

Express: Lowering the Cost of Metadata-hiding Communication with Cryptographic Privacy

Saba Eskandarian, Henry Corrigan-Gibbs, Matei Zaharia, Dan Boneh

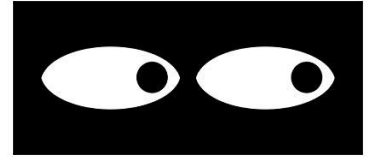
Our Story



Our Story



Our Story



How to Communicate Privately?

Option 1:

End to end encrypted messaging apps

E.g. Signal, WhatsApp

Problem: **metadata**



How to Communicate Privately?

Option 1:

End to end encrypted messaging apps

E.g. Signal, WhatsApp

Problem: **metadata**

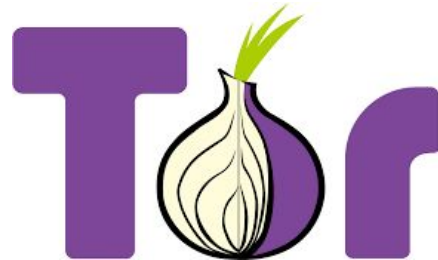


Option 2:

Anonymizing proxy

E.g. Tor, SecureDrop

Problem: **global adversaries**



How to Communicate Privately?

Option 3: Metadata-hiding communication systems with cryptographic privacy

How to Communicate Privately?

Option 3: Metadata-hiding communication systems with cryptographic privacy

E.g. Riposte, Pung, Talek, Karaoke, Atom, XRD, Verdict, Dissent,

How to Communicate Privately?

Option 3: Metadata-hiding communication systems with cryptographic privacy

E.g. Riposte, Pung, Talek, Karaoke, Atom, XRD, Verdict, Dissent,

Drawback: **heavy requirements placed on clients**

- Requirement to run in synchronized rounds
- High communication costs

How to Communicate Privately?

Option 3: Metadata-hiding communication systems with cryptographic privacy

E.g. Riposte, Pung, Talek, Karaoke, Atom, XRD, Verdict, Dissent,

Drawback: **heavy requirements placed on clients**

- Requirement to run in synchronized rounds
- High communication costs

Can we make metadata-hiding communication work for whistleblowing?

Introducing Express

Communication system designed for practical metadata-hiding whistleblowing

Introducing Express

Communication system designed for practical metadata-hiding whistleblowing

Journalists can register mailboxes for sources to send messages/documents

Introducing Express

Communication system designed for practical metadata-hiding whistleblowing

Journalists can register mailboxes for sources to send messages/documents

Whistleblowers do not need to access the system in synchronized rounds

Introducing Express

Communication system designed for practical metadata-hiding whistleblowing

Journalists can register mailboxes for sources to send messages/documents

Whistleblowers do not need to access the system in synchronized rounds

Asymptotic improvements:

client computation costs $O(1)$

communication costs $O(1)$

(both previously $O(\sqrt{N})$)

Introducing Express

Communication system designed for practical metadata-hiding whistleblowing

Journalists can register mailboxes for sources to send messages/documents

Whistleblowers do not need to access the system in synchronized rounds

Asymptotic improvements:

client computation costs $O(1)$

communication costs $O(1)$

(both previously $O(\sqrt{N})$)

Practical improvements:

6x improvement in server computation time

8x improvement in client computation time

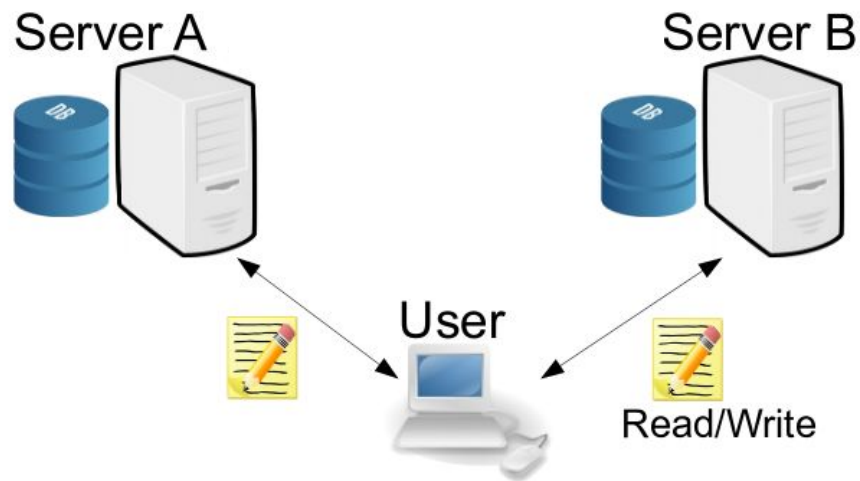
>10x improvement in communication costs

6x reduction in dollar cost to run system

Express Overview

2 server system, secure against:

- Arbitrarily many corrupt users
- Up to one corrupt server



Express Overview

2 server system, secure against:

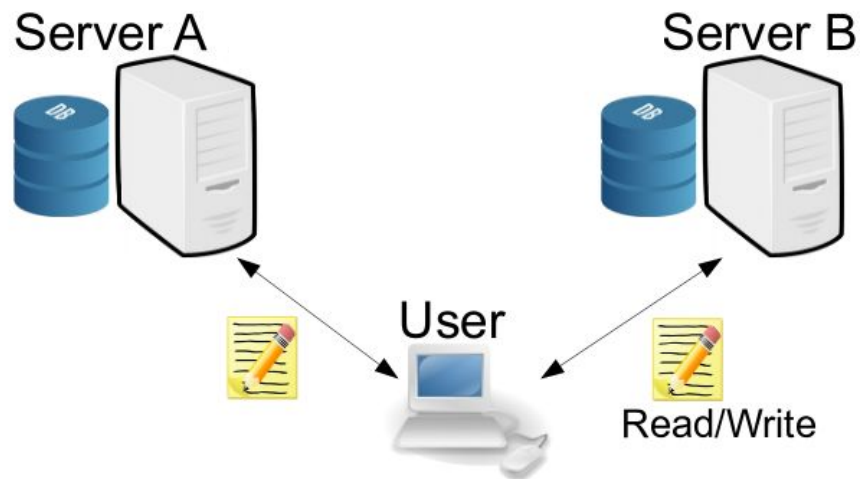
- Arbitrarily many corrupt users
- Up to one corrupt server

Supported operations:

Register mailbox

(Private) write to mailbox

Read from mailbox



Express Overview

2 server system, secure against:

- Arbitrarily many corrupt users
- Up to one corrupt server

Supported operations:

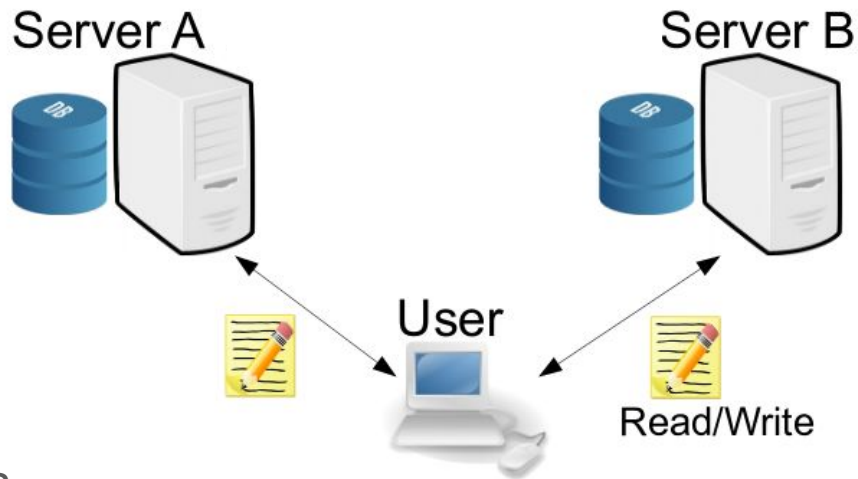
Register mailbox

(Private) write to mailbox

Read from mailbox

Security:

can't tell who the *recipient* of a message is



Tool: Private Writing with Distributed Point Functions

Point function: a function that is zero everywhere, except at one point

Tool: Private Writing with Distributed Point Functions

Point function: a function that is zero everywhere, except at one point

x	f(x)
0	0
1	0
2	0
3	"Hi!"
4	0

Tool: Private Writing with Distributed Point Functions

Point function: a function that is zero everywhere, except at one point

x	$f_1(x)$
0	“abc”
1	“xf\$”
2	“^tg”
3	“!7≈”
4	“jhV”

\oplus

x	$f_2(x)$
0	“abc”
1	“xf\$”
2	“^tg”
3	“2!”
4	“jhV”

$=$

x	f(x)
0	0
1	0
2	0
3	“Hi!”
4	0

Tool: Private Writing with Distributed Point Functions

Point function: a function that is zero everywhere, except at one point

Distributed point function: technique for efficiently splitting a point function into two pieces, each a (non-point) function whose XOR is the original point function

x	$f_1(x)$
0	“abc”
1	“xf\$”
2	“^tg”
3	“!7≈”
4	“jhV”

\oplus

x	$f_2(x)$
0	“abc”
1	“xf\$”
2	“^tg”
3	“2!”
4	“jhV”

=

x	f(x)
0	0
1	0
2	0
3	“Hi!”
4	0

Key features:

- concise representation
- fast to generate

Tool: Private Writing with Distributed Point Functions



I want to write
"Hi!" to address 3

Addr	Data
0	0
1	0
2	0
3	0
4	0



Addr	Data
0	0
1	0
2	0
3	0
4	0

Tool: Private Writing with Distributed Point Functions



x	f(x)
0	0
1	0
2	0
3	"Hi!"
4	0

Addr	Data
0	0
1	0
2	0
3	0
4	0



Addr	Data
0	0
1	0
2	0
3	0
4	0

Tool: Private Writing with Distributed Point Functions



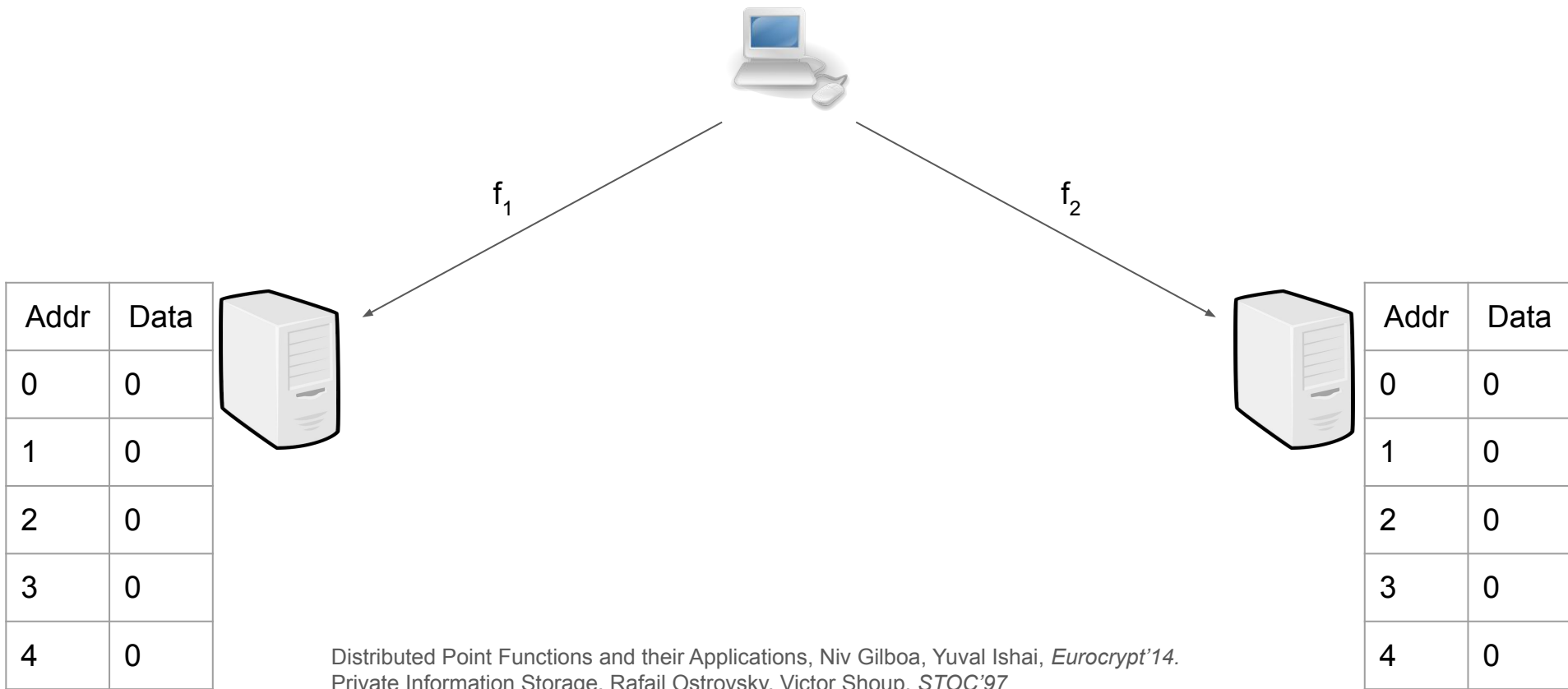
x	$f_1(x)$	x	$f_2(x)$
0	"abc"	0	"abc"
1	"xf\$"	1	"xf\$"
2	"^tg"	2	"^tg"
3	"!7≈"	3	"!2!)"
4	"jhV"	4	"jhV"

Addr	Data
0	0
1	0
2	0
3	0
4	0

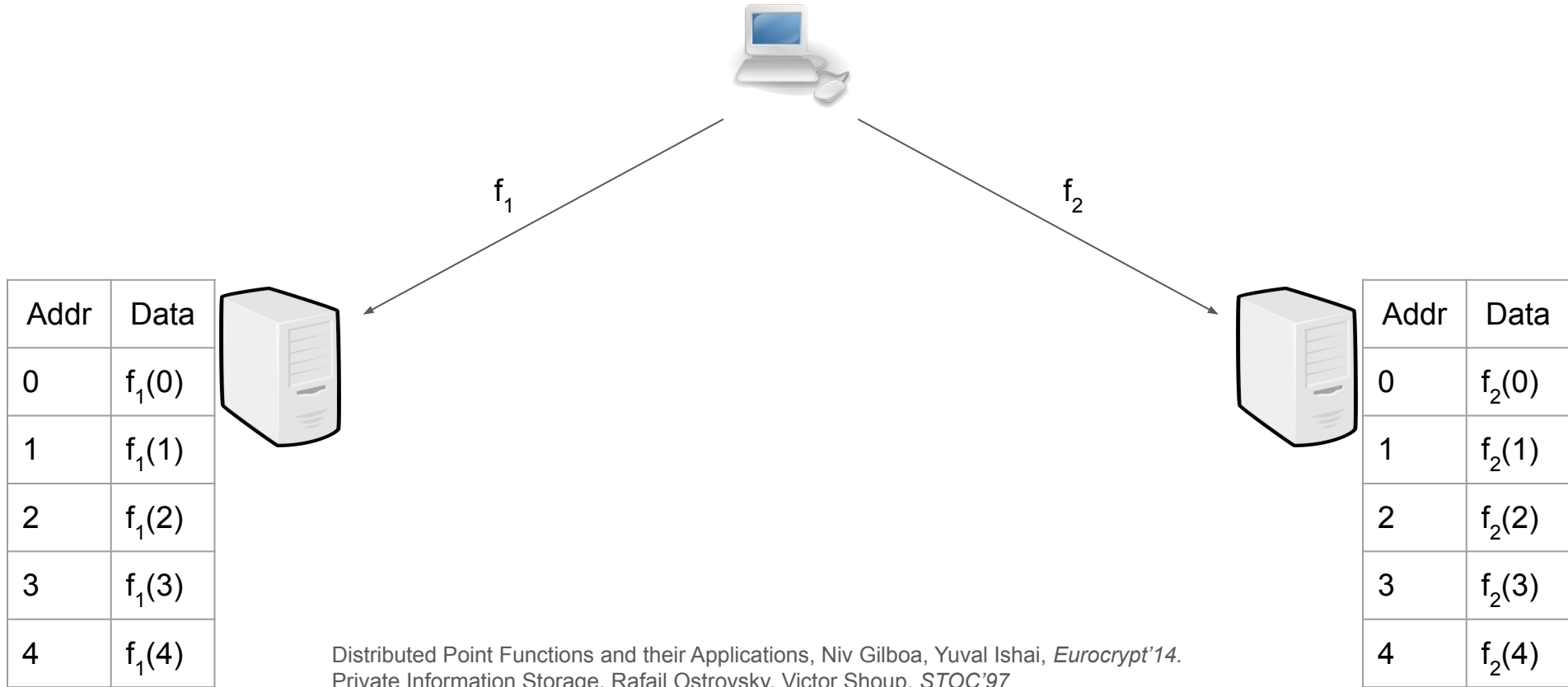


Addr	Data
0	0
1	0
2	0
3	0
4	0

Tool: Private Writing with Distributed Point Functions



Tool: Private Writing with Distributed Point Functions



Tool: Private Writing with Distributed Point Functions



f_1

f_2



Addr	Data
0	"abc"
1	"xf\$"
2	"^tg"
3	"!7~"
4	"jhV"

Addr	Data
0	"abc"
1	"xf\$"
2	"^tg"
3	"2!)"
4	"jhV"

Tool: Private Writing with Distributed Point Functions



f_1

f_2



\oplus

“Hi!”

Addr	Data
0	“abc”
1	“xf\$”
2	“^tg”
3	“!7~”
4	“jhV”

Addr	Data
0	“abc”
1	“xf\$”
2	“^tg”
3	“2!”
4	“jhV”

Hiding Data

How to prevent curious clients from reading others' mailboxes?

Addr	Data
0	"abc"
1	"xf\$"
2	"^tg"
3	"!7≈"
4	"jhV"



Addr	Data
0	"abc"
1	"xf\$"
2	"^tg"
3	"“2!”"
4	"jhV"

Hiding Data

How to prevent curious clients from reading others' mailboxes?

Encrypt each row with a different key held by the owner of the mailbox

Addr	Data	Key
0	"abc"	k_{NYT}
1	"xf\$"	k_{WaPo}
2	"^tg"	k_{WSJ}
3	"!7≈"	k_{Buzzfeed}
4	"jhV"	k_{Inquirer}



Addr	Data	Key
0	"abc"	k_{NYT}
1	"xf\$"	k_{WaPo}
2	"^tg"	k_{WSJ}
3	"“2!”"	k_{Buzzfeed}
4	"jhV"	k_{Inquirer}

Hiding Data

How to prevent curious clients from reading others' mailboxes?

Encrypt each row with a different key held by the owner of the mailbox

Different key sent to each server

Addr	Data	Key
0	"abc"	k_{NYT1}
1	"xf\$"	k_{WaPo1}
2	"^tg"	k_{WSJ1}
3	"!7≈"	$k_{\text{Buzzfeed1}}$
4	"jhV"	$k_{\text{Inquirer1}}$



Addr	Data	Key
0	"abc"	k_{NYT2}
1	"xf\$"	k_{WaPo2}
2	"^tg"	k_{WSJ2}
3	"“2!”"	$k_{\text{Buzzfeed2}}$
4	"jhV"	$k_{\text{Inquirer2}}$

Hiding *Metadata*

Construction thus far vulnerable to polling attack:

Attacker reads every row after each write to see which one was changed

Hiding *Metadata*

Construction thus far vulnerable to polling attack:

Attacker reads every row after each write to see which one was changed

Solution: servers non-interactively re-randomize every row after each write

Additional cost is low since they already write to each row

Plausible Deniability

How to protect privacy of whistleblowers if *all users* are whistleblowers?

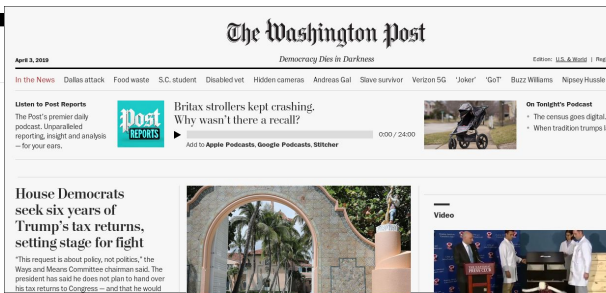
Plausible Deniability

How to protect privacy of whistleblowers if *all users* are whistleblowers?

Idea: Cooperative web sites embed JS that sends dummy write requests



The Wall Street Journal homepage features a navigation bar with the site name and various utility links. The main content area includes a 'What's News' section with a prominent article titled 'Ghosh Is Arrested on New Allegations' about former Monocor Chairman Carvin Ghosh. Below this, there are sections for 'U.S. Demand to Keep Tariffs Impedes Deal With China on Trade' and 'Lord Loughlin, Felicity Huffman Appear in Court in College Admissions Case'. A 'Markets' section displays a line chart for the S&P 500 and a table of market data. The page also includes a 'Trending the Mueller Report' section.



The Washington Post homepage features a navigation bar with the site name and various utility links. The main content area includes a 'Democracy Dies in Darkness' headline. Below this, there are sections for 'Britlux strollers kept crashing. Why wasn't there a recall?' and 'House Democrats seek six years of Trump's tax returns, setting stage for fight'. A 'Post Reports' section is also visible. The page includes a 'Video' section and a 'Trending the Mueller Report' section.



The New York Times homepage features a navigation bar with the site name and various utility links. The main content area includes a 'Some on Mueller's Team Say Report Was More Damaging Than Barr Revealed' headline. Below this, there is a diagram titled 'Here are the criminal inquiries that sprouted from the special counsel's investigation.' showing a flow from 'Trump (Special)' to 'Possible partners' and 'Foreign lobbying'. The page also includes a 'Video' section and a 'Trending the Mueller Report' section.

Plausible Deniability

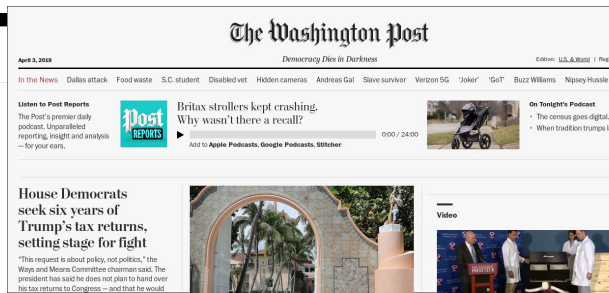
How to protect privacy of whistleblowers if *all users* are whistleblowers?

Idea: Cooperative web sites embed JS that sends dummy write requests

- Incentives properly aligned for news organizations
- Metadata-hiding means we only need 1 recipient mailbox for dummy writes
- Client-side costs low enough to not affect browsing experience



The Wall Street Journal homepage screenshot. The main headline is "Ghosh is Arrested on New Allegations" with a sub-headline "Former Morgan Chairman Carlos Ghosn was arrested again over new allegations of financial misconduct less than a month after he was released on bail." Other headlines include "U.S. Demand to Keep Tariffs Impedes Deal With China on Trade" and "Lord Loughlin, Felicity Huffman Appear in Court in College Admissions Case". A "Markets" section shows a line graph for the S&P 500 and a table of market data.



The Washington Post homepage screenshot. The main headline is "House Democrats seek six years of Trump's tax returns, setting stage for fight" with a sub-headline "This request is about policy, not politics," the Ways and Means Committee chairman said. The president has said he does not plan to hand over his tax returns to Congress — and that he would



The New York Times homepage screenshot. The main headline is "Some on Mueller's Team Say Report Was More Damaging Than Barr Revealed" with a sub-headline "A number of former Mueller's team have said their findings are more troubling than President Trump's attorney General William Barr had indicated." Other headlines include "Year-Weekend Evening Briefing" and "Listen: Modern Love Podcast". A diagram titled "Here are the criminal inquiries that sprouted from the special counsel's investigations." shows a flow from "Trump (criminal)" to "Position (political)" and "Foreign (foreign)".

Handling Disruptive Users

Any number of users can act maliciously in arbitrary ways

Handling Disruptive Users

Any number of users can act maliciously in arbitrary ways

Two kinds of attacks:

1. Disruptive user writes to others' mailbox
2. Disruptive user sends malformed DPF to write to many mailboxes

Handling Disruptive Users

Any number of users can act maliciously in arbitrary ways

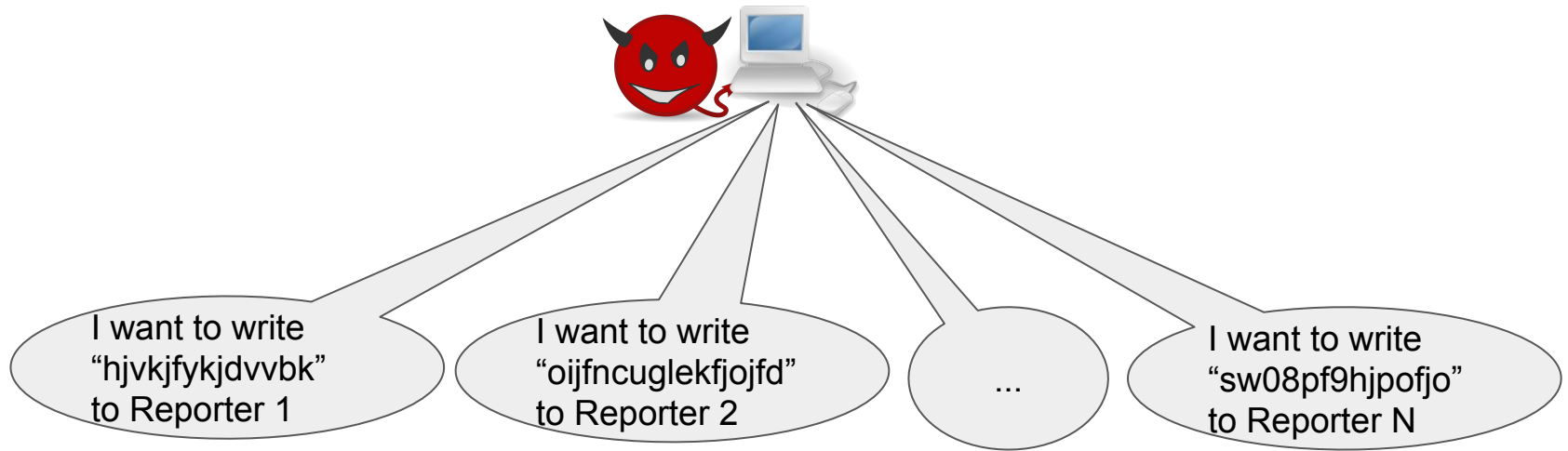
Two kinds of attacks:

1. Disruptive user writes to others' mailbox
2. Disruptive user sends malformed DPF to write to many mailboxes

Mechanism for preventing disruption can't compromise privacy

Handling Disruptive Users

Problem: disruptive user writes to others' mailboxes



Virtual Addresses

Problem: disruptive user writes to others' mailboxes

Solution: hide mailboxes in exponentially large address space

Addr	Data
0	"abc"
1	"xf\$"
2	"^tg"
...	...
...	...
...	...
$2^{128}-2$	"!7≈"
$2^{128}-1$	"jhV"

Virtual Addresses

Problem: disruptive user writes to others' mailboxes

Solution: hide mailboxes in exponentially large address space

New problem: too many addresses, bad performance

Addr	Data
0	"abc"
1	"xf\$"
2	"^tg"
...	...
...	...
...	...
$2^{128}-2$	"!7≈"
$2^{128}-1$	"jhV"

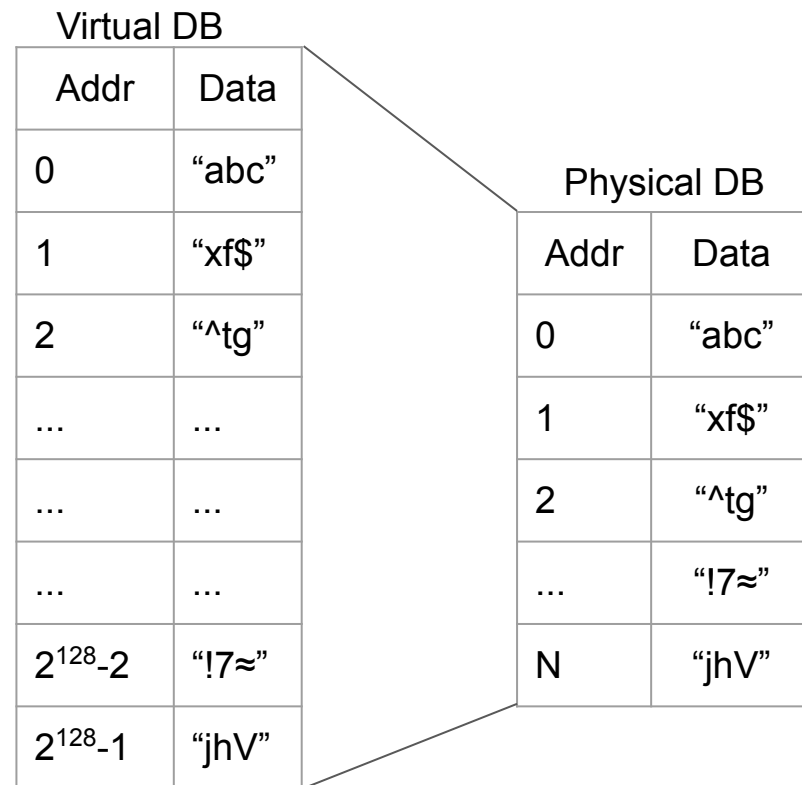
Virtual Addresses

Problem: disruptive user writes to others' mailboxes

Solution: hide mailboxes in exponentially large address space

New problem: too many addresses, bad performance

Solution: virtual addresses



Auditing

Problem: disruptive user sends malformed DPF to write to many mailboxes



x	f(x)
0	989f4
1	dDf73
...	
$2^{128}-2$	08dji3
$2^{128}-1$	89hfif

Auditing

Problem: disruptive user sends malformed DPF to write to many mailboxes

Solution: servers blindly *audit* all incoming write requests

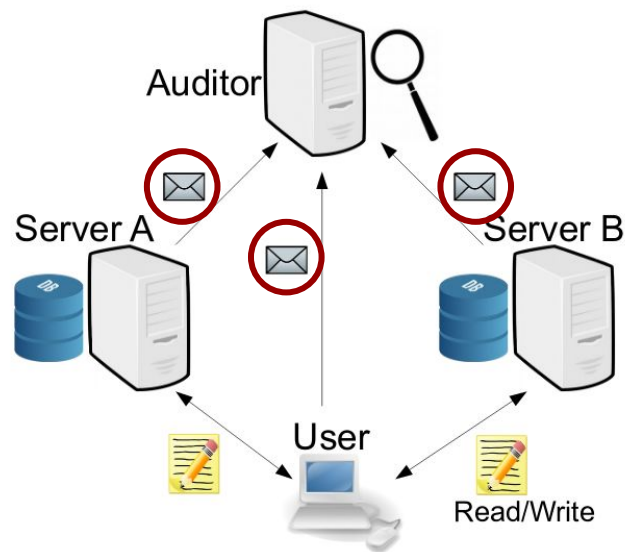
Auditing

Problem: disruptive user sends malformed DPF to write to many mailboxes

Solution: servers blindly *audit* all incoming write requests

Prior work: third server audits requests

- $O(\sqrt{N})$ communication
- $O(\sqrt{N})$ client/auditor computation



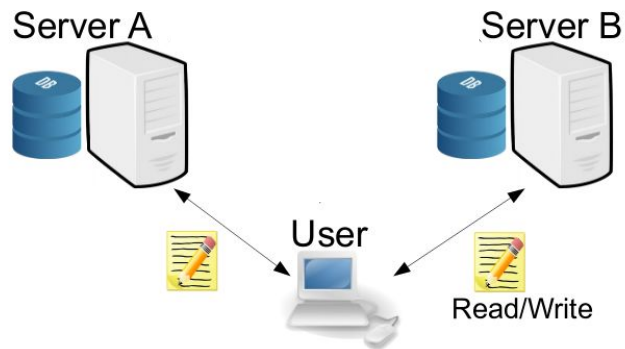
Auditing

Problem: disruptive user sends malformed DPF to write to many mailboxes

Solution: servers blindly *audit* all incoming write requests

New auditing protocol:

- $O(1)$ communication
- $O(1)$ client computation
- No additional server!

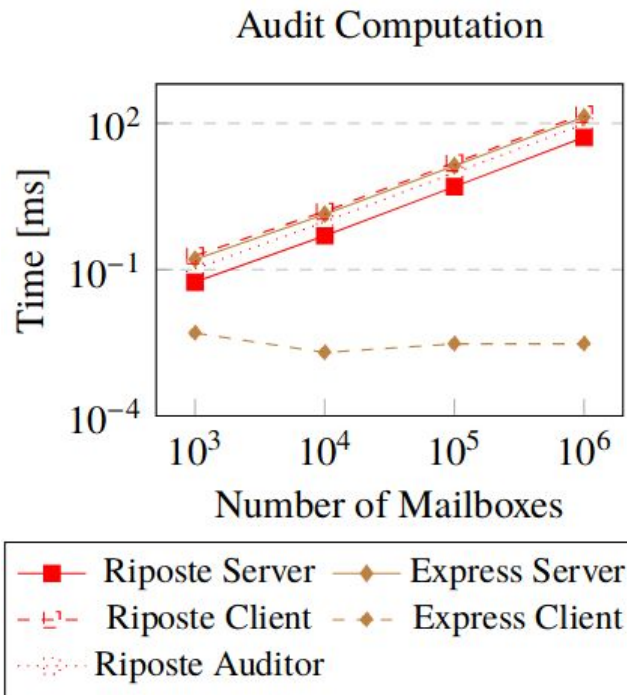


Evaluation

Evaluation

Auditing Protocol

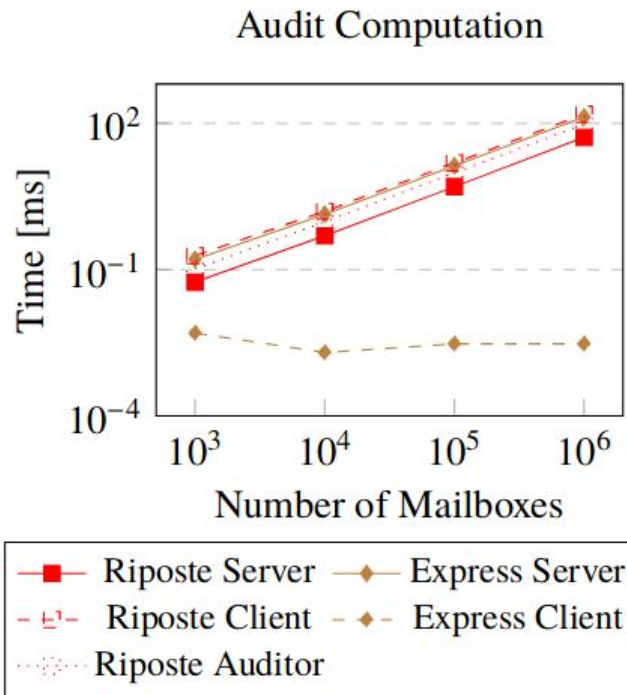
- Client runs in under 5 *microseconds* always



Evaluation

Auditing Protocol

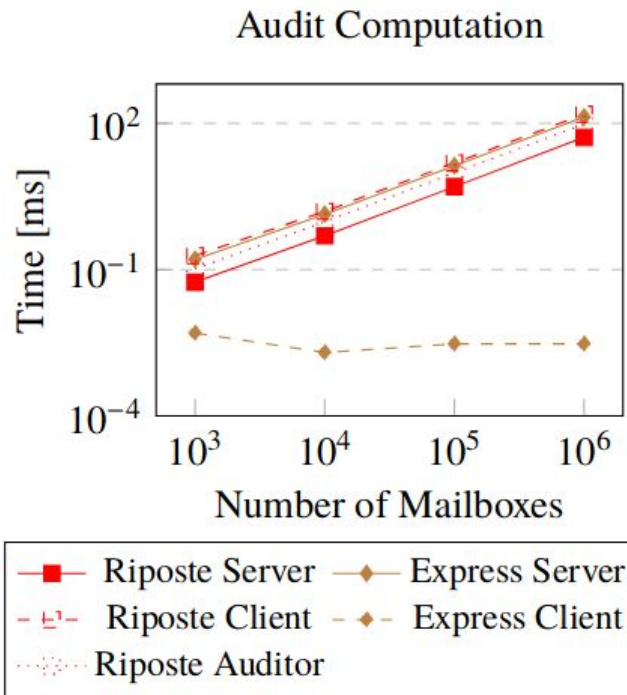
- Client runs in under 5 *microseconds* always
- 55,000x faster than Riposte for 1m mailboxes



Evaluation

Auditing Protocol

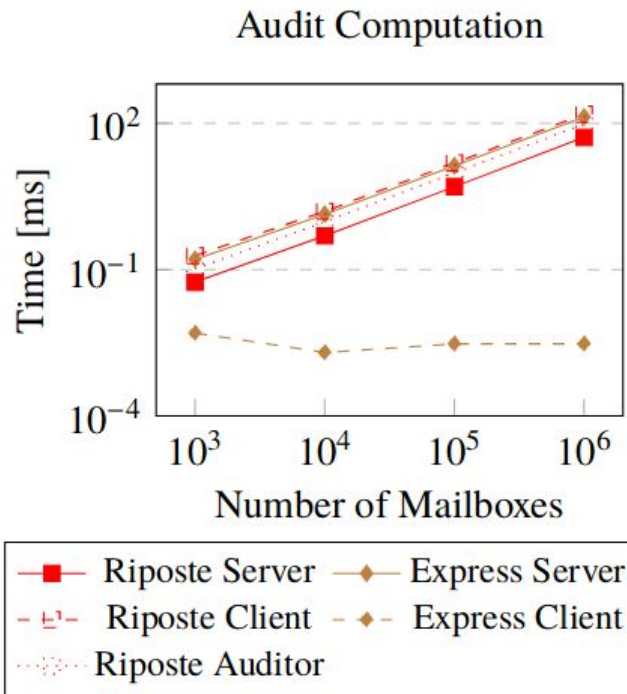
- Client runs in under 5 *microseconds* always
- 55,000x faster than Riposte for 1m mailboxes
- Enables 8x reduction in overall client computation (now 20ms)



Evaluation

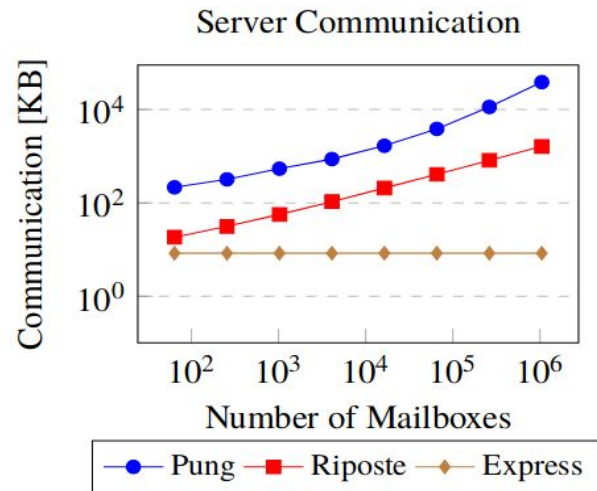
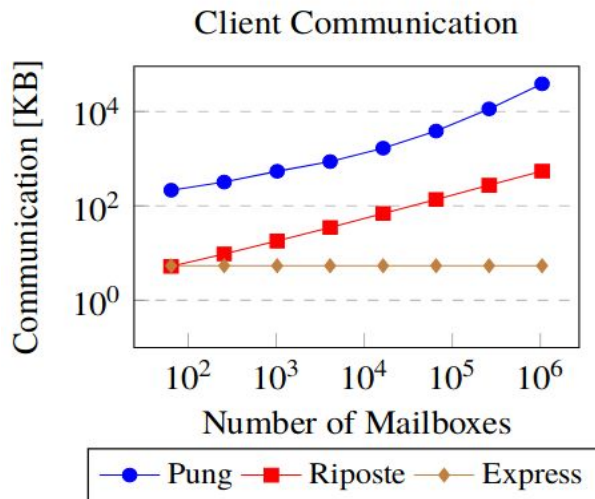
Auditing Protocol

- Client runs in under 5 *microseconds* always
- 55,000x faster than Riposte for 1m mailboxes
- Enables 8x reduction in overall client computation (now 20ms)
- Comparable on server, where auditing is not the bottleneck



Evaluation

Communication Costs

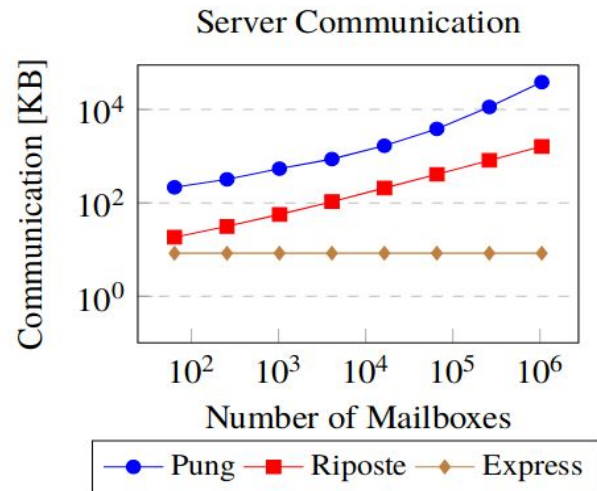
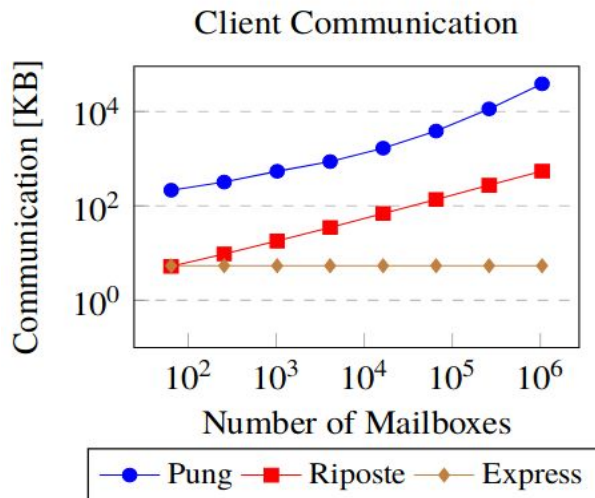


(Sending 160B messages)

Evaluation

Communication Costs

For 2^{14} mailboxes:
13x improvement on
client, 25x on server



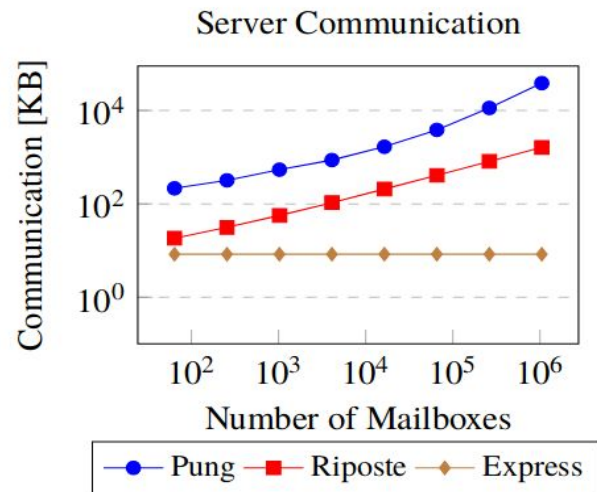
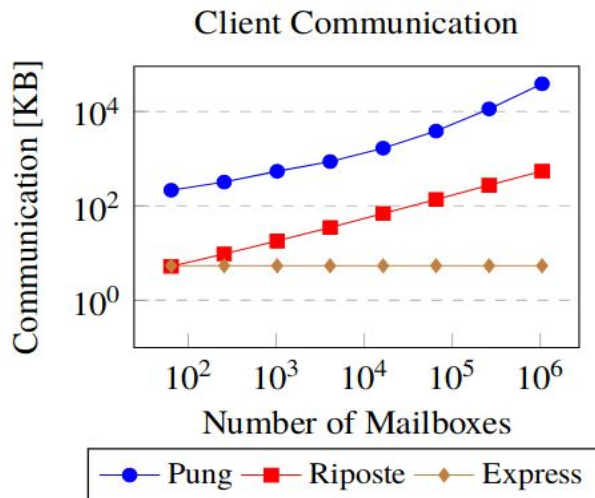
(Sending 160B messages)

Evaluation

Communication Costs

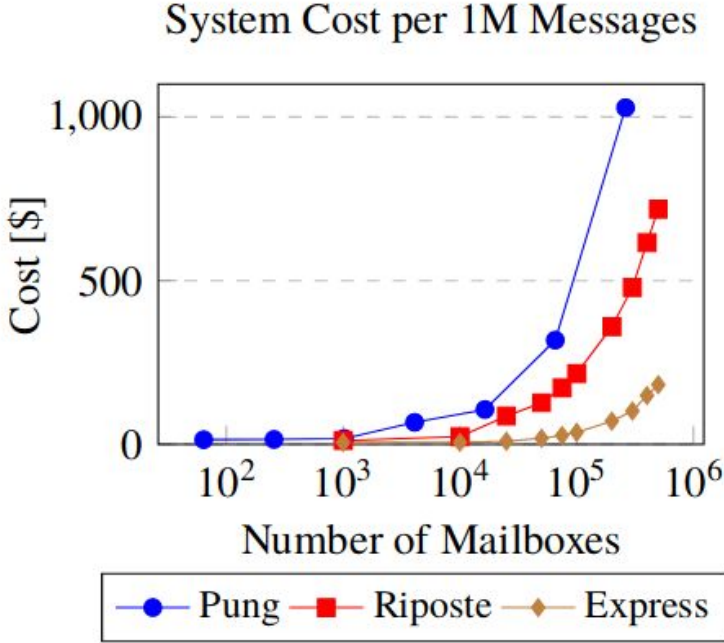
For 2^{14} mailboxes:
13x improvement on
client, 25x on server

For 2^{20} mailboxes:
101x improvement on
client, 195x on server



(Sending 160B messages)

Evaluation



(Based on GCP prices)

Express

Metadata-hiding communication system with application to private whistleblowing

Asymptotic speedup from $O(\sqrt{N})$ to $O(1)$ for auditing

Practical speedup up of 6x on server, 8x on client (compared to Riposte)

6x lower dollar costs to operate servers

13-7,000x or more reduction in communication costs

Contact: saba@cs.stanford.edu