The X3DH Key Agreement Protocol

Moxie Marlinspike Trevor Perrin (editor)

Revision 1, 2016-11-04

Contents

1. Introduction

This document describes the "X3DH" (or "Extended Triple Diffie-Hellman") key agreement protocol. X3DH establishes a shared secret key between two parties who mutually authenticate each other based on public keys. X3DH provides forward secrecy and cryptographic deniability.

X3DH is designed for asynchronous settings where one user ("Bob") is offline but has published some information to a server. Another user ("Alice") wants to use that information to send encrypted data to Bob, and also establish a shared secret key for future communication.

2. Preliminaries

2.1. X3DH parameters

An application using X3DH must decide on several parameters:

Name	Definition
curve	X25519 or X448
hash	A 256 or 512-bit hash function (e.g. SHA-256 or SHA-512)
info	An ASCII string identifying the application

For example, an application could choose *curve* as X25519, *hash* as SHA-512, and *info* as "MyProtocol".

An application must additionally define an encoding function *Encode(PK)* to encode an X25519 or X448 public key *PK* into a byte sequence. The recommended encoding consists of some single-byte constant to represent the type of curve, followed by little-endian encoding of the u-coordinate as specified in [1].

2.2. Cryptographic notation

X3DH will use the following notation:

- The concatenation of byte sequences $\boldsymbol{X}$ and $\boldsymbol{Y}$ is $\boldsymbol{X} \mid\mid \boldsymbol{Y}$.
- ***DH(PK1, PK2)*** represents a byte sequence which is the shared secret output from an Elliptic Curve Diffie-Hellman function involving the key pairs represented by public keys *PK1* and *PK2*. The Elliptic Curve Diffie-Hellman function will be either the X25519 or X448 function from [1], depending on the *curve* parameter.

- ***Sig(PK, M)*** represents a byte sequence that is an XEdDSA signature on the byte sequence M and verifies with public key PK, and which was created by signing M with PK's corresponding private key. The signing and verification functions for XEdDSA are specified in[2].
- ***KDF(KM)*** represents 32 bytes of output from the HKDF algorithm [3] with inputs:
  - *HKDF input key material* = $F \mid\mid KM$, where KM is an input byte sequence containing secret key material, and F is a byte sequence containing 32 0xFF bytes if *curve* is X25519, and 57 0xFF bytes if *curve* is X448. F is used for cryptographic domain separation with XEdDSA [2].
  - *HKDF salt* = A zero-filled byte sequence with length equal to the *hash* output length.
  - *HKDF info* = The *info* parameter from Section 2.1.

2.3. Roles

The X3DH protocol involves three parties: **Alice**, **Bob**, and a **server**.

- **Alice** wants to send Bob some initial data using encryption, and also establish a shared secret key which may be used for bidirectional communication.
- **Bob** wants to allow parties like Alice to establish a shared key with him and send encrypted data. However, Bob might be offline when Alice attempts to do this. To enable this, Bob has a relationship with some server.
- The **server** can store messages from Alice to Bob which Bob can later retrieve. The server also lets Bob publish some data which the server will provide to parties like Alice. The amount of trust placed in the server is discussed in Section 4.7.

In some systems the server role might be divided amongst multiple entities, but for simplicity we assume a single server that provides the above functions for Alice and Bob.

2.4. Keys

X3DH uses the following elliptic curve public keys:

Name	Definition
IK_A	Alice's identity key
EK_A	Alice's ephemeral key
IK_B	Bob's identity key
SPK_B	Bob's signed prekey
OPK_B	Bob's one-time prekey

All public keys have a corresponding private key, but to simplify description we will focus on the public keys.

The public keys used within an X3DH protocol run must either all be in X25519 form, or they must all be in X448 form, depending on the *curve* parameter [1].

Each party has a long-term identity public key (IK_A for Alice, IK_B for Bob).

Bob also has a signed prekey SPK_B, which he will change periodically, and a set of one-time prekeys OPK_B, which are each used in a single X3DH protocol run. ("Prekeys" are so named because they are essentially protocol messages which Bob publishes to the server *prior* to Alice beginning the protocol run).

During each protocol run, Alice generates a new ephemeral key pair with public key EK_A.

After a successful protocol run Alice and Bob will share a 32-byte secret key SK. This key may be used within some post-X3DH secure communication protocol, subject to the security considerations in Section 4.

3. The X3DH protocol

3.1. Overview

X3DH has three phases:

1. Bob publishes his identity key and prekeys to a server.
2. Alice fetches a "prekey bundle" from the server, and uses it to send an initial message to Bob.
3. Bob receives and processes Alice's initial message.

The following sections explain these phases.

3.2. Publishing keys

Bob publishes a set of elliptic curve public keys to the server, containing:

- Bob's identity key IK_B
- Bob's signed prekey SPK_B
- Bob's prekey signature $Sig(IK_B, Encode(SPK_B))$
- A set of Bob's one-time prekeys ($OPK_B{}^1$, $OPK_B{}^2$, $OPK_B{}^3$, ...)

Bob only needs to upload his identity key to the server once. However, Bob may upload new one-time prekeys at other times (e.g. when the server informs Bob that the server's store of one-time prekeys is getting low).

Bob will also upload a new signed prekey and prekey signature at some interval (e.g. once a week, or once a month). The new signed prekey and prekey signature will replace the previous values.

After uploading a new signed prekey, Bob may keep the private key corresponding to the previous signed prekey around for some period of time, to handle messages using it that have been delayed in transit. Eventually, Bob should delete this private key for forward secrecy (one-time prekey private keys will be deleted as Bob receives messages using them; see Section 3.4).

3.3. Sending the initial message

To perform an X3DH key agreement with Bob, Alice contacts the server and fetches a "prekey bundle" containing the following values:

- Bob's identity key IK_B
- Bob's signed prekey SPK_B
- Bob's prekey signature $Sig(IK_B, Encode(SPK_B))$
- (Optionally) Bob's one-time prekey OPK_B

The server should provide one of Bob's one-time prekeys if one exists, and then delete it. If all of Bob's one-time prekeys on the server have been deleted, the bundle will not contain a one-time prekey.

Alice verifies the prekey signature and aborts the protocol if verification fails. Alice then generates an ephemeral key pair with public key EK_A.

If the bundle does not contain a one-time prekey, she calculates:

$\mathrm{DH1} = \mathrm{DH}(\mathrm{IK_A}, \mathrm{SPK_B})$
$\mathrm{DH2} = \mathrm{DH}(\mathrm{EK_A}, \mathrm{IK_B})$
$\mathrm{DH3} = \mathrm{DH}(\mathrm{EK_A}, \mathrm{SPK_B})$
$\mathrm{SK} = \mathrm{KDF}(\mathrm{DH1} \,||\, \mathrm{DH2} \,||\, \mathrm{DH3})$

If the bundle *does* contain a one-time prekey, the calculation is modified to include an additional DH:

$\text{DH4} = \text{DH}(\text{EK}_\text{A}, \text{OPK}_\text{B})$
$\text{SK} = \text{KDF}(\text{DH1} \mid\mid \text{DH2} \mid\mid \text{DH3} \mid\mid \text{DH4})$

The following diagram shows the *DH* calculations between keys. Note that *DH1* and *DH2* provide mutual authentication, while *DH3* and *DH4* provide forward secrecy.

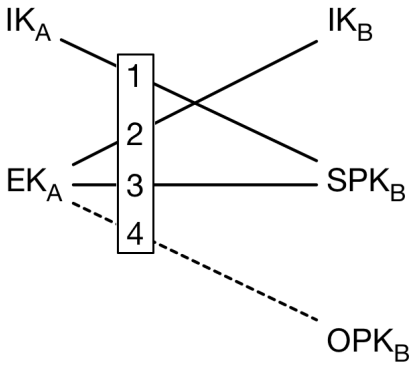

Figure 1: DH1. . . DH4

After calculating *SK*, Alice deletes her ephemeral private key and the *DH* outputs.

Alice then calculates an "associated data" byte sequence *AD* that contains identity information for both parties:

$\text{AD} = \text{Encode}(\text{IK}_\text{A}) \mid\mid \text{Encode}(\text{IK}_\text{B})$

Alice may optionally append additional information to *AD*, such as Alice and Bob's usernames, certificates, or other identifying information.

Alice then sends Bob an initial message containing:

- Alice's identity key IK_A
- Alice's ephemeral key EK_A
- Identifiers stating which of Bob's prekeys Alice used
- An initial ciphertext encrypted with some AEAD encryption scheme [4] using *AD* as associated data and using an encryption key which is either *SK* or the output from some cryptographic PRF keyed by *SK*.

The initial ciphertext is typically the first message in some post-X3DH communication protocol. In other words, this ciphertext typically has two roles, serving as the first message within some post-X3DH protocol, and as part of Alice's X3DH initial message.

After sending this, Alice may continue using *SK* or keys derived from *SK* within the post-X3DH protocol for communication with Bob, subject to the security considerations in Section 4.

3.4. Receiving the initial message

Upon receiving Alice's initial message, Bob retrieves Alice's identity key and ephemeral key from the message. Bob also loads his identity private key, and the private key(s) corresponding to whichever signed prekey and one-time prekey (if any) Alice used.

Using these keys, Bob repeats the *DH* and *KDF* calculations from the previous section to derive *SK*, and then deletes the *DH* values.

Bob then constructs the *AD* byte sequence using IK_A and IK_B, as described in the previous section. Finally, Bob attempts to decrypt the initial ciphertext using *SK* and *AD*. If the initial ciphertext fails to decrypt, then Bob aborts the protocol and deletes *SK*.

If the initial ciphertext decrypts successfully the protocol is complete for Bob. Bob deletes any one-time prekey private key that was used, for forward secrecy. Bob may then continue using *SK* or keys derived from *SK* within the post-X3DH protocol for communication with Alice, subject to the security considerations in Section 4.

4. Security considerations

4.1. Authentication

Before or after an X3DH key agreement, the parties may compare their identity public keys IK_A and IK_B through some authenticated channel. For example, they may compare public key fingerprints manually, or by scanning a QR code. Methods for doing this are outside the scope of this document.

If authentication is not performed, the parties receive no cryptographic guarantee as to who they are communicating with.

4.2. Protocol replay

If Alice's initial message doesn't use a one-time prekey, it may be replayed to Bob and he will accept it. This could cause Bob to think Alice had sent him the same message (or messages) repeatedly.

To mitigate this, a post-X3DH protocol may wish to quickly negotiate a new encryption key for Alice based on fresh random input from Bob. This is the typical behavior of Diffie-Hellman based ratcheting protocols [5].

Bob could attempt other mitigations, such as maintaining a blacklist of observed messages, or replacing old signed prekeys more rapidly. Analyzing these mitigations is beyond the scope of this document.

4.3. Replay and key reuse

Another consequence of the replays discussed in the previous section is that a successfully replayed initial message would cause Bob to derive the same *SK* in different protocol runs.

For this reason, any post-X3DH protocol MUST randomize the encryption key before Bob sends encrypted data. For example, Bob could use a DH-based ratcheting protocol to combine *SK* with a freshly generated *DH* output to get a randomized encryption key [5].

Failure to randomize Bob's encryption key may cause catastrophic key reuse.

4.4. Deniability

X3DH doesn't give either Alice or Bob a publishable cryptographic proof of the contents of their communication or the fact that they communicated.

Like in the OTR protocol [6], in some cases a third party that has compromised legitimate private keys from Alice or Bob could be provided a communication transcript that appears to be between Alice and Bob and that can only have been created by some other party that also has access to legitimate private keys from Alice or Bob (i.e. Alice or Bob themselves, or someone else who has compromised their private keys).

If either party is collaborating with a third party during protocol execution, they will be able to provide proof of their communication to such a third party. This limitation on "online" deniability appears to be intrinsic to the asynchronous setting [7].

4.5. Signatures

It might be tempting to observe that mutual authentication and forward secrecy are achieved by the *DH* calculations, and omit the prekey signature. However, this would allow a "weak forward secrecy" attack: A malicious server could provide Alice a prekey bundle with forged prekeys, and later compromise Bob's IK_B to calculate *SK*.

Alternatively, it might be tempting to replace the DH-based mutual authentication (i.e. *DH1* and *DH2*) with signatures from the identity keys. However, this reduces deniability, increases the size of initial messages, and increases the damage done if ephemeral or prekey private keys are compromised, or if the signature scheme is broken.

4.6. Key compromise

Compromise of a party's private keys has a disastrous effect on security, though the use of ephemeral keys and prekeys provides some mitigation.

Compromise of a party's identity private key allows impersonation of that party to others. Compromise of a party's prekey private keys may affect the security of older or newer SK values, depending on many considerations.

A full analysis of all possible compromise scenarios is outside the scope of this document, however a partial analysis of some plausible scenarios is below:

- If one-time prekeys are used for a protocol run then a compromise of Bob's identity key and prekey private keys at some future time will not compromise the older SK, assuming the private key for OPK_B was deleted.
- If one-time prekeys were *not* used for a protocol run, then a compromise of the private keys for IK_B and SPK_B from that protocol run would compromise the SK that was calculated earlier. Frequent replacement of signed prekeys mitigates this, as does using a post-X3DH ratcheting protocol which rapidly replaces SK with new keys to provide fresh forward secrecy [5].
- Compromise of prekey private keys may enable attacks that extend into the future, such as passive calculation of SK values, and impersonation of arbitrary other parties to the compromised party ("key-compromise impersonation"). These attacks are possible until the compromised party replaces his compromised prekeys on the server (in the case of passive attack); or deletes his compromised signed prekey's private key (in the case of key-compromise impersonation).

4.7. Server trust

A malicious server could cause communication between Alice and Bob to fail (e.g. by refusing to deliver messages).

If Alice and Bob authenticate each other as in Section 4.1, then the only additional attack available to the server is to refuse to hand out one-time prekeys, causing forward secrecy for SK to depend on the signed prekey's lifetime (as analyzed in the previous section).

This reduction in initial forward secrecy could also happen if one party maliciously drains another party's one-time prekeys, so the server should attempt to prevent this, e.g. with rate limits on fetching prekey bundles.

4.8. Identity binding

Authentication as in Section 4.1 does not necessarily prevent an "identity misbinding" or "unknown key share" attack.

This results when an attacker ("Charlie") falsely presents Bob's identity key fingerprint to Alice as his (Charlie's) own, and then either forwards Alice's initial message to Bob, or falsely presents Bob's contact information as his own.

The effect of this is that Alice thinks she is sending an initial message to Charlie when she is actually sending it to Bob.

To make this more difficult the parties can include more identifying information into AD, or hash more identifying information into the fingerprint, such as usernames, phone numbers, real names, or other identifying information. Charlie would be forced to lie about these additional values, which might be difficult.

However, there is no way to reliably prevent Charlie from lying about additional values, and including more identity information into the protocol often brings trade-offs in terms of privacy, flexibility, and user interface. A detailed analysis of these trade-offs is beyond the scope of this document.

5. IPR

6. Acknowledgements

The X3DH protocol was developed by Moxie Marlinspike and Trevor Perrin.

The underlying "Triple DH" key agreement was proposed by Caroline Kudla and Kenny Paterson in [8], extending the earlier "Double DH" (aka "Protocol 4") key agreement from Simon Blake-Wilson et al [9]. Using signatures in combination with implicitly-authenticated key agreement has been discussed in works like [10] and [11].

Thanks to Mike Hamburg for discussions about identity binding and elliptic curve public keys.

Thanks to Nik Unger and Matthew Green for discussions about deniability.

Thanks to Matthew Green, Tom Ritter, Joseph Bonneau, and Benedikt Schmidt for editorial feedback.

7. References

[1] A. Langley, M. Hamburg, and S. Turner, "Elliptic Curves for Security." Internet Engineering Task Force; RFC 7748 (Informational); IETF, Jan-2016. http://www.ietf.org/rfc/rfc7748.txt

[2] T. Perrin, "The XEdDSA and VXEdDSA Signature Schemes," 2016. https://whispersystems.org/docs/specifications/xeddsa/

[3] H. Krawczyk and P. Eronen, "HMAC-based Extract-and-Expand Key Derivation Function (HKDF)." Internet Engineering Task Force; RFC 5869 (Informational); IETF, May-2010. http://www.ietf.org/rfc/rfc5869.txt

[4] P. Rogaway, "Authenticated-encryption with Associated-data," in Proceedings of the 9th ACM Conference on Computer and Communications Security, 2002. http://web.cs.ucdavis.edu/~rogaway/papers/ad.pdf

[5] T. Perrin, "The Double Ratchet Algorithm (work in progress)," 2016.

[6] N. Borisov, I. Goldberg, and E. Brewer, "Off-the-record Communication, or, Why Not to Use PGP," in Proceedings of the 2004 acm workshop on privacy in the electronic society, 2004. http://doi.acm.org/10.1145/1029179.1029200

[7] N. Unger and I. Goldberg, "Deniable Key Exchanges for Secure Messaging," in Proceedings of the 22Nd acm sigsac conference on computer and communications security, 2015. http://doi.acm.org/10.1145/2810103.2813616

[8] C. Kudla and K. G. Paterson, "Modular Security Proofs for Key Agreement Protocols," in Advances in Cryptology - ASIACRYPT 2005: 11th International Conference on the Theory and Application of Cryptology and Information Security, 2005. http://www.isg.rhul.ac.uk/~kp/ModularProofs.pdf

[9] S. Blake-Wilson, D. Johnson, and A. Menezes, "Key agreement protocols and their security analysis," in Crytography and Coding: 6th IMA International Conference Cirencester, UK, December 17–19, 1997 Proceedings, 1997. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.387

[10] C. Cremers and M. Feltz, "One-round Strongly Secure Key Exchange with Perfect Forward Secrecy and Deniability." Cryptology ePrint Archive, Report 2011/300, 2011. http://eprint.iacr.org/2011/300

[11] J. P. Degabriele, A. Lehmann, K. G. Paterson, N. P. Smart, and M. Strefler, "On the Joint Security of Encryption and Signature in EMV." Cryptology ePrint Archive, Report 2011/615, 2011. http://eprint.iacr.org/2011/615