

Anonymous asynchronous messaging

Various techniques & why they are not common

Steyn Hommes, Nick Marx, Vincent Dankbaar

Agenda & scope

- Motivation
- Background
 - Legal
 - Technical
 - Societal
- Approaches
 - Tor & Mix-nets
 - Private information retrieval (PIR)
- Conclusion

Motivation



We can use End-to-end Encryption to hide the content

But the service provider still knows...

- Sender & receiver
- Sender & receiver location
- Send & receive time
- Message length



© 2013 Geek Culture



joyoftech.com

Deduce other properties (data mining)

- Social graph
- Work schedule, schedule of other activities
- What type of message

Motivation: Snowden revelations



- In 2013: Revelations about extensive global internet mass surveillance
- AIVD also participated (intelligence and security agency of the Netherlands)
- Both primary data and metadata



“We Kill People Based on Metadata”

General Michael Hayden, former director of the CIA and NSA

<https://www.youtube.com/watch?v=kV2HDM86Xgl>

Usage of Metadata

- Law enforcement: Members of complex criminal networks
- Behavioural studies
- Dictatorship: Friends of people who are against Government
- Prejudices/profiling

Good or Bad: Depends on context

Background

- Legal basis
- General concepts/terminology
- Technical definitions

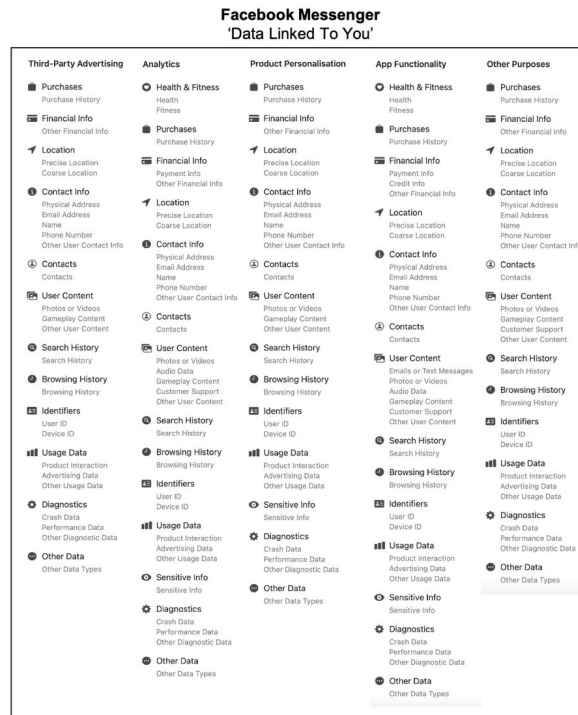
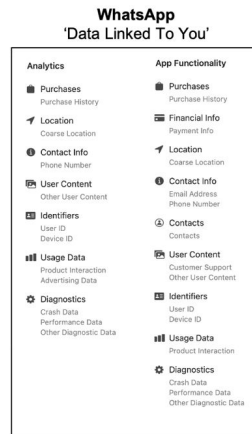
Legal basis

- Article 8 of the European Convention on Human Rights
 - Right to respect for private and family life
- General Data Protection Regulation (GDPR)
 - Article 4: 'personal data' means any information relating to an identified or identifiable natural person
 - Metadata is also protected

Asynchronous messaging apps

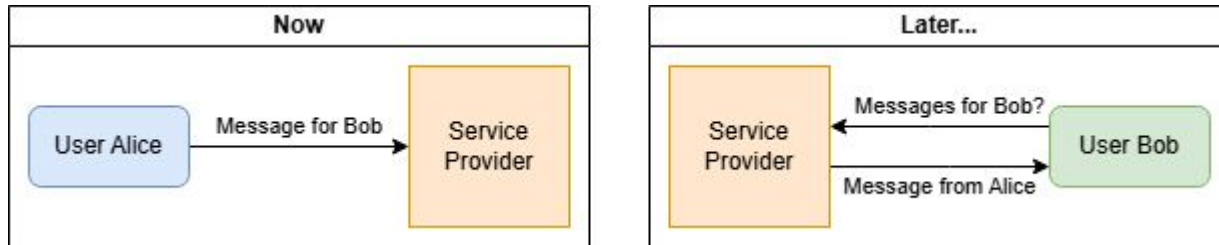
Common examples:

- Signal
- Whatsapp
- Messenger
- iMessage
- Line
- WeChat



Technical definition: Asynchronous messaging

- Asynchronous?
 - Service provider keeps messages stored
 - Messages can be sent or retrieved when the user wants
- Pro: Users don't have to coordinate communication
- Cons:
 - Servers need to operate continuously
 - Traditional asynchronous messaging systems 'require' metadata
- True asynchronous messaging systems do not use synchronized rounds!



Technical definition: Trust models

No-trust
(Hard PET)

Anytrust
(Soft PET)

Threshold trust
(Soft PET)

General concept: The definition of privacy

No clear/unanimous definition -
we adopt the *Privacy Topology*

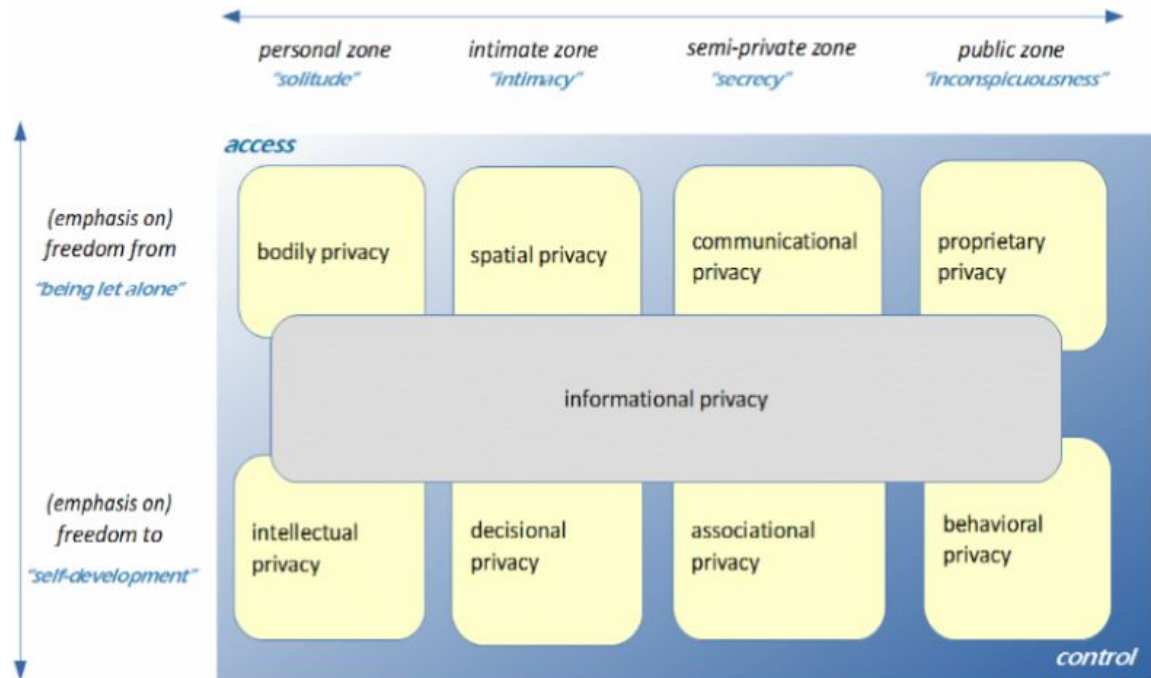
Our interest:

- *Communicational* privacy
- *Informational* privacy
- Freedom from “being let alone”

Concretely:

Hide messages from those
considered unauthorized to read
them

- Is this enough?



Privacy topology by Koops et al. [23]

Technical definition: Anonymity

The anonymity of a subject can be defined¹ as:

‘A subject is not identifiable - not uniquely characterized - within a set of subjects (anonymity set)’

- **Sender anonymity:** identity of sender is hidden
- **Receiver anonymity:** identity of receiver is hidden
- **Unlinkability of sender/receiver:** sender & receiver communication cannot be identified
- **Sender unobservability:** adversary cannot learn which sender sent a message

Technical definition: Threat models

Global Passive Adversary (GPA)

- Monitor all incoming/outgoing traffic
- Perform traffic analysis (correlation etc.)
- Does not control nodes

Global Active Adversary (GAA)

- Can monitor and manipulate traffic
- May control/compromise network
- Control traffic

Local Passive Adversary (LPA)

- Monitor incoming/outgoing traffic of a single node
- Does not control node

Local Active Adversary (LAA)

- Can monitor and manipulate traffic of a single node
- May control/compromise node

Approaches

- What we are not talking about
- Tor & Mix-nets
- Private information retrieval (PIR)

What we are not talking about

- Dialing protocols
- DC-Nets
- Multi party computation
- Distributed point functions
- Identity based encryption

How could you hide the receiver of a message?

Broadcast!

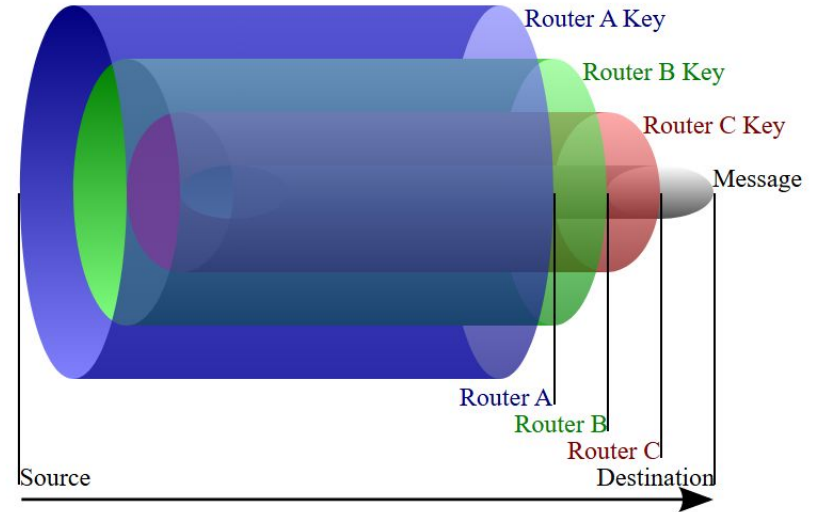
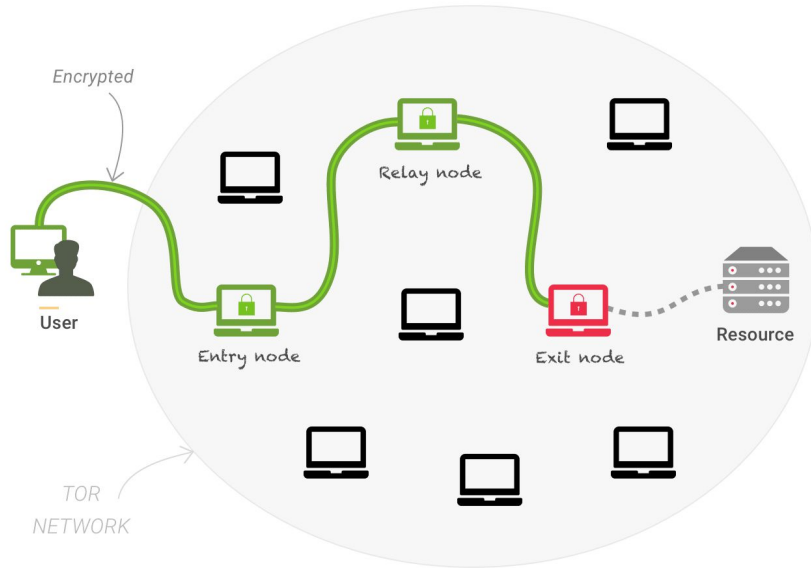
enc(m):

b753b8cc3059717c50288ff822514f8598fc3c606d3b0f02cdb7f778b5a0bbc14908
df2a640f02a8b674640c5b0b57eba0396fcd1642e6744eaf597adee18a447a8fa45e

Inefficient!

Different systems are built on that idea, like P5

Tor



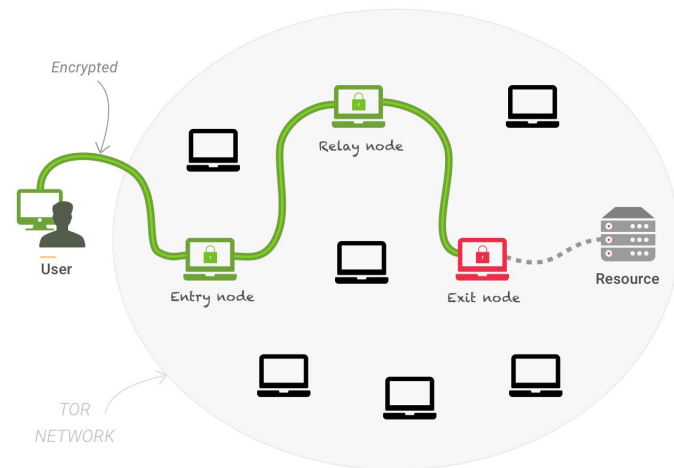
Pros and cons- Tor

Pros:

- Provides privacy and anonymity by preventing linkability for a single-node observer
- Scalable (horizontally → adding more nodes)
- Low latency
- Congestion control

Cons:

- No protection to traffic tampering (e.g. delay traffic)
- Tradeoff for low latency implies choosing best circuit → increases probability of linkability
- **Not resilient against GPA & GAA**
 - Recall our assumption; mainstream service provider could acts as GAA

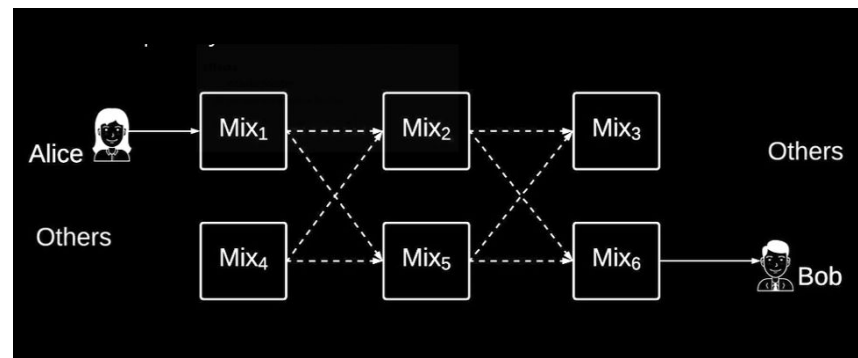


Mixnets

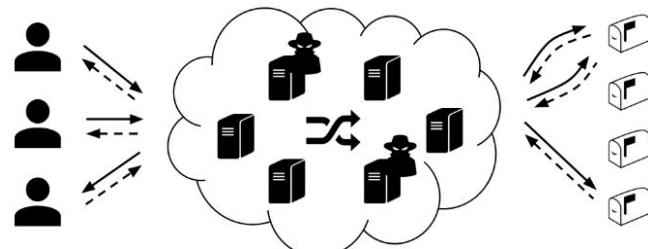
Mixnets

Mixnet characteristics:

- Use *relay* servers for anonymous communication
 - Layered/Onion (TOR-like) encrypted
- Batching & shuffling
 - Break incoming-outgoing traffic link
 - Sending with random delays
 - Shuffle to randomize message order
- Add noise/cover traffic



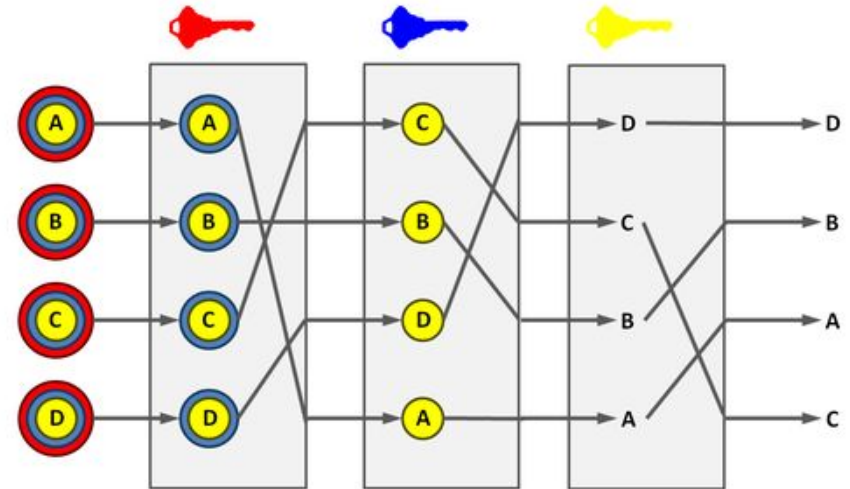
From: <https://constructiveproof.com/posts/2020-02-17-a-simple-introduction-to-mixnets/>



Layered/Onion encryption

- Recall Tor intuition
- Resist against LPA analysis

Issue?

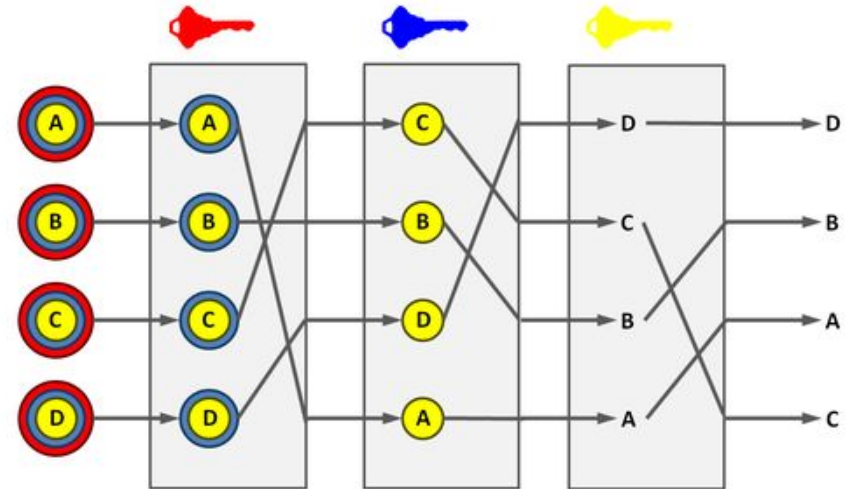


Layered/Onion encryption

- Recall Tor intuition
- Resist against LPA analysis

Issue?

- Latency



Batching and shuffling

Batching

- Send messages at the same time → resist timing incoming/outgoing traffic
 - Create a larger anonymity set
- Send message batches in rounds

Batching and shuffling

Batching

- Send messages at the same time → resist timing incoming/outgoing traffic
 - Create a larger anonymity set
- Send message batches in rounds

Shuffling

- Change order of queued/batched messages → mitigate analysis

Batching and shuffling

Batching

- Send messages at the same time → resist timing incoming/outgoing traffic
 - Create a larger anonymity set
- Send message batches in rounds

Shuffling

- Change order of queued/batched messages → mitigate analysis

Is this asynchronous? Recall our definition

- Users don't have to coordinate communication (can be offline)
- System operates continuously
- System does not depend on synchronized rounds for security*

Batching and shuffling

Batching

- Send messages at the same time → resist timing incoming/outgoing traffic
 - Create a larger anonymity set
- Send message batches in rounds

Shuffling

- Change order of queued/batched messages → mitigate analysis

Is this asynchronous? Recall our definition

- Users don't have to coordinate communication (can be offline)
 - System operates continuously
 - System does not depend on synchronized rounds for security*
-
- Prior work often operates in synchronized rounds (e.g. Vuvuzela) → offline users cannot receive messages

Solution?

- Have a service provider (SP) take care of clients accessing the network (e.g. Loopix, Groove)

Adding noise/cover traffic

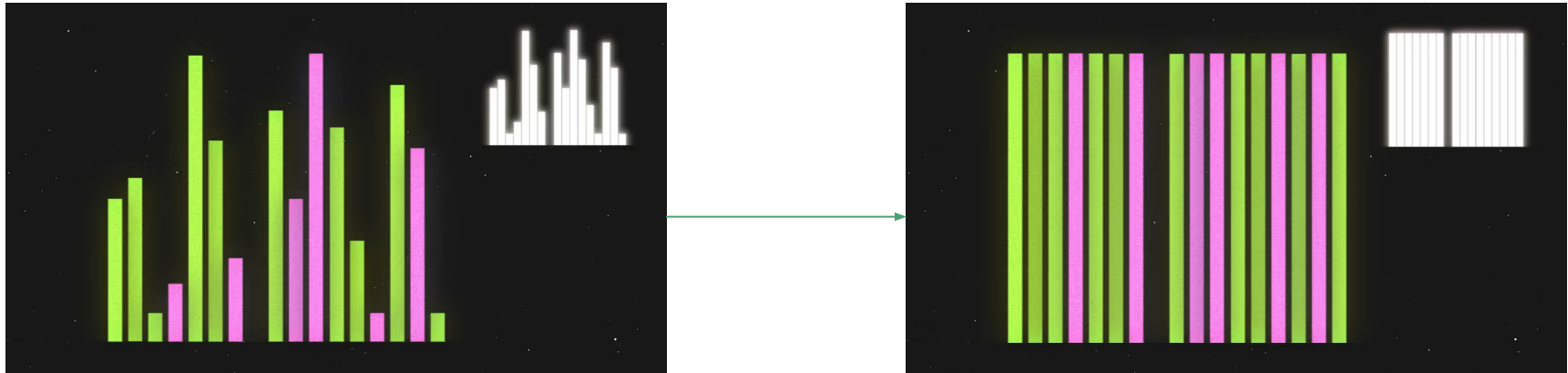
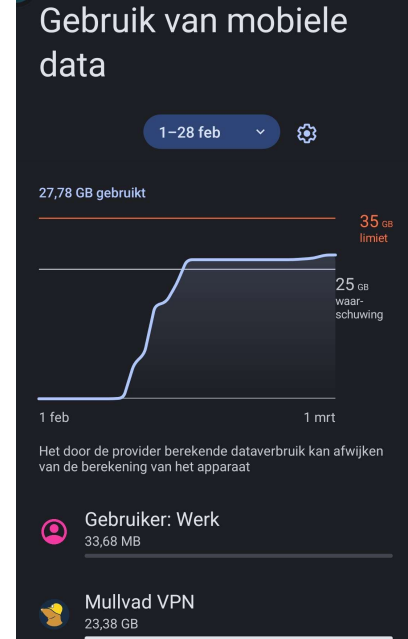
Noise/cover traffic...

- May improve sender/receiver unlinkability

But also...

- Provides limited protection against GPA over time
- Costs a lot of bandwidth

From: <https://mullvad.net/en/vpn/daita>



Proposal highlight:



Groove: Flexible Metadata-Private Messaging

Ludovic Barman, *EPFL*; Moshe Kol, *Hebrew University of Jerusalem*;
David Lazar, *EPFL*; Yossi Gilad, *Hebrew University of Jerusalem*;
Nickolai Zeldovich, *MIT CSAIL*

<https://www.usenix.org/conference/osdi22/presentation/barman>

Groove characteristics

- Strong threat model - GAA
- Flexible; support multiple devices
 - Portable devices (mobile phones)
- *Oblivious delegation*
- Forward secrecy
 - GAA can't learn from past communications

Goals:

- Achieve *differential privacy*
- Support millions of users with each, many contacts
- Availability of other service providers remain, if one fails

Groove characteristics

Strong threat model

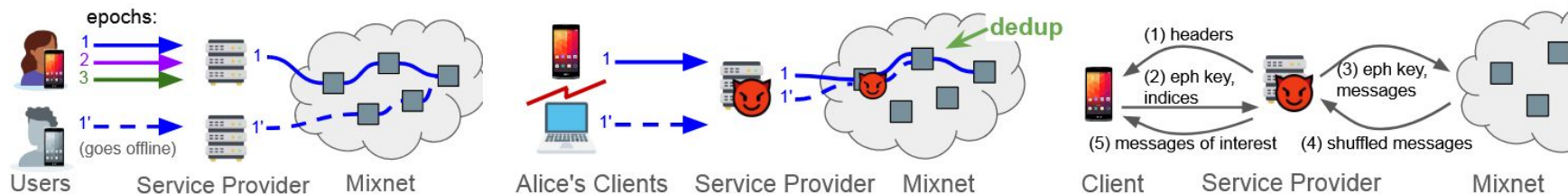
- Global Active Adversary (GAA)
 - Controls all network links
 - Observes when client (dis)connects
 - Can run arbitrary many clients
 - Can observe IP & geographic locations

Groove - design

- Send/receive messages over a *circuit*
 - Fixed route in the mixnet
 - Persists for an epoch
 - In rounds (30s-60s)
- Users exchange messages at *dead drop*
 - “Meeting point” - ephemeral address to deposit messages
- Add 2 types of noise
 - *Doubles*: “simulate” dead drop → mask number of relationships
 - *Singles*: mask when a client does not create a circuit (e.g. being offline)

Groove - basic communication flow

1. Alice and Bob add each other to address book
2. Establish fresh shared secret
 - a. Authenticate user, E2EE, agree on dead drop
 - b. Sharing this secret is out of scope (see dialing protocol)
3. *Oblivious delegation* to SP
 - a. Choose an SP to store messages
 - b. Participate on behalf of client
4. SP send message to mixnet periodically
5. Mixnet shuffles and exchanges message at dead drop
6. Mixnet forwards message back to SP
7. User retrieves message from SP (from 1 device)



Groove - highlights

- Trust models:
 - *No trust* for SP
 - Send loopback message to check for malfunctioning (based on Loopix)
 - *Threshold trust* for mixnet
 - Resist $f = 20\%$ malicious servers (by default)
- Client may choose schedule interval; good for low power-devices (vs Vuzela recommending being always online)
- Multi device support by: synch user's contact through SP on start
 - Mixnet server removes duplicate messages
- Groove protects against rogue SP
 - Address book is padded to fixed length
 - Onion encrypted messages contain epoch number; honest server drops wrong number
 - Tagging messages in circuits → honest server drops duplicates → no unusual amount of accessing dead drop

Groove - performance comparison

Vuvuzela

- 1 million users, 1 contact, 37s latency

Stadium

- Latency in order of minutes (Zero-knowledge proofs)

Karaoke

- 1 million users, 1 contact, 7s of latency

Yodel*

- 1 million users, 750 ms
 - But connects contacts to dead drop without mixnet

Groove

- 1 million users, 50 contacts, 32s latency

Private information retrieval (PIR)

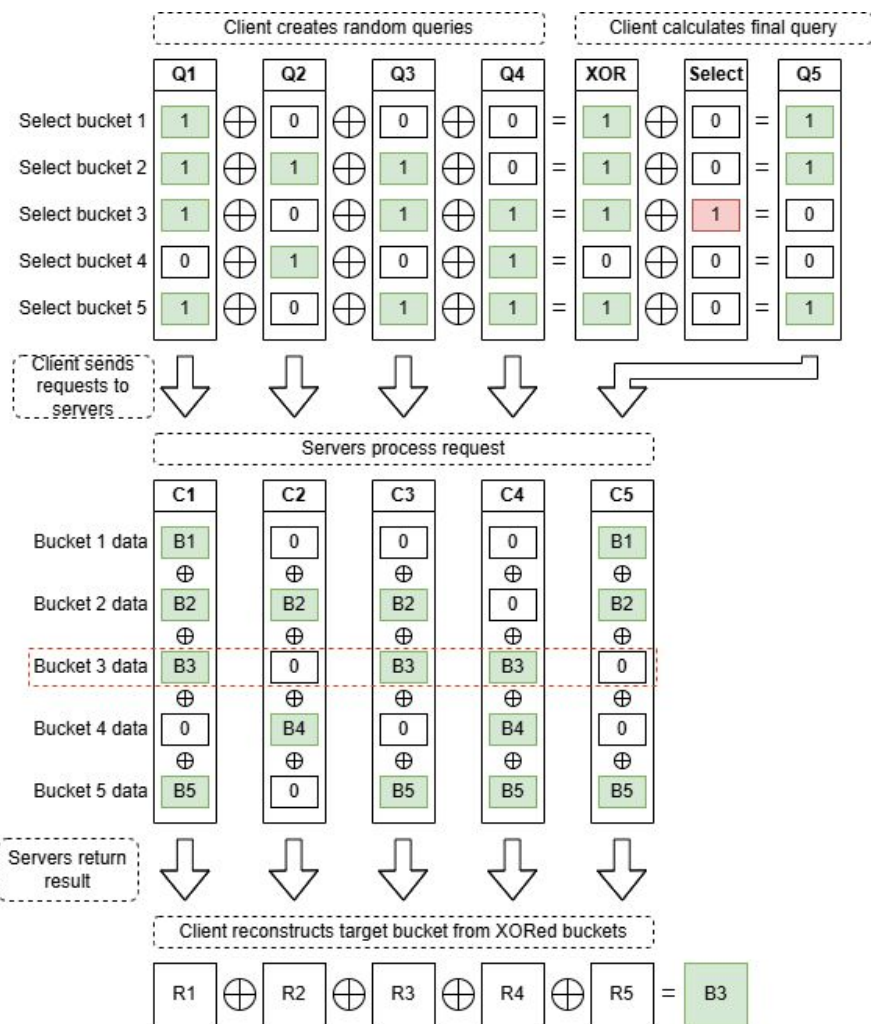
Privacy is expensive

Private information retrieval (PIR)

- Fetching data from a server
- Information theoretic PIR (IT-PIR)
- Computational PIR (C-PIR)
- Receiver anonymity
 - Cannot link data to user
- Generally expensive
- Pseudonymous mailbox architecture

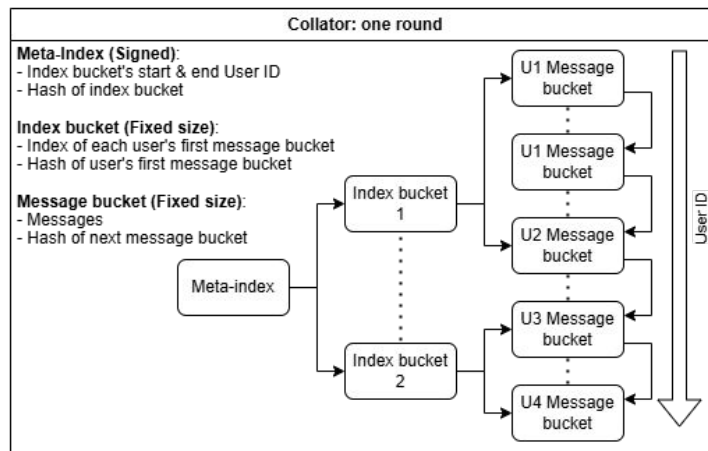
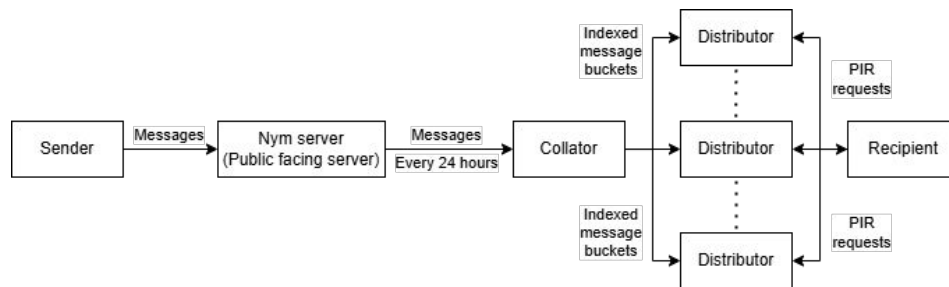
Information Theoretic PIR (IT-PIR)

- Anytrust model
- Resists computationally unbounded adversary
- Cheap XOR operations
- Requires multiple servers
- Redundancy weakens trust model



IT-PIR example: Pynchon gate

- Sender anonymity: Mix net
- Receiver anonymity: IT-PIR
- Mailbox architecture
- Round based
- Not perfectly asynchronous



Computational PIR (C-PIR)

- No-trust model
- Homomorphic encryption
- Only one server needed
- Very expensive

C-PIR example: Pung

- No sender anonymity
- Receiver anonymity: C-PIR
- 1 send & receive request per round
- Optimizations: BST and batch-codes

Actual usage of these technologies

No actual messaging app uses PIR or mix-nets



Biar: Tor



Session: custom onion-network



SimpleX Chat: optional Tor
(using Signal protocol: PFS, post-quantum encryption)

Conclusion

Costs of protecting metadata is quite high - tradeoffs

- Latency
- Potentially no multi-device support

Transport encryption only really took off after people felt the disadvantages

Maybe it will become more relevant in the future

References

[1] How it works - Briar.

[2] A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management.

[3] AGUILAR MELCHOR, C., BARRIER, J., FOUSSE, L., AND KILLIAN, M.-O. XPIR : Private Information Retrieval for Everyone. *Proceedings on Privacy Enhancing Technologies april 2016* (Apr. 2016), 155–174. Publisher: Privacy Enhancing Technologies Symposium.

[4] ANGEL, S., AND SETTY, S. Unobservable Communication over Fully Untrusted Infrastructure. pp. 551–569.

[5] BARMAN, L., KOL, M., LAZAR, D., GILAD, Y., AND ZELDOVICH, N. Groove: Flexible Metadata-Private Messaging.

[6] BEN-ZE'EV, A. Privacy, emotional closeness, and openness in cyberspace. *Computers in Human Behavior* 19, 4 (July 2003), 451–467.

[7] BORISOV, N., DANEZIS, G., AND GOLDBERG, I. DP5: A Private Presence Service. *Proceedings on Privacy Enhancing Technologies 2015 2015*, 2 (June 2015), 4–24. Number: 2.

[8] CHAUM, D. L. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM* 24, 2 (Feb. 1981), 84–90.

[9] CHENG, R., SCOTT, W., MASSEROVA, E., ZHANG, L., GOYAL, V., ANDERSON, T., KRISHNAMURTHY, A., AND PARNO, B. Talek: Private Group Messaging with Hidden Access Patterns. In *Proceedings of the 36th Annual Computer Security Applications Conference* (New York, NY, USA, Dec. 2020), ACSAC '20, Association for Computing Machinery, pp. 84–99.

[10] COLE, D. 'We Kill People Based on Metadata', May 2014.

[11] CORRIGAN-GIBBS, H., BONEH, D., AND MAZIERES, D. Riposte: An Anonymous Messaging System Handling Millions of Users. In *2015 IEEE Symposium on Security and Privacy* (May 2015), pp. 321–338.

[12] CORRIGAN-GIBBS, H., AND FORD, B. Dissent: Accountable Anonymous Group Messaging. Apr. 2010. Issue: arXiv:1004.3057 arXiv:1004.3057 [cs].

[13] DANEZIS, G., DINGLELINE, R., AND MATHEWSON, N. Mixminion: design of a type III anonymous remailer protocol. In *Proceedings 19th International Conference on Data Engineering (Cat. No. 03CH37405)* (Berkeley, CA, USA, 2003), IEEE Comput. Soc., pp. 2–15.

[14] DAS, D., MEISER, S., MOHAMMADI, E., AND KATE, A. Anonymity Trilemma: Strong Anonymity, Low Bandwidth Overhead, Low Latency - Choose Two. In *2018 IEEE Symposium on Security and Privacy (SP)* (May 2018), pp. 108–126. ISSN: 2375-1207.

[15] ESKANDARIAN, S., AND BONEH, D. Clarion: Anonymous Communication from Multiparty Shuffling Protocols. In *Proceedings 2022 Network and Distributed System Security Symposium* (San Diego, CA, USA, 2022), Internet Society.

[16] ESKANDARIAN, S., CORRIGAN-GIBBS, H., ZAHARIA, M., AND BONEH, D. Express: Lowering the Cost of Metadata-hiding Communication with Cryptographic Privacy. Sept. 2020. Issue: arXiv:1911.09215 arXiv:1911.09215 [cs].

[17] GILBOA, N., AND ISHAI, Y. Distributed Point Functions and Their Applications. In *Advances in Cryptology - EUROCRYPT 2014*, D. Hutchison, T. Kanade, J. Kirtler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, P. Q. Nguyen, and E. Oswald, Eds., vol. 8441. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 640–658. Series Title: Lecture Notes in Computer Science.

[18] GOEL, S., ROBSON, M., POLTE, M., AND SIRER, E. Herbivore: A Scalable and Efficient Protocol for Anonymous Communication. Tech. rep., Cornell University, Jan. 2003.

[19] GOLDBERG, I. Privacy-Enhancing Technologies for the Internet, II: Five Years Later. In *Privacy Enhancing Technologies*, R. Dingledine, P. Syverson, G. Goos, J. Hartmanis, and J. Van Leeuwen, Eds., vol. 2482. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 1–12. Series Editors: —426 Series Title: Lecture Notes in Computer Science.

[20] HOEPMAN, J. Opinions - Privacy is Hard and Seven Other Myths. *European Data Protection Law Review* 9, 2 (2023), 104–111.

[21] HOEPMAN, J.-H. Privately (and Unlinkably) Exchanging Messages Using a Public Bulletin Board. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society* (Denver Colorado USA, Oct. 2015), ACM, pp. 85–94.

[22] KIAYIAN, A., LEONARDOS, N., LIPMAA, H., PAVLYK, K., AND TANG, Q. Optimal Rate Private Information Retrieval from Homomorphic Encryption. *Water Treatment Technology* 2015, 2 (June 2015). Publisher: Sciendo.

[23] KOOPS, B.-J., NEWELL, B. C., TIMAN, T., CHOKREVSII, T., AND GALIC, M. A TYPOLOGY OF PRIVACY.

[24] KWON, A., CORRIGAN-GIBBS, H., DEVADAS, S., AND FORD, B. Atom: Horizontally Scaling Strong Anonymity. In *Proceedings of the 26th Symposium on Operating Systems Principles* (New York, NY, USA, Oct. 2017), SOSP '17, Association for Computing Machinery, pp. 406–422.

[25] KWON, A., LU, D., AND DEVADAS, S. XRD: Scalable Messaging System with Cryptographic Privacy. Jan. 2019. Issue: arXiv:1901.04368 arXiv:1901.04368 [cs].

[26] KWON, A. H., LAZAR, D., DEVADAS, S., AND FORD, B. Rifle: An Efficient Communication System With Strong Anonymity. *De Gruyter* (Dec. 2015). Accepted: 2020-12-10T14:56:13Z Publisher: Walter de Gruyter GmbH.

[27] LAZAR, D., GILAD, Y., AND ZELDOVICH, N. Karaoke: Distributed Private Messaging Immune to Passive Traffic Analysis. pp. 711–725.

[28] LAZAR, D., GILAD, Y., AND ZELDOVICH, N. Yodel: strong metadata security for voice calls. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles* (New York, NY, USA, Oct. 2019), SOSP '19, Association for Computing Machinery, pp. 211–224.

[29] LAZAR, D., AND ZELDOVICH, N. Alpenhorn: bootstrapping secure communication without leaking metadata. In *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation* (USA, Nov. 2016), OSDI'16, USENIX Association, pp. 571–586.

[30] LU, D., YUREK, T., KULSHRESHTHA, S., GOVIND, R., KATE, A., AND MILLER, A. Honey-BudgetMPC and AsynchroMix: Practical Asynchronous MPC and its Application to Anonymous Communication. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, Nov. 2019), CCS '19, Association for Computing Machinery, pp. 887–903.

[31] NISSENBAUM, H. Privacy as Contextual Integrity. *Washington Law Review* 79, 1 (Feb. 2004), 119.

[32] PETZTMANN, A., AND KOHNSTOPP, M. Anonymity, Unobservability, and Pseudonymity — A Proposal for Terminology. In *Designing Privacy Enhancing Technologies*, G. Goos, J. Hartmanis, J. Van Leeuwen, and H. Fodorath, Eds., vol. 2009. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 1–9. Series Title: Lecture Notes in Computer Science.

[33] PIOTROWSKA, A. M., HAYES, J., ELABI, T., MEISER, S., AND DANEZIS, G. The Loopix Anonymity System. pp. 1199–1216.

[34] PULLS, T., AND WITWER, E. Maybenot: A Framework for Traffic Analysis Defenses. In *Proceedings of the 22nd Workshop on Privacy in the Electronic Society* (Copenhagen Denmark, Nov. 2023), ACM, pp. 75–89.

[35] REITER, M. K., AND RUBIN, A. D. Crowds: anonymity for Web transactions. *ACM Transactions on Information and System Security* 1, 1 (Nov. 1998), 66–92. Number: 1.

[36] SASSAMAN, L., COHEN, B., AND MATHEWSON, N. The pynchon gate: a secure method of pseudonymous mail retrieval. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (New York, NY, USA, Nov. 2005), WPES '05, Association for Computing Machinery, pp. 1–9.

[37] SHERWOOD, R., BHATTACHARJEE, B., AND SRINIVASAN, A. P5: A protocol for scalable anonymous communication. *Journal of Computer Security* 13, 6 (Jan. 2005), 839–876. Publisher: IOS Press.

[38] SOLOVE, D. J. "I'VE GOT NOTHING TO HIDE," AND OTHER MISUNDERSTANDINGS OF PRIVACY.

[39] TAVANI, H. T. Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy. *Metaphilosophy* 38, 1 (2007), 1–22. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9973.2006.00474.x>.

[40] TYAGI, N., GILAD, Y., LEUNG, D., ZAHARIA, M., AND ZELDOVICH, N. Stadium: A Distributed Metadata-Private Messaging System. In *Proceedings of the 26th Symposium on Operating Systems Principles* (New York, NY, USA, Oct. 2017), SOSP '17, Association for Computing Machinery, pp. 423–440.

[41] VADAPALLI, A., STORRIER, K., AND HENRY, R. Sabre: Sender-Anonymous Messaging with Fast Audits. In *2022 IEEE Symposium on Security and Privacy (SP)* (May 2022), pp. 1953–1970. ISSN: 2375-1207.

[42] VAN DEN HOOFF, J., LAZAR, D., ZAHARIA, M., AND ZELDOVICH, N. Vuvuzela: scalable private messaging resistant to traffic analysis. In *Proceedings of the 25th Symposium on Operating Systems Principles* (New York, NY, USA, Oct. 2015), SOSP '15, Association for Computing Machinery, pp. 137–152.

[43] WOLINSKY, D. I., CORRIGAN-GIBBS, H., FORD, B., AND JOHNSON, A. Dissent in Numbers: Making Strong Anonymity Scale. pp. 179–182.

Chronological paper graph

