Protecting sensitive metadata so it can’t be used for surveillance

-MIT News

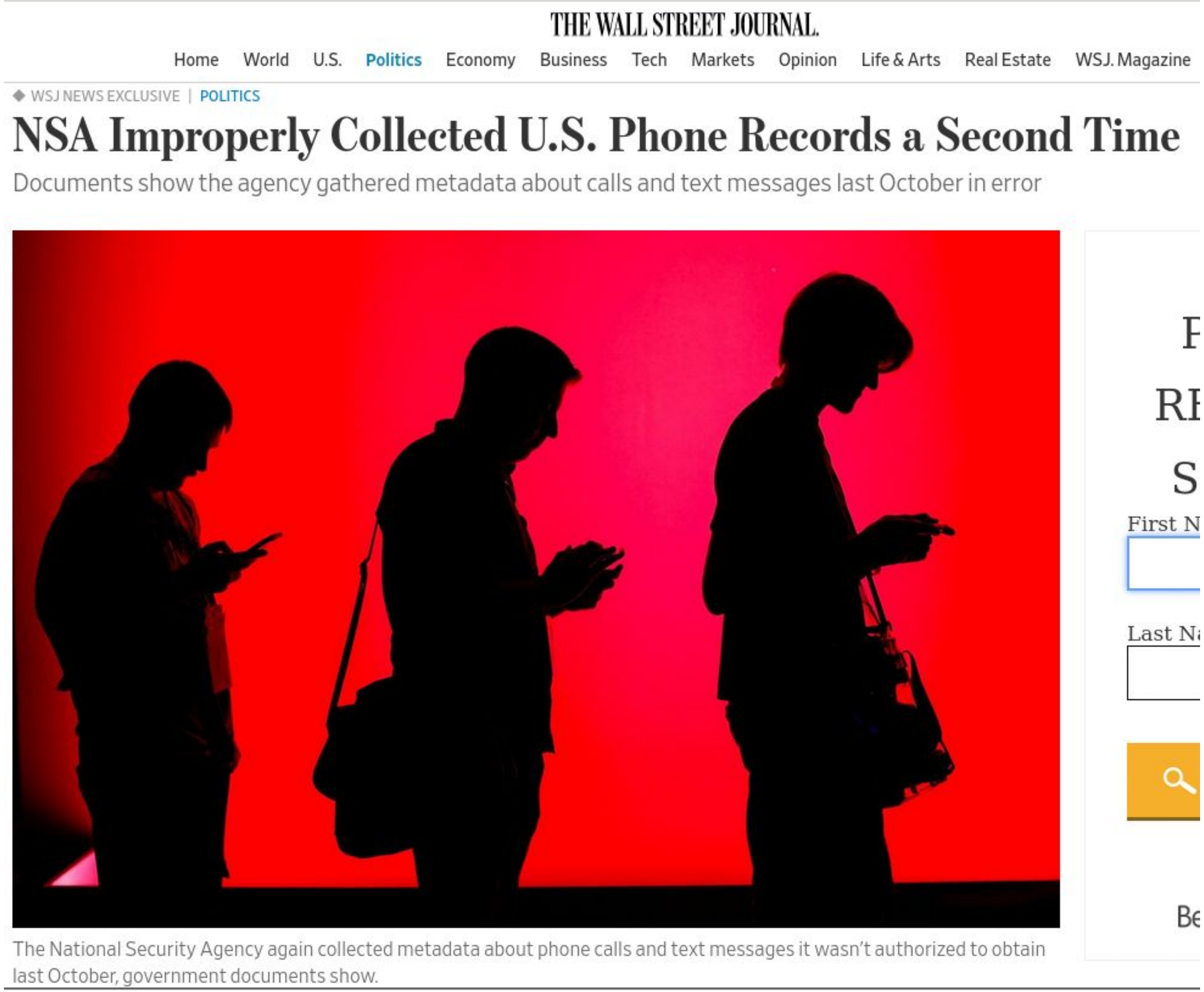
Researchers:

Albert Kwon, PhD thesis
Srinivas Devadas, Advisor
David Lu

Massachusetts
Institute of
Technology

MDCCLXXVI

26/6/2019



NSA Improperly Collected U.S. Phone Records a Second Time

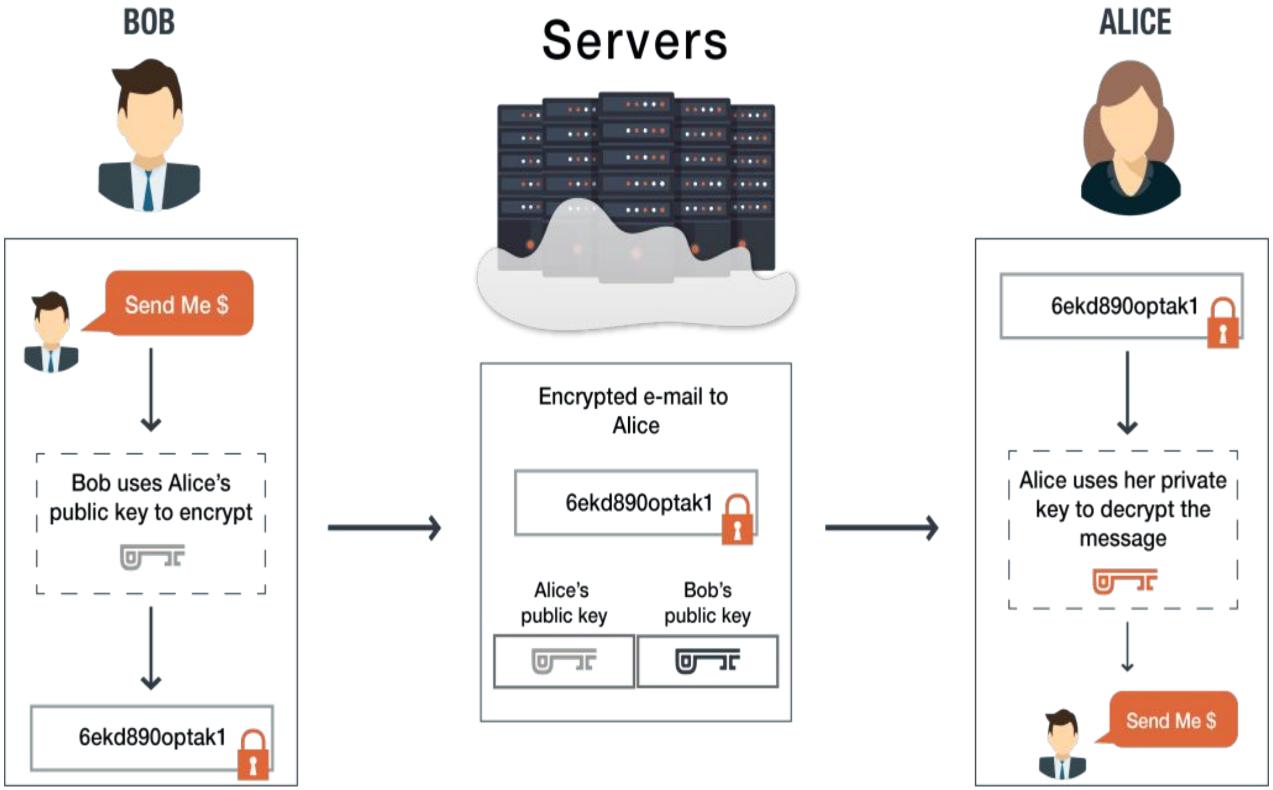
Documents show the agency gathered metadata about calls and text messages last October in error

The National Security Agency again collected metadata about phone calls and text messages it wasn't authorized to obtain last October, government documents show.

End-to-End Encryption (E2EE)

Applications:
HTTPS, Signal, etc.

Protocols:
SSL, TSL, etc.


BOB
Servers
ALICE
Send Me $
Bob uses Alice's public key to encrypt
6ekd890optak1
Encrypted e-mail to Alice
6ekd890optak1
Alice's public key
Bob's public key
6ekd890optak1
Alice uses her private key to decrypt the message
Send Me $

Metadata

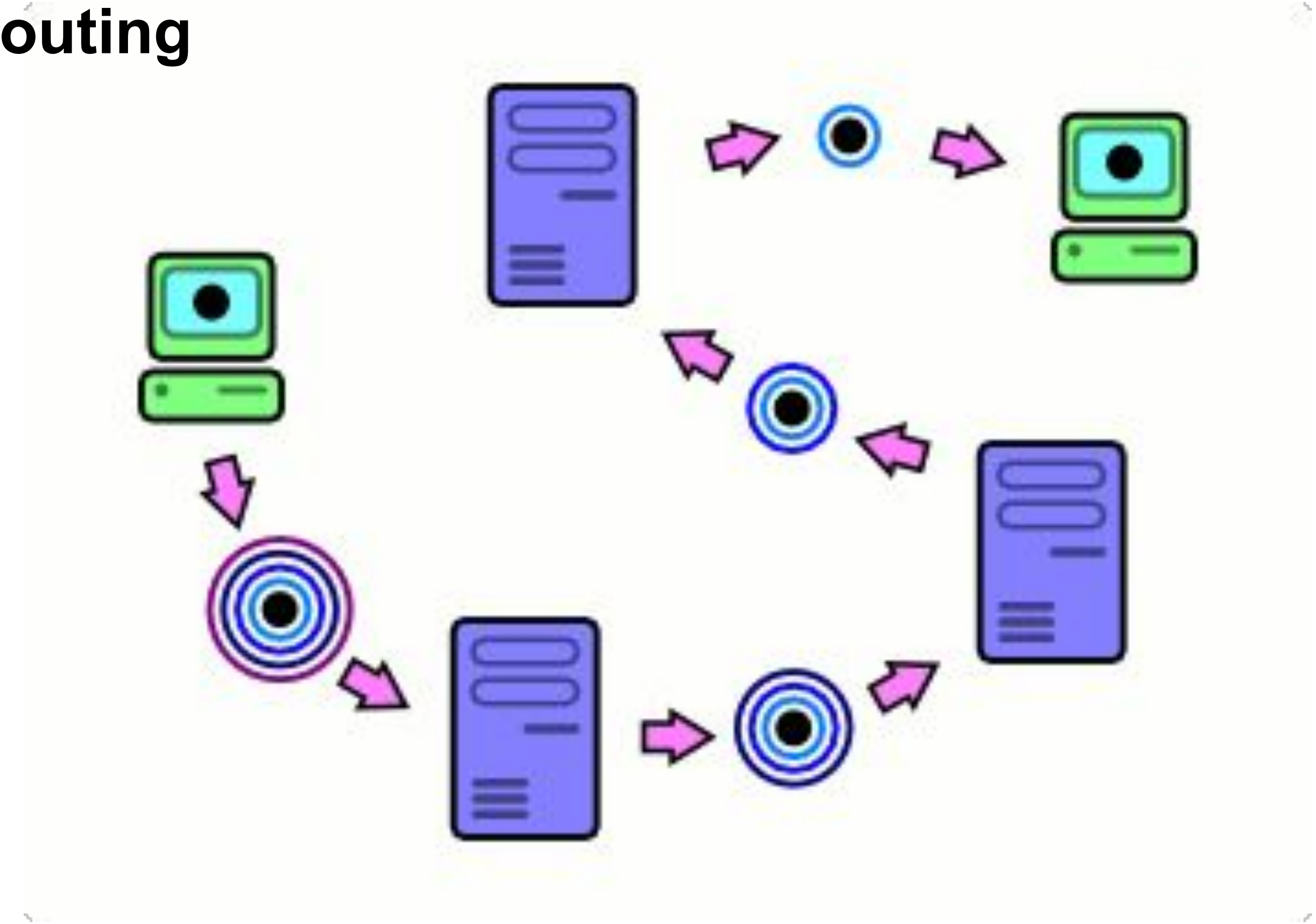
a set of data that describes and gives information about other data.

Examples:

Phone calls and messaging:
    time, duration, correspondents, location
Internet traffic:
    IP addresses (location/identity), time,
    size of information transmitted, protocol (type of information)

While the data itself is often encrypted, the metadata almost never is.
This is mainly due to latency concerns affecting the speed of communication.

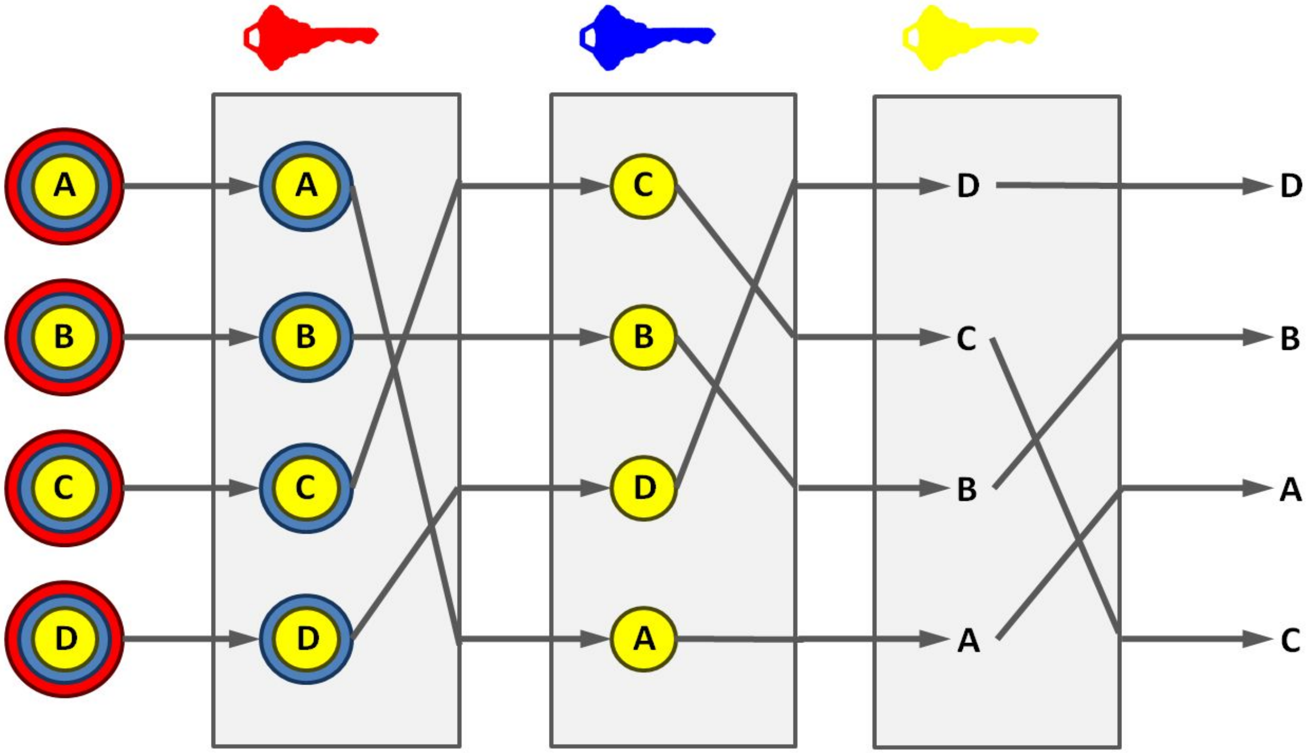
Onion routing



Mixnets

Created in the ‘80s, but have suffered from 3 main flaws:

- Overhead
- Scalability
- Security



XRD (Crossroads)

A fast(er), scalable metadata private messaging system that provides cryptographic privacy. Implemented using a novel technique called “aggregate hybrid shuffle.”

Much faster than other mixnet implementations:
Using 100 servers supports 2 million users with 223 seconds of latency.
This is equivalent to 4x to 13x faster than other comparable systems.

Protects the privacy and integrity of the communication (including metadata), even if all or some of the servers are compromised by an attacker.

How it works, at a high level:

- Every user has a unique mailbox, equivalent to an email address

- In each round of communication, a series of servers, or “mix-chains” is selected based on the user’s public key

- Onion encrypt and wait for other messages

- Shuffle and send along to the next mix-chain

- Mix-chain verifies integrity of the data

- Decrypt outer layer and shuffle, send, repeat until the destination is reached

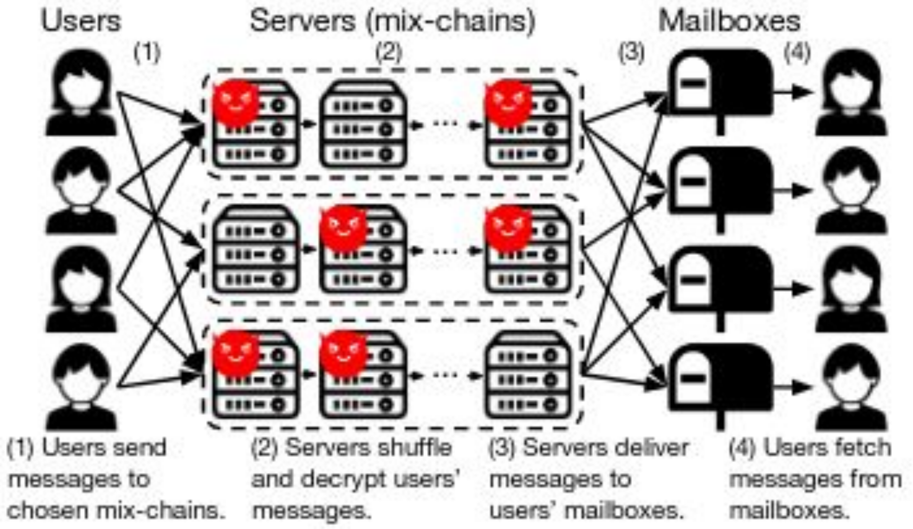

Users
Servers (mix-chains)
Mailboxes
(1)
(2)
(3)
(4)
(1) Users send messages to chosen mix-chains.
(2) Servers shuffle and decrypt users' messages.
(3) Servers deliver messages to users' mailboxes.
(4) Users fetch messages from mailboxes.

Figure 1: Overview of XRD operation.

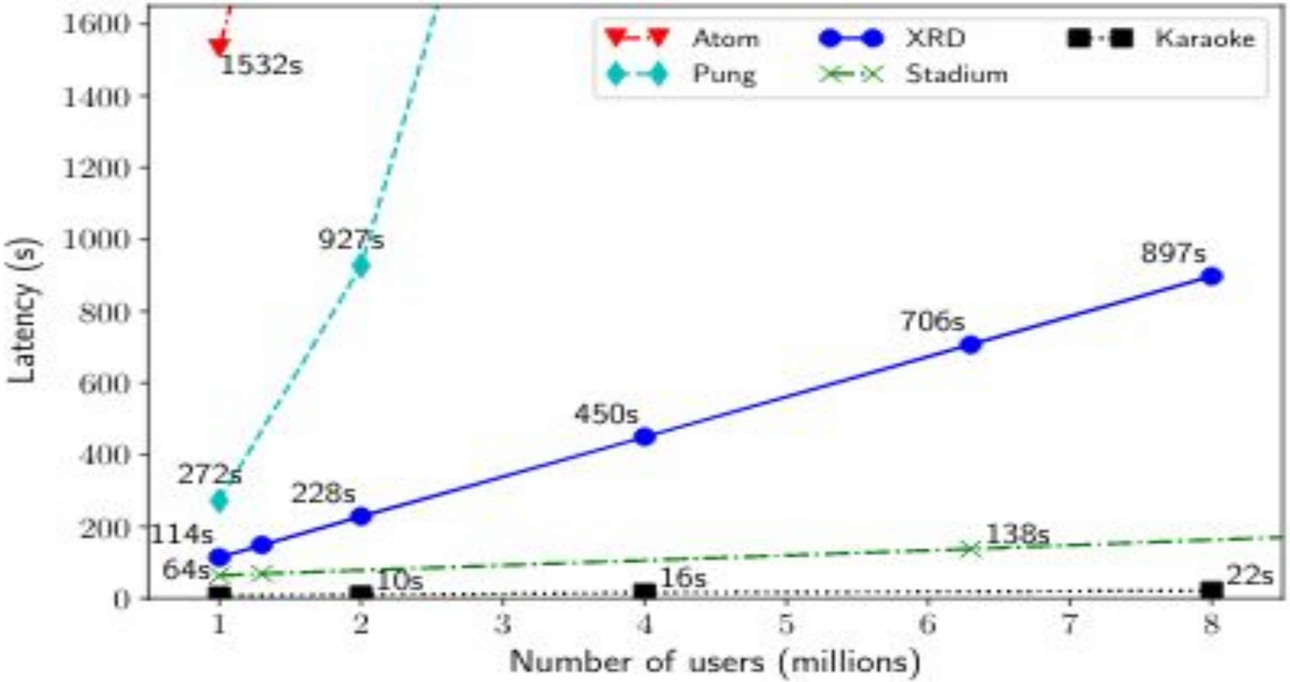


Figure 4: End-to-end latency of XRD and prior systems with varying numbers of users with 100 servers.

Drawbacks

- Non-constant latency increase for increase in number of servers

- Vulnerable to malicious entry server

- Let n and N be the number of chains and servers in the network, respectively. Each user would need to send sqrt(n) messages to ensure all users intersect.

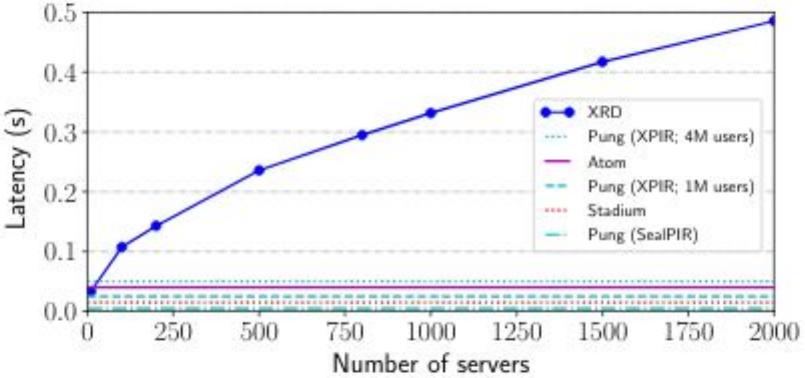


Figure 3: Required user computation as a function of number of servers with a single core. The computation could easily be parallelized with more cores for XRD.

Conclusion

- Not a perfect system, but one of the most secure and scalable systems yet

- Could be used to create a secure method of communication for whistleblowers, journalists, activists, or anyone who doesn’t want to be illegally surveilled

Quextracreditions?