Real-World Deniability in Messaging

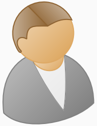
Daniel Collins, Simone Colombo and Loïs Huguenin-Dumittan

PETS 2025

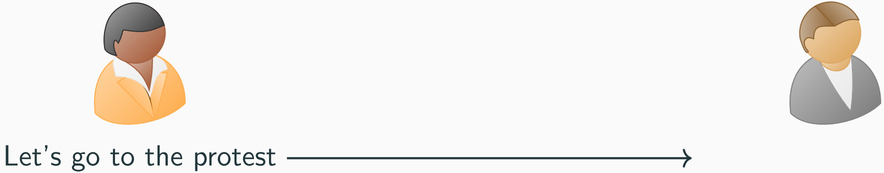
Deniability

> *If someone receives a [. . .] message from you, they can be absolutely sure you sent it (rather than having been forged by some third party),* **but can't prove to anyone else that it was a message you wrote***.*
>
> — *Moxie Marlinspike [Mar13] (emphasis added)*

Is deniability practical?

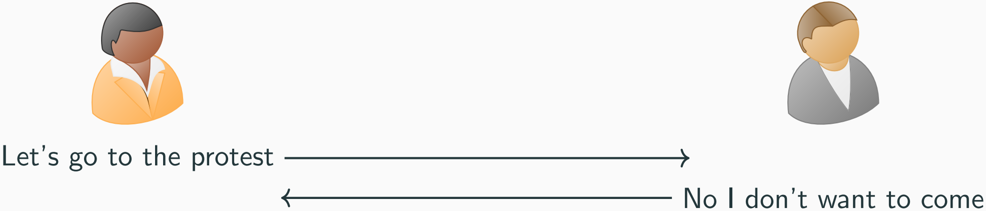


Is deniability practical?



Let's go to the protest

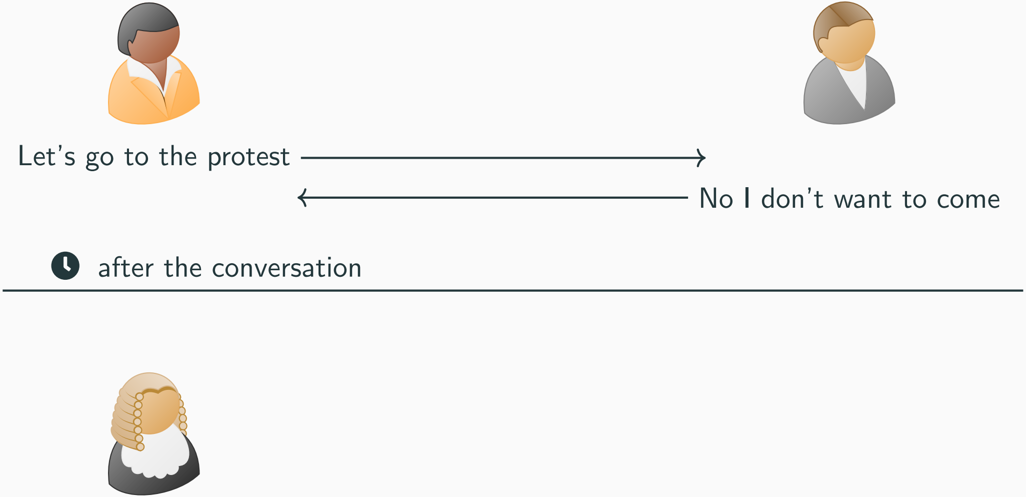
Is deniability practical?



Let's go to the protest →

← No I don't want to come

Is deniability practical?

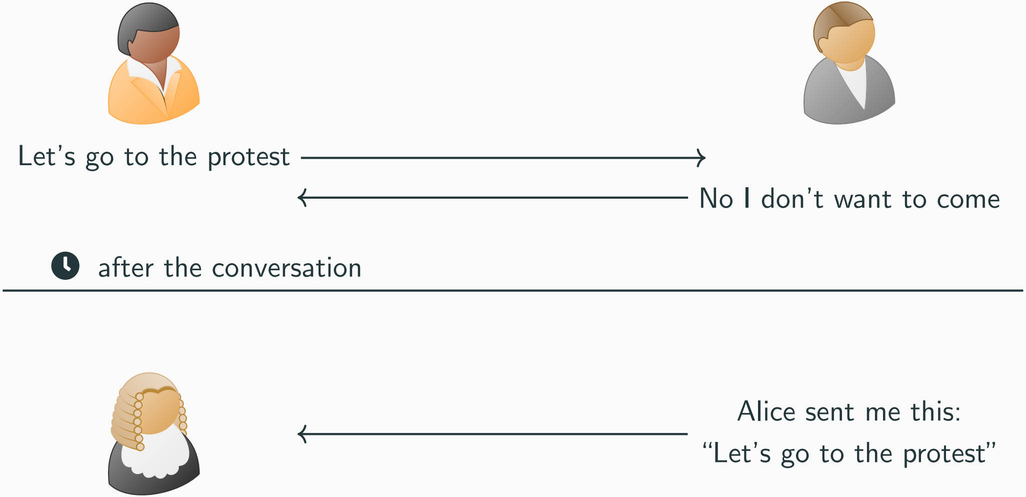


Let's go to the protest

No I don't want to come

after the conversation

Is deniability practical?



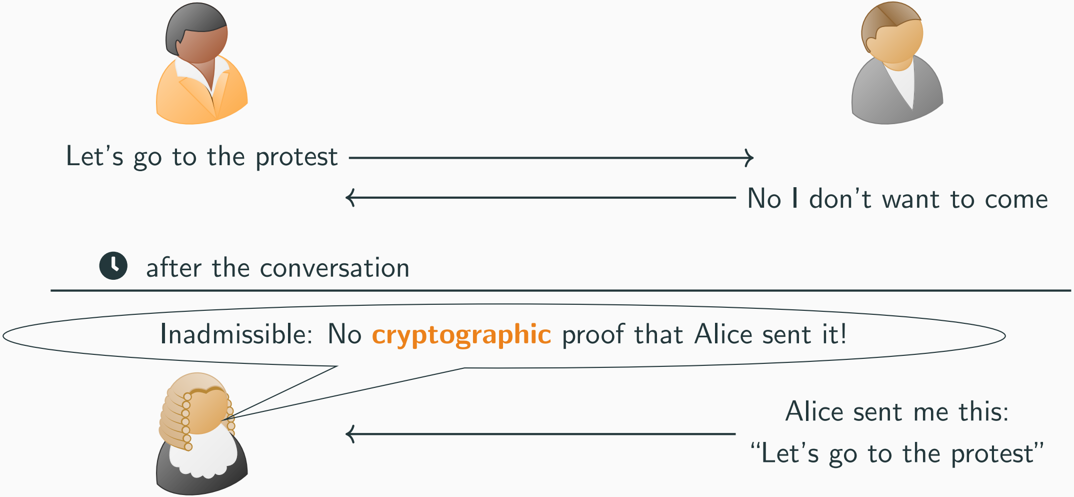
Let's go to the protest

No I don't want to come

after the conversation

Alice sent me this:
"Let's go to the protest"

Is deniability practical?



Let's go to the protest

No I don't want to come

after the conversation

Inadmissible: No **cryptographic** proof that Alice sent it!

Alice sent me this:
"Let's go to the protest"

Does this work in practice? If not, can we make it?

Outline

1. Technical case study
2. Legal case study
3. A possible solution
4. Conclusion

Technical case study

Technical case study: Signal

Signal claims to provide deniability and recent works show it achieves some form of **cryptographic** deniability [VGIK20, FJ24, KNTW25].

On the Cryptographic Deniability of the Signal Protocol

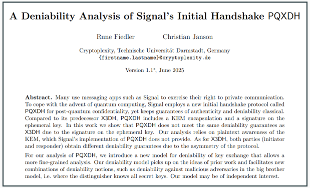
Nihal Vatandas[1], Rosario Gennaro[1], Bertrand Ithurburn[1], and Hugo Krawczyk[2]

[1] Center for Algorithms and Interactive Scientific Software. The City College of New York. Email: nihal.vatandas@gmail.com (contact author), rosario@ccny.cuny.edu, bithurburn@gmail.com
[2] Algorand Foundation. Email: hugokraw@gmail.com

Abstract. *Offline deniability* is the ability to *a posteriori* deny having participated in a particular communication session. This property has been widely assumed for the Signal messaging application, yet no formal proof has appeared in the literature. In this paper, we present what we believe is the first formal study of the offline deniability of the Signal protocol. Our analysis shows that building a deniability proof for Signal is non-trivial and requires strong assumptions on the underlying mathematical groups where the protocol is run.

To do so, we study various implicitly authenticated key exchange protocols including MQV, HMQV and 3DH/X3DH, the latter being the core key agreement protocol in Signal. We first present examples of mathematical groups where running MQV results in a provably non-deniable interaction. While the concrete attack applies only to MQV, it also exemplifies the problems in attempting to prove the deniability of other implicitly authenticated protocols, such as 3DH. In particular, it shows that the intuition that the minimal transcript produced by these protocols suffices for ensuring deniability does not hold. We then provide a characterization of the groups where deniability holds, defined in terms of a knowledge assumption that extends the Knowledge of Exponent Assumption (KEA).

We conclude the paper by showing two additional positive results. The first is a general theorem that links the deniability of a communication session to the deniability of the key agreement protocol starting the session. This allows us to extend our results on the deniability of 3DH/X3DH to the entire Signal communication session.



A Deniability Analysis of Signal's Initial Handshake PQXDH

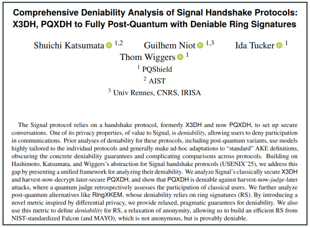
Rune Fiedler   Christian Janson

Cryptoplexity, Technische Universität Darmstadt, Germany
{firstname.lastname}@cryptoplexity.de

Version 1.1*, June 2025

Abstract. Many use messaging apps such as Signal to exercise their right to private communication. To cope with the advent of quantum computing, Signal employs a new initial handshake protocol called PQXDH for post-quantum confidentiality, yet keeps guarantees of authenticity and deniability classical. Compared to its predecessor X3DH, PQXDH includes a KEM encapsulation and a signature on the ephemeral key. In this work we show that PQXDH does not meet the same deniability guarantees as X3DH due to the signature on the ephemeral key. Our analysis relies on plaintext awareness of the KEM, which Signal's implementation of PQXDH does not provide. As for X3DH, both parties (initiator and responder) obtain different deniability guarantees due to the asymmetry of the protocol.

For our analysis of PQXDH, we introduce a new model for deniability of key exchange that allows a more fine-grained analysis. Our deniability model picks up on the ideas of prior work and facilitates new combinations of deniability notions, such as deniability against malicious adversaries in the big brother model, i.e. where the distinguisher knows all secret keys. Our model may be of independent interest.



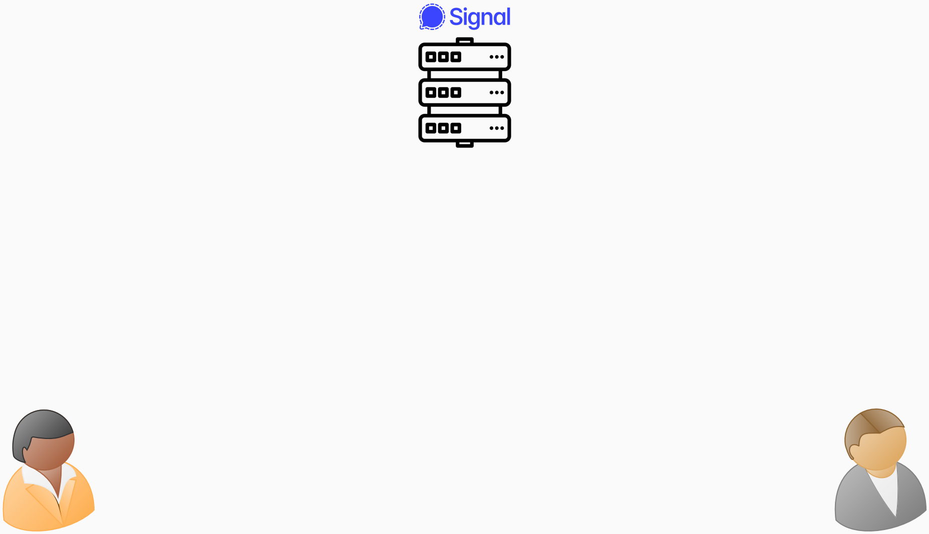
Comprehensive Deniability Analysis of Signal Handshake Protocols: X3DH, PQXDH to Fully Post-Quantum with Deniable Ring Signatures

Shuichi Katsumata [1,2]   Guilhem Niot [1,3]   Ida Tucker [1]
Thom Wiggers [1]

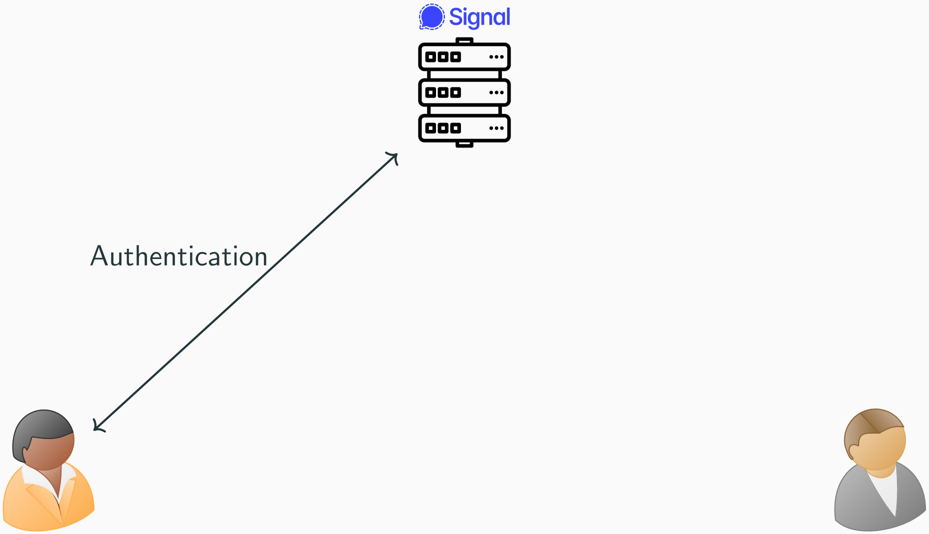
[1] PQShield
[2] AIST
[3] Univ Rennes, CNRS, IRISA

The Signal protocol relies on a handshake protocol, formerly X3DH and now PQXDH, to set up secure conversations. One of its privacy properties, of value to Signal, is *deniability*, allowing users to deny participation in communications. Prior analyses of deniability for these protocols, including post-quantum variants, use models highly tailored to the individual protocols and generally make ad-hoc adaptations to "standard" AKE definitions, obscuring the concrete deniability guarantees and complicating comparisons across protocols. Building on Hashimoto, Katsumata, and Wiggers's abstraction for Signal handshake protocols (USENIX'25), we address this gap by presenting a unified framework for analyzing their deniability. We analyze Signal's classically secure X3DH and harvest-now-decrypt-later-secure PQXDH, and show that PQXDH is deniable against harvest-now-*judge*-later attacks, where a quantum judge retrospectively assesses the participation of classical users. We further analyze post-quantum alternatives like RingXKEM, whose deniability relies on ring signatures (RS). By introducing a novel metric inspired by differential privacy, we provide relaxed, pragmatic guarantees for deniability. We also use this metric to define *deniability* for RS, a relaxation of anonymity, allowing us to build an efficient RS from NIST-standardized Falcon (and MAYO), which is not anonymous, but is provably deniable.

Signal with classic authentication

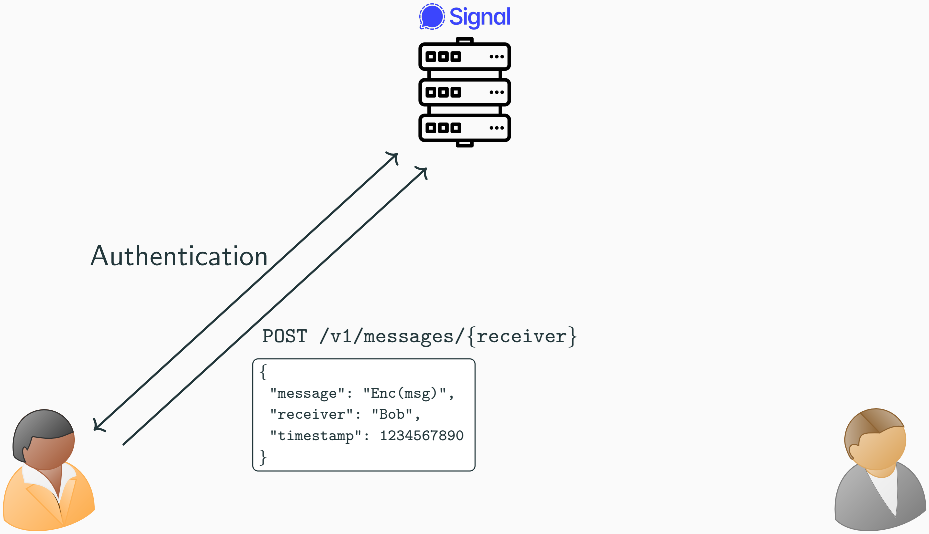


Signal with classic authentication



Authentication

Signal with classic authentication

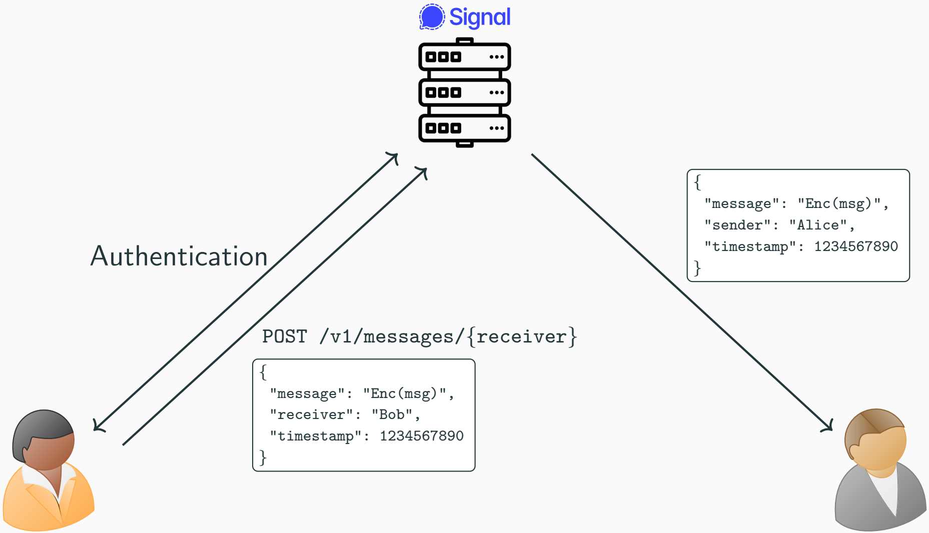


Authentication

POST /v1/messages/{receiver}

```
{
 "message": "Enc(msg)",
 "receiver": "Bob",
 "timestamp": 1234567890
}
```

Signal with classic authentication

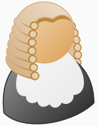


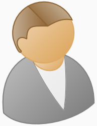
POST /v1/messages/{receiver}

```
{
 "message": "Enc(msg)",
 "receiver": "Bob",
 "timestamp": 1234567890
}
```

```
{
 "message": "Enc(msg)",
 "sender": "Alice",
 "timestamp": 1234567890
}
```

Classic authentication hinders deniability





Classic authentication hinders deniability

Look at my **phone**,
Alice sent me this message

Classic authentication hinders deniability

Look at my **phone**,
Alice sent me this message

Unless something bad happened, if Bob's device contains Alice's message, then

Classic authentication hinders deniability

Look at my **phone**,
Alice sent me this message

Unless something bad happened, if Bob’s device contains Alice’s message, then

- either Alice really sent it after authenticating with the server, or

Classic authentication hinders deniability

Look at my **phone**,
Alice sent me this message

Unless something bad happened, if Bob's device contains Alice's message, then

- either Alice really sent it after authenticating with the server, or
- Bob tampered with the phone to insert Alice's message.

Classic authentication hinders deniability

Look at my **phone**,
Alice sent me this message

Unless something bad happened, if Bob's device contains Alice's message, then

- either Alice really sent it after authenticating with the server, or
- Bob tampered with the phone to insert Alice's message.

If the server stores logs the situation is even worse.

Classic authentication hinders deniability

Look at my **phone**,
Alice sent me this message

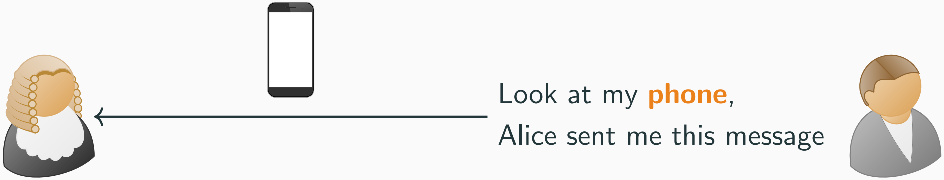
Unless something bad happened, if Bob's device contains Alice's message, then

- either Alice really sent it after authenticating with the server, or
- Bob tampered with the phone to insert Alice's message.

If the server stores logs the situation is even worse.

Signal is undeniable unless Bob knows how to tamper with the phone.

Classic authentication hinders deniability



Look at my **phone**,
Alice sent me this message

Unless something bad happened, if Bob's device contains Alice's message, then

- either Alice really sent it after authenticating with the server, or
- Bob tampered with the phone to insert Alice's message.

If the server stores logs the situation is even worse.

Signal is undeniable unless Bob knows how to tamper with the phone.
What about the legal impact of deniability?

Legal case study

Legal case study methodology

Manual analysis of 341 penal cases in Switzerland that mention “WhatsApp”.

Legal case study methodology

Manual analysis of 341 penal cases in Switzerland that mention “WhatsApp”.

Research questions:

Legal case study methodology

Manual analysis of 341 penal cases in Switzerland that mention “WhatsApp”.

Research questions:

- Do judges in Swiss courts use WhatsApp as evidence?

Legal case study methodology

Manual analysis of 341 penal cases in Switzerland that mention “WhatsApp”.

Research questions:

- Do judges in Swiss courts use WhatsApp as evidence?
- When they do, is their usage contested by any of the parties involved?

Legal case study methodology

Manual analysis of 341 penal cases in Switzerland that mention “WhatsApp”.

Research questions:

- Do judges in Swiss courts use WhatsApp as evidence?
- When they do, is their usage contested by any of the parties involved?
- What are the reasons used to dispute the legal validity of such messages?

Legal case study methodology

Manual analysis of 341 penal cases in Switzerland that mention “WhatsApp”.

Research questions:

- Do judges in Swiss courts use WhatsApp as evidence?
- When they do, is their usage contested by any of the parties involved?
- What are the reasons used to dispute the legal validity of such messages?
- How do judges respond to these disputes?

Legal case study results

Total Cases	N/A	Evidence	Contested	Rejected
341	201 (59%)	140 (41%)	2	0

Legal case study results

Total Cases	N/A	Evidence	Contested	Rejected
341	201 (59%)	140 (41%)	**2**	0

- Deniability is not invoked in the **contested cases**;

Legal case study results

Total Cases	N/A	Evidence	Contested	Rejected
341	201 (59%)	140 (41%)	2	0

- Deniability is not invoked in the contested cases;
- Yadav et al. [YGS23] obtain similar results in an analysis of US court cases.

Legal case study results

Total Cases	N/A	Evidence	Contested	Rejected
341	201 (59%)	140 (41%)	2	0

- Deniability is not invoked in the contested cases;
- Yadav et al. [YGS23] obtain similar results in an analysis of US court cases.

Cryptographic deniability fails technically and (likely) legally

Legal case study results

Total Cases	N/A	Evidence	Contested	Rejected
341	201 (59%)	140 (41%)	2	0

- Deniability is not invoked in the contested cases;
- Yadav et al. [YGS23] obtain similar results in an analysis of US court cases.

Cryptographic deniability fails technically and (likely) legally: what to do?

A possible solution

A possible solution. . .

Let users **edit any sent or received** message in the user interface.

A possible solution. . .

Let users **edit any sent or received** message in the user interface.

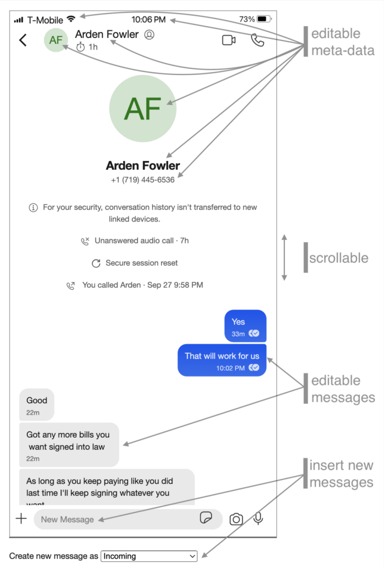
- either Alice really sent it after authenticating with the server, or
- Bob tampered with the phone to insert Alice’s message.

A possible solution. . .

Let users **edit any sent or received** message in the user interface.

- either Alice really sent it after authenticating with the server, or
- **Bob tampered with the phone to insert Alice's message.**

. . . backed by two user studies



For your security, conversation history isn't transferred to new linked devices.

Unanswered audio call · 7h

Secure session reset

You called Arden · Sep 27 9:58 PM

Yes
33m

That will work for us
10:02 PM

Good
22m

Got any more bills you want signed into law
22m

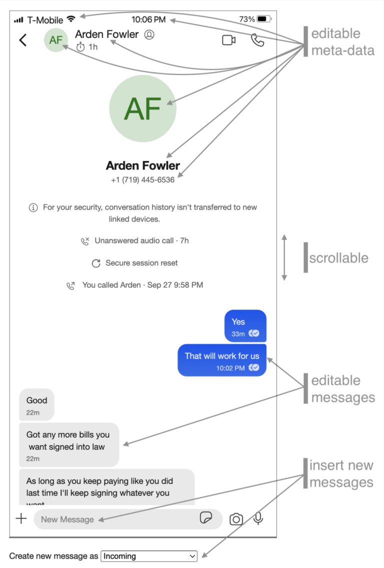
As long as you keep paying like you did last time I'll keep signing whatever you want

New Message

Create new message as Incoming

- Reitinger et al. independently suggest this could improve deniability [RMA+23] (source of image).

. . . backed by two user studies



AF

Arden Fowler

+1 (719) 445-6536

For your security, conversation history isn't transferred to new linked devices.

Unanswered audio call · 7h

Secure session reset

You called Arden · Sep 27 9:58 PM

Yes
33m

That will work for us
10:02 PM

Good
22m

Got any more bills you want signed into law
22m

As long as you keep paying like you did last time I'll keep signing whatever you want

New Message

Create new message as Incoming

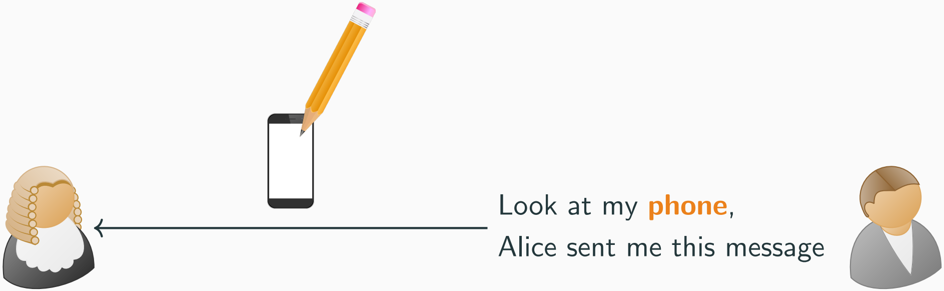
- Reitinger et al. independently suggest this could improve deniability [RMA+23] (source of image).
- Rajendran et al. implement the solution and conduct a user study that reports positive results [RYA+24].

Conclusion

Conclusion

Look at my **phone**,
Alice sent me this message

Conclusion



Look at my **phone**,
Alice sent me this message

Conclusion

Look at my **phone**,
Alice sent me this message

In the paper (`https://ia.cr/2023/403`) we also

Conclusion

Look at my **phone**,
Alice sent me this message

In the paper (`https://ia.cr/2023/403`) we also

- propose a model to analyze real-world deniability,

Conclusion

Look at my **phone**,
Alice sent me this message

In the paper (`https://ia.cr/2023/403`) we also

- propose a model to analyze real-world deniability,
- analyze real-world deniability of Signal with sealed sender,

Conclusion

Look at my **phone**,
Alice sent me this message

In the paper (`https://ia.cr/2023/403`) we also

- propose a model to analyze real-world deniability,
- analyze real-world deniability of Signal with sealed sender,
- analyze real-world deniability of DKIM-protected email and KeyForge [SPG21],

Conclusion

Look at my **phone**,

Alice sent me this message

In the paper (`https://ia.cr/2023/403`) we also

- propose a model to analyze real-world deniability,
- analyze real-world deniability of Signal with sealed sender,
- analyze real-world deniability of DKIM-protected email and KeyForge [SPG21],
- discuss how to design systems with real-world deniability.

References i

- Rune Fiedler and Christian Janson.
  A deniability analysis of Signal's initial handshake PQXDH.
  PoPETs, 2024(4):907–928, October 2024.

- Shuichi Katsumata, Guilhem Niot, Ida Tucker, and Thom Wiggers.
  Comprehensive deniability analysis of signal handshake protocols: X3DH, PQXDH to fully post-quantum with deniable ring signatures.
  Cryptology ePrint Archive, Paper 2025/1090, 2025.

- Moxie Marlinspike.
  Simplifying OTR deniability.
  `https://signal.org/blog/simplifying-otr-deniability/`, 2013.
  Last visited on 19-06-2025.

References ii

- Nathan Reitinger, Nathan Malkin, Omer Akgul, Michelle L. Mazurek, and Ian Miers.
  Is cryptographic deniability sufficient? non-expert perceptions of deniability in secure messaging.
  In *2023 IEEE Symposium on Security and Privacy*, pages 274–292. IEEE Computer Society Press, May 2023.

- Anamika Rajendran, Tarun Kumar Yadav, Malek Al-Jbour, Francisco Manuel Mares Solano, Kent E. Seamons, and Joshua Reynolds.
  Deniable encrypted messaging: User understanding after hands-on social experience.
  In *EuroUSEC*, pages 155–171. ACM, 2024.

- Michael A. Specter, Sunoo Park, and Matthew Green.
KeyForge: Non-attributable email from forward-forgeable signatures.
In Michael Bailey and Rachel Greenstadt, editors, *USENIX Security 2021*, pages 1755–1773. USENIX Association, August 2021.

- 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 2020, Part II*, volume 12147 of *LNCS*, pages 188–209. Springer, Cham, October 2020.

References iv

Tarun Kumar Yadav, Devashish Gosain, and Kent E. Seamons.
Cryptographic deniability: A multi-perspective study of user perceptions and expectations.
In Joseph A. Calandrino and Carmela Troncoso, editors, *USENIX Security 2023*, pages 3637–3654. USENIX Association, August 2023.