}}(\mathcal{A}).$$

If $ss'_i \neq z_j$ (for $ss' = \mathsf{G}_{256}(m'_i \| h_i)$), again the reduction $\mathcal{B}_2$ from Game $\mathcal{G}_{\text{A.0}}$ that outputs $ss'_i \| \mathsf{H}(ct_i)$ and $z_j \| \mathsf{H}(ct_j)$ produces a collision under J. Otherwise, we can build a reduction $\mathcal{B}_3$ that guesses $j \in [n]$ and embeds an image of G_{256} in z_j, producing a preimage under G_{256} for randomly sampled images ($\mathsf{PR\text{-}r}$) if it guesses j correctly. Hence,

$$\mathsf{Adv}^{\mathcal{G}_{\text{C.0}}}_{\mathsf{Kyber}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{J}}(\mathcal{B}_2) + n \cdot \mathsf{Adv}^{\mathsf{PR\text{-}r}}_{\mathsf{G}_{256}}(\mathcal{B}_3).$$

Case D (Both decapsulations accept, distinct public keys: $A_i \wedge A_j \wedge pk_i \neq pk_j$).

Here, we bound the probability of the adversary producing ciphertexts that both get accepted during decapsulation and decapsulate to the same shared secret under distinct public keys.

Game D.0. This is the game conditioned on $A_i \wedge A_j \wedge pk_i \neq pk_j$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\text{D.0}}}_{\mathsf{Kyber}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0[A_i \wedge A_j \wedge pk_i \neq pk_j]}_{\mathsf{Kyber}}(\mathcal{A}).$$

Now, $\mathsf{J}(\mathsf{G}_{256}(m_i \| h_i) \| \mathsf{H}(ct_i)) = ss_i = ss_j = \mathsf{J}(\mathsf{G}_{256}(m_j \| h_j) \| \mathsf{H}(ct_j))$ for $pk_i \neq pk_j$. It must be that $\mathcal{A}$ provides us with a collision either in J, G_{256}, or H. Once more, we can distinguish the cases:

1. In case $\mathsf{G}_{256}(m_i \| h_i) \| \mathsf{H}(ct_i) \neq \mathsf{G}_{256}(m_j \| h_j) \| \mathsf{H}(ct_j)$, then reduction $\mathcal{B}_2$ outputs $\mathsf{G}_{256}(m_i \| h_i) \| \mathsf{H}(ct_i)$, $\mathsf{G}_{256}(m_j \| h_j) \| \mathsf{H}(ct_j)$ as collision under J.

2. Otherwise, if $m_i \| h_i \neq m_j \| h_j$, then we can build a reduction $\mathcal{B}_4$ that outputs $m_i \| h_i, m_j \| h_j$ as collision under G_{256}.

3. Otherwise, we must have $m_i \| h_i = m_j \| h_j$ and $\mathsf{H}(pk_i) = h_i = h_j = \mathsf{H}(pk_j)$. By event D, $pk_i \neq pk_j$ and we can build a reduction $\mathcal{B}_5$ that outputs pk_i, pk_j as collision under H.

Jointly,

$$\mathsf{Adv}^{\mathcal{G}_{\text{C.0}}}_{\mathsf{Kyber}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{J}}(\mathcal{B}_2) + \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{G}_{256}}(\mathcal{B}_4) + \mathsf{Adv}^{\mathsf{CR}}_{\mathsf{H}}(\mathcal{B}_5).$$

Collecting the bounds yields the claim. □

[5] This is the same reduction $\mathcal{B}_2$ as in Game $\mathcal{G}_{\text{A.0}}$, which outputs $x_i \| \mathsf{H}(ct_i), x_j \| \mathsf{H}(ct_j)$, where x_i, x_j are z_i, z_j for implicitly rejecting decapsulations that we consider here.

4 Security Model

We analyze security of the PQXDH protocol in a computational, game-based security model for authenticated key exchange protocols, following the tradition of Bellare and Rogaway (BR) [BR94], but refined towards the maximum-exposure security that Signal's handshakes (both X3DH and PQXDH) aim at. Our model is based on prior Signal and Signal-like security models by CCDGS [CCD$^+$17] and particularly BFGJS [BFG$^+$22, BFG$^+$21], the latter introducing syntax and modeling for the particular class of asynchronous key exchange protocols that the Signal handshakes belong to.

In BR-style models, a computational adversary interacts with multiple users across multiple sessions of the key exchange protocol, fully controlling the network and being able to reroute, modify, inject, and drop messages at will (through a Send oracle). For maximum-exposure, the model allows the adversary to compromise a user's long-term key (via a CorruptLTKey oracle) and semi-static keys (CorruptSSKey), as well as reveal a session's key (RevealSessKey) and ephemeral randomness (RevealRand).[6] The targeted security property finally is key indistinguishability, asking that keys established in so-called "fresh" sessions are indistinguishable from random keys. "Freshness" here encodes that a session is not trivially compromised and that the protocol under analysis aims to protect against the involved key compromises (defined through a set of "clean" predicates). We will detail those technical bits below.

For the analysis of PQXDH, we extend the prior models in several ways:

- Prior models [CCD$^+$17, BFG$^+$22, BFG$^+$21] assumed semi-static keys being authentically distributed. However, only long-term keys can be verified out-of-band in Signal, and semi-static keys are signed with those keys instead. We are the first to capture these *signed semi-static keys* in a traditional game-based model for a computational security analysis. In both the model and the security analysis (cf. Section 5), capturing these signatures leads to notable added complexity.

- In contrast to Signal's classical X3DH handshake, PQXDH also involves *signed ephemeral (KEM) keys*. We capture this, too, expanding on the set of attack vectors our model covers.

- PQXDH adding KEM keys means that *multiple semi-static keys* (DH and KEM) can now be involved in a handshake. We capture this by accordingly modified identification of semi-static keys.

Note that CCDGS [CCD$^+$17] also analyzed the Double Ratchet protocol, working in a multi-stage key exchange model [FG14]; PQXDH does not affect the ratcheting part, hence we focus on the initial handshake. BFGJS [BFG$^+$22, BFG$^+$21] in turn also studied deniability, which we do not consider in this work.

In the following, we will first review the syntax required to express both the PQXDH protocol and the security model formally; we mostly follow BFGJS [BFG$^+$22, BFG$^+$21] here, in parts verbatim, and focus on highlighting (with light-gray background) the main differences introduced in this work to account for signed semi-static and ephemeral keys as well as multiple identifiers for semi-static keys.

4.1 Syntax and Notation

Key exchange syntax. We define a two-party key exchange protocol via the following probabilistic algorithms:

- $\mathsf{KGenLT}() \xrightarrow{\$} (ltpk, ltsk)$: The *long-term key generation* algorithm that outputs a party's public-key/secret-key pair.

[6] This in particular covers forward secrecy (compromise of long-term and semi-static keys after a handshake took place) and a form of post-compromise security [CCG16] (compromise of long-term and semi-static keys before a handshake with uncompromised ephemeral randomness takes place) for the handshake. Note that our model only considers the initial handshake; Signal's subsequent Double Ratchet protocol [MP16a] provides further (classical) forward security and post-compromise security.

- $\mathsf{KGenSS}(\mathit{ltsk}) \ \$\!\!\rightarrow ((\mathit{sspk}_1, \mathit{sssk}_1, \sigma_1), \dots)$: The *semi-static key generation* algorithm that takes as input a long-term secret key *ltsk* and outputs a vector of public-key/secret-key pairs and corresponding signatures.[7]
- $\mathsf{Run}(\mathit{ltsk}, \vec{\mathit{sssk}}, \vec{\mathit{ltpk}}, \pi, m) \ \$\!\!\rightarrow (\pi', m')$: The *session execution algorithm* that takes as input a party's long-term secret key *ltsk*, that party's semi-static secret keys $\vec{\mathit{sssk}}$, all parties' long-term public keys $\vec{\mathit{ltpk}}$, a session state π, and an incoming message m, and outputs an updated session state π' and a (possibly empty) outgoing message m'. The session sending the first message is set up by calling Run with a distinguished message $m = (\mathsf{create}, (\vec{\mathsf{ssid}}, \mathsf{type}))$, where $\vec{\mathsf{ssid}}$ indicates the semi-static keys[8] to be used and type whether a full ($\mathsf{type} = \mathsf{full}$) or reduced ($\mathsf{type} = \mathsf{reduced}$) handshake should be performed.

Note that an explicit ephemeral key generation algorithm is not mandatory and can happen inside of Run.

Parties and sessions. In our model, *parties* $P \in [n_p]$, each holding a long-term public-key/secret-key pair generated by KGenLT may run multiple instances of the protocol (simultaneously or sequentially); we denote the ith such *session* of party P by π^i_P. Each session maintains the following information:

- $\mathsf{oid} \in [n_p]$: The identity of the session owner.
- $\mathsf{pid} \in [n_p] \cup \{\star\}$: The identity of the intended peer, which may initially be unknown (indicated by $\star$).
- $\mathsf{role} \in \{\mathsf{initiator}, \mathsf{responder}\}$: The role of the party.
- $\mathsf{st}_{\mathsf{exec}} \in \{\bot, \mathsf{running}, \mathsf{accepted}, \mathsf{rejected}\}$: The status of this session's execution.
- $\mathsf{sid} \in \{0,1\}^* \cup \{\bot\}$: A session identifier defining partnering.
- $\mathsf{cid} \in \{0,1\}^* \cup \{\bot\}$: A contributive identifier, defining a preliminary form of partnering (often as a substring or prefix of the session identifier) for the case the session is not yet bound to an authenticated peer [DFGS15].
- $\mathsf{K} \in \mathcal{K}_{\mathsf{KE}} \cup \{\bot\}$: The session key established in this session, initialized to $\bot$.
- $\mathsf{type} \in \{\mathsf{full}, \mathsf{reduced}\}$: Indicator whether an ephemeral pre-key was used ($\mathsf{type} = \mathsf{full}$) for key establishment, or not ($\mathsf{type} = \mathsf{reduced}$).
- $\mathsf{coins} \in \mathcal{R}_{\mathsf{KE}}$: The random coins from the randomness space $\mathcal{R}_{\mathsf{KE}}$ used in the execution of Run; set by the game and read-only thereafter.
- $\mathsf{sspks} \in (\{0,1\}^*)^* \cup \{\bot\}$: The semi-static public keys used in this session.

For bookkeeping in the security game we additionally introduce the following flags, which are not accessible by the protocol sessions:

- $\mathsf{revrand} \in \{\mathsf{true}, \mathsf{false}\}$ indicates whether the random coins $\pi.\mathsf{coins}$ have been revealed via a RevealRand query. The default value is false.
- $\mathsf{pcorr} \in \{\mathsf{true}, \mathsf{false}\}$ indicates whether the peer's long-term key was corrupted at the point in time when this session accepted. The default value is false.

[7] The length of this vector is protocol-dependent; e.g., in X3DH, semi-static keys consist of one DH key pair, and PQXDH adds a second semi-static key (for a KEM scheme). For simplicity, we have a set of semi-static keys be generated in one operation, yet in the protocol each key can be used independently.

[8] In the model, semi-static keys are identified by some value $\mathsf{ssid} = {}''s, n''$ where s indicates the KGenSS call through which they were generated and n the key's position in that call's output. In practice, the latter signifies the type of semi-static key if there are several; for PQXDH, $n = 1$ for DH keys and $n = 2$ for KEM keys.

Session partnering. We say two sessions π_U^i and π_V^j are *partnered* if they agree on the session identifier: $\pi_U^i.\mathsf{sid} = \pi_V^j.\mathsf{sid} \neq \bot$. *Contributive* identifiers (cid) indicate when sessions may eventually derive the same key but are not fully partnered (yet); we use these to model security of initiator's keys in incomplete handshakes.

4.2 Security Game

The security property of an authenticated key exchange protocol KE we consider in this work is *indistinguishability of session keys* (KI), formalized through the game $\mathcal{G}_{\mathsf{KE}}^{\mathsf{KI}}(\mathcal{A})$ in Figure 8 played by an adversary $\mathcal{A}$. At the start of the game, a random challenge bit $b_{\mathsf{test}} \leftarrow_{\$} \{0,1\}$ is fixed and long-term public-key/secret-key pairs are generated for all n_p honest parties and their public keys $\vec{ltpk}$ provided to the adversary. The adversary is then able to interact with honest parties via the following queries:

- $\textsc{Send}(U, i, m)$: Sends message m to session π_U^i, which corresponds to executing $\mathsf{Run}(ltsk_U, \vec{sssk}_U, \vec{ltpk}, \pi_U^i, m)$, saving the updated session state π' as π_U^i, and returning the outgoing message m' to the adversary.
- $\textsc{CorruptLTKey}(U)$: Returns party U's long-term secret key $ltsk_U$ to the adversary; recorded through the flag $\mathsf{corrltk}_U$.
- $\textsc{CorruptSSKey}(U, \mathsf{ssid})$: Returns party U's semi-static secret key $sssk_U^{\mathsf{ssid}}$ to the adversary; recorded through the flag $\mathsf{corrssk}_{sspk}$, where $sspk$ is the public key corresponding to $sssk_U^{\mathsf{ssid}}$.[9]
- $\textsc{RevealRand}(U, i)$: Returns the random coins of session π_U^i to the adversary, recorded in the session through the flag $\mathsf{revrand}$.
- $\textsc{RevealSessKey}(U, i)$: If session π_U^i has accepted, return its session key $\pi_U^i.\mathsf{K}$ to the adversary.
- $\textsc{Test}(U, i)$: If the $\textsc{Test}$ query has been called before or session π_U^i has not accepted, then return $\bot$. Otherwise; if $b_{\mathsf{test}} = 0$, return $\pi_U^i.\mathsf{K}$, otherwise return a randomly sampled session key from the protocol's key space $\mathcal{K}_{\mathsf{KE}}$. Record the test session as $\pi^* \leftarrow \pi_U^i$.

At the end of the game, the adversary outputs a bit b'. The adversary is said to win if $b' = b_{\mathsf{test}}$ and the test session π^* is fresh. Formally, if the test session is fresh, the experiment outputs 1 if $b' = b_{\mathsf{test}}$ and 0 otherwise; if the test session is not fresh, then the experiment outputs a random bit. The adversary's advantage in the key indistinguishability game measured as the experiment outputting 1 minus the adversary's guessing chane, $\frac{1}{2}$.

Definition 4.1 (Key indistinguishability)**.** *Let* KE *be a key exchange protocol and* $\mathcal{A}$ *an adversary against the key indistinguishability (*KI*) game* $\mathcal{G}_{\mathsf{KE}}^{\mathsf{KI}}(\mathcal{A})$ *in Figure 8. We say that* KE *achieves* $(t, \epsilon, (q_{\text{Snd}}, q_{\text{CorrLT}}, q_{\text{CorrSS}}, q_{\text{RevR}}, q_{\text{RevSK}}))$–key indistinguishability, *if for any adversary* $\mathcal{A}$ *against* $\mathcal{G}_{\mathsf{KE}}^{\mathsf{KI}}(\mathcal{A})$ *with running time at most* t *and making at most* q_{Snd}*,* q_{CorrLT}*,* q_{CorrSS}*,* q_{RevR}*, resp.* q_{RevSK} *queries to its* $\textsc{Send}$*,* $\textsc{CorruptLTKey}$*,* $\textsc{CorruptSSKey}$*,* $\textsc{RevealRand}$*, resp.* $\textsc{RevealSessKey}$ *oracles, we have that*

$$\mathsf{Adv}_{\mathsf{KE}}^{\mathsf{KI}}(\mathcal{A}) = \Pr\left[\mathcal{G}_{\mathsf{KE}}^{\mathsf{KI}}(\mathcal{A})\right] - \frac{1}{2} \leq \epsilon.$$

Note that the model restricts the adversary to a single query to the $\textsc{Test}$ *oracle.*

[9] Since we model that semi-static keys are not authentically distributed (but signed), sessions only know the public key received ($sspk$), but not the game label ssid, which was used in prior models [CCD+17, BFG+22, BFG+21] to track compromise more easily.

```
$\mathcal{G}^{\mathsf{KI}}_{\mathsf{KE}}(\mathcal{A})$:
$b_{\mathsf{test}} \leftarrow\!\$\ \{0,1\}$ // sample challenge bit
$\pi^* \leftarrow \perp$ // variable for test session
for $U \in [n_p]$ // generate long-term keys
    $(pk_U, sk_U) \leftarrow\!\$\ \mathsf{KGenLT}()$
    for $s \in [n_{ss}]$ // generate semi-static keys
        $\big((sspk_U^{s,1}, sssk_U^{s,1}, \sigma_U^{s,1}), \dots\big) \leftarrow\!\$\ \mathsf{KGenSS}(ltsk_U)$
    $\vec{sssk}_U \leftarrow \big\{sssk_U^{s,1}, \dots\big\}^{s\in[n_{ss}]}$
$\vec{ltpk} \leftarrow \big\{ltpk_U\big\}_{U\in[n_p]}$
$\vec{sspk} \leftarrow \big\{(sspk_U^{s,1}, \sigma_U^{s,1}), \dots\big\}^{s\in[n_{ss}]}_{U\in[n_p]}$
$b' \leftarrow\!\$\ \mathcal{A}\big(\vec{ltpk}, \vec{sspk}\big)$ // run adversary
if $\mathsf{sound}() = \mathsf{false}$: return true // adversary wins if it breaks soundness
if $\mathsf{fresh}(\pi^*) = \mathsf{false}$ : $b' \leftarrow 0$ // attack invalid if test session is not fresh
return $[\![b' = b_{\mathsf{test}}]\!]$ // determine win or loss
```

```
$\mathsf{fresh}(\pi^*)$:
if $\pi^*.\mathsf{revealed} = \mathsf{true}$ : return false // test session is revealed
if $\exists\, \pi_V^j \neq \pi^* : (\pi_V^j.\mathsf{sid} = \pi^*.\mathsf{sid} \wedge \pi_V^j.\mathsf{revealed} = \mathsf{true})$ : return false
    // test session's partner is revealed
return $\mathsf{clean}^{\mathsf{KE}}_{\pi^*.\mathsf{type}}(\pi^*)$
    // test session is clean wrt. to its handshake type (full resp. reduced)
```

```
$\mathsf{sound}()$:
return $\forall$ distinct $\pi, \pi', \pi''\big($
    $(\pi.\mathsf{sid} = \pi'.\mathsf{sid} \neq \perp \implies \pi.\mathsf{K} = \pi'.\mathsf{K} \wedge \pi.\mathsf{type} = \pi'.\mathsf{type} \wedge \pi.\mathsf{cid} = \pi'.\mathsf{cid})$
    // same session identifiers imply same key, type, contributive identifiers
    and $(\pi.\mathsf{sid} = \pi'.\mathsf{sid} \neq \perp \wedge \pi.\mathsf{role} = \mathsf{initiator} \implies \pi'.\mathsf{role} = \mathsf{responder})$
    // session identifiers of two initiator sessions never collide
    and $(\pi.\mathsf{sid} = \pi'.\mathsf{sid} = \pi''.\mathsf{sid} \neq \perp \implies \pi.\mathsf{type} = \mathsf{reduced})\big)$
    // session identifiers of three sessions only collide in reduced mode
```

```
$\textsc{Send}(U, i, m)$:
if $\pi_U^i = \perp$ // initiate session: for responders, we have $m = (\mathsf{create}, (\vec{\mathsf{ssid}}, \mathsf{type}))$
    $\pi_U^i.\mathsf{oid} \leftarrow U$ // set owner identity
    if $m = (\mathsf{create}, \dots)$: $\pi_U^i.\mathsf{role} \leftarrow \mathsf{responder}$ // set responder role
    else $\pi_U^i.\mathsf{role} \leftarrow \mathsf{initiator}$ // set initiator role ($m$ is first message)
    $\pi_U^i.\mathsf{coins} \leftarrow\!\$\ \mathcal{R}_{\mathsf{KE}}$ // sample session randomness
    $\pi_U^i.\mathsf{st}_{\mathsf{exec}} \leftarrow \mathsf{running}$
$(\pi_U^i, m') \leftarrow \mathsf{Run}(sk_U, \vec{sssk}_U, \vec{ltpk}, \pi_U^i, m)$
    // run session, random coins in $\pi_U^i$
if $\pi_U^i.\mathsf{st}_{\mathsf{exec}} = \mathsf{accepted}$: // flag if peer was corrupted upon acceptance
    $\pi_U^i.\mathsf{pcorr} \leftarrow \mathsf{corrltk}_{\pi_U^i.\mathsf{pid}}$
return $(m', \pi_U^i.\mathsf{st}_{\mathsf{exec}})$ // return message and session state
```

```
$\textsc{Test}(U, i)$:
if $\pi_U^i = \perp$ or $\pi_U^i.\mathsf{st}_{\mathsf{exec}} \neq \mathsf{accepted}$ or $\pi^* \neq \perp$: return $\perp$
    // session does not exist, has not accepted yet, or test already asked
$\pi^* \leftarrow \pi_U^i$ // record test session
$\mathsf{K}_0 \leftarrow \pi_U^i.\mathsf{K}$
$\mathsf{K}_1 \leftarrow\!\$\ \mathcal{K}_{\mathsf{KE}}$
return $\mathsf{K}_{b_{\mathsf{test}}}$ // return real-or-random challenge key
```

```
$\textsc{CorruptLTKey}(U)$:
$\mathsf{corrltk}_U \leftarrow \mathsf{true}$ // mark long-term key corrupted
return $sk_U$ // return long-term secret key
```

```
$\textsc{CorruptSSKey}(U, \mathsf{ssid})$:
$\mathsf{corrssk}_{sspk_U^{\mathsf{ssid}}} \leftarrow \mathsf{true}$ // mark semi-static key corrupted
return $sssk_U^{\mathsf{ssid}}$ // return semi-static secret key
```

```
$\textsc{RevealRand}(U, i)$:
if $\pi_U^i = \perp$: return $\perp$ // session does not exist
$\pi_U^i.\mathsf{revrand} \leftarrow \mathsf{true}$ // mark randomness revealed
return $\pi_U^i.\mathsf{coins}$ // return session's random coins
```

```
$\textsc{RevealSessKey}(U, i)$:
if $\pi_U^i = \perp$ or $\pi_U^i.\mathsf{st}_{\mathsf{exec}} \neq \mathsf{accepted}$: return $\perp$
    // session does not exist or has not yet derived session key
$\pi_U^i.\mathsf{revealed} \leftarrow \mathsf{true}$ // mark session key revealed
return $\pi_U^i.\mathsf{K}$ // return session key
```

Figure 8: Key indistinguishability (KI) game for key exchange protocol KE (top), in which adversary $\mathcal{A}$ has access to oracles Send, Test, CorruptLTKey, CorruptSSKey, RevealRand, and RevealSessKey (bottom), and wrt. to (protocol-specific) clean predicates $\mathsf{clean}^{\mathsf{KE}}$ for KE. Highlighted code reflects the main changes compared to the model of BFGJS [BFG+22, BFG+21]. The clean predicates for PQXDH protocol capturing classic security ($\mathsf{clean}^{\mathsf{PQXDH}}$) and post-quantum security ($\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$) are given in Figure 9.

Predicates $\mathsf{clean}^{\mathsf{PQXDH}}$ capturing classic security

```
$\mathsf{clean}^{\mathsf{PQXDH}}_{\mathsf{full}}(\pi^*)$:
    return $\mathsf{clean}^{\mathsf{PQXDH}}_{\mathsf{reduced}}(\pi^*)$ or $\mathsf{clean}_{\mathsf{EE}}(\pi^*)$

$\mathsf{clean}^{\mathsf{PQXDH}}_{\mathsf{reduced}}(\pi^*)$:
    return $\mathsf{clean}_{\mathsf{LTSS}}(\pi^*)$ or $\mathsf{clean}_{\mathsf{ELT}}(\pi^*)$ or $\mathsf{clean}_{\mathsf{ESS}}(\pi^*)$

$\mathsf{clean}_{\mathsf{EE}}(\pi^*)$:
    return $\neg\pi^*.\mathsf{revrand}$ and $\big(\mathsf{clean}_{\mathsf{peerE}}(\pi^*)$ or $\mathsf{clean}_{\mathsf{sigE}}(\pi^*)\big)$
        // test session randomness is unrevealed and peer's ephemeral contribution is clean or cleanly signed

$\mathsf{clean}_{\mathsf{peerE}}(\pi^*)$:
    return
        $\big(\pi^*.\mathsf{role} = \mathsf{initiator}$ and $\exists \pi \neq \pi^*$ :
          $(\pi.\mathsf{role} = \mathsf{responder}$ and $\pi^*.\mathsf{cid} = \pi.\mathsf{cid}$ and $\neg\pi.\mathsf{revrand})\big)$
        // contributively-partnered responder session's randomness is unrevealed
    or $\big(\pi^*.\mathsf{role} = \mathsf{responder}$ and $\exists \pi \neq \pi^*$ :
          $(\pi.\mathsf{role} = \mathsf{initiator}$ and $\pi^*.\mathsf{sid} = \pi.\mathsf{sid}$ and $\neg\pi.\mathsf{revrand})\big)$
        // partnered initiator session's randomness is unrevealed

$\mathsf{clean}_{\mathsf{sigE}}(\pi^*)$:
    return $\big(\pi^*.\mathsf{role} = \mathsf{initiator}$ and $\neg\pi^*.\mathsf{pcorr}$
          and $\forall \pi : ((\pi.\mathsf{role} = \mathsf{responder}$ and $\pi^*.\mathsf{cid} = \pi.\mathsf{cid}) \implies \neg\pi.\mathsf{revrand})$
          and $(\pi^*.\mathsf{cid}[4][2], \cdot) \notin \vec{sspk}\big)$
        // long-term (signing) secret key of the responder peer was uncompromised upon acceptance,
          and if a partner exists then that partner's randomness is unrevealed,
          and the received ephemeral KEM public key was not replaced with a semi-static one
```

```
$\mathsf{clean}_{\mathsf{LTSS}}(\pi^*)$:
    return
        $\big(\pi^*.\mathsf{role} = \mathsf{initiator}$ and $\neg\mathsf{corrltk}_{\pi^*.\mathsf{oid}}$ and $\neg\mathsf{corrssk}_{\pi^*.\mathsf{sspks}[1]}$ and $\neg\pi^*.\mathsf{pcorr}\big)$
        // tested initiator's long-term and responder's semi-static DH secret keysare uncompromised
          and responder long-term (signing) secret key was uncompromised upon acceptance
    or $\big(\pi^*.\mathsf{role} = \mathsf{responder}$ and $\neg\mathsf{corrltk}_{\pi^*.\mathsf{pid}}$ and $\neg\mathsf{corrssk}_{\pi^*.\mathsf{sspks}[1]}\big)$
        // initiator long-term and tested responder's semi-static DH secret keysare uncompromised

$\mathsf{clean}_{\mathsf{ELT}}(\pi^*)$:
    return
        $\big(\pi^*.\mathsf{role} = \mathsf{initiator}$ and $\neg\pi^*.\mathsf{revrand}$ and $\neg\mathsf{corrltk}_{\pi^*.\mathsf{pid}}\big)$
        // tested initiator's randomness is unrevealed and responder long-term secret key is uncompromised
    or $\big(\pi^*.\mathsf{role} = \mathsf{responder}$ and $\mathsf{clean}_{\mathsf{peerE}}(\pi^*)$ and $\neg\mathsf{corrltk}_{\pi^*.\mathsf{oid}}\big)$
        // initiator's ephemeral contribution is clean and tested responder's long-term secret key is uncompromised

$\mathsf{clean}_{\mathsf{ESS}}(\pi^*)$:
    return
        $\big(\pi^*.\mathsf{role} = \mathsf{initiator}$ and $\neg\pi^*.\mathsf{revrand}$
          and $\forall sspk \in \pi^*.\mathsf{sspks} : (\neg\mathsf{corrssk}_{sspk})$ and $\neg\pi^*.\mathsf{pcorr}$
          and $\pi^*.\mathsf{type} = \mathsf{reduced} \implies (\pi^*.\mathsf{sspks}[2], \cdot) \in \vec{sspk}\ \big)$
        // tested initiator's randomness is unrevealed and responder semi-static secret keys
          are uncompromised and responder long-term (signing) secret key was uncompromised upon acceptance
          and in reduced mode the received semi-static KEM public keywas honestly generated as semi-static key
    or $\big(\pi^*.\mathsf{role} = \mathsf{responder}$ and $\mathsf{clean}_{\mathsf{peerE}}(\pi^*)$
          and $\forall sspk \in \pi^*.\mathsf{sspks} : (\neg\mathsf{corrssk}_{sspk})\big)$
        // initiator's ephemeral contribution is clean and tested responder's semi-static secret keys are uncompromised
```

Predicates $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ capturing post-quantum security

```
$\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}_{\mathsf{full}}(\pi^*)$:
    return $\mathsf{clean}^{Q}_{\mathsf{EE}}(\pi^*)$

$\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}_{\mathsf{reduced}}(\pi^*)$:
    return $\mathsf{clean}^{Q}_{\mathsf{ESS}}(\pi^*)$

$\mathsf{clean}^{Q}_{\mathsf{EE}}(\pi^*)$:
    return $\neg\pi^*.\mathsf{revrand}$ and $\mathsf{clean}_{\mathsf{peerE}}(\pi^*)$
        // test session randomness is unrevealed and peer's ephemeral contribution is clean
```

```
$\mathsf{clean}^{Q}_{\mathsf{ESS}}(\pi^*)$:
    return
        $\big(\pi^*.\mathsf{role} = \mathsf{initiator}$ and $\neg\pi^*.\mathsf{revrand}$ and
          $\exists s \in [n_{ss}] :\ \pi^*.\mathsf{sspks}[2] = sspk^{s,2}_{\pi^*.\mathsf{pid}}$ and $\neg\mathsf{corrssk}_{\pi^*.\mathsf{sspks}[2]}\big)$
        // tested initiator's randomness is unrevealed and
          responder's semi-static KEM secret key was honestly generated and is uncompromised
    or $\big(\pi^*.\mathsf{role} = \mathsf{responder}$ and $\mathsf{clean}_{\mathsf{peerE}}(\pi^*)$ and $\neg\mathsf{corrssk}_{\pi^*.\mathsf{sspks}[2]}\big)$
        // initiator's ephemeral contribution is clean and tested responder's semi-static KEM secret key is uncompromised
```

Figure 9: Clean predicates for the PQXDH protocol capturing classic security ($\mathsf{clean}^{\mathsf{PQXDH}}$, top) and post-quantum security ($\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$, bottom). For post-quantum security, the elliptic curve–related clean predicates ($\mathsf{clean}_{\mathsf{LTSS}}$, $\mathsf{clean}_{\mathsf{ELT}}$, $\mathsf{clean}_{\mathsf{sigE}}$) become moot; security is expected from the ephemeral ($\mathsf{clean}^{Q}_{\mathsf{EE}}$) resp. semi-static ($\mathsf{clean}^{Q}_{\mathsf{ESS}}$) KEM component for full resp. reduced handshakes, upon non-tampered transmission. Lighter gray code reflects the main changes in the classic security predicates compared to the model of BFGJS [BFG+22, BFG+21]. Highlighted code reflects changes in the post-quantum security predicate $\mathsf{clean}^{Q}_{\mathsf{ESS}}$ compared to the classic-security version $\mathsf{clean}_{\mathsf{ESS}}$. Dashed-boxed code prevents trivial confusion attacks between ephemeral and semi-static KEM public keys in PQXDH, since their signatures lack domain separation (cf. discussion in Section 4.3).

Soundness. The model also captures soundness (via the predicate sound), i.e., that session identifiers appropriately reflect correct protocol executions. Concretely, soundness demands that an adversary cannot create one of the following situations (or else it will win the game immediately):

(i) Two sessions accept with the same session identifier, but derive different session keys, indicate different handshake types (full vs. $\mathsf{reduced}$), or do not agree on their contributive identifiers (Fig. 8, line 18).

(ii) Two initiator sessions accept with the same session identifier (Fig. 8, line 19).

(iii) Three sessions accept with the same session identifier in full handshake type (Fig. 8, line 20).

Freshness. Unrestricted access to the game oracles allows for trivial wins, e.g., by testing a session key and also revealing it (or its partner). The freshness predicate fresh rules out such trivial wins, and further encodes that the adversary has not obtained sufficiently many secrets, via CorruptLTKey and/or CorruptSSKey and/or RevealRand queries, to derive the session key of the test session itself and/or substituted signed keys in its execution; this is encoded through a set of so-called clean predicates (Figure 8, line 16 and lines 46 ff.).

4.3 Clean Predicates for PQXDH: Capturing Classic and Post-quantum Security

Following the terminology of CCDGS [CCD+17], the specific security properties targeted by a key exchange protocol KE are expressed through a set of *clean predicates* $\mathsf{clean}^{\mathsf{KE}}$. Figure 9 formalizes the maximum-exposure security properties for PQXDH, both for classic security ($\mathsf{clean}^{\mathsf{PQXDH}}$) as well as for post-quantum security ($\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$).

Classic security. The core classic security properties in $\mathsf{clean}^{\mathsf{PQXDH}}$ (Figure 9, top) closely follow those for the classic X3DH handshake [CCD+17, BFG+22]: in handshakes where the long-term/semi-static, ephemeral/long-term, or ephemeral/semi-static (or, for $\mathsf{type} = \mathsf{full}$ handshakes, ephemeral/ephemeral) secrets combination between initiator/responder is not revealed, we expect security of the derived session key. Jumping ahead, for those combinations where PQXDH employs both DH and KEM keys, we will obtain hybrid/combiner security terms in our analysis (see Section 5.1), capturing that an adversary has to break *both* primitives to successfully attack the protocol via these compromise angles.

In addition to prior work [CCD+17, BFG+22, BFG+21], our clean predicates further model the signatures on the responder's (semi-static and ephemeral public) keys and allow the adversary to replace semi-static and ephemeral keys sent over the network. Instead of assuming authentic distribution of semi-static public keys, we merely preclude the adversary from compromising the (long-term) signing key prior to the targeted session having accepted (cf. Figure 8, lines 54 and 60). Treating signatures explicitly, we also capture the additional guarantee that signing KEM keys gives: for the ephemeral/ephemeral combination to contribute security, it is now (also) sufficient to have the peer's contribution be cleanly signed (cf. Figure 8, line 48). To that end we set a flag pcorr in a session if its peer was corrupted upon acceptance (Figure 8, line 29). For PQXDH, this part of the model ensures that a quantum-later adversary cannot, in addition to a "harvest now, decrypt later" attack, replace the ephemeral KEM public key with its own without being noticed.

Trivial KEM key confusion due to lack of domain separation. Unfortunately, PQXDH signs ephemeral and semi-static KEM public keys without any domain separation (e.g., labels indicating the key type). This leads to a trivial *KEM key confusion attack*: an active adversary can swap out an ephemeral for a semi-static KEM public key (or vice versa), without Alice being able to distinguish the two types, resulting in reduced forward security guarantees. Note that this is different from the key confusion attack noticed by BJKS [BJKS24], where the adversary exchanges the semi-static DH and the semi-static KEM key.[10]

The formal analysis of BJKS [BJKS24] treated all KEM public keys as semi-static, side-stepping the issue. Here, we model both ephemeral and semi-static KEM keys and capture their distinct security properties *if* the adversary does not mount such trivial key confusion attack. We highlight the code ruling out this trivial attack with dashed boxes in Figure 9 (lines 52 and 60). We emphasize that this attack would be easily avoided by domain separating PQXDH's signature, allowing stronger security guarantees (cf. our proof details), a fix suggested also by BJKS [BJKS24] and envisioned by the Signal team for the next protocol upgrade.

Post-quantum security. Via a separate set $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ (Figure 9, bottom) of clean predicates for post-quantum security, we capture the security that remains against a quantum adversary: PQXDH is expected to provide key indistinguishability through a non-tampered KEM encapsulation, either ephemeral (in a full handshake) or semi-static (in a reduced handshake). The authentic transmission of KEM keys captures an adversary that can *not yet* forge signatures (i.e., the signatures still protect the authenticity of the KEM keys) or a quantum adversary that does not tamper with the transmission of the KEM keys. Note that since a quantum adversary can efficiently break discrete logarithms, the DH shares cannot contribute to security anymore.

Clean predicates details. Let us go through the clean predicates in formal detail now. We start with the classic security predicates, which are in parts also used for defining post-quantum security. Let π^* denote the test session. Depending on whether an ephemeral pre-key was used in the key derivation of π^* or not, we apply either the $\mathsf{clean}_{\mathsf{full}}$ or the $\mathsf{clean}_{\mathsf{reduced}}$ predicate to π^*.

Since $\mathsf{clean}_{\mathsf{reduced}}$ is part of the description of $\mathsf{clean}_{\mathsf{full}}$, we first discuss the case $\pi^*.\mathsf{type} = \mathsf{reduced}$. Intuitively, a session key derived in such a session remains unknown to the adversary, if one of the four shared secrets (i.e., excluding the unused DH_4) that constitute the master secret is "clean", i.e., cannot be computed by the adversary. This is the case if one of the following three clean predicates holds for the test session π^*:

$\mathsf{clean}_{\mathsf{LTSS}}$**:** This predicate indicates whether the combination of the long-term key of the initiator and the semi-static DH key of the responder are unknown to the adversary. If the test session is an initiator, the responder's signing key must further be uncompromised upon acceptance.

$\mathsf{clean}_{\mathsf{ELT}}$**:** This predicate indicates whether the combination of the ephemeral contribution of the initiator and the long-term key of the responder is unknown to the adversary.

$\mathsf{clean}_{\mathsf{ESS}}$**:** This predicate indicates whether the combination of the ephemeral contribution of the initiator and the semi-static keys of the responder are unknown to the adversary. If the test session is an initiator, the responder's signing key must further be uncompromised upon acceptance and in reduced mode the received semi-static KEM public key was honestly generated as semi-static key. The latter, dashed-boxed part rules out the trivial attack due to lack of domain separation in PQXDH's signatures discussed above. If such domain separation were added, PQXDH could be proven secure wrt. stronger clean predicates not ruling out such confusion as trivial.

[10] This is not a trivial attack against Signal's implementation of PQXDH since the sizes of DH and KEM public keys are distinct. Hence, we do not exclude it in the model but explicitly rely on this distinction in our analysis.

If the test session π^* is a responder session, the evaluation of $\mathsf{clean}_{\mathsf{ELT}}$ and $\mathsf{clean}_{\mathsf{ESS}}$ necessitates a further predicate called $\mathsf{clean}_{\mathsf{peerE}}$ (in all other cases, it is sufficient to consider the compromise of keys and randomness via the flags $\mathsf{corrltk}, \mathsf{corrssk}$, $\mathsf{revrand}$, and pcorr).

$\mathsf{clean}_{\mathsf{peerE}}$**:** This (sub)predicate indicates that the randomness used within any (sid- or cid-)partnered session is unknown to the adversary.

For test sessions in full handshake mode, i.e., where $\pi^*.\mathsf{type} = \mathsf{full}$, it must either hold that $\mathsf{clean}_{\mathsf{reduced}}$ is true or that the additional input to the master secret computation is clean. The latter is captured by the following predicate:

$\mathsf{clean}_{\mathsf{EE}}$**:** This predicate indicates that the ephemeral contribution of the test session is unknown to the adversary, and that the ephemeral contribution of the peer is unknown to the adversary (captured by $\mathsf{clean}_{\mathsf{peerE}}$) or was "cleanly" signed (captured by the new $\mathsf{clean}_{\mathsf{sigE}}$ predicate we introduce). It is the second part of this "or" statement by which our model captures that KEM ephemeral keys are protected by signatures against replacement.

Again, the predicate $\mathsf{clean}_{\mathsf{peerE}}$ helps to determine within $\mathsf{clean}_{\mathsf{EE}}$ whether the randomness of the test session's (contributive) partners is unrevealed. The following predicate precludes that the adversary substitutes the responder's ephemeral key and signs it with the signing key obtained from corrupting the long-term key.

$\mathsf{clean}_{\mathsf{sigE}}$**:** This (sub)predicate indicates that for initiator test sessions the peer was not corrupted at the time of acceptance and that, if the test session has a contributively partnered session then the ephemeral contribution of that session is not known to the adversary, and that the received ephemeral KEM public key was not replaced with a semi-static one. The latter, dashed-boxed part again rules out the trivial key confusion attack due to lack of domain separation in PQXDH's signatures discussed above.

Finally, in the post-quantum case, security relies solely on the ephemeral/ephemeral KEM contribution for full handshakes resp. on the ephemeral/semi-static KEM contribution for reduced handshakes, captured via the following two predicates:

$\mathsf{clean}^{Q}_{\mathsf{EE}}$**:** As for classic security, this predicate requires that the ephemeral contribution of both the test session and its peer session are unknown to the adversary. Note that since PQXDH's signatures are only classically secure, they cannot contribute to post-quantum security and the $\mathsf{clean}_{\mathsf{sigE}}$ subpredicate is accordingly omitted.

$\mathsf{clean}^{Q}_{\mathsf{ESS}}$**:** Similar to classic security, this predicate requires that the ephemeral contribution of the initiator and the semi-static contribution of the responder are unknown to the adversary. Just that here, only the KEM secret key is relevant for security and must have been honestly generated, since again we cannot hope for tampering protection from the (classically-secure) signatures. Requiring honestly generated keys for post-quantum security also excludes the trivial KEM key confusion attack.

5 PQXDH Analysis

We formalize the PQXDH handshake in Figure 10, highlighting the changes compared to X3DH to achieve post-quantum security in dark gray. Ours is the first reductionist analysis to model the signatures on

```
$\underline{\mathsf{KGenLT}():}$
$(ltpk^{\mathsf{DH}}, ltsk^{\mathsf{DH}}) \leftarrow\$ \mathsf{DH.KGen}()$
$(ltpk^{\mathsf{SIG}}, ltsk^{\mathsf{SIG}}) \leftarrow\$ \mathsf{SIG.KGen}()$
$\mathbf{return}\ ((ltpk^{\mathsf{DH}}, ltpk^{\mathsf{SIG}}),\ (ltsk^{\mathsf{DH}}, ltsk^{\mathsf{SIG}}))$
```

```
$\underline{\mathsf{KGenSS}(ltsk):}$
$(sspk^{\mathsf{DH}}, sssk^{\mathsf{DH}}) \leftarrow\$ \mathsf{DH.KGen}()$ ; $(sspk^{\mathsf{KEM}}, sssk^{\mathsf{KEM}}) \leftarrow\$ \mathsf{KEM.KGen}()$
$(ltsk^{\mathsf{DH}}, ltsk^{\mathsf{SIG}}) \leftarrow ltsk$
$\sigma^{\mathsf{DH}} \leftarrow\$ \mathsf{SIG.Sig}(ltsk^{\mathsf{SIG}}, sspk_B^{\mathsf{DH}})$ ; $\boxed{\sigma^{\mathsf{KEM}} \leftarrow\$ \mathsf{SIG.Sig}(ltsk^{\mathsf{SIG}}, sspk_B^{\mathsf{KEM}})}$
$sssk^{\mathsf{DH}*} \leftarrow (sssk^{\mathsf{DH}}, sspk^{\mathsf{DH}}, \sigma^{\mathsf{DH}})$ ; $sssk^{\mathsf{KEM}*} \leftarrow (sssk^{\mathsf{KEM}}, sspk^{\mathsf{KEM}}, \sigma^{\mathsf{KEM}})$
$\mathbf{return}\ ((sspk^{\mathsf{DH}}, sssk^{\mathsf{DH}*}, \sigma^{\mathsf{DH}}), (sspk^{\mathsf{KEM}}, sssk^{\mathsf{KEM}*}, \sigma^{\mathsf{KEM}}))$
```

Alice **Bob**

```
Bob:
$\underline{\mathsf{Run}(ltsk_B, \vec{ltpk}, \vec{sssk}_B, \pi_B, (\mathsf{create}, (\vec{\mathsf{ssid}}, \mathsf{type})))}$
$(r_1, r_2, r_3, \bot) \leftarrow \pi_B.\mathsf{coins}$
$(sssk_B^{\mathsf{DH}}, sspk_B^{\mathsf{DH}}, \sigma_B^{\mathsf{DH}}) \leftarrow sssk_B^{\vec{\mathsf{ssid}}[1]}$
$\mathbf{if}$ $\mathsf{type} = \mathsf{full}$ // full handshake
    $(epk_B^{\mathsf{DH}}, esk_B^{\mathsf{DH}}) \leftarrow \mathsf{DH.KGen}(; r_1)$
    $(epk_B^{\mathsf{KEM}}, esk_B^{\mathsf{KEM}}) \leftarrow \mathsf{KEM.KGen}(; r_2)$
    $(ltsk_B^{\mathsf{DH}}, ltsk_B^{\mathsf{SIG}}) \leftarrow ltsk_B$
    $\boxed{\sigma^{\mathsf{KEM}} \leftarrow \mathsf{SIG.Sig}(ltsk_B^{\mathsf{SIG}}, epk_B^{\mathsf{KEM}}; r_3)}$
    $epk_B \leftarrow (epk_B^{\mathsf{DH}}, epk_B^{\mathsf{KEM}})$
    $sspk \leftarrow (sspk_B^{\mathsf{DH}}, \bot)$
$\mathbf{else}$ // reduced handshake
    $(sspk_B^{\mathsf{KEM}}, sssk_B^{\mathsf{KEM}}, \sigma_B^{\mathsf{KEM}}) \leftarrow sssk_B^{\vec{\mathsf{ssid}}[2]}$
    $epk_B \leftarrow \bot$
    $sspk \leftarrow (sspk_B^{\mathsf{DH}}, sspk_B^{\mathsf{KEM}})$
    $\boxed{\sigma^{\mathsf{KEM}} \leftarrow \sigma_B^{\mathsf{KEM}}}$
$\pi_B.\mathsf{pid} \leftarrow \star$
$\pi_B.\mathsf{sspks} \leftarrow sspk$
$\pi_B.\mathsf{type} \leftarrow \mathsf{type}$
$\pi_B.\mathsf{cid} \leftarrow (B, ltpk_B, sspk, epk_B)$
$\mathbf{return}\ (\pi_B, m = (B, sspk, epk_B, \sigma_B^{\mathsf{DH}}, \boxed{\sigma^{\mathsf{KEM}}}))$
```

m (Bob → Alice)

```
Alice:
$\underline{\mathsf{Run}(ltsk_A, \vec{sssk}_A, \vec{ltpk}, \pi_A, m)}$
$(r_4, \bot, \bot, r_5) \leftarrow \pi_A.\mathsf{coins}$
$(ltsk_A^{\mathsf{DH}}, ltsk_A^{\mathsf{SIG}}) \leftarrow ltsk_A$
$(epk_A^{\mathsf{DH}}, esk_A^{\mathsf{DH}}) \leftarrow \mathsf{DH.KGen}(; r_4)$
$(B, sspk, epk_B, \sigma_B^{\mathsf{DH}}, \boxed{\sigma^{\mathsf{KEM}}}) \leftarrow m$
$(ltpk_B^{\mathsf{DH}}, ltpk_B^{\mathsf{SIG}}) \leftarrow ltpk_B$
$(sspk_B^{\mathsf{DH}}, sspk_B^{\mathsf{KEM}}) \leftarrow sspk$
$\mathbf{if}$ $\mathsf{SIG.Vf}(ltpk_B^{\mathsf{SIG}}, sspk_B^{\mathsf{DH}}, \sigma_B^{\mathsf{DH}}) = \mathsf{false}$
    $\mathbf{return}\ (\pi_A, \epsilon)$
$DH_1 \leftarrow \mathsf{DH}(ltsk_A^{\mathsf{DH}}, sspk_B^{\mathsf{DH}})$
$DH_2 \leftarrow \mathsf{DH}(esk_A^{\mathsf{DH}}, ltpk_B^{\mathsf{DH}})$
$DH_3 \leftarrow \mathsf{DH}(esk_A^{\mathsf{DH}}, sspk_B^{\mathsf{DH}})$
$\mathbf{if}$ $epk_B \neq \bot$ // full handshake
    $(epk_B^{\mathsf{DH}}, epk_B^{\mathsf{KEM}}) \leftarrow epk_B$
    $DH_4 \leftarrow \mathsf{DH}(esk_A^{\mathsf{DH}}, epk_B^{\mathsf{DH}})$
    $\boxed{\mathbf{if}\ \mathsf{SIG.Vf}(ltpk_B^{\mathsf{SIG}}, epk_B^{\mathsf{KEM}}, \sigma^{\mathsf{KEM}}) = \mathsf{false}}$
    $\boxed{\quad \mathbf{return}\ (\pi_A, \epsilon)}$
    $(ct, ss) \leftarrow \mathsf{KEM.Enc}(epk_B^{\mathsf{KEM}}; r_5)$
    $\pi_A.\mathsf{type} \leftarrow \mathsf{full}$
$\mathbf{else}$ // reduced handshake
    $DH_4 \leftarrow \epsilon$
    $\boxed{\mathbf{if}\ \mathsf{SIG.Vf}(ltpk_B^{\mathsf{SIG}}, sspk_B^{\mathsf{KEM}}, \sigma^{\mathsf{KEM}}) = \mathsf{false}}$
    $\boxed{\quad \mathbf{return}\ (\pi_A, \epsilon)}$
    $(ct, ss) \leftarrow \mathsf{KEM.Enc}(sspk_B^{\mathsf{KEM}}; r_5)$
    $\pi_A.\mathsf{type} \leftarrow \mathsf{reduced}$
$\pi_A.\mathsf{K} \leftarrow \mathsf{KDF}(DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss)$
$\pi_A.\mathsf{sid} \leftarrow \begin{pmatrix} A, B, ltpk_A, ltpk_B, \\ sspk, epk_B, epk_A^{\mathsf{DH}}, ct \end{pmatrix}$
$\pi_A.\mathsf{pid} \leftarrow B$
$\pi_A.\mathsf{sspks} \leftarrow sspk$
$\pi_A.\mathsf{cid} \leftarrow (B, ltpk_B, sspk, epk_B)$
$\pi_A.\mathsf{st}_{\mathsf{exec}} \leftarrow \mathsf{accepted}$
$\mathbf{return}\ (\pi_A, m' = (A, epk_A^{\mathsf{DH}}, ct))$
```

m' (Alice → Bob)

```
Bob:
$\underline{\mathsf{Run}(ltsk_B, \vec{sssk}_B, \vec{ltpk}, \pi_B, m'; \mathsf{coins})}$
$(A, epk_A^{\mathsf{DH}}, ct) \leftarrow m'$
$(ltpk_A^{\mathsf{DH}}, ltpk_A^{\mathsf{SIG}}) \leftarrow ltpk_A$
$DH_1 \leftarrow \mathsf{DH}(ltpk_A^{\mathsf{DH}}, sssk_B^{\mathsf{DH}})$
$DH_2 \leftarrow \mathsf{DH}(epk_A^{\mathsf{DH}}, ltsk_B^{\mathsf{DH}})$
$DH_3 \leftarrow \mathsf{DH}(epk_A^{\mathsf{DH}}, sssk_B^{\mathsf{DH}})$
$\mathbf{if}$ $\pi_B.\mathsf{type} = \mathsf{full}$ // full handshake
    $DH_4 \leftarrow \mathsf{DH}(epk_A^{\mathsf{DH}}, esk_B^{\mathsf{DH}})$
    $ss \leftarrow \mathsf{KEM.Dec}(esk_B^{\mathsf{KEM}}, ct)$
$\mathbf{else}$ // reduced handshake
    $DH_4 \leftarrow \epsilon$
    $ss \leftarrow \mathsf{KEM.Dec}(sssk_B^{\mathsf{KEM}}, ct)$
$\pi_B.\mathsf{K} \leftarrow \mathsf{KDF}(DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss)$
$\pi_B.\mathsf{sid} \leftarrow \begin{pmatrix} A, B, ltpk_A, ltpk_B, \\ sspk, epk_B, epk_A^{\mathsf{DH}}, ct \end{pmatrix}$
$\pi_B.\mathsf{pid} \leftarrow A$
$\pi_B.\mathsf{st}_{\mathsf{exec}} \leftarrow \mathsf{accepted}$
$\mathbf{return}\ (\pi_B, \epsilon)$
```

Figure 10: The PQXDH protocol. Darker gray code are additions compared to the original X3DH protocol to add post-quantum security. Lighter gray code are signature elements of the X3DH and PQXDH not covered in the prior analyses of [CCD+17].

public keys, which we accent with light gray. Long-term key pairs consist of a DH key pair and a signing key pair.[11] Semi-static key pairs consist of a DH key pair and a KEM key pair, each signed under the long-term signing key. The protocol then works as follows.

First, Bob produces a pre-key bundle with semi-static and ephemeral keys as indicated by the $\vec{\mathsf{ssid}}$ vector and type: The pre-key bundle always includes the semi-static DH key and for $\mathsf{reduced}$ handshakes a semi-static KEM key. Only for full handshakes, the pre-key bundle includes ephemeral DH and KEM keys.[12] Signatures on the semi-static DH key and the KEM key (regardless if it is semi-static or ephemeral) are always included.

Second, Alice verifies both signatures and samples an ephemeral DH key pair of her own. She computes several DH secret combinations, namely long-term/semi-static, ephemeral/long-term, ephemeral/semi-static, and ephemeral/ephemeral (the last one only for full handshakes), and encapsulates against Bob's KEM public key. She derives the session key via a key derivation function KDF on input all three/four DH shared secrets and the KEM shared secret and sends her sampled DH ephemeral public key and the KEM ciphertext to Bob.

Third, Bob computes the same DH shared secrets and decapsulates the KEM ciphertext. He derives the session key in the same manner and accepts.

5.1 Classic Security of PQXDH

We first show that PQXDH achieves key indistinguishability against classic adversaries wrt. the maximum-exposure attack vectors formalized through the clean predicates $\mathsf{clean}^{\mathsf{PQXDH}}$ in our model (cf. Figure 9). This result relies on the GapDH assumption in the DH group, the KEM providing one-way CCA and $LEAK$-BIND-SS-$\{CT, PK\}$ security as well as low key-collision probability and correctness errors, the signature scheme being existentially unforgeable, and the key derivation function behaving like a random oracle. Note that we further explicitly assume the representations of KEM and Diffie–Hellman public keys to be disjoint. This is the case for Signal's PQXDH implementation, employing Kyber as KEM and X25519 or X448 for elliptic-curve DH [LHT16].

As for the prior analysis of X3DH [CCD+17] whose proof structure we follow in parts, the bound is highly non-tight; we still give it in concrete terms for clarity.

Theorem 5.1 (Classic key indistinguishability of PQXDH)**.** *The* PQXDH *protocol given in Figure 10 with randomness space* $\mathcal{R}_{\mathsf{KE}} = \mathcal{R}_{\mathsf{DH.KGen}} \times \mathcal{R}_{\mathsf{KEM.KGen}} \times \mathcal{R}_{\mathsf{KEM.Enc}} \times \mathcal{R}_{\mathsf{SIG.Sig}}$ *achieves* $(t, \epsilon, (q_{\mathrm{Snd}}, q_{\mathrm{CorrLT}}, q_{\mathrm{CorrSS}}, q_{\mathrm{RevR}}, q_{\mathrm{RevSK}}))$*–key indistinguishability wrt. the classic security clean predicates* $\mathsf{clean}^{\mathsf{PQXDH}}$ *(cf. Figure 9), assuming (for* $t' \approx t$*) the GapDH problem in the group* $(\mathbb{G}, g, q)$ *is* $(t', \epsilon_{\mathsf{GDH}}, 1.5(q_{\mathrm{RO}} + q_{\mathrm{Snd}})^2)$*-hard,* KEM *is a* $(t', \epsilon_{\mathsf{CCA}}, n_s)$*–*OW-CCA*-secure and* $(t', \epsilon_{LEAK^{+r}}, n_p \cdot n_{ss} + n_s)$*–*$LEAK^{+r}$-$\mathsf{BIND}$-$SS$-$\{CT, PK\}$*-secure KEM with public-key collision probability* γ_{coll} *and correctness error* δ_{corr}*,* SIG *is a* $(t', \epsilon_{\mathsf{SIG}}, 2n_{ss} + n_s)$*-unforgeable signature scheme,* KDF *behaves like a (programmable) random oracle, and the group* $(\mathbb{G}, g, q)$ *and the public*

[11] The PQXDH description [KS24] suggests to use a single key pair for both DH and the signature scheme XEdDSA [Per16], modeling which we leave to future work.

[12] Ephemeral DH and KEM pre-keys are used as long as there are some left on the Signal server. In practice, ephemeral DH and KEM pre-keys may run out at different points in time. We do not model this to improve accessibility of the proof.

key space of KEM *have disjoint representations. Concretely, for any efficient classic adversary* $\mathcal{A}$ *we have*

$$\begin{aligned}\mathsf{Adv}^{\mathsf{KI}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \frac{(n_p + n_p \cdot n_{ss} + n_s)^2}{q} &+ \gamma_{\mathsf{coll}}(n_p \cdot n_{ss} + n_s) + n_s \cdot \delta_{\mathsf{corr}} + \epsilon_{LEAK^{+r}} \\ &+ n_p \cdot (\epsilon_{\mathsf{SIG}} + n_p \cdot n_{ss} \cdot \epsilon_{\mathsf{GDH}}) && // \ \mathsf{clean}_{\mathsf{LTSS}} \\ &+ n_s \cdot n_p \cdot \epsilon_{\mathsf{GDH}} && // \ \mathsf{clean}_{\mathsf{ELT}} \\ &+ n_p \cdot (\epsilon_{\mathsf{SIG}} + n_{ss} \cdot n_s \cdot \epsilon_{\mathsf{GDH}}) && // \ \mathsf{clean}_{\mathsf{ESS}} \wedge \mathsf{type} = \mathsf{full} \\ &+ n_p \cdot (\epsilon_{\mathsf{SIG}} + n_{ss} \cdot n_s \cdot \min(\epsilon_{\mathsf{GDH}}, \epsilon_{\mathsf{CCA}})) && // \ \mathsf{clean}_{\mathsf{ESS}} \wedge \mathsf{type} = \mathsf{reduced} \\ &+ n_s^2 \cdot \min(\epsilon_{\mathsf{GDH}}, \epsilon_{\mathsf{CCA}}) && // \ \mathsf{clean}_{\mathsf{EE}} \wedge \mathsf{clean}_{\mathsf{peerE}} \\ &+ n_p \cdot \big(\epsilon_{\mathsf{SIG}} + n_s^2 \cdot q_{\mathrm{RO}} \cdot \epsilon_{\mathsf{CCA}}\big)\,. && // \ \mathsf{clean}_{\mathsf{EE}} \wedge \mathsf{clean}_{\mathsf{sigE}}\end{aligned}$$

For easier accessibility, we first give a summary of the proof; the full proof follows below.

Proof summary. The proof proceeds via a series of game hops. First, we exclude collisions in the DH and KEM public keys, yielding the terms $\frac{(n_p+n_p \cdot n_{ss}+n_s)^2}{q} + \gamma_{\mathsf{coll}}(n_p \cdot n_{ss} + n_s)$. Second, we exclude KEM correctness errors when decapsulating in sessions ($n_s \cdot \delta_{\mathsf{corr}}$). Now, soundness holds, as the session identifiers carrying non-colliding DH/KEM keys make them non-colliding themselves and sessions agreeing on session identifiers in particular derive the same session keys. Third, we ensure that KEM shared secrets bind the public keys and ciphertexts, introducing the $\epsilon_{LEAK^{+r}}$ term; this in particular rules out re-encapsulation attacks where sessions that are not partnered possibly derive the same session key.[13]

At this stage we separate the proof into several cases based on the clean predicate (cf. the annotated sum in the bound of Theorem 5.1). Let us illustrate the fourth case ($\mathsf{clean}_{\mathsf{ESS}} \wedge \mathsf{type} = \mathsf{reduced}$) here since it covers the most interesting proof techniques, including handling the signatures and a hybrid DH/KEM argument; the other cases follow in a similar fashion. In this case, the $\mathsf{clean}_{\mathsf{ESS}}$ predicate is satisfied and the adversary is testing a session running a reduced ($\mathsf{type} = \mathsf{reduced}$) handshake. In the proof, we first guess the responder identity involved in the test session, introducing a n_p loss. Second, we ensure that the guessed responder's semi-static KEM and DH keys are not tampered with in the test session, reducing to the unforgeability of the signature scheme (ϵ_{SIG}). Third and fourth, we guess the two semi-static key identifiers of the responder as well as the initiator session, with a loss of $n_{ss}^2 \cdot n_s$. Fifth, we abort if the adversary queries the key derivation function KDF, modeled as random oracle, on the master secret of the test session. To bound this step, we embed *both* a GapDH challenge *and* a OW-CCA challenge into the test session (concretely, into the ephemeral DH and KEM encapsulation contribution of the initiator and the semi-static DH and KEM keys of the responder, which we both guessed before). If the adversary detects this change, the reduction can win *both* games, yielding the minimum of the advantages as the *hybrid* bound, i.e., $\min(\epsilon_{\mathsf{GDH}}, \epsilon_{\mathsf{CCA}})$. Finally, we can replace the session key of the test session with a random key, making it independent of the challenge bit b_{test}. Concluding that the session key of the test session is now independent of b_{test} since non-partnered sessions derive distinct keys (in particular due to the KEM binding hop in the beginning), this completes the proof case. □

We get hybrid guarantees for the fourth and fifth proof cases ($\mathsf{clean}_{\mathsf{ESS}}$ in reduced mode and $\mathsf{clean}_{\mathsf{EE}}$ in full mode with a clean peer, i.e., $\mathsf{clean}_{\mathsf{peerE}}$), showing that for these attack vectors, an adversary would need to break *both* GapDH in the DH group and OW-CCA security of the KEM. The other attack vectors either involve only DH secrets (and hence do not provide post-quantum security), or, for the sixth case ($\mathsf{clean}_{\mathsf{EE}}$ in full mode with clean signatures), the adversary can manipulate the (unsigned) ephemeral DH key, hence only the KEM secret ensures security here. All in all, PQXDH *maintains* the classical guarantees of X3DH for the first three cases (as intended), extends the guarantees for the fourth and fifth case to provide *hybrid*

[13] As we will detail in the proof (Game 3) and discuss in Section 6, if PQXDH included the session context (e.g., the values in the session identifier) in the key derivation, as is good practice, this problem would not occur: different KEM public keys or ciphertexts would lead to different session keys (by being part of the KDF input as session context).

guarantees, and *adds* a (post-quantum) guarantee for the last case due to the signed ephemeral KEM key (compared to no guarantee under this attack vector for X3DH).

We now give the full proof for Theorem 5.1.

Proof. **Game 0.** We start with Game $\mathcal{G}_0$ being the original key indistinguishability game $\mathcal{G}^{\mathsf{KI}}_{\mathsf{PQXDH}}(\mathcal{A})$, i.e.,

$$\mathsf{Adv}^{\mathsf{KI}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_0}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game 1 (DH and KEM public key collisions). We let $\mathcal{G}_0$ abort (overwriting the adversary's output with 0) if any two honestly generated DH public keys or KEM public keys coincide. There are n_p many long-term DH keys, $n_p \cdot n_{ss}$ many semi-static DH keys, and at most n_s many ephemeral DH keys, so by the birthday bound the probability for any two DH keys colliding can be upper-bounded by $\frac{(n_p + n_p \cdot n_{ss} + n_s)^2}{q}$. As for KEM keys, there are $n_p \cdot n_{ss}$ semi-static ones and at most n_s ephemeral ones, which collide with probability at most $\gamma_{\mathsf{coll}}(n_p \cdot n_{ss} + n_s)$; cf. Definition 3.1. Put together, we have

$$\mathsf{Adv}^{\mathcal{G}_0}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \frac{(n_p + n_p \cdot n_{ss} + n_s)^2}{q} + \gamma_{\mathsf{coll}}(n_p \cdot n_{ss} + n_s) + \mathsf{Adv}^{\mathcal{G}_1}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game 2 (KEM correctness). We next let $\mathcal{G}_1$ abort if any honestly created KEM encapsulation fails to decapsulate correctly. This introduces a KEM correctness error term for each of the n_s many sessions:

$$\mathsf{Adv}^{\mathcal{G}_1}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s \cdot \delta_{\mathsf{corr}} + \mathsf{Adv}^{\mathcal{G}_2}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Soundness. At this point soundness holds unconditionally; we consider the three sub-conditions of the sound predicate:

- *Agreement on shared key, type, contributive identifiers* (Figure 8, line 18): The DH keys and the KEM key and the KEM ciphertext in the session identifier determine all inputs to the key derivation function KDF. By $\mathcal{G}_2$, KEM ciphertexts decapsulate correctly and both parties derive the same session key. The session identifier further determines the type by epk_B being empty ($\mathsf{type} = \mathsf{reduced}$) or not ($\mathsf{type} = \mathsf{full}$). Finally, the contributive identifier is a subset of the session identifier, hence matching if the latter match.

- *No initiator session identifiers collide* (Figure 8, line 19): Every initiator samples a fresh ephemeral DH key epk^{DH}_A included in the session identifier. These DH keys do not collide by Game $\mathcal{G}_1$. Hence, initiator session identifiers do not collide.

- *No three session identifiers collide in full mode* (Figure 8, line 20): Three colliding session identifiers implies colliding identifiers for two initiator sessions or for two responder sessions. Collisions of initiator session are ruled out above already. Collisions of responder sessions imply colliding ephemeral DH keys epk^{DH}_B, which are ruled out by Game $\mathcal{G}_1$.

Game 3 (KEM shared secret collisions). We will need later that two sessions that use distinct (semi-static or ephemeral) KEM public keys or distinct ciphertexts do not end up using the same shared secret ss (and hence possibly the same session key, despite not being partnered).[14] In this game, we abort if any two sessions derive the same shared secret ss while using different KEM public keys (either key may be an ephemeral key in a full session or a semi-static key in a reduced session) or different KEM ciphertexts.

We bound the probability of this abort happening by the advantage of a reduction $\mathcal{B}_1$ in breaking KCR security of the KEM scheme for $n = n_p \cdot n_{ss} + n_s$. Let $\mathcal{B}_1$ use the first $n_p \cdot n_{ss}$ challenge keys as semi-static KEM keys in KGenSS instead of generating them itself, and the remaining challenge keys in place of the (at most n_s) ephemeral KEM keys generated by responder sessions (letting the $\mathsf{KEM.KGen}$ randomness obtained in the KCR game replace r_2 in the session's random coins). If any two sessions π_i, π_j obtain the same KEM shared secret $ss_i = ss_j$ (as the output of encapsulation for initiator sessions, or as the output of decapsulation for responder sessions) involving two distinct public keys pk_i, pk_j or two distinct ciphertexts ct_i, ct_j, $\mathcal{B}_1$ outputs (i, ct_i, j, ct_j) and wins. Hence,

$$\mathsf{Adv}^{\mathcal{G}_2}_{\mathsf{PQXDH}}(\mathcal{A}) \le \epsilon_{LEAK+r} + \mathsf{Adv}^{\mathcal{G}_3}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Separating the clean cases. At this point, we will divide the proof into six sub-cases following the structure of the $\mathsf{clean}^{\mathsf{PQXDH}}_{\mathsf{full}}$ resp. $\mathsf{clean}^{\mathsf{PQXDH}}_{\mathsf{reduced}}$ predicates evaluated on the test session π^*. We can bound the advantage of the adversary in Game $\mathcal{G}_3$ by the sum of its advantages in Games $\mathcal{G}_3[\mathsf{c}]$ which are $\mathcal{G}_3$ conditioned on a sub-predicate c:

$$\mathsf{Adv}^{\mathcal{G}_3}_{\mathsf{PQXDH}}(\mathcal{A}) \le \sum_{\substack{\mathsf{c} \in \{ \mathsf{clean_{LTSS}}(\pi^*),\ \mathsf{clean_{ELT}}(\pi^*), \\ \mathsf{clean_{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full},\ \mathsf{clean_{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{reduced}, \\ \mathsf{clean_{EE}}(\pi^*) \wedge \mathsf{clean_{peerE}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full}, \\ \mathsf{clean_{EE}}(\pi^*) \wedge \mathsf{clean_{sigE}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full} \}}} \mathsf{Adv}^{\mathcal{G}_3[\mathsf{c}]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

In the remainder of the proof, we will bound each of these cases, numbered A–F, separately.

Case A ($\mathsf{clean_{LTSS}}(\pi^*)$).

In this proof case, the $\mathsf{clean_{LTSS}}$ predicate ensures for the test session π^* that

1. the initiator's long-term key is uncompromised, and
2. the responder's semi-static key is uncompromised, and
3. if π^* is an initiator, the responder's long-term (signing) key was uncompromised upon acceptance.

Via the last point, we can guarantee that the initiator uses a semi-static key that was honestly generated by the responder, given signatures are unforgeable. Then, similar to the classical Signal proof [CCD+17], the uncompromised long-term/semi-static Diffie–Hellman combination then ensures key indistinguishability for the test session.

Game A.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean_{LTSS}}(\pi^*)$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{A.0}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean_{LTSS}}(\pi^*)]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

[14] If PQXDH included the session context (e.g., the values in the session identifier) in the key derivation, we would not need this game hop, since different KEM public keys or ciphertexts would lead to different session keys. See also the discussion in Section 6.

Game A.1 (Guess responder identity V^*). We first guess the identity V^* of the responder involved in the test session, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of users n_p:

$$\mathsf{Adv}^{\mathcal{G}_{A.0}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_p \cdot \mathsf{Adv}^{\mathcal{G}_{A.1}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.2 (Signature unforgeability). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the test session π^* is an initiator session and accepts using a semi-static DH public key $sspk^{\mathsf{DH}}_{V^*}$ that was not generated through a KGenSS run for V^*. Hence, the test session accepts with $\pi^*.\mathsf{sspks} = (sspk^{\mathsf{DH}}_{V^*}, \cdot)$ corresponding to a DH key pair of which the adversary does not know the secret key, since $\mathsf{clean}_{\mathsf{LTSS}}(\pi^*)$ guarantees that the semi-static secret corresponding to $sspk^{\mathsf{DH}}_{V^*}$ is not compromised. The probability of such an abort can be bounded by the advantage of the following reduction $\mathcal{B}_2$ against the $(t', \epsilon_{\mathsf{SIG}}, 2n_{ss} + n_s)$-unforgeability of SIG.

The reduction $\mathcal{B}_2$ samples all key components itself except for the signature key of V^*: In place of the long-term public signature key $ltpk^{\mathsf{SIG}}_{V^*}$ of V^* it uses the public key pk obtained in its unforgeability game. In its simulation of Game $\mathcal{G}_{A.1}$, $\mathcal{B}_2$ uses its signing oracle to obtain signatures under $ltsk^{\mathsf{SIG}}_{V^*}$: two per semi-static key (for DH and KEM public keys) and up to n_s signatures on ephemeral KEM keys. Since $\mathsf{clean}_{\mathsf{LTSS}}(\pi^*) = \mathsf{true}$, we know that $\pi^*.\mathsf{pcorr} = \mathsf{false}$, i.e., the long-term key of V^* was not corrupted when the test session accepted and so $\mathcal{B}_2$ does not have to answer a $\textsc{CorruptLTKey}(V^*)$ query prior to the abort event. Hence, $\mathcal{B}_2$ provides a perfect simulation of Game $\mathcal{G}_{A.1}$ up to when the abort would happen, and if π^* receives a signature σ_{V^*} on a semi-static DH public key that V^* did not generate, then $\mathcal{B}_2$ can output this as its forgery and wins. Note that the adversary cannot masquerade a signed KEM public key as signed DH public key since public keys are distinct for DH and KEM by assumption. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{A.1}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{SIG}} + \mathsf{Adv}^{\mathcal{G}_{A.2}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.3 (Guess initiator identity U^*). We guess the identity U^* of the initiator to the test session, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step again loses at most a factor of the number of users n_p:

$$\mathsf{Adv}^{\mathcal{G}_{A.2}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_p \cdot \mathsf{Adv}^{\mathcal{G}_{A.3}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.4 (Guess semi-static DH key identifier $\mathsf{ssid}^*_{\mathsf{DH}}$ of V^*). We now guess the identifier ssid of the responder V^*'s (uncorrupted) semi-static DH key $sspk^{\mathsf{DH}}_{V^*}$. Note that depending on the role of π^* this is either the test session's own key (if $\pi^*.\mathsf{role} = \mathsf{responder}$), or of the intended peer (if $\pi^*.\mathsf{role} = \mathsf{initiator}$). We denote the guessed identifier by $\mathsf{ssid}^*_{\mathsf{DH}}$, and abort, setting the adversary's output bit to 0, if this guess is incorrect, losing at most a factor of the number of semi-static keys per user n_{ss}:

$$\mathsf{Adv}^{\mathcal{G}_{A.3}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_{ss} \cdot \mathsf{Adv}^{\mathcal{G}_{A.4}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.5 (GapDH). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the adversary queries the random oracle on a value formatted like a master secret and beginning with $\mathsf{CDH}(ltpk^{\mathsf{DH}}_{U^*}, sspk^{\mathsf{DH}}_{V^*})$, i.e., the shared DH key between $ltpk^{\mathsf{DH}}_{U^*}$ and $sspk^{\mathsf{DH}}_{V^*}$, where $sspk^{\mathsf{DH}}_{V^*}$ is the semi-static DH key with identifier $\mathsf{ssid}^*_{\mathsf{DH}}$. The probability of such an abort can be bounded by the advantage of the following reduction $\mathcal{B}_3$ against the $(t', \epsilon_{\mathsf{GDH}}, q_{\mathrm{DDH}})$-hardness of the GapDH problem in $(\mathbb{G}, g, q)$ for $q_{\mathrm{DDH}} \leq 1.5(q_{\mathrm{RO}} + q_{\mathrm{Snd}})^2$.

```
$RO'(x)$:
if $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss = x$ // if we can parse the query string $x$ as three or four DH shared secrets ($DH_4$ may be empty) and a KEM shared secret...
    if $\mathrm{DDH}(ltpk^{\mathsf{DH}}_{U^*}, sspk^{\mathsf{DH}}_{V^*}, DH_1)$
        $\mathcal{B}_3$ halts and returns $DH_1$ as its GapDH solution // change for Game $\mathcal{G}_{\mathrm{A.5}}$
    return $\mathsf{patch}(DH_1, DH_2, DH_3, DH_4, ss)$ // extra routine to ensure consistency
return $RO(x)$
```

```
$\mathsf{patch}(PoS_1, PoS_2, PoS_3, DH_4, ss)$:
    // the first three arguments may pairs of DH public keys or DH shared secrets
foreach $(e_1, e_2, e_3, e_4, e_5, y)$ in $L$ // iterate over list entries
    if $\mathsf{DDH\text{-}eq}(PoS_1, e_1) \land \mathsf{DDH\text{-}eq}(PoS_2, e_2) \land \mathsf{DDH\text{-}eq}(PoS_3, e_3) \land DH_4 = e_4 \land ss = e_5$
        return $y$ // query matches a prior one, respond with recorded value $y$
$y \leftarrow\!\!\$\ \{0,1\}^{256}$ // new query; sample value at random
$L \leftarrow L \cup \{(PoS_1, PoS_2, PoS_3, DH_4, ss, y)\}$ // record response $y$ on this query
return $y$
```

```
$\mathsf{DDH\text{-}eq}(PoS_1, PoS_2)$:
if $(pk_1, pk_2) = PoS_1 \land$ if $(pk_3, pk_4) = PoS_2$ // if both $PoS_1$ and $PoS_2$ parse as two public keys...
    return $\{pk_1, pk_2\} = \{pk_3, pk_4\}$
if $(pk_1, pk_2) = PoS_1$ // if $PoS_1$ parses as two public keys...
    return $\mathrm{DDH}(pk_1, pk_2, PoS_2)$
if $(pk_1, pk_2) = PoS_2$ // if $PoS_2$ parses as two public keys...
    return $\mathrm{DDH}(pk_1, pk_2, PoS_1)$
return $PoS_1 = PoS_2$ // else, both are DH shared secrets
```

```
SEND$(W, i, m)$ substitutes $\mathsf{KDF}$ call in $\mathsf{Run}$ where GapDH challenge keys are involved:
...
$U \leftarrow \pi^i_W.\mathsf{initiator}$;    $V \leftarrow \pi^i_W.\mathsf{responder}$;    $sspk^{DH}_V \leftarrow \pi^i_W.\mathsf{sspks}[1]$
$(\ldots, epk^{\mathsf{DH}}_U, \ldots) \leftarrow \pi^i_W.\mathsf{sid}$
if $U = U^* \land sspk^{DH}_V = sspk^{\mathsf{DH}}_{V^*}$
    compute $DH_2, DH_4, ss$ with the corresponding secret keys as specified in $\mathsf{Run}$
    $\pi^i_W.\mathsf{K} \leftarrow \mathsf{patch}((ltpk^{\mathsf{DH}}_{U^*}, sspk^{\mathsf{DH}}_{V^*}), DH_2, (epk^{\mathsf{DH}}_U, sspk^{\mathsf{DH}}_{V^*}), DH_4, ss)$
else if $V = U^*$
    compute $DH_1, DH_3, DH_4, ss$ with the corresponding secret keys as specified in $\mathsf{Run}$
    $\pi^i_W.\mathsf{K} \leftarrow \mathsf{patch}(DH_1, (epk^{\mathsf{DH}}_U, ltpk^{\mathsf{DH}}_{U^*}), DH_3, DH_4, ss)$
else if $U \neq U^* \land sspk^{DH}_V = sspk^{\mathsf{DH}}_{V^*}$
    compute $DH_1, DH_2, DH_4, ss$ with the corresponding secret keys as specified in $\mathsf{Run}$
    $\pi^i_W.\mathsf{K} \leftarrow \mathsf{patch}(DH_1, DH_2, (epk^{\mathsf{DH}}_U, sspk^{\mathsf{DH}}_{V^*}), DH_4, ss)$
else // $\mathcal{B}_3$ can compute all secrets
    compute $DH_1, DH_2, DH_3, DH_4, ss$ with the corresponding secret keys as specified in $\mathsf{Run}$
    $\pi^i_W.\mathsf{K} \leftarrow \mathsf{patch}(DH_1, DH_2, DH_3, DH_4, ss)$
...
```

Figure 11: Algorithmic description of how $\mathcal{B}_3$ simulates the random oracle and the SEND oracle in game $\mathcal{G}_{\mathrm{A.5}}$. To ensure consistency of queries we use an additional (global) list L, which is initialized as empty. This list L contains six-tuples consisting of three pairs of DH public keys or DH shared secrets (PoS), a DH shared secret, a KEM shared secret, and the associated RO output. The patch routine checks if incoming queries match list entries with the DDH-eq routine: DDH-eq returns true if both PoS argument are the same two public keys, if the two PoS arguments are a valid DH tuple according to the DDH oracle, or if the two PoS arguments are identical DH shared secrets. We patch the SEND oracle to use the patch routine in lieu of computing the KDF, whenever there are challenge secret keys involved that $\mathcal{B}_3$ does not know.

Reduction $\mathcal{B}_3$ samples all key components itself except for replacing the long-term DH key $ltpk^{\mathsf{DH}}_{U^*}$ of U^* and the semi-static DH key $sspk^{\mathsf{DH}}_{V^*}$ of V^* with its own challenge values g^a and g^b. Since $\mathsf{clean}_{\mathsf{LTSS}}(\pi^*) = \mathsf{true}$, $\mathcal{B}_3$ never has to answer the queries $\textsc{CorruptLTKey}(U^*)$ or $\textsc{CorruptSSKey}(V^*, \mathsf{ssid}^*_{\mathsf{DH}})$ in its simulation of Game $\mathcal{G}_{\text{A.4}}$. Though, $\mathcal{B}_3$ may have to answer for $\textsc{RevealSessKey}$ and $\textsc{Send}$ queries that involve the substituted keys. Hence, the reduction patches the random oracle and the $\textsc{Send}$ oracle to generate the session keys in a consistent manner. In consequence, the reduction answers queries to the $\textsc{RevealSessKey}$ oracle consistently as well.

Figure 11 gives an algorithmic description of the changes. In particular, the reduction gives special treatment to $\textsc{Send}$ queries that involve any of the challenge secrets, i.e., for $\textsc{Send}$ queries where the initiator is the initiator of the test session U^* partnered with the responder of the test session V^* using the semi-static key $sspk^{\mathsf{DH}}_{V^*}$ of the test session, the responder is U^*, or the responder is V^* with semi-static key $sspk^{\mathsf{DH}}_{V^*}$ and the initiator is not U^*. In these cases the reduction cannot compute all DH shared secrets entering the key derivation function and chooses a session key via the patch routine. Specifically, the patch routine takes as arguments three $\underline{p}$airs of DH public keys $\underline{o}$r DH shared $\underline{s}$ecrets PoS (since DH_1, DH_2, DH_3 are the DH shared secrets that the reduction can potentially not compute), a DH shared secret, and a KEM shared secret, and returns a RO output that is consistent with previous queries. To this end, patch saves all queries together with their RO response in a list L and checks new queries against L. Note here that two PoS are equivalent if the pair of public keys in one of them forms a valid Diffie–Hellman triple with the shared secret in the other one according to the DDH oracle or if the PoS components are equivalent, as described by the $\mathsf{DDH\text{-}eq}$ routine. If a query does not match any previous query, the patch routine randomly chooses a new session key and records the query with the response in L.

Since the $\textsc{Send}$ oracle computes consistent session keys for each session, $\mathcal{B}_3$ also consistently answers to the $\textsc{RevealSessKey}$ oracle queries. To ensure consistency with queries to the random oracle, the reduction also uses the patch routine for RO queries. Finally, in case the adversary queries the random oracle on a value beginning with $\mathsf{CDH}(ltpk^{\mathsf{DH}}_{U^*}, sspk^{\mathsf{DH}}_{V^*})$, the reduction $\mathcal{B}_3$ halts and returns this value to its own challenger.

Note that for each RO query the reduction makes up to $1 + 2|L| \leq 3|L|$ queries to the DDH oracle (line 2 and up to two times per entry in line 7). For each query to the $\textsc{Send}$ oracle,[15] the reduction calls the patch routine once, which makes up to $3|L|$ queries to the DDH oracle. Each call to patch adds at most one entry to L, so $|L|$ grows from 1 to $q_{\text{RO}} + q_{\text{Snd}}$. Using the Gaussian sum we get

$$q_{\text{DDH}} \leq \sum_{i=1}^{q_{\text{RO}}+q_{\text{Snd}}} 3 \cdot (i-1) = 3 \cdot \frac{(q_{\text{RO}} + q_{\text{Snd}} - 1) \cdot (q_{\text{RO}} + q_{\text{Snd}})}{2} \leq 1.5(q_{\text{RO}} + q_{\text{Snd}})^2.$$

Unless $\mathcal{A}$ queries the random oracle on $\mathsf{CDH}(ltpk^{\mathsf{DH}}_{U^*}, sspk^{\mathsf{DH}}_{V^*})$ in the DH_1 position, $\mathcal{B}_3$ provides a perfect simulation of game $\mathcal{G}_{\text{A.4}}$. If $\mathcal{A}$ makes such a random oracle query, then $\mathcal{B}_3$ will detect this and win its GapDH game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\text{A.4}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{GDH}} + \mathsf{Adv}^{\mathcal{G}_{\text{A.5}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.6 (Replacing the session key). We now replace the session key of the test session π^* with a uniformly sampled key. Since Game $\mathcal{G}_{\text{A.5}}$ has ruled out that the adversary queries the RO on the master secret $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss$ of the test session π^*, the adversary has no chance of detecting this change. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\text{A.5}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_{\text{A.6}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

[15] If PQXDH would include the session context (e.g., the values in the session identifier) in the key derivation, as is good practice, the reduction would need less DDH oracle queries: Using the context, the reduction can detect when a challenge DH key is combined with a maliciously generated (ephemeral) key. In the modified $\textsc{Send}$ oracle in Figure 11, only the first PoS in line 24 remains a pair of public keys, while the other three occurrences are solved via context. In consequence, at maximum PoS per entry in L contains two public keys and we get a maximum of $q_{\text{RO}} + q_{\text{Snd}}$ DDH queries.

To conclude this proof case, observe that in Game $\mathcal{G}_{\mathrm{A.6}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. Furthermore, $\mathcal{A}$ cannot reveal the session key of the test session π^* via a RevealSessKey query on π^* or any partnered session which might hold the same key. Finally, any non-partnered session will derive a different session key: any difference in the identities A, B yields different DH public keys by Game $\mathcal{G}_1$, any difference in the DH shares implies a difference in the $\mathsf{DH}[1]$–$\mathsf{DH}[4]$ inputs, while differing KEM public keys or ciphertexts yield a different ss input by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.6}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq 0.$$

Case B ($\mathsf{clean}_{\mathsf{ELT}}(\pi^*)$).

In this proof case, the $\mathsf{clean}_{\mathsf{ELT}}$ predicate ensures for the test session π^* that

1. the initiator session's randomness is not revealed, and
2. the responder's long-term key is uncompromised, and
3. if π^* is a responder, then there exists a partnered initiator session.

Via the last point, we can guarantee that the responder uses an ephemeral key that was honestly generated by the initiator. Similar to the classical Signal proof [CCD+17], the uncompromised ephemeral/long-term Diffie–Hellman combination ensures key indistinguishability for the test session.

Game B.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}_{\mathsf{ELT}}(\pi^*)$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean}_{\mathsf{ELT}}(\pi^*)]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.1 (Guess initiator session). We guess the initiator session π^*_{i} contributing to the test session π^* (i.e., either the test session itself if it is an initiator session, or the initiator session partnered to the test session), overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{B.1}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.2 (Guess responder identity V^*). We guess the identity V^* of the responder involved in the test session, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of users n_p:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.1}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_p \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{B.2}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.3 (GapDH). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the adversary queries the random oracle on a value formatted like a master secret and beginning with $\mathsf{CDH}(epk^{\mathsf{DH}}, ltpk^{\mathsf{DH}}_{V^*})$, i.e., the shared DH key between the ephemeral key of the test session's initiator epk^{DH} and $ltpk^{\mathsf{DH}}_{V^*}$. The probability of such an abort can be bounded by the advantage of the following reduction $\mathcal{B}_4$ against the $(t', \epsilon_{\mathsf{GDH}}, q_{\mathrm{Snd}} + q_{\mathrm{RO}})$-hardness of the GapDH problem in $(\mathbb{G}, g, q)$.

The reduction follows the idea of Game $\mathcal{G}_{A.5}$, only that now we embed the GapDH challenge keys as the ephemeral key in the initiator session π^*_{i} and as the long-term key of V^*. The reduction does not need to answer corresponding RevealRand queries on π^*_{i} or CorruptLTKey queries on V^* due to $\mathsf{clean}_{\mathsf{ELT}}$. It can use a similar strategy to the one in Game $\mathcal{G}_{A.5}$ to patch the random oracle, yet it has an easier time ensuring consistency of the Send oracle: It can always compute DH_1 and DH_3 by using the responder's semi-static secret key, and DH_4 (if it is a full handshake) with the honest party's ephemeral secret key. Outside of the test session $\mathcal{B}_4$ can use the initiator's ephemeral secret key to compute DH_2 if the ephemeral key was honestly generated. Hence, $\mathcal{B}_4$ can compute all inputs to the KDF for all Send queries itself, except for the Send queries to V^* with a malicious ephemeral key on the initiator side. In consequence, $\mathcal{B}_4$ queries the DDH oracle at most once per Send query and once when checking RO queries for the DH_2 position. All in all, $\mathcal{B}_4$ needs a maximum of $q_{\mathrm{Snd}} + q_{\mathrm{RO}}$ queries to the DDH oracle, which is well below the number in Game $\mathcal{G}_{A.5}$.

Unless $\mathcal{A}$ queries the random oracle on $\mathsf{CDH}(epk^{\mathsf{DH}}, ltpk^{\mathsf{DH}}_{V^*})$ in the DH_2 position, $\mathcal{B}_4$ provides a perfect simulation of $\mathcal{G}_{B.2}$. If $\mathcal{A}$ does make such a random oracle query, then $\mathcal{B}_4$ will detect this and win its GapDH game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{B.2}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{GDH}} + \mathsf{Adv}^{\mathcal{G}_{B.3}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.4 (Replacing the session key). We now replace the session key of the test session π^* with a uniformly sampled key. Since Game $\mathcal{G}_{B.3}$ has ruled out that the adversary queries the RO on the master secret $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss$ of the test session π^*, the adversary has no chance of detecting this change. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{B.3}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_{B.4}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{B.4}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}^{\mathcal{G}_{B.4}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq 0.$$

Case C ($\mathsf{clean}_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{full}$).

In this proof case, the test session performs a full handshake and the $\mathsf{clean}_{\mathsf{ESS}}$ predicate ensures for the test session π^* that

1. the initiator session's randomness is not revealed, and
2. the responder's semi-static key is uncompromised, and
3. if π^* is an initiator, then the responder's long-term (signing) key was uncompromised upon acceptance, and
4. if π^* is a responder, then there exists a partnered initiator session.

Via the second last point, similarly to Case A, we can guarantee that the initiator uses a semi-static key that was honestly generated by the responder, given signatures are unforgeable. Via the last point, similarly to Case B, we can guarantee that the responder uses an ephemeral key that was honestly generated by the initiator. Then, similar to the classical Signal proof [CCD+17], the uncompromised ephemeral/semi-static Diffie–Hellman combination ensures key indistinguishability for the test session.

Game C.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{full}$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{C.0}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean}_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full}]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game C.1 (Guess responder identity V^*). We first guess the identity V^* of the responder to the test session, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of users n_p:

$$\mathsf{Adv}^{\mathcal{G}_{C.0}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_p \cdot \mathsf{Adv}^{\mathcal{G}_{C.1}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game C.2 (Signature unforgeability). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the test session π^* is an initiator session and accepts using a semi-static DH public key $sspk^{\mathsf{DH}}_{V^*}$ that was not generated through a KGenSS run for V^*. Note that DH public keys and KEM public keys are distinct by assumption and cannot be confused with each other. Hence, the test session accepts with $\pi^*.\mathsf{sspks} = (sspk^{\mathsf{DH}}_{V^*}, \cdot)$ corresponding to a DH key pair of which the adversary does not know the secret key, since $\mathsf{clean}_{\mathsf{ESS}}(\pi^*)$ guarantees that the semi-static secret belong to $sspk^{\mathsf{DH}}_{V^*}$ is not compromised. Similar to Game $\mathcal{G}_{A.2}$, the probability of such an abort can be bounded by the advantage of a reduction against the $(t', \epsilon_{\mathsf{SIG}}, 2n_{ss}+n_s)$-unforgeability of SIG. The argument is identical to the Game $\mathcal{G}_{A.2}$. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{C.1}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{SIG}} + \mathsf{Adv}^{\mathcal{G}_{C.2}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game C.3 (Guess semi-static DH key identifier $\mathsf{ssid}^*_{\mathsf{DH}}$ of V^*). We now guess the identifier ssid of the responder V^*'s (uncorrupted) semi-static DH key $sspk^{\mathsf{DH}}_{V^*}$. Note that depending on the role of π^* this is either the test session's own key (if $\pi^*.\mathsf{role} = \mathsf{responder}$), or of the intended peer (if $\pi^*.\mathsf{role} = \mathsf{initiator}$). We denote the guessed identifier by $\mathsf{ssid}^*_{\mathsf{DH}}$, and abort, setting the adversary's output bit to 0, if this guess is incorrect, losing at most a factor n_{ss} of the number of semi-static keys per user:

$$\mathsf{Adv}^{\mathcal{G}_{C.2}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_{ss} \cdot \mathsf{Adv}^{\mathcal{G}_{C.3}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game C.4 (Guess initiator session). We guess the initiator session π^*_{i} involved in the test session π^* (i.e., either the test session itself if it is an initiator session, or the initiator session partnered to the test session), overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s:

$$\mathsf{Adv}^{\mathcal{G}_{C.3}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s \cdot \mathsf{Adv}^{\mathcal{G}_{C.4}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game C.5 (GapDH). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the adversary queries the random oracle on a value formatted like a master secret, beginning with $\mathsf{CDH}(epk^{\mathsf{DH}}, sspk^{\mathsf{DH}}_{V^*})$, i.e., the shared DH key between the ephemeral key of the test session's initiator epk^{DH} and $sspk^{\mathsf{DH}}_{V^*}$, where $sspk^{\mathsf{DH}}_{V^*}$ is the semi-static DH key with identifier $\mathsf{ssid}^*_{\mathsf{DH}}$. The probability of such an abort can be bounded by the advantage of the following reduction $\mathcal{B}_5$ against the $(t', \epsilon_{\mathsf{GDH}}, q_{\mathrm{Snd}} + q_{\mathrm{RO}})$-hardness of the GapDH problem in $(\mathbb{G}, g, q)$.

The proof follows the idea of Games $\mathcal{G}_{A.5}$ and $\mathcal{G}_{B.3}$. Though, now we embed the GapDH challenge keys as the ephemeral key in the initiator session π^*_{i} and as semi-static key of V^* with semi-static id $\mathsf{ssid}^*_{\mathsf{DH}}$. The reduction does not need to answer corresponding RevealRand queries on π^*_{i} or CorruptSSKey$(V^*, \mathsf{ssid}^*_{\mathsf{DH}})$ queries due to $\mathsf{clean}_{\mathsf{ESS}}$. It can use a similar strategy to patch the random oracle: It can always compute DH_1 and DH_2 by using the initiator's and responder's long-term secret key, respectively, and DH_4 with the honest party's ephemeral secret key. Outside of the test session $\mathcal{B}_5$ can use the initiator's ephemeral secret key to compute DH_3 if the ephemeral key was honestly generated. Hence, $\mathcal{B}_5$ can compute all inputs to the KDF for all Send queries itself, except for the Send queries to the test session or to V^* with $\mathsf{ssid}^*_{\mathsf{DH}}$ and a malicious ephemeral key on the initiator side. In consequence, $\mathcal{B}_5$ queries the DDH oracle at most once per Send query and once when checking RO queries for the DH_3 position. All in all, $\mathcal{B}_5$ needs a maximum of $q_{\mathrm{Snd}} + q_{\mathrm{RO}}$ queries to the DDH oracle, which is well below the number in Game $\mathcal{G}_{\mathrm{A.5}}$.

Unless $\mathcal{A}$ queries the random oracle on $\mathsf{CDH}(epk^{\mathsf{DH}}, ltpk^{\mathsf{DH}}_{V^*})$ in the DH_2 position, $\mathcal{B}_5$ provides a perfect simulation of $\mathcal{G}_{\mathrm{C.4}}$. If $\mathcal{A}$ does make such a random oracle query, then $\mathcal{B}_5$ will detect this and win its GapDH game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{C.4}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{GDH}} + \mathsf{Adv}^{\mathcal{G}_{\mathrm{C.5}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game C.6 (Replacing the session key). We now replace the session key of the test session π^* with a uniformly sampled key. Since Game $\mathcal{G}_{\mathrm{C.5}}$ has ruled out that the adversary queries the RO on the master secret $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss$ of the test session π^*, the adversary has no chance of detecting this change. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{C.5}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_{\mathrm{C.6}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{\mathrm{C.6}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{C.6}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq 0.$$

Case D ($\mathsf{clean}_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{reduced}$).

In this proof case, the test session performs a reduced handshake and the $\mathsf{clean}_{\mathsf{ESS}}$ predicate ensures for the test session π^* that

1. the initiator session's randomness is not revealed, and
2. the responder's semi-static key is uncompromised, and
3. if π^* is an initiator, the responder's long-term (signing) key was uncompromised upon acceptance
4. and the received semi-static KEM public key was honestly generated as semi-static key (ensured via the dashed-box trivial attack check), and
5. if π^* is a responder, then there exists a partnered initiator session.

Via points 3 and 4, we can guarantee that the initiator uses a semi-static key that was honestly generated by the responder, given signatures are unforgeable. Via the last point, we can guarantee that the responder uses an ephemeral key that was honestly generated by the initiator. The difference to Case C is that, in a reduced handshake, the semi-static KEM public key of the responder is used (and there is no ephemeral/ephemeral Diffie–Hellman combination). This means we get the following *hybrid* guarantee: both the ephemeral/semi-static Diffie–Hellman combination being uncompromised *and* the semi-static

KEM key as well as the ephemeral encapsulation randomness being uncompromised are, on their own, sufficient to ensure key indistinguishability for the test session.

Game D.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{reduced}$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean}_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{reduced}]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game D.1 (Guess responder identity V^*). We first guess the identity V^* of the responder to the test session, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of users n_p:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_p \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{D.1}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game D.2 (Signature unforgeability). In this game case, we are interested in the authenticity of *both* semi-static DH and KEM keys, and hence now abort the game (again, returning 0 as the adversary's bit guess) in the event that the test session π^* is an initiator session and accepts using a semi-static DH public key $sspk^{\mathsf{DH}}_{V^*}$ *or* a semi-static KEM public key $sspk^{\mathsf{KEM}}_{V^*}$ that was not generated through a KGenSS run for V^*. Since for reduced handshakes both public keys are recorded in $\pi^*.\mathsf{sspks} = (sspk^{\mathsf{DH}}_{V^*}, sspk^{\mathsf{KEM}}_{V^*})$, this ensures that the test session accepts with semi-static DH and KEM public keys of which the adversary does not know the secret, since $\mathsf{clean}_{\mathsf{ESS}}(\pi^*)$ guarantees that the semi-static secrets corresponding to $sspk^{\mathsf{DH}}_{V^*}$ or to $sspk^{\mathsf{KEM}}_{V^*}$ are not compromised.

Again similar to Game $\mathcal{G}_{\mathrm{A.2}}$, we can reduce this hop to the $(t', \epsilon_{\mathsf{SIG}}, 2n_{ss} + n_s)$-unforgeability of SIG. Note that DH public keys and KEM public keys have disjoint representations by assumption and cannot be confused with each other. In addition, the dashed-box trivial attack check (Figure 9, line 60) ensures that the KEM public key must be an honestly generated semi-static key from $\overrightarrow{sspk}$ and hence cannot be confused with a signed ephemeral KEM public key (note that keys do not collide by Game $\mathcal{G}_1$).[16]

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.1}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{SIG}} + \mathsf{Adv}^{\mathcal{G}_{\mathrm{D.2}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game D.3 (Guess semi-static key identifiers $\mathsf{ssid}^*_{\mathsf{DH}}$, $\mathsf{ssid}^*_{\mathsf{KEM}}$ of V^*). We now guess the ssid identifiers of the responder V^*'s (uncorrupted) semi-static DH and KEM keys, $sspk^{\mathsf{DH}}_{V^*}$ resp. $sspk^{\mathsf{KEM}}_{V^*}$. Note that depending on the role of π^* this is either the test session's own key (if $\pi^*.\mathsf{role} = \mathsf{responder}$), or of the intended peer (if $\pi^*.\mathsf{role} = \mathsf{initiator}$). We denote the guessed identifiers by $\mathsf{ssid}^*_{\mathsf{DH}}$ (for the DH key) and $\mathsf{ssid}^*_{\mathsf{KEM}}$ (for the KEM key), and abort, setting the adversary's output bit to 0, if this guess is incorrect, losing at most a factor n_{ss} of the number of semi-static keys of each type per user, for each of the two guesses:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.2}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_{ss}^2 \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{D.3}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game D.4 (Guess initiator session). We guess the initiator session π^*_{i} involved in the test session π^* (i.e., either the test session itself if it is an initiator session, or the initiator session partnered to the test session), overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.3}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{D.4}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

[16] If signatures in PQXDH were properly domain-separated, this trivial key confusion attack against PQXDH would be avoided, enabling a proof in a stronger model not forbidding the trivial key confusion attack; cf. the discussion in Section 4.3.

Game D.5 (GapDH + OW-CCA). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the adversary queries the random oracle on a value formatted like a master secret, beginning with $\mathsf{CDH}(epk^{\mathsf{DH}}, sspk^{\mathsf{DH}}_{V^*})$, i.e., the shared DH key between the ephemeral key of the test session's initiator epk^{DH} and $sspk^{\mathsf{DH}}_{V^*}$, and ending with the KEM shared secret ss^* resulting from π^*_{i} encapsulating against $sspk^{\mathsf{KEM}}_{V^*}$, where $sspk^{\mathsf{DH}}_{V^*}$ and $sspk^{\mathsf{KEM}}_{V^*}$ are the semi-static DH and KEM keys with identifiers $\mathsf{ssid}^*_{\mathsf{DH}}$ and $\mathsf{ssid}^*_{\mathsf{KEM}}$, respectively. The probability of such an abort can be bounded by the *minimum* of the advantages of the following reduction $\mathcal{B}_6$ playing *simultaneously* against the $(t', \epsilon_{\mathsf{GDH}}, q_{\mathrm{Snd}} + q_{\mathrm{RO}})$-hardness of the GapDH problem in $(\mathbb{G}, g, q)$ and $(t', \epsilon_{\mathsf{CCA}}, n_s)$–$\mathsf{OW\text{-}CCA}$ security of KEM.

The reduction embeds the GapDH challenge as the ephemeral DH share in the initiator session π^*_{i} and as semi-static DH key of V^* with semi-static id $\mathsf{ssid}^*_{\mathsf{DH}}$ (as in Game $\mathcal{G}_{\mathrm{C.5}}$), and the KEM public key pk_{KEM} from the OW-CCA game as the semi-static key $sspk^{\mathsf{KEM}}_{V^*}$ of V^* with semi-static id $\mathsf{ssid}^*_{\mathsf{KEM}}$. Furthermore, it uses the KEM challenge ciphertext ct^* in π^*_{i}. Similar to Game $\mathcal{G}_{\mathrm{C.5}}$, the reduction does not need to answer corresponding RevealRand on π^*_{i} or CorruptSSKey$(V^*, \mathsf{ssid}^*_{\mathsf{DH}}/\mathsf{ssid}^*_{\mathsf{KEM}})$ queries due to $\mathsf{clean}_{\mathsf{ESS}}$. Furthermore, as for Game $\mathcal{G}_{\mathrm{C.5}}$, patching the Send oracle is relatively easy: Except for queries to the test session or to V^* using $sspk^{\mathsf{DH}}_{V^*}$ and a malicious ephemeral DH key on the initiator side, the reduction can compute all DH shared secrets with the DH secret keys and it can decapsulate the KEM shared secret with the KEM secret key or, for sessions using $sspk^{\mathsf{KEM}}_{V^*}$, the Decaps oracle of the OW-CCA game. Otherwise (and in particular in the test session), the reduction uses a freshly sampled key as session key. If the DDH oracle accepts a query, then the reduction returns the DH_3 and ss^* from the corresponding RO query to its GapDH and OW-CCA games after finishing the simulation for $\mathcal{A}$.

The reduction ensures consistency with the random oracle by checking RO queries against the above mentioned patches. In consequence, the remaining queries to the DDH oracle are once per Send query and when checking RO queries for the DH_3 position. All in all, $\mathcal{B}_6$ needs a maximum of $q_{\mathrm{Snd}} + q_{\mathrm{RO}}$ queries to the DDH oracle, which is well below the number in $\mathcal{G}_{\mathrm{A.5}}$; it further makes at most n_s many Decaps query in the OW-CCA game.

Unless $\mathcal{A}$ queries the random oracle on $\mathsf{CDH}(epk^{\mathsf{DH}}, sspk^{\mathsf{DH}}_{V^*})$ in the DH_3 position *and* ss^* in the ss position, $\mathcal{B}_6$ provides a perfect simulation of $\mathcal{G}_{\mathrm{D.4}}$. If $\mathcal{A}$ does make such a random oracle query, then $\mathcal{B}_6$ will detect this and win *both* its GapDH game and its OW-CCA game. We can thus bound the game hop by the *minimum* advantage against the two games:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.4}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \min(\epsilon_{\mathsf{GDH}}, \epsilon_{\mathsf{CCA}}) + \mathsf{Adv}^{\mathcal{G}_{\mathrm{D.5}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game D.6 (Replacing the session key). We now replace the session key of the test session π^* with a uniformly sampled key. Since Game $\mathcal{G}_{\mathrm{D.5}}$ has ruled out that the adversary queries the RO on the master secret $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss$ of the test session π^*, the adversary has no chance of detecting this change. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.5}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_{\mathrm{D.6}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{\mathrm{D.6}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{D.6}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq 0.$$

Case E ($\mathsf{clean}_{\mathsf{EE}}(\pi^*) \wedge \mathsf{clean}_{\mathsf{peerE}}(\pi^*) \wedge \mathsf{type}(\pi^*) = \mathsf{full}$).

In this proof case, the test session performs a full handshake and the $\mathsf{clean}_{\mathsf{EE}}$ predicate and its $\mathsf{clean}_{\mathsf{peerE}}$ sub-predicate ensure for the test session π^* that

1. the test session's randomness is not revealed, and
2. there exists a (contributively) partnered session whose randomness is not revealed.

Via the last point, we can guarantee that initiator and responder agree on the ephemeral contributions. Similar to Case D, we get a *hybrid* guarantee here: both the ephemeral/ephemeral Diffie–Hellman combination being uncompromised *and* the ephemeral KEM key as well as the ephemeral encapsulation randomness being uncompromised are, on their own, sufficient to ensure key indistinguishability for the test session.

Game E.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}_{\mathsf{EE}}(\pi^*) \wedge \mathsf{clean}_{\mathsf{peerE}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{full}$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{E.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean}_{\mathsf{EE}}(\pi^*) \wedge \mathsf{clean}_{\mathsf{peerE}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full}]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game E.1 (Guess initiator and responder sessions). We guess the initiator and responder sessions π^*_{i} resp. π^*_{r} involved in the test session π^* (i.e., both the test session itself and its sid- resp. cid-partner), overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s squared:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{E.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s^2 \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{E.1}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game E.2 (GapDH + OW-CCA). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the adversary queries the random oracle on a value formatted like a master secret, beginning with $\mathsf{CDH}(epk^{\mathsf{DH}}_{U^*}, epk^{\mathsf{DH}}_{V^*})$, i.e., the shared DH key between the ephemeral keys $epk^{\mathsf{DH}}_{U^*}$ and $epk^{\mathsf{DH}}_{V^*}$ of the guessed initiator session π^*_{i}, resp. responder session π^*_{r}, and ending with the KEM shared secret ss^* resulting from π^*_{i} encapsulating against $epk^{\mathsf{KEM}}_{V^*}$. The probability of such an abort can be bounded by the *minimum* of the advantages of the following reduction $\mathcal{B}_7$ playing *simultaneously* against the $(t', \epsilon_{\mathsf{GDH}}, q_{\mathrm{RO}} + 1)$-hardness of the GapDH problem in $(\mathbb{G}, g, q)$ and $(t', \epsilon_{\mathsf{CCA}}, 1)$–OW-CCA security of KEM.

The reduction embeds the GapDH challenge as the ephemeral keys in the sessions π^*_{i} and π^*_{r}, and the KEM public key pk_{KEM} from the OW-CCA game as ephemeral key $epk^{\mathsf{KEM}}_{V^*}$ in π^*_{r}. It uses the KEM challenge ciphertext ct^* in π^*_{i}. Similar to Game $\mathcal{G}_{\mathrm{D.5}}$, the reduction does not need to answer corresponding RevealRand queries on π^*_{i} or π^*_{r} due to $\mathsf{clean}_{\mathsf{EE}} \wedge \mathsf{clean}_{\mathsf{peerE}}$. Furthermore, as for Game $\mathcal{G}_{\mathrm{D.5}}$, patching the Send oracle is relatively easy: Except for queries to the test session or to π^*_{r} with a modified initiator ephemeral key,[17] the reduction can compute all DH shared secrets itself. For the latter query, $\mathcal{B}_7$ uses a single call to the DDH oracle of GapDH resp. Decaps oracle of OW-CCA to simulate. In the test session, the reduction uses a freshly sampled key as session key. If the DDH oracle accepts a query, then the reduction returns the DH_3 and ss from the corresponding RO query to its GapDH and OW-CCA games after finishing the simulation for $\mathcal{A}$.

The reduction ensures consistency with the random oracle by checking RO queries against the above mentioned patches. It queries the DDH oracle possibly once for π^*_{r} and when checking RO queries for the DH_4 position, so overall a maximum of $q_{\mathrm{RO}} + 1$ queries, which is well below the number in Game $\mathcal{G}_{\mathrm{A.5}}$. It further makes at most one Decaps query in the OW-CCA game.

[17] This can happen if the test session is a responder session and the adversary mauls the initiator public key on the wire.

Unless $\mathcal{A}$ queries the random oracle on $\mathsf{CDH}(epk_{U^*}^{\mathsf{DH}}, epk_{V^*}^{\mathsf{DH}})$ in the DH_4 position and ss^* in the ss position, $\mathcal{B}_7$ provides a perfect simulation of $\mathcal{G}_{\mathrm{E.1}}$. If $\mathcal{A}$ does make such a random oracle query, then $\mathcal{B}_7$ will detect this and win *both* its GapDH game and its OW-CCA game. We can thus bound the game hop by the *minimum* advantage against the two games:

$$\mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{E.1}}}(\mathcal{A}) \leq \min(\epsilon_{\mathsf{GDH}}, \epsilon_{\mathsf{CCA}}) + \mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{E.2}}}(\mathcal{A}).$$

Game E.3 (Replacing the session key). We now replace the session key of the test session π^* with a uniformly sampled key. Since Game $\mathcal{G}_{\mathrm{E.2}}$ has ruled out that the adversary queries the RO on the master secret $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss$ of the test session π^*, the adversary has no chance of detecting this change. Thus,

$$\mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{E.2}}}(\mathcal{A}) = \mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{E.3}}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{\mathrm{E.3}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{E.3}}}(\mathcal{A}) \leq 0.$$

Case F ($\mathsf{clean}_{\mathsf{EE}}(\pi^*) \land \mathsf{clean}_{\mathsf{sigE}}(\pi^*) \land \mathsf{type}(\pi^*) = \mathsf{full}$).

In this proof case, the test session performs a full handshake and the $\mathsf{clean}_{\mathsf{EE}}$ predicate and its $\mathsf{clean}_{\mathsf{sigE}}$ sub-predicate ensure for the test session π^* that

1. π^* is an initiator, and
2. the test session's randomness is not revealed, and
3. the responder's long-term (signing) key was uncompromised upon acceptance of π^*, and
4. a potential responder partner session's randomness was not revealed, and
5. the received ephemeral KEM public key was not replaced with a semi-static one (ensured via the dashed-box trivial attack check).

Via points 3 and 5, we can guarantee that the initiator uses an ephemeral KEM key that was honestly generated by the responder, given signatures are unforgeable. The KEM encapsulation then ensures key indistinguishability for the test session.

Game F.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}_{\mathsf{EE}}(\pi^*) \land \mathsf{clean}_{\mathsf{sigE}}(\pi^*) \land \mathsf{type}(\pi^*) = \mathsf{full}$ being satisfied.

$$\mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{F.0}}}(\mathcal{A}) = \mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_3[\mathsf{clean}_{\mathsf{EE}}(\pi^*) \land \mathsf{clean}_{\mathsf{sigE}}(\pi^*) \land \mathsf{type}(\pi^*)=\mathsf{full}]}(\mathcal{A}).$$

Game F.1 (Guess responder identity V^*). We first guess the identity V^* of the responder to the test session, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of users n_p:

$$\mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{F.0}}}(\mathcal{A}) \leq n_p \cdot \mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{F.1}}}(\mathcal{A}).$$

Game F.2 (Signature unforgeability). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the initiator test session π^* accepts using an ephemeral KEM public key $epk^{\mathsf{KEM}}_{V^*}$ that was not generated by V^*. Note again that DH public keys and KEM public keys are distinct by assumption and cannot be confused with each other. Similar to Game $\mathcal{G}_{\text{D.2}}$, the dashed-box trivial attack check (Figure 9, line 52) ensures that the received ephemeral KEM public key cannot be confused with a signed semi-static KEM public key from $\overrightarrow{sspk}$ (note that keys do not collide by Game $\mathcal{G}_1$).[18] This ensures that the test session accepts with an ephemeral KEM public key of which the adversary does not know the secret key, since the adversary cannot reveal V^*'s session randomness. Similar to Game $\mathcal{G}_{\text{A.2}}$, the probability of such an abort can be bounded by the advantage of a reduction against the $(t', \epsilon_{\mathsf{SIG}}, 2n_{ss}+n_s)$-unforgeability of SIG. The proof is identical to the one for Game $\mathcal{G}_{\text{A.2}}$. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\text{F.1}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{SIG}} + \mathsf{Adv}^{\mathcal{G}_{\text{F.2}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game F.3 (Guess initiator and responder sessions). We guess the initiator test session and the partnered responder sessions π^* resp. π^*_{r}, overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s squared:

$$\mathsf{Adv}^{\mathcal{G}_{\text{F.2}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s^2 \cdot \mathsf{Adv}^{\mathcal{G}_{\text{F.3}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game F.4 (OW-CCA). We now abort the game (again, returning 0 as the adversary's bit guess) in the event that the adversary queries the random oracle on a value formatted like a master secret and ending with the KEM shared secret ss^* resulting from π^* encapsulating against $epk^{\mathsf{KEM}}_{V^*}$. The probability of such an abort can be bounded by the advantage of the following reduction $\mathcal{B}_8$ playing against $(t', \epsilon_{\mathsf{CCA}}, 1)$–OW-CCA security of KEM.

The reduction embeds the challenge pk_{KEM}, ct^* from the OW-CCA game as ephemeral key $epk^{\mathsf{KEM}}_{V^*}$ of V^*, and as KEM ciphertext in π^*. The reduction does not need to answer a RevealRand query on π^*_{r} due to $\mathsf{clean}_{\mathsf{EE}}$. If π^*_{r} receives a different ciphertext than ct^*, the reduction uses its Decaps oracle (once) to compute the required shared secret. When $\mathcal{A}$ halts, the reduction guesses $i \in [q_{\text{RO}}]$ and returns the ss position of the ith RO query (assuming this query is formatted like a master secret) as the target shared secret.

Unless $\mathcal{A}$ queries the random oracle on ss^* in the ss position, $\mathcal{B}_8$ provides a perfect simulation of Game $\mathcal{G}_{\text{F.3}}$. If $\mathcal{A}$ does make such a random oracle query, then $\mathcal{B}_8$ wins its OW-CCA game if it guesses the ith RO query correctly. We can thus bound the game hop by the advantage against OW-CCA times a loss of q_{RO}:

$$\mathsf{Adv}^{\mathcal{G}_{\text{F.3}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq q_{\text{RO}} \cdot \epsilon_{\mathsf{CCA}} + \mathsf{Adv}^{\mathcal{G}_{\text{F.4}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game F.5 (Replacing the session key). We now replace the session key of the test session π^* with a uniformly sampled key. Since Game $\mathcal{G}_{\text{F.4}}$ has ruled out that the adversary queries the RO on the master secret $DH_1 \| DH_2 \| DH_3 \| DH_4 \| ss$ of the test session π^*, the adversary has no chance of detecting this change. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\text{F.4}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_{\text{F.5}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{\text{F.5}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test

[18] Again, this trivial key confusion attack against PQXDH would be avoided if its signatures were properly domain-separated; cf. the discussion in Section 4.3.

session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}_{\mathsf{PQXDH}}^{\mathcal{G}_{\mathrm{F.5}}}(\mathcal{A}) \leq 0.$$

Collecting the bounds from all cases yields the overall theorem bound. □

5.2 Post-quantum Security of PQXDH

Next, we show that PQXDH achieves key indistinguishability against quantum adversaries wrt. the maximum-exposure attack vectors formalized formalized through the clean predicates $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ in our model (cf. Figure 9). In particular, these clean predicates model authentic distribution of the KEM keys (e.g., due to the adversary being passive or *not yet* able to forge signatures) and do not rely on the contribution of the DH shared secrets. This result is in the standard model and relies on the KEM providing IND-CCA and $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ security, as well as low key-collision probability and correctness errors, and the KDF providing PRF security. The proof follows the same structure as the previous proof, but is shorter following the fewer cases of the $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ predicate in which one can expect security against a quantum adversary.

Theorem 5.2 (Post-quantum key indistinguishability of PQXDH)**.** *The* PQXDH *protocol given in Figure 10 with randomness space $\mathcal{R}_{\mathsf{KE}} = \mathcal{R}_{\mathsf{DH.KGen}} \times \mathcal{R}_{\mathsf{KEM.KGen}} \times \mathcal{R}_{\mathsf{KEM.Enc}} \times \mathcal{R}_{\mathsf{SIG.Sig}}$ achieves $(t, \epsilon, (q_{\mathrm{Snd}}, q_{\mathrm{CorrLT}}, q_{\mathrm{CorrSS}}, q_{\mathrm{RevR}}, q_{\mathrm{RevSK}}))$–key indistinguishability wrt. the post-quantum security clean predicates $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ (cf. Figure 9), assuming (for $t' \approx t$)* KEM *is a $(t', \epsilon_{\mathsf{CCA}}, n_s)$–*IND-CCA*-secure and $(t', \epsilon_{LEAK^{+r}}, n_p \cdot n_{ss} + n_s)$–$LEAK^{+r}$-*BIND-*$SS$-$\{CT, PK\}$-secure KEM with public-key collision probability γ_{coll} and correctness error δ_{corr}, and* KDF *is $(t', \epsilon_{\mathsf{PRF}}, n_s)$–PRF-secure when keyed via the last 256 input bits. Concretely, for any efficient quantum adversary $\mathcal{A}$ we have*

$$\begin{aligned}\mathsf{Adv}_{\mathsf{PQXDH}\text{-}Q}^{\mathsf{KI}}(\mathcal{A}) \leq{}& \frac{(n_p + n_p \cdot n_{ss} + n_s)^2}{q} + \gamma_{\mathsf{coll}}(n_p \cdot n_{ss} + n_s) + n_s \cdot \delta_{\mathsf{corr}} + \epsilon_{LEAK^{+r}} \\ &+ n_{ss} \cdot n_s \cdot (\epsilon_{\mathsf{CCA}} + \epsilon_{\mathsf{PRF}}) \quad // \, \mathsf{clean}_{\mathsf{ESS}}^{Q} \\ &+ n_s^2 \cdot (\epsilon_{\mathsf{CCA}} + \epsilon_{\mathsf{PRF}}), \quad // \, \mathsf{clean}_{\mathsf{EE}}^{Q}\end{aligned}$$

where we write PQXDH-Q *to indicate that we assess the post-quantum security of* PQXDH *wrt. to the* $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ *clean predicates.*

Proof. The proof against quantum adversaries is structurally similar to the proof against classic adversaries. We first exclude collisions in the public keys and KEM decapsulation errors. Now, soundness holds. Next, we preclude the KEM re-encapsulation attack with a $LEAK^{+r}$ game hop. Again, we branch out over the cases of the clean predicate – now in the quantum version with only two cases. In either case we guess the involved sessions and (ephemeral or semi-static) KEM public keys, then embed an IND-CCA challenge in the test session's KEM encapsulation, and conclude with a PRF security hop from the now random KEM shared secret. Note that all of these game hops rely on standard-model assumptions, i.e., we do not need to rely on a (quantum) random oracle.

Game 0. We start with Game $\mathcal{G}_0$ being the original key indistinguishability game $\mathcal{G}_{\mathsf{PQXDH}\text{-}Q}^{\mathsf{KI}}(\mathcal{A})$, i.e.,

$$\mathsf{Adv}_{\mathsf{PQXDH}\text{-}Q}^{\mathsf{KI}}(\mathcal{A}) = \mathsf{Adv}_{\mathsf{PQXDH}\text{-}Q}^{\mathcal{G}_0}(\mathcal{A}).$$

We proceed with Games 1, 2, and 3 as in the proof of Theorem 5.1, excluding collisions in DH and KEM public keys, KEM decapsulation errors, and collisions of KEM shared secrets. Similarly to the proof of Theorem 5.1, these game hops ensure soundness. With these three game hops we have

$$\mathsf{Adv}^{\mathcal{G}_0}_{\mathsf{PQXDH}\text{-}Q}(\mathcal{A}) \le \frac{(n_p + n_p \cdot n_{ss} + n_s)^2}{q} + \gamma_{\mathsf{coll}}(n_p \cdot n_{ss} + n_s) + n_s \cdot \delta_{\mathsf{corr}} + \epsilon_{LEAK^{+r}} + \mathsf{Adv}^{\mathcal{G}_3}_{\mathsf{PQXDH}\text{-}Q}(\mathcal{A}).$$

Separating the clean^Q cases. At this point, we will divide the proof into two sub-cases following the $\mathsf{clean}^{\mathsf{PQXDH}\text{-}Q}$ predicate with its sub-predicates $\mathsf{clean}^Q_{\mathsf{EE}}$ for full sessions and $\mathsf{clean}^Q_{\mathsf{ESS}}$ for $\mathsf{reduced}$ sessions. We can bound the advantage of the adversary in Game $\mathcal{G}_3$ by the sum of its advantages in Games $\mathcal{G}_3[\mathsf{c}]$ which are $\mathcal{G}_3$ conditioned on a sub-predicate c:

$$\mathsf{Adv}^{\mathcal{G}_3}_{\mathsf{PQXDH}\text{-}Q}(\mathcal{A}) \le \sum_{\mathsf{c} \in \left\{ \substack{\mathsf{clean}^Q_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{reduced}, \\ \mathsf{clean}^Q_{\mathsf{EE}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full}} \right\}} \mathsf{Adv}^{\mathcal{G}_3[\mathsf{c}]}_{\mathsf{PQXDH}\text{-}Q}(\mathcal{A}).$$

In the remainder of the proof, we will bound both of these cases, called A and B, separately.

Case A ($\mathsf{clean}^Q_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{reduced}$).

In this proof case, the test session performs a reduced handshake and the $\mathsf{clean}^Q_{\mathsf{ESS}}$ predicate ensures for the test session π^* that

1. the initiator session's randomness is not revealed, and
2. the responder's semi-static KEM key is uncompromised, and
3. if π^* is an initiator, then the semi-static KEM key was honestly generated, and
4. if π^* is a responder, then there exists a partnered initiator session.

Hence, the semi-static KEM shared secret ensures key indistinguishability for the test session.

Game A.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}^Q_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{reduced}$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{A.0}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean}^Q_{\mathsf{ESS}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{reduced}]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.1 (Guess semi-static KEM key identifier $\mathsf{ssid}^*_{\mathsf{KEM}}$ of V^*). We now guess the ssid identifier of the responder V^*'s (uncorrupted) semi-static KEM key, $sspk^{\mathsf{KEM}}_{V^*}$. Note that depending on the role of π^* this is either the test session's own key (if $\pi^*.\mathsf{role} = \mathsf{responder}$), or of the intended peer (if $\pi^*.\mathsf{role} = \mathsf{initiator}$). In the latter case, the highlighted code of $\mathsf{clean}^Q_{\mathsf{ESS}}$ (Figure 9, line 66) ensures that $sspk^{\mathsf{KEM}}_{V^*}$ is indeed an honestly generated key. We denote the guessed identifier by $\mathsf{ssid}^*_{\mathsf{KEM}}$ and abort, setting the adversary's output bit to 0, if this guess is incorrect, losing at most a factor n_{ss} of the number of semi-static keys of each type per user:

$$\mathsf{Adv}^{\mathcal{G}_{A.0}}_{\mathsf{PQXDH}}(\mathcal{A}) \le n_{ss} \cdot \mathsf{Adv}^{\mathcal{G}_{A.1}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.2 (Guess initiator session). We guess the initiator session π^*_{i} involved in the test session π^* (i.e., either the test session itself if it is an initiator session, or the initiator session partnered to the test session), overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.1}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq n_s \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{A.2}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.3 (IND-CCA). We replace the KEM shared secret ss with a uniformly random value ss^* in the initiator session π^*_{i} and test session π^*, as well as any session of V^* using the same semi-static KEM key $sspk^{\mathsf{KEM}}_{V^*}$ and receiving the same ciphertext. The probability of the adversary detecting this change can be bounded by the the advantage of the following reduction $\mathcal{B}_1$ playing against the $(t', \epsilon_{\mathsf{CCA}}, n_s)$–$\mathsf{IND\text{-}CCA}$ security of KEM.

The reduction embeds the KEM public key pk_{KEM} from the IND-CCA game as the semi-static key $sspk^{\mathsf{KEM}}_{V^*}$ of V^* with semi-static id $\mathsf{ssid}^*_{\mathsf{KEM}}$. Furthermore, it uses the KEM challenge ct^*, ss^*_b in π^*_{i} and ss^*_b in π^*. The reduction does not need to answer corresponding RevealRand queries on π^*_{i} or CorruptSSKey$(V^*, \mathsf{ssid}^*_{\mathsf{KEM}})$ queries due to $\mathsf{clean}^{Q}_{\mathsf{ESS}}$. For Send queries to further sessions of V^* with $\mathsf{ssid}^*_{\mathsf{KEM}}$, the reduction can use ss^*_b as decapsulation of ct^* and queries the Decaps oracle of the IND-CCA game otherwise. Hence, the reduction makes at most n_s many Decaps queries in the IND-CCA game.

If the IND-CCA game provides the real KEM shared secret as challenge, then $\mathcal{B}_1$ perfectly simulates $\mathcal{G}_{\mathrm{A.2}}$. Otherwise, the IND-CCA challenge is a randomly sampled KEM shared secret and $\mathcal{B}_1$ perfectly simulates $\mathcal{G}_{\mathrm{A.3}}$. If $\mathcal{A}$ can distinguish these two games, then $\mathcal{B}_1$ can win its IND-CCA game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.2}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{CCA}} + \mathsf{Adv}^{\mathcal{G}_{\mathrm{A.3}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game A.4 (KDF pseudorandomness). We now replace KDF, when keyed with ss^* as per $\mathcal{G}_{\mathrm{A.3}}$ through the last 256 input bits, with a random function. This in particular effectively replaces the session key of the test session π^* with a uniformly sampled key. We bound the advantage difference of this change with a reduction to $(t', \epsilon_{\mathsf{PRF}}, n_s)$–PRF security of KDF, where KDF is keyed in its last 256-bit input ss and treats the remaining inputs DH_1, DH_2, DH_3, DH_4 as label.

The reduction $\mathcal{B}_2$ does not sample ss^* itself but instead queries its PRF oracle in the sessions where it would need to compute KDF on input ss^*; in particular in the test session π^*.

Depending on the secret bit in the PRF game, the reduction perfectly simulates $\mathcal{G}_{\mathrm{A.3}}$ or $\mathcal{G}_{\mathrm{A.4}}$. If $\mathcal{A}$ can distinguish these two games, then $\mathcal{B}_2$ can win its pseudorandomness game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.3}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq \epsilon_{\mathsf{PRF}} + \mathsf{Adv}^{\mathcal{G}_{\mathrm{A.4}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{\mathrm{A.4}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{A.4}}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq 0.$$

Case B ($\mathsf{clean}^{Q}_{\mathsf{EE}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{full}$).

In this proof case, the test session performs a full handshake and the $\mathsf{clean}^{Q}_{\mathsf{EE}}$ predicate ensures for the test session π^* that

1. the test session's randomness is not revealed, and

2. there exists a (contributively) partnered session whose randomness is not revealed.

Via the last point, we can guarantee that initiator and responder agree on the ephemeral KEM key. Neither the ephemeral KEM key nor the KEM encapsulation randomness is compromised. Hence, the ephemeral KEM shared secret ensures key indistinguishability for the test session.

Game B.0. This case begins with Game $\mathcal{G}_3$ conditioned on $\mathsf{clean}^{Q}_{\mathsf{EE}}(\pi^*) \wedge \pi^*.\mathsf{type} = \mathsf{full}$ being satisfied.

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) = \mathsf{Adv}^{\mathcal{G}_3[\mathsf{clean}^{Q}_{\mathsf{EE}}(\pi^*) \wedge \pi^*.\mathsf{type}=\mathsf{full}]}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.1 (Guess initiator and responder sessions). We guess the initiator and responder sessions π^*_{i} resp. π^*_{r} involved in the test session π^* (i.e., both the test session itself and its sid- resp. cid-partner), overwriting the adversary's bit guess with 0 if this guess was incorrect. This step loses at most a factor of the number of sessions n_s squared:

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.0}}}_{\mathsf{PQXDH}}(\mathcal{A}) \le n_s^2 \cdot \mathsf{Adv}^{\mathcal{G}_{\mathrm{B.1}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.2 (IND-CCA). We replace the KEM shared secret ss with a uniformly random value ss^* in the test session π^*. Furthermore, we replace ss with ss^* in π^*_{i} if π^* is a responder, and in π^*_{r} if π^* is an initiator and π^*_{r} received the KEM ciphertext sent by π^*. The probability of the adversary detecting this change can be bounded by the the advantage of the following reduction $\mathcal{B}_3$ playing against the $(t', \epsilon_{\mathsf{CCA}}, 1)$–$\mathsf{IND\text{-}CCA}$ security of KEM.

The reduction embeds the KEM public key pk_{KEM} from the IND-CCA game as the ephemeral KEM key $epk^{\mathsf{KEM}}_{V^*}$ of V^* in the responder test session π^*_{r}. Furthermore, it uses the KEM challenge ct^*, ss^*_b in π^*_{i} and ss^*_b in π^*. The reduction does not need to answer corresponding RevealRand queries on π^*_{i} or π^*_{r} due to $\mathsf{clean}^{Q}_{\mathsf{EE}}$. If π^*_{r} receives ct^*, the reduction uses ss^* there as well; otherwise, the reduction uses its Decaps oracle (once) to compute the required shared secret. Hence, the reduction makes at most 1 Decaps query in the IND-CCA game.

If the IND-CCA game provides the real KEM shared secret as challenge, then $\mathcal{B}_3$ perfectly simulates $\mathcal{G}_{\mathrm{B.1}}$. Otherwise, the IND-CCA challenge is a randomly sampled KEM shared secret and $\mathcal{B}_3$ perfectly simulates $\mathcal{G}_{\mathrm{B.2}}$. If $\mathcal{A}$ can distinguish these two games, then $\mathcal{B}_3$ can win its IND-CCA game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.1}}}_{\mathsf{PQXDH}}(\mathcal{A}) \le \epsilon_{\mathsf{CCA}} + \mathsf{Adv}^{\mathcal{G}_{\mathrm{B.2}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

Game B.3 (KDF pseudorandomness). We now replace KDF, when keyed with ss^* as per $\mathcal{G}_{\mathrm{B.2}}$ through the last 256 input bits, with a random function. This in particular effectively replaces the session key of the test session π^* with a uniformly sampled key. We bound the advantage difference of this change with a reduction to $(t', \epsilon_{\mathsf{PRF}}, 2)$–PRF security of KDF, where KDF is keyed in its last 256-bit input ss and treats the remaining inputs DH_1, DH_2, DH_3, DH_4 as label.

The reduction $\mathcal{B}_4$ does not sample ss^* itself but instead queries its PRF oracle in the sessions where it would need to compute KDF on input ss^*; i.e., in the test session π^* and possibly its partner session.

Depending on the secret bit in the PRF game, the reduction perfectly simulates $\mathcal{G}_{\mathrm{B.2}}$ or $\mathcal{G}_{\mathrm{B.3}}$. If $\mathcal{A}$ can distinguish these two games, then $\mathcal{B}_4$ can win its pseudorandomness game. Thus,

$$\mathsf{Adv}^{\mathcal{G}_{\mathrm{B.2}}}_{\mathsf{PQXDH}}(\mathcal{A}) \le \epsilon_{\mathsf{PRF}} + \mathsf{Adv}^{\mathcal{G}_{\mathrm{B.3}}}_{\mathsf{PQXDH}}(\mathcal{A}).$$

To conclude this proof case, observe that in Game $\mathcal{G}_{\mathrm{B.3}}$ the session key of the test session π^* is now a uniformly random key, independent of b_{test}. As before, $\mathcal{A}$ can neither reveal the session key of the test session π^* nor of any partnered session, nor does any other session derive the same session key by Game $\mathcal{G}_3$. Thus, $\mathcal{A}$ cannot gain any information about the test bit b_{test} and can do no better than to guess:

$$\mathsf{Adv}^{\mathcal{G}_{B.3}}_{\mathsf{PQXDH}}(\mathcal{A}) \leq 0.$$

Collecting the bounds from all cases yields the overall theorem bound. □

6 Conclusion and Discussion

In this work, we provided a reductionist security analyis of Signal's PQXDH in a "maximum-exposure", game-based security model, augmented compared to prior versions [CCD+17, BFG+22, BFG+21] to capture the KEM component but also key signing, and gave fully-parameterized, concrete bounds for PQXDH's classical and post-quantum security.

PQXDH design discussion. Our bound relies (among other things) on the $LEAK^{+r}$-BIND-SS-$\{CT, PK\}$ binding property of the KEM. While we show that this property is satisfied by both Kyber and ML-KEM, future KEMs deployed in PQXDH or similar protocols may not satisfy this property; so achieving PQXDH-like security without relying on a binding property is of general interest. We can indeed forgo this assumption (and the corresponding advantage term), if the key derivation of PQXDH takes the KEM public key and ciphertext (or, ideally, the whole session context as is good practice) as additional arguments. This supports and complements a proposal discussed in the tool-based formal verification of PQXDH by BJKS [BJKS24], to make sure the key agreement in the initial handshake does not rely on the follow-up AEAD encryption (which would indeed violate key indistinguishability as we prove it). As a side note, including context into the key derivation also supports tighter security proofs [CCG+19].

The lack of domain separation in PQXDH when signing ephemeral and semi-static KEM public keys allows for unnecessary, trivial key confusion attacks. Our analysis further confirms and pinpoints the stronger security guarantees achievable if PQXDH were to properly domain-separate these keys.

Our classical security reduction for PQXDH has an additional case (compared to the X3DH analysis [CCD+17]), where key indistinguishability hinges solely on the signed ephemeral KEM key. This guarantees security against an *active* adversary who later gets quantum powers or against a *passive* quantum adversary. Hence, PQXDH protects against an even stronger class of attacks than the "harvest now, decrypt later" attack which motivated the design.

Limitations and future work. Our analysis is focused on the PQXDH handshake, analyzing both (full and reduced) modes in isolation. We hence do not treat attacks across protocol modes or versions (X3DH vs. PQXDH), but note that BJKS [BJKS24] highlighted the importance of domain separating the KDF calls to avoid confusion between X3DH in full mode and PQXDH in reduced mode.

Prior works on Signal's initial handshake have assumed authentic distribution of public keys instead of modelling signatures [CCD+17, BFG+22], treated the signing keys informally [VGIK20], or, like our work, assumed that long-term signing keys and long-term DH keys are distinct [FJ24, BJKS24, CHDN+24]. We leave it to future work to more accurately reflect the implementation which re-uses the same keys for both purposes.

It remains open to find the minimal changes necessary to PQXDH that guarantee security against active quantum adversaries, i.e., showing security without assuming honest distribution of KEM keys. In Section 1 we discussed related work [HKKP22, BFG+22, DG22, CHDN+24] that proposes post-quantum *replacements* for X3DH and PQXDH. Preferably, these minimal changes would also preserve deniability of the protocol.

Acknowledgments

We thank Ehren Kret and Rolfe Schmidt for discussions on the PQXDH protocol, Franziskus Kiefer for insights on the potential KEM re-encapsulation attack on PQXDH, and the anonymous reviewers for their insightful comments. R.F. was supported by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[AHK+23] Joël Alwen, Dominik Hartmann, Eike Kiltz, Marta Mularczyk, and Peter Schwabe. Post-quantum multi-recipient public key encryption. In Weizhi Meng, Christian Damsgaard Jensen, Cas Cremers, and Engin Kirda, editors, *ACM CCS 2023: 30th Conference on Computer and Communications Security*, pages 1108–1122, Copenhagen, Denmark, November 26–30, 2023. ACM Press. (Cited on page 4.)

[BC] Bruno Blanchet and Vincent Cheval. Proverif: Cryptographic protocol verifier in the formal model. https://bblanche.gitlabpages.inria.fr/proverif/. (Cited on page 2.)

[BCD+24] Manuel Barbosa, Deirdre Connolly, João Diogo Duarte, Aaron Kaiser, Peter Schwabe, Karoline Varner, and Bas Westerbaan. X-wing. *IACR Communications in Cryptology (CiC)*, 1(1):21, 2024. (Cited on page 4.)

[BDPV08] Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. On the indifferentiability of the sponge construction. In Nigel P. Smart, editor, *Advances in Cryptology – EUROCRYPT 2008*, volume 4965 of *Lecture Notes in Computer Science*, pages 181–197, Istanbul, Turkey, April 13–17, 2008. Springer, Berlin, Heidelberg, Germany. (Cited on page 6.)

[BFG+20] Jacqueline Brendel, Marc Fischlin, Felix Günther, Christian Janson, and Douglas Stebila. Towards post-quantum security for Signal's X3DH handshake. In Orr Dunkelman, Michael J. Jacobson Jr., and Colin O'Flynn, editors, *SAC 2020: 27th Annual International Workshop on Selected Areas in Cryptography*, volume 12804 of *Lecture Notes in Computer Science*, pages 404–430, Halifax, NS, Canada (Virtual Event), October 21-23, 2020. Springer, Cham, Switzerland. (Cited on page 4.)

[BFG+21] Jacqueline Brendel, Rune Fiedler, Felix Günther, Christian Janson, and Douglas Stebila. Post-quantum asynchronous deniable key exchange and the Signal handshake. Cryptology ePrint Archive, Report 2021/769, 2021. (Cited on pages 2, 3, 17, 19, 20, 21, 22, and 46.)

[BFG+22] Jacqueline Brendel, Rune Fiedler, Felix Günther, Christian Janson, and Douglas Stebila. Post-quantum asynchronous deniable key exchange and the Signal handshake. In Goichiro Hanaoka, Junji Shikata, and Yohei Watanabe, editors, *PKC 2022: 25th International Conference on Theory and Practice of Public Key Cryptography, Part II*, volume 13178 of *Lecture Notes in Computer Science*, pages 3–34, Virtual Event, March 8–11, 2022. Springer, Cham, Switzerland. (Cited on pages 2, 3, 4, 17, 19, 20, 21, 22, and 46.)

[BJK23] Karthikean Barghavan, Charlie Jacomme, and Franziskus Kiefer. Formal analysis of the PQXDH protocol, 2023. https://github.com/Inria-Prosecco/pqxdh-analysis. (Cited on pages 2 and 3.)

[BJKS23] Karthikean Barghavan, Charlie Jacomme, Franziskus Kiefer, and Rolfe Schmidt. An analysis of Signal's PQXDH, October 2023. https://cryspen.com/post/pqxdh/. (Cited on pages 2 and 3.)

[BJKS24] Karthikeyan Bhargavan, Charlie Jacomme, Franziskus Kiefer, and Rolfe Schmidt. Formal verification of the PQXDH post-quantum key agreement protocol for end-to-end secure messaging. In Davide Balzarotti and Wenyuan Xu, editors, *33rd USENIX Security Symposium, USENIX Security 2024, Philadelphia, PA, USA, August 14-16, 2024*, Philadelphia, PA, August 2024. USENIX Association. (Cited on pages 2, 3, 4, 8, 23, and 46.)

[Bla] Bruno Blanchet. Cryptoverif: Cryptographic protocol verifier in the computational model. https://bblanche.gitlabpages.inria.fr/CryptoVerif/. (Cited on page 2.)

[BR94] 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, Santa Barbara, CA, USA, August 22–26, 1994. Springer, Berlin, Heidelberg, Germany. (Cited on page 17.)

[CCD$^+$17] Katriel Cohn-Gordon, Cas J. F. Cremers, Benjamin Dowling, Luke Garratt, and Douglas Stebila. A formal security analysis of the Signal messaging protocol. In *IEEE European Symposium on Security and Privacy, EuroS&P 2017*, pages 451–466, 2017. (Cited on pages 2, 3, 6, 17, 19, 22, 25, 26, 29, 33, 34, and 46.)

[CCG16] Katriel Cohn-Gordon, Cas J. F. Cremers, and Luke Garratt. On post-compromise security. In Michael Hicks and Boris Köpf, editors, *CSF 2016: IEEE 29th Computer Security Foundations Symposium*, pages 164–178, Lisbon, Portugal, June 27–1, 2016. IEEE Computer Society Press. (Cited on page 17.)

[CCG$^+$19] Katriel Cohn-Gordon, Cas Cremers, Kristian Gjøsteen, Håkon Jacobsen, and Tibor Jager. Highly efficient key exchange protocols with optimal tightness. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology – CRYPTO 2019, Part III*, volume 11694 of *Lecture Notes in Computer Science*, pages 767–797, Santa Barbara, CA, USA, August 18–22, 2019. Springer, Cham, Switzerland. (Cited on page 46.)

[CD23] Wouter Castryck and Thomas Decru. An efficient key recovery attack on SIDH. In Carmit Hazay and Martijn Stam, editors, *Advances in Cryptology – EUROCRYPT 2023, Part V*, volume 14008 of *Lecture Notes in Computer Science*, pages 423–447, Lyon, France, April 23–27, 2023. Springer, Cham, Switzerland. (Cited on page 4.)

[CDM24] Cas Cremers, Alexander Dax, and Niklas Medinger. Keeping up with the KEMs: Stronger security notions for kems and automated analysis of KEM-based protocols. In Bo Luo, Xiaojing Liao, Jun Xu, Engin Kirda, and David Lie, editors, *ACM CCS 2024: 31st Conference on Computer and Communications Security*, Salt Lake City, UT, USA, October 14–18 2024. ACM Press. To appear. (Cited on pages 3, 8, and 9.)

[CHDN$^+$24] Daniel Collins, Loïs Huguenin-Dumittan, Ngoc Khanh Nguyen, Nicolas Rolin, and Serge Vaudenay. K-Waay: Fast and deniable Post-Quantum X3DH without ring signatures. In Davide Balzarotti and Wenyuan Xu, editors, *33rd USENIX Security Symposium, USENIX Security 2024, Philadelphia, PA, USA, August 14-16, 2024*, Philadelphia, PA, August 2024. USENIX Association. (Cited on pages 4 and 46.)

[DFGS15] Benjamin Dowling, Marc Fischlin, Felix Günther, and Douglas Stebila. A cryptographic analysis of the TLS 1.3 handshake protocol candidates. In Indrajit Ray, Ninghui Li, and Christopher Kruegel, editors, *ACM CCS 2015: 22nd Conference on Computer and Communications Security*, pages 1197–1210, Denver, CO, USA, October 12–16, 2015. ACM Press. (Cited on page 18.)

[DG22] Samuel Dobson and Steven D. Galbraith. Post-quantum signal key agreement from SIDH. In Jung Hee Cheon and Thomas Johansson, editors, *Post-Quantum Cryptography - 13th International Workshop, PQCrypto 2022*, pages 422–450, Virtual Event, September 28–30, 2022. Springer, Cham, Switzerland. (Cited on pages 4 and 46.)

[FG14] Marc Fischlin and Felix Günther. Multi-stage key exchange and the case of Google's QUIC protocol. In Gail-Joon Ahn, Moti Yung, and Ninghui Li, editors, *ACM CCS 2014: 21st Conference on Computer and Communications Security*, pages 1193–1204, Scottsdale, AZ, USA, November 3–7, 2014. ACM Press. (Cited on page 17.)

[FJ24] Rune Fiedler and Christian Janson. A deniability analysis of Signal's initial handshake PQXDH. *Proc. Priv. Enhancing Technol.*, 2024(4):907–928, 2024. (Cited on pages 2 and 46.)

[FO99] Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In Michael J. Wiener, editor, *Advances in Cryptology – CRYPTO'99*, volume 1666 of *Lecture Notes in Computer Science*, pages 537–554, Santa Barbara, CA, USA, August 15–19, 1999. Springer, Berlin, Heidelberg, Germany. (Cited on page 11.)

[HKKP22] Keitaro Hashimoto, Shuichi Katsumata, Kris Kwiatkowski, and Thomas Prest. An efficient and generic construction for Signal's handshake (X3DH): Post-quantum, state leakage secure, and deniable. *Journal of Cryptology*, 35(3):17, July 2022. (Cited on pages 4 and 46.)

[JK18] Tibor Jager and Rafael Kurek. Short digital signatures and ID-KEMs via truncation collision resistance. In Thomas Peyrin and Steven Galbraith, editors, *Advances in Cryptology – ASIACRYPT 2018, Part II*, volume 11273 of *Lecture Notes in Computer Science*, pages 221–250, Brisbane, Queensland, Australia, December 2–6, 2018. Springer, Cham, Switzerland. (Cited on page 5.)

[JKN21] Tibor Jager, Rafael Kurek, and David Niehues. Efficient adaptively-secure IB-KEMs and VRFs via near-collision resistance. In Juan Garay, editor, *PKC 2021: 24th International Conference on Theory and Practice of Public Key Cryptography, Part I*, volume 12710 of *Lecture Notes in Computer Science*, pages 596–626, Virtual Event, May 10–13, 2021. Springer, Cham, Switzerland. (Cited on page 5.)

[KS23] Ehren Kret and Rolfe Schmidt. The PQXDH key agreement protocol, September 2023. Revision 1, `https://signal.org/docs/specifications/pqxdh/`. (Cited on pages 1 and 2.)

[KS24] Ehren Kret and Rolfe Schmidt. The PQXDH key agreement protocol, January 2024. Revision 3, `https://signal.org/docs/specifications/pqxdh/`. (Cited on pages 2, 3, and 26.)

[LHT16] A. Langley, M. Hamburg, and S. Turner. Elliptic Curves for Security. RFC 7748 (Informational), January 2016. (Cited on page 26.)

[LT11] Gaëtan Leurent and Søren S. Thomsen. Practical near-collisions on the compression function of BMW. In Antoine Joux, editor, *Fast Software Encryption – FSE 2011*, volume 6733 of *Lecture Notes in Computer Science*, pages 238–251, Lyngby, Denmark, February 13–16, 2011. Springer, Berlin, Heidelberg, Germany. (Cited on page 5.)

[MMP$^+$23] Luciano Maino, Chloe Martindale, Lorenz Panny, Giacomo Pope, and Benjamin Wesolowski. A direct key recovery attack on SIDH. In Carmit Hazay and Martijn Stam, editors, *Advances in Cryptology – EUROCRYPT 2023, Part V*, volume 14008 of *Lecture Notes in Computer Science*, pages 448–471, Lyon, France, April 23–27, 2023. Springer, Cham, Switzerland. (Cited on page 4.)

[MP16a] Moxie Marlinspike and Trevor Perrin. The Double Ratchet algorithm, November 2016. https://www.signal.org/docs/specifications/doubleratchet/. (Cited on page 17.)

[MP16b] Moxie Marlinspike and Trevor Perrin. The X3DH key agreement protocol, November 2016. https://signal.org/docs/specifications/x3dh/. (Cited on page 1.)

[MvV97] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography*. The CRC Press series on discrete mathematics and its applications. CRC Press, 2000 N.W. Corporate Blvd., Boca Raton, FL 33431-9868, USA, 1997. (Cited on page 5.)

[Nat24] National Institute of Standards and Technology. Module-lattice-based key-encapsulation mechanism standard. Technical Report Federal Information Processing Standards Publications (FIPS PUBS) 203, U.S. Department of Commerce, Washington, D.C., 2024. (Cited on pages 3, 7, 8, and 11.)

[OP01] Tatsuaki Okamoto and David Pointcheval. The gap-problems: A new class of problems for the security of cryptographic schemes. In Kwangjo Kim, editor, *PKC 2001: 4th International Workshop on Theory and Practice in Public Key Cryptography*, volume 1992 of *Lecture Notes in Computer Science*, pages 104–118, Cheju Island, South Korea, February 13–15, 2001. Springer, Berlin, Heidelberg, Germany. (Cited on page 6.)

[Per16] Trevor Perrin. The XEdDSA and VXEdDSA signature schemes, October 2016. https://signal.org/docs/specifications/xeddsa/. (Cited on page 26.)

[PS14] Inna Polak and Adi Shamir. Using random error correcting codes in near-collision attacks on generic hash-functions. In Willi Meier and Debdeep Mukhopadhyay, editors, *Progress in Cryptology - INDOCRYPT 2014: 15th International Conference in Cryptology in India*, volume 8885 of *Lecture Notes in Computer Science*, pages 219–236, New Delhi, India, December 14–17, 2014. Springer, Cham, Switzerland. (Cited on page 5.)

[Rob23] Damien Robert. Breaking SIDH in polynomial time. In Carmit Hazay and Martijn Stam, editors, *Advances in Cryptology – EUROCRYPT 2023, Part V*, volume 14008 of *Lecture Notes in Computer Science*, pages 472–503, Lyon, France, April 23–27, 2023. Springer, Cham, Switzerland. (Cited on page 4.)

[Rog06] Phillip Rogaway. Formalizing human ignorance. In Phong Q. Nguyen, editor, *Progress in Cryptology - VIETCRYPT 06: 1st International Conference on Cryptology in Vietnam*, volume 4341 of *Lecture Notes in Computer Science*, pages 211–228, Hanoi, Vietnam, September 25–28, 2006. Springer, Berlin, Heidelberg, Germany. (Cited on pages 5 and 51.)

[SAB$^+$] Peter Schwabe, Roberto Avanzi, Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M. Schanck, Gregor Seiler, Damien Stehlé, and Jintai Ding. CRYSTALS-Kyber. https://pq-crystals.org/kyber/. (Cited on pages 3, 7, 8, and 11.)

[Sch24] Sophie Schmieg. Unbindable kemmy schmidt: ML-KEM is neither MAL-BIND-K-CT nor MAL-BIND-K-PK. Cryptology ePrint Archive, Paper 2024/523, 2024. https://eprint.iacr.org/2024/523. (Cited on pages 4, 8, and 10.)

[Sig] Signal: Technical information. https://signal.org/docs/. (Cited on page 1.)

[VGIK20] Nihal Vatandas, Rosario Gennaro, Bertrand Ithurburn, and Hugo Krawczyk. On the cryptographic deniability of the Signal protocol. In Mauro Conti, Jianying Zhou, Emiliano Casalicchio, and Angelo Spognardi, editors, *ACNS 20: 18th International Conference on Applied Cryptography and Network Security, Part II*, volume 12147 of *Lecture Notes in Computer Science*, pages 188–209, Rome, Italy, October 19–22, 2020. Springer, Cham, Switzerland. (Cited on pages 2 and 46.)

[WY05] Xiaoyun Wang and Hongbo Yu. How to break MD5 and other hash functions. In Ronald Cramer, editor, *Advances in Cryptology – EUROCRYPT 2005*, volume 3494 of *Lecture Notes in Computer Science*, pages 19–35, Aarhus, Denmark, May 22–26, 2005. Springer, Berlin, Heidelberg, Germany. (Cited on page 5.)

[YCW14] Hongbo Yu, Jiazhe Chen, and Xiaoyun Wang. Partial-collision attack on the round-reduced compression function of Skein-256. In Shiho Moriai, editor, *Fast Software Encryption – FSE 2013*, volume 8424 of *Lecture Notes in Computer Science*, pages 263–283, Singapore, March 11–13, 2014. Springer, Berlin, Heidelberg, Germany. (Cited on page 5.)

[Zha15] Mark Zhandry. A note on the quantum collision and set equality problems. *Quantum Inf. Comput.*, 15(7&8):557–567, 2015. (Cited on page 12.)

A Summary of Major Changes

- **version 1.0 – May 2024**: Initial release
- **version 2.0 – August 2024**:
  - added post-quantum(-only) analysis (Section 5.2) via separate clean predicates (Figure 9) and theorem statement (Theorem 5.2) for quantum adversaries
  - discuss key confusion between DH and KEM keys as well as ephemeral and semi-static KEM keys in Sections 1 and 4.3 and incorporate them into Theorems 5.1 and 5.2
  - define and use hash functions with the "human ignorance" approach [Rog06] (cp. Definitions 2.4 and 2.5, Theorems 3.6 to 3.11)
  - rename the adversarial model $LEAK^{+}$ for the KEM binding notion to $LEAK^{+r}$