Metadata Protection in Instant Messaging Apps

A review

Florian Maury - @x_cli@infosec.exchange

Talk Non-Goals

- Determine the best instant messaging application; many valid criteria:
  - availability in your country
  - what app/networks your contacts use
  - usability/user friendliness
  - battery usage
  - availability/resilience
  - encryption quality
  - whether was independently reviewed
  - metadata protection
  - author political stance
- Being exhaustive
- Break cryptography

Foreword on conspiracy theories

We are only considering what is possible to do

No accusation of evil intent from any software developer or hosting provider

Selected Applications/Protocols

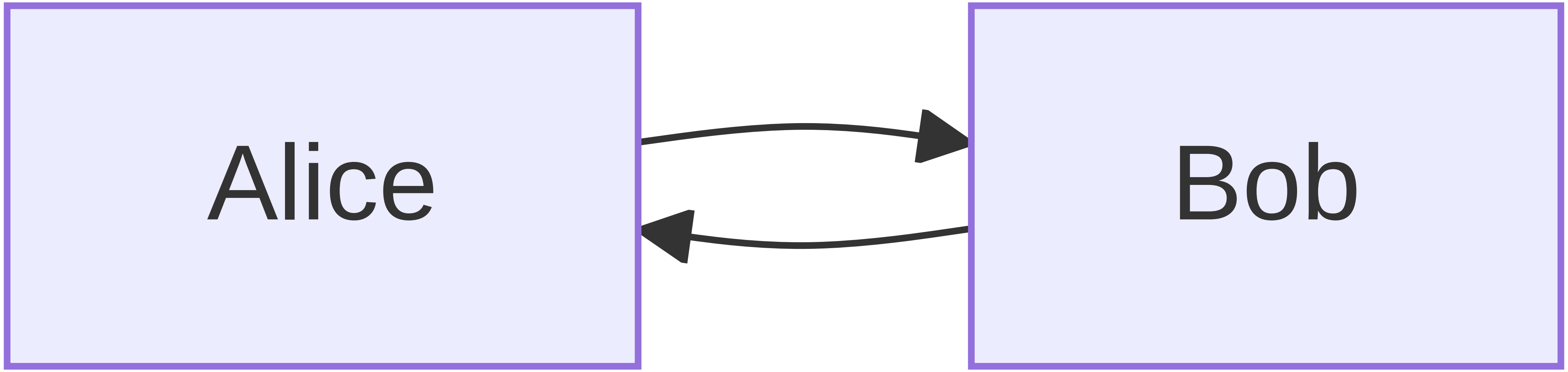
Featured in this talk:

- Signal
- Olvid
- Matrix
- SimpleX Chat

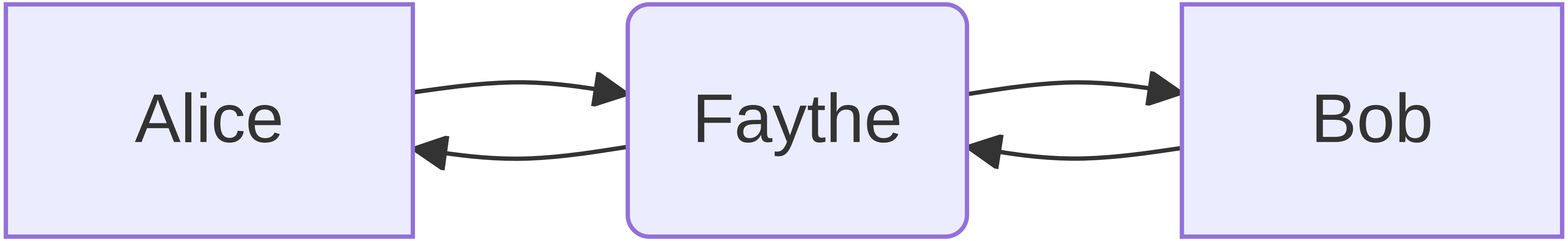
These are apps that I use/fund

Architecture

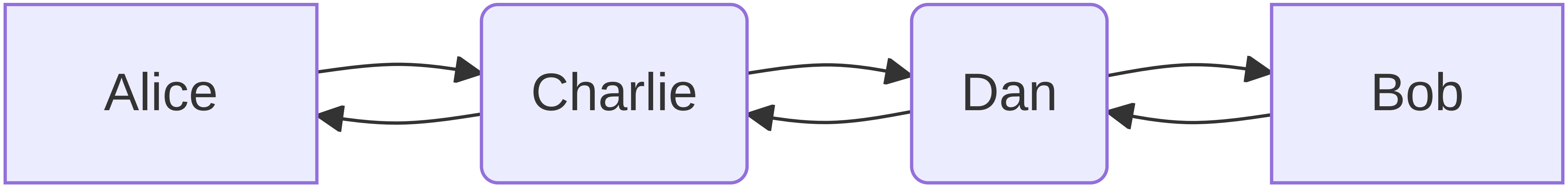
"Everything is encrypted; who cares?"



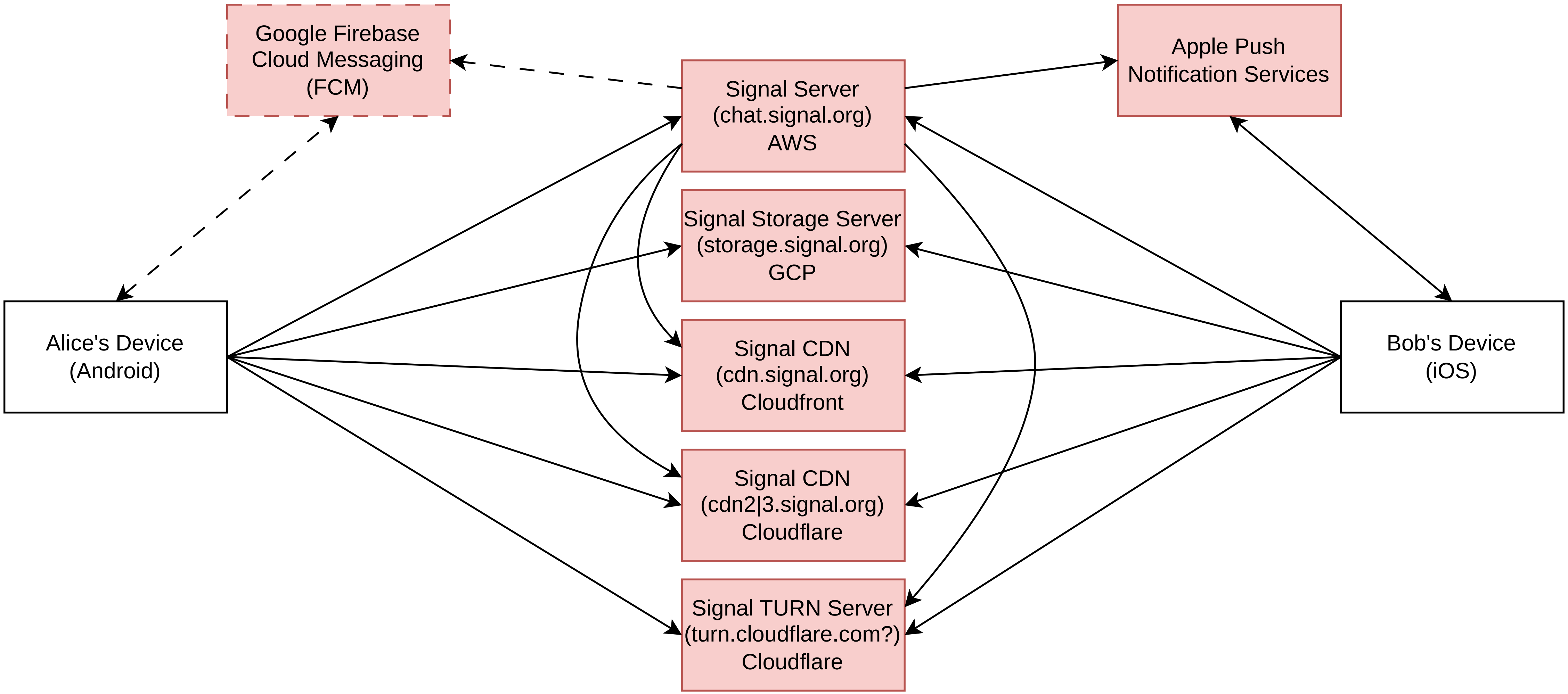
Fantasized centralized network



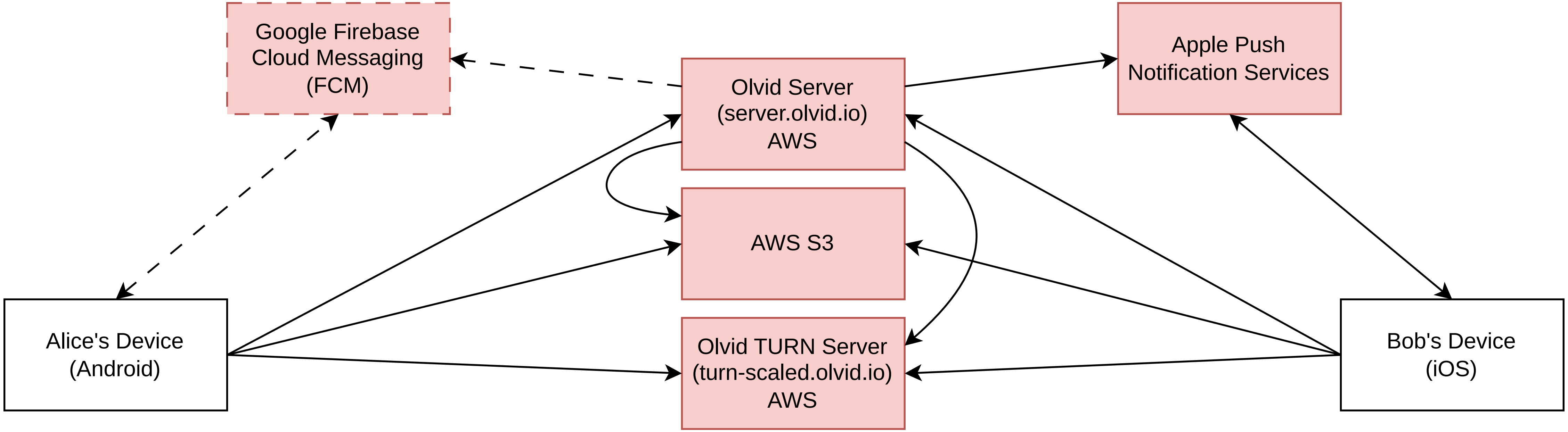
Fantasized decentralized network



How Signal works


Google Firebase Cloud Messaging (FCM)
Signal Server (chat.signal.org) AWS
Apple Push Notification Services
Signal Storage Server (storage.signal.org) GCP
Alice's Device (Android)
Signal CDN (cdn.signal.org) Cloudfront
Bob's Device (iOS)
Signal CDN (cdn2|3.signal.org) Cloudflare
Signal TURN Server (turn.cloudflare.com?) Cloudflare

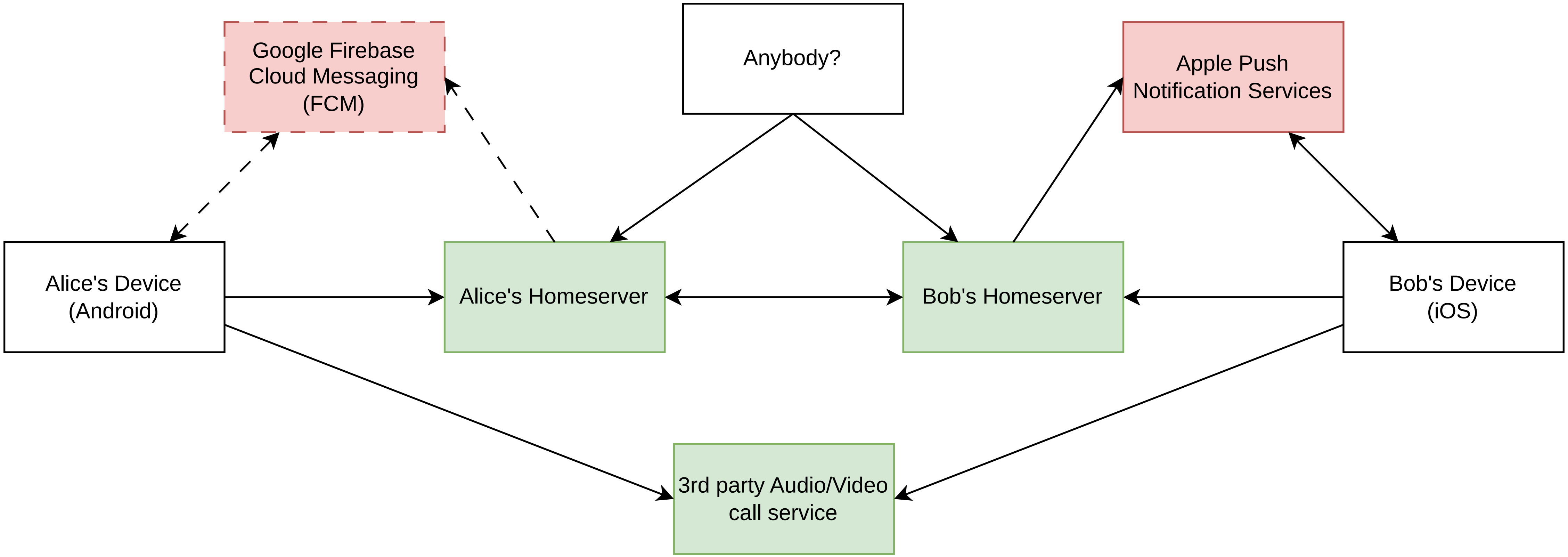
How Olvid works


Google Firebase Cloud Messaging (FCM)
Olvid Server (server.olvid.io) AWS
Apple Push Notification Services
AWS S3
Alice's Device (Android)
Olvid TURN Server (turn-scaled.olvid.io) AWS
Bob's Device (iOS)

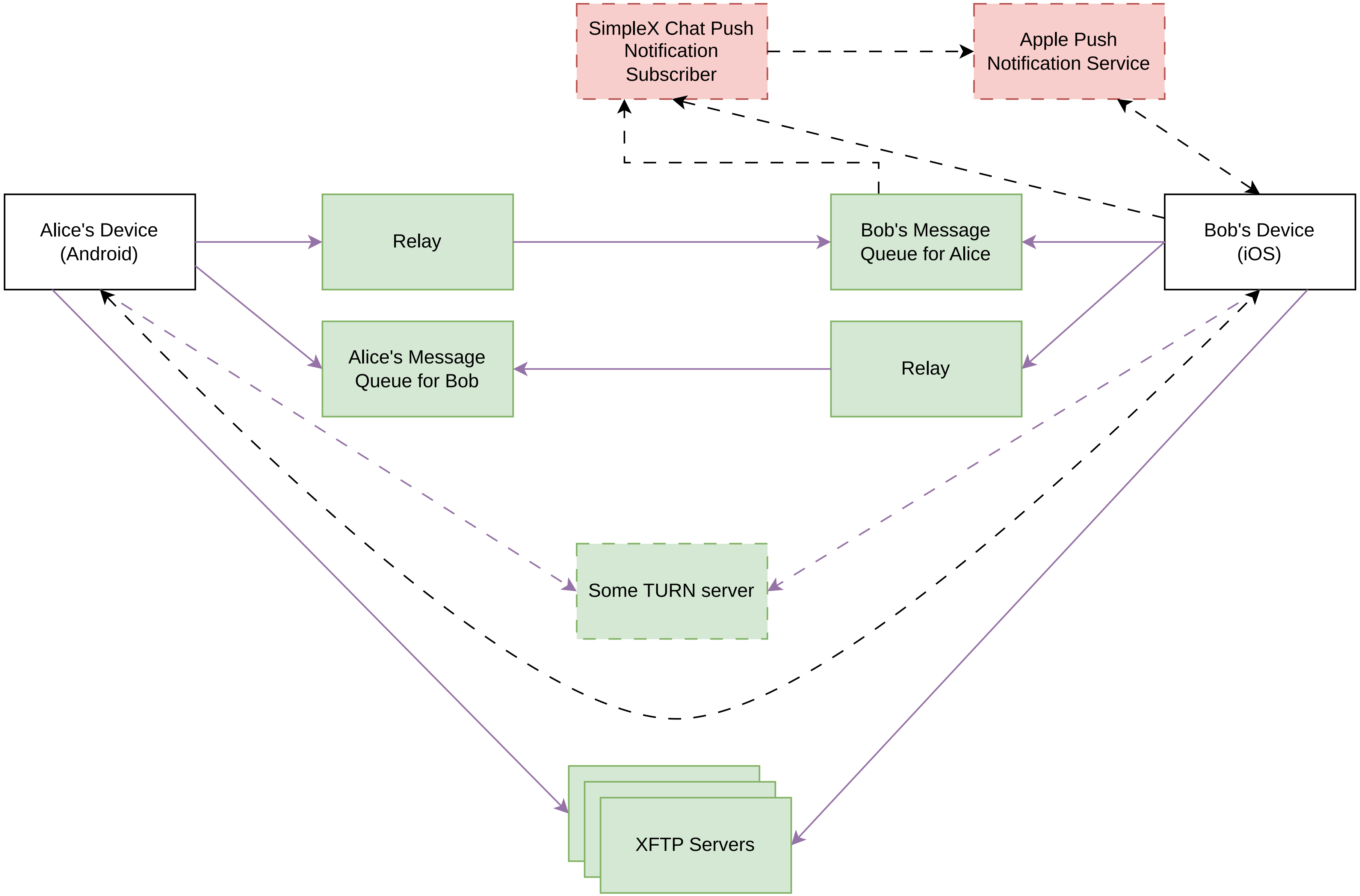
Olvid has provisions and plans to go decentralized

No release date

How Matrix works


Google Firebase Cloud Messaging (FCM)
Anybody?
Apple Push Notification Services
Alice's Device (Android)
Alice's Homeserver
Bob's Homeserver
Bob's Device (iOS)
3rd party Audio/Video call service

How SimpleX Chat works


SimpleX Chat Push Notification Subscriber
Apple Push Notification Service
Alice's Device (Android)
Relay
Bob's Message Queue for Alice
Bob's Device (iOS)
Alice's Message Queue for Bob
Relay
Some TURN server
XFTP Servers

Why does the infrastructure/architecture matter?

Decentralization == Fragmentation of the user space

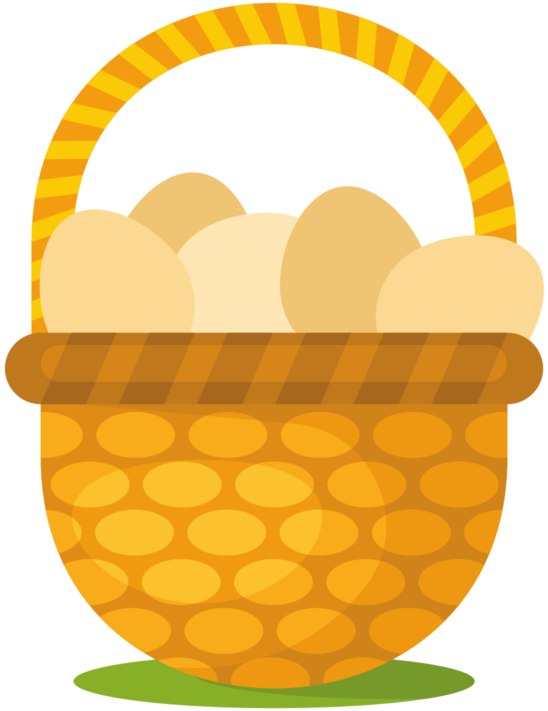
Advantages:

- reduced vision
- different jurisdictions

Drawbacks:

- server-to-server API exposure
- reduced anonymity set

Anonymity set: a group whose members are indistinguishable



About empty privacy claims

Vendors make privacy claims and sometimes publish transparency reports:

- "you cannot disclose what you don't have"

Legitimate questions:

- **What about information collected under a gag order?**
- **What about your subcontractors?**
- **What about cross-referenced information?**

Identities

Identities during registration

Signal requires a phone number.

Phone numbers can often be linked to real world identities.

Olvid, Matrix and SimpleX Chat have no such requirements.

Identifiers to get in touch with someone else

- Signal: usernames and optional phone number
- Olvid: identifier composed of an URL and two public keys
- Matrix: matrix ID (email-like) and optionally an email address or a phone number
- SimpleX Chat: a temporary message queue URL

Frog put the cookies in SGX "There," he said. "Now we will not eat any more cookies."

"But we can use side channels," said Toad.

"That is true," said Frog.

Multiple identity support

- Signal: no
- Olvid: yes
- Matrix: yes (e.g. via FluffyChat)
- SimpleX Chat: strong yes (unique ID per contact)

User profiles

- Matrix: stored in clear on the homeserver
- Signal: encrypted and stored on storage server
- Olvid: encrypted and stored on Olvid's servers
- SimpleX Chat: encrypted and sent during session handshake

Server account

Used to authenticate some API calls (e.g. "give me my new messages")

- Signal: login/password
- Matrix: bearer token
- Olvid: public key signature (from identity)
- SimpleX Chat: public key signature

Attachments

- Signal: login/password for upload; no authentication on download
- Olvid: public key signature (from identity) for upload; no authentication on download
- Matrix: bearer token on their respective homeservers
- SimpleX Chat: public key signature (one throwaway key per chunk for the sender and one throwaway key per chunk and per recipient)

Push Notification and linkability to a real world identity

FCM registration ID and APNS application ID:

- tied to a Google/Apple account
- must be stored by the Signal and Olvid servers, by the recipient homeserver with Matrix and the push notification subscriber with SimpleX Chat

Deanonymization can happen by cross referencing the messaging identity and the Google/Apple account.

The unavoidable common identifier: the source IP address

The source IP address:

- common identifier observed by all servers and technical partners
- unavoidable

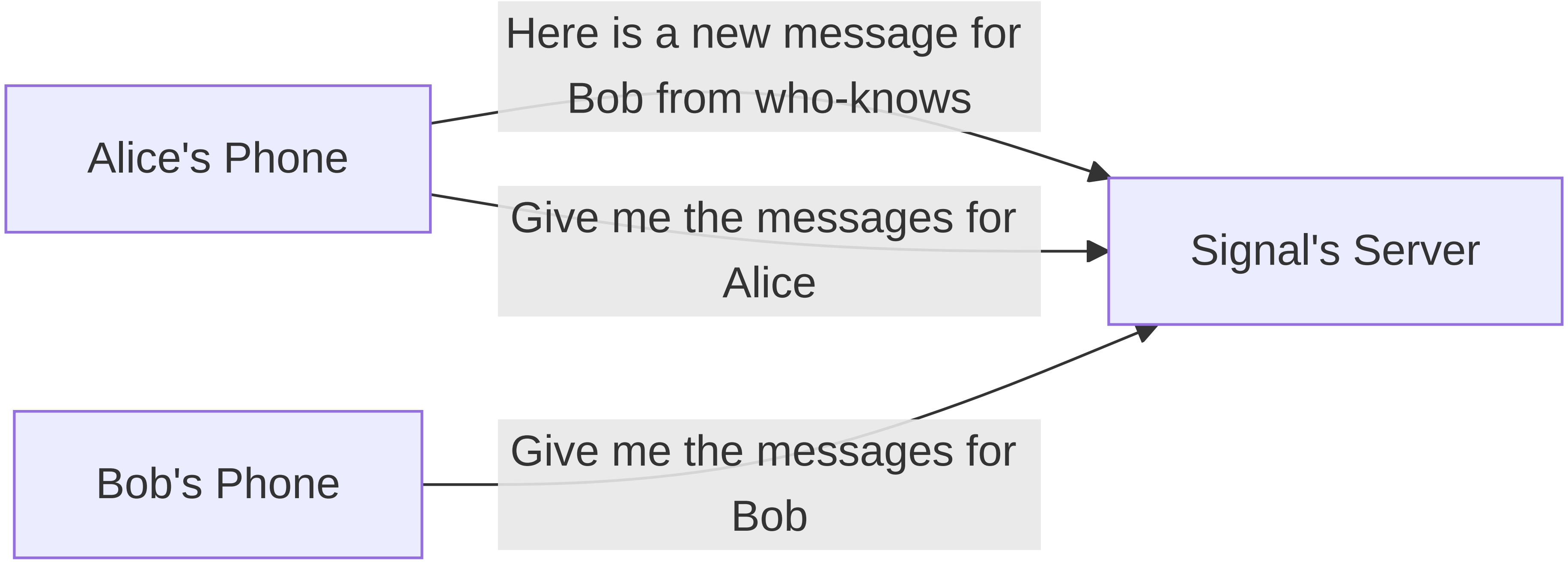
Anonymity set:

- in a household: 1 to 10 members
- behind a VPN/Tor: unknown; from 1 to thousands

Smoke and mirrors: unauthenticated API endpoints

Signal and Olvid tries to hide message sender identities.

Signal calls it "Sealed Senders".

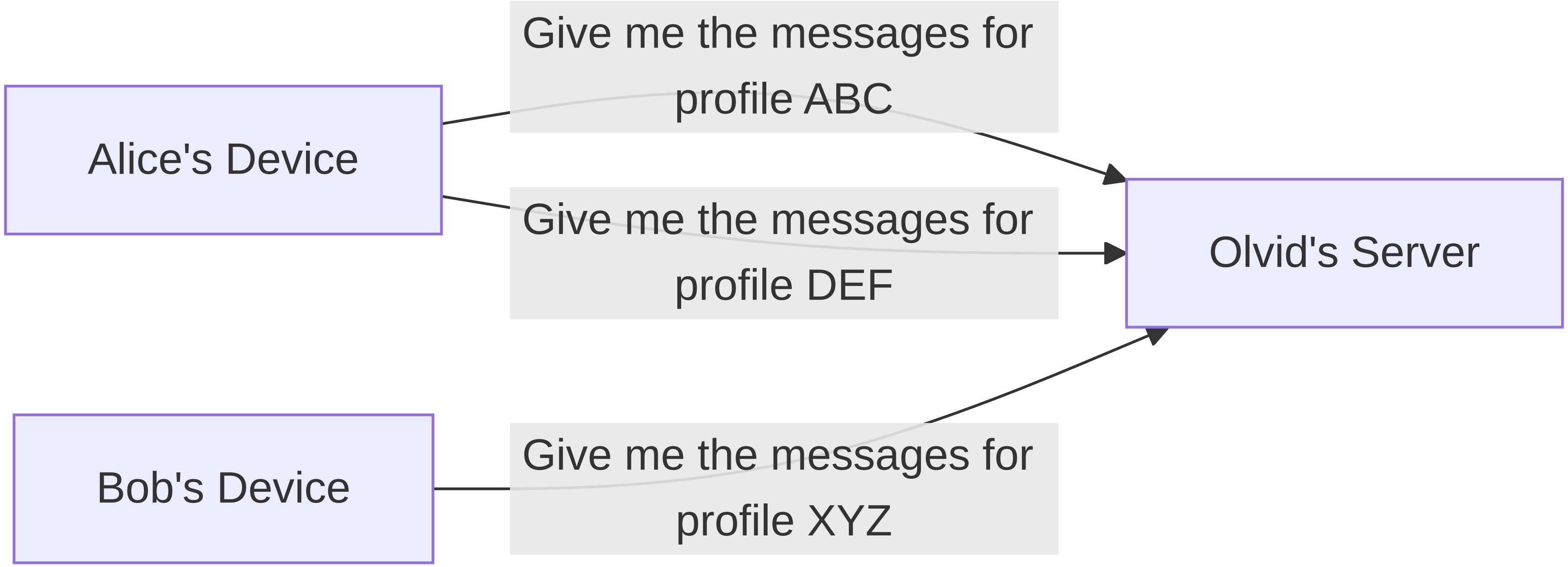

Alice's Phone
Here is a new message for Bob from who-knows
Give me the messages for Alice
Signal's Server
Bob's Phone
Give me the messages for Bob

Linking "anonymous" API calls to authenticated API calls is trivial using the source IP address.

Even more smoke and mirrors: multiple profiles

Olvid and Matrix multiple profiles are trivially linked by the source IP address.

These multiple profiles do not fragment the social graph.


Alice's Device
Give me the messages for profile ABC
Give me the messages for profile DEF
Olvid's Server
Bob's Device
Give me the messages for profile XYZ

Transport Isolation and Relays

SimpleX Chat implements transport isolation:

- uses different sockets per chat profile, session, server or contact
- uses Tor SOCKS proxy extension for stream isolation

Relays can be used to hide the sender IP address from the recipient queue server

Message Metadata

Padding

- Matrix: no padding
- Signal, Olvid and SimpleX Chat pad messages, attachments, etc.

Reactions

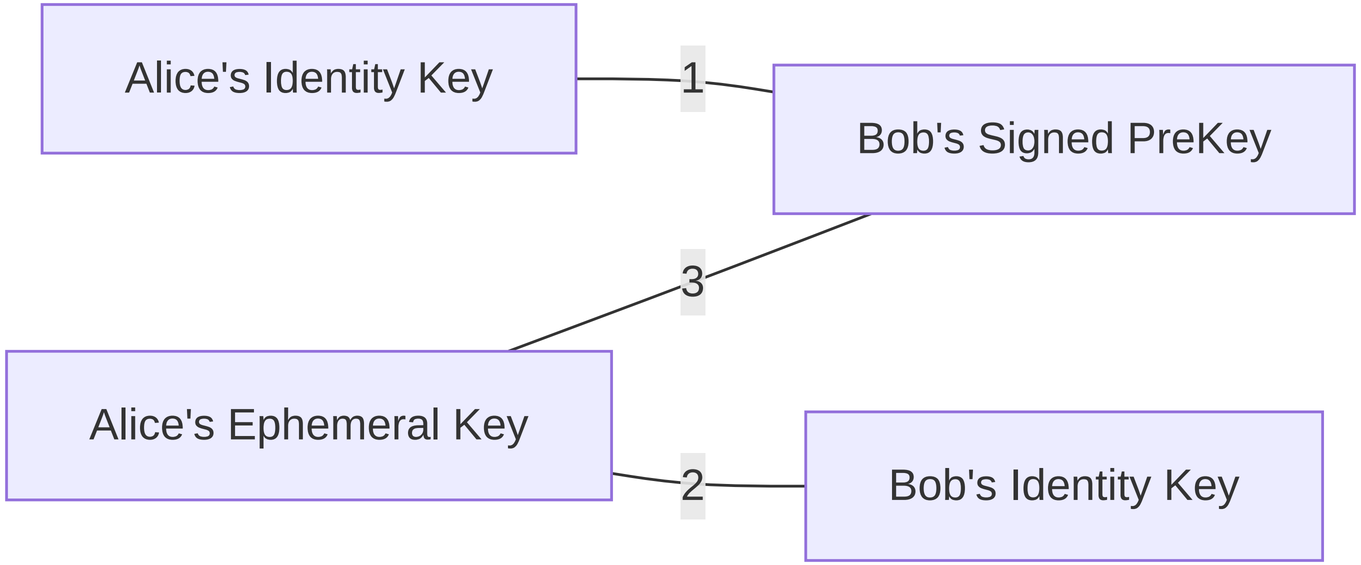
Matrix reactions (emojis) to a message are stored in clear text

Reactions are E2EE standard messages for Signal, Olvid and SimpleX Chat

More advanced inference and corrolation attacks

Who talks to who? (1) - Prekeys

Prekeys/One-time keys (X3DH, PQXDH…) are fetched during session handshake.



Prekeys are stored on Signal, Olvid, and Matrix homeservers.

Different prekeys per device.

SimpleX Chat prekeys are 100% out-of-band.

Who talks to who? (2) - Traffic corrolation and receipts

- Matrix: encodes the sender in cleartext
- Signal and Olvid: delivery receipts may be used to deanonymize senders. Delivery receipts cannot be deactivated
- SimpleX Chat: uses separate one-way queues and delivery receipt can be disabled

Who talks to who? (3) - Push Notifications

Matrix Push Notifications contains the roomId in clear text

Olvid only shares an "identity mask" in clear.

Signal encrypts everything.

SimpleX Chat may send nothing.

Who talks to who? (4) - Support servers

Attachment servers (Signal, Olvid and SimpleX Chat) and TURN servers (Signal, Olvid and optionally SimpleX Chat) can see both participants

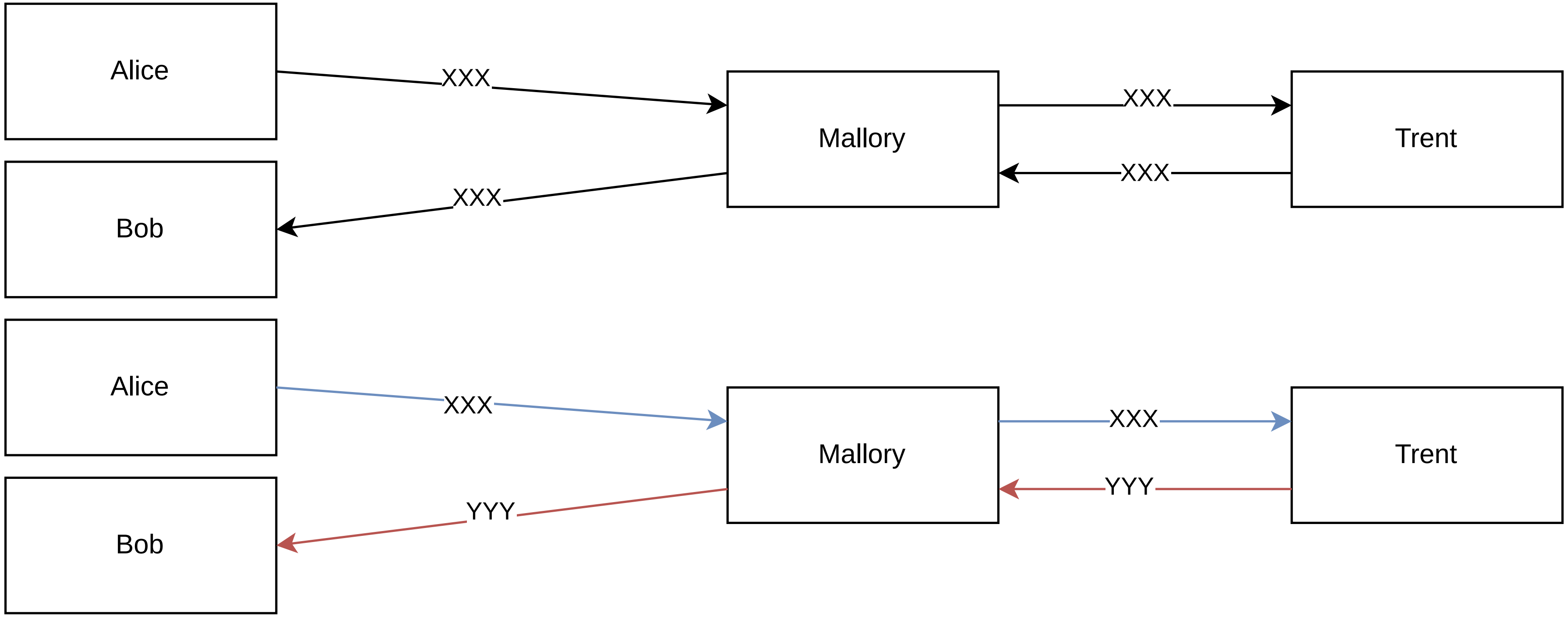
Signal's Storage server can observe who downloads which (encrypted) profiles

=> gives away which conversations were accepted

Who talks to who? (5) - Ciphertext linkability

Signal, Olvid and Matrix have implicit trust in their TLS stack.

Next Heartbleed: ciphertext linkability issues



SimpleX Chat superencrypts E2EE messages for each recipient

Who talks to who? (6) - Groups

Matrix groups:

- cleartext structures on homeservers

Signal and Olvid groups:

- encrypted blob stored on the server
- members must download it
- admins can update it

SimpleX Chat:

- decentralized; similar to Signal and Olvid group v1
- no different than standard messages

Conclusion

Marketing bullshit and security theater

Many "privacy features" only holds in a vacuum

- source IP address
- centralized architecture allowing observability and corrolation
- support servers (attachments, storage, notifications, TURN...)
- no defense in depth

Example: Signal's sealed senders

- source IP address corrolation
- delivery receipt corrolation
- can be silently disabled per user by the server

Metadata protection

- encrypt everything
- decentralized
- no permanent identity
- user-configurable support servers
- one-way channels
- limited server storage
- traffic isolation
- no (centralized) push notification services
- optional receipts
- ciphertext unlinkability

Matrix << Signal < Olvid << SimpleX Chat

Questions?

Thank you for your attention.

Bad questions:

- What is THE best messaging application?
  - Depends. What is your threat model?
- What about application XXX?
  - I don't know (yet).
- What is your threat model?
  - Mine does not matter.