Protocols for Private Messaging: Beyond TLS

Lena
mirren@ioc.exchange

April 15,2023

Messengers

Messengers

Messengers

Messengers

Messengers

Messengers

Contact Discovery

user

server

Contact Discovery

user

server

Contact Discovery

user

server

How to discover contacts?

Naïvely:

- unique identifier
  e.g. phone number

How to discover contacts?

Naïvely:

- unique identifier
  e.g. phone number
- hash it! (e.g. SHA256)

How to discover contacts?

Naïvely:

- unique identifier
  e.g. phone number
- hash it! (e.g. SHA256)
- add server-side secret (salt)

How do they do Contact Discovery?

How do they do Contact Discovery?

What now?

Private Information Retrieval (PIR)

Private information retrieval allows a user to anonymously retrieve an entry from a database.

Ideas?

Private Information Retrieval (PIR)

Private information retrieval allows a user to anonymously retrieve an entry from a database.

Ideas?

send the entire database

Private Information Retrieval (PIR)

Private information retrieval allows a user to anonymously retrieve an entry from a database.

Ideas?

send the entire database

information-theoretic secure!

Private Information Retrieval (PIR)

Private information retrieval allows a user to anonymously retrieve an entry from a database.

Ideas?

send the entire database

- information-theoretic secure!
- a lot of traffic...
- ~~user is potentially malicious~~

Private Information Retrieval (PIR)

Private information retrieval allows a user to anonymously retrieve an entry from a database.

Ideas?

- send the entire database
  - information-theoretic secure!
  - a lot of traffic...
  - ~~user is potentially malicious~~
- active field of research

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

k-anonymity

request *all* passwords starting with these five characters

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

k-anonymity

> request *all* passwords starting with these five characters

Can anybody identify an issue with this?

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

k-anonymity

> request *all* passwords starting with these five characters

Can anybody identify an issue with this?

database needs to be large enough

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

k-anonymity

- request *all* passwords starting with these five characters

Can anybody identify an issue with this?

database needs to be large enough

- one query is usually 800-1000 responses

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

k-anonymity

- request *all* passwords starting with these five characters

Can anybody identify an issue with this?

database needs to be large enough

- one query is usually 800-1000 responses

size of response leaks requested hashes.

PIR use-case: haveibeenpwned

as seen on `https://haveibeenpwned.com/Passwords`

online lookup service for leaked passwords

sending passwords = bad

but leaking a partial hash is not ~~that~~ problematic!

k-anonymity

- request *all* passwords starting with these five characters

Can anybody identify an issue with this?

database needs to be large enough

- one query is usually 800-1000 responses

size of response leaks requested hashes.

- *Random Padding*

Interlude: Bloom/Cuckoo Filters

compact representation

probabilistic data structure

Interlude: Bloom/Cuckoo Filters

compact representation

probabilistic data structure

tolerates some false positives, no false negatives

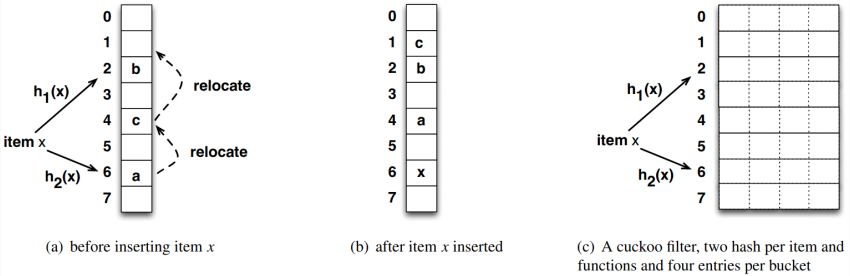


Figure 1: Cuckoo filter, taken from [?]

PIR and Signal

Signal should scale to billions of users

computational example:
10 million users, in a bloom filter: $\approx$ 40MB

PIR and Signal

Signal should scale to billions of users

computational example:
10 million users, in a bloom filter: $\approx$ 40MB

refresh once a day per user: 116 requests/s

use sharding?

PIR and Signal

Signal should scale to billions of users

computational example:
10 million users, in a bloom filter: $\approx$ 40MB

refresh once a day per user: 116 requests/s

use sharding?

- privacy vs. network overhead

What if we use different hashes?

Oblivious Pseudorandom Function

Generate a high-entropy cryptographic object, such as a key or a token, from some low-entropy input

Oblivious Pseudorandom Function

Generate a high-entropy cryptographic object, such as a key or a token, from some low-entropy input

- timestamp
- password
- username

- filename
- identifier
- ...

Oblivious Pseudorandom Function

client

input $x \in \mathcal{X}$

compute

$$F(k, x)$$

server

key $k \in \mathcal{K}$

Oblivious Pseudorandom Function

client

server

$b(x_1), \cdots, b(x_n)$

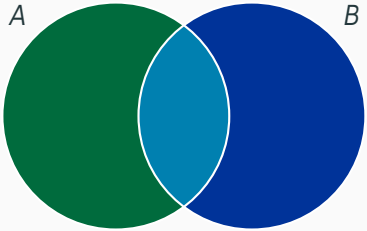
$f_k(b(x_1)), \cdots, f_k(b(x_n))$

Set $X = [x_1, \cdots, x_n]$
blinding function $b(x)$
unblinding
$b^{-1}(f_k(b(x))) = f_k(x)$

key $k \in \mathcal{K}$

Private Set Intersection

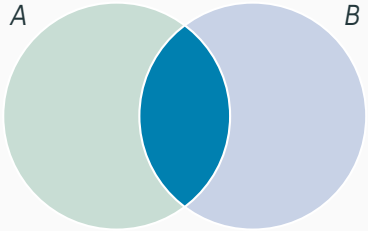
two sets of data



Private Set Intersection

two sets of data

only learn $A \cap B$



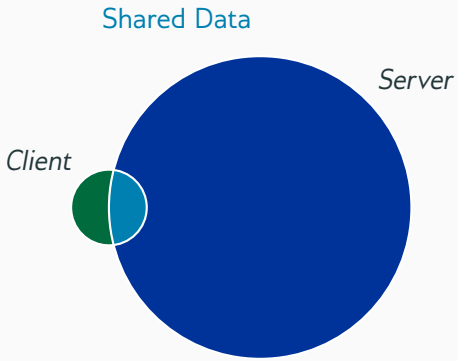
Unbalanced Private Set Intersection

two sets of data

only learn $A \cap B$

server dataset is much larger

private contact discovery [**?**]



PSI and Signal

That Academic Private Set Intersection Stuff

PSI and Signal

That Academic Private Set Intersection Stuff

largely again a representation problem

Secure Enclave

- encrypt some private memory area
- should be isolated within process
- provide *attestation* that the correct code runs

Secure Enclave

- encrypt some private memory area
- should be isolated within process
- provide *attestation* that the correct code runs
  $\rightarrow$ Secure even with a malicious host OS!

But what does Signal do?

- 2014: *range of options that do not work* ... PIR, PSI

But what does Signal do?

- 2014: *range of options that do not work* ... PIR, PSI
- old process: hash$\rightarrow$ server lookup$\rightarrow$return hashes

But what does Signal do?

- 2014: *range of options that do not work* ... PIR, PSI
- old process: hash$\rightarrow$ server lookup$\rightarrow$return hashes
- Beta Contact Discovery in 2017:

But what does Signal do?

- 2014: *range of options that do not work* ... PIR, PSI
- old process: hash→ server lookup→return hashes
- Beta Contact Discovery in 2017:

  hash local numbers
  attestate correct code runs
  encrypt and send to server
  server does Set Intersection in SGX
  server returns encrypted result

`https://signal.org/blog/private-contact-discovery/`

Wait, wasn't there something?

Wait, wasn't there something?

- LVI (2018 & 2020)
- SGAxe (2021)
- ÆPIC(2020 & 2021)

Wait, wasn't there something?

- LVI (2018 & 2020)
- SGAxe (2021)
- ÆPIC(2020 & 2021)

So what now?

- Paper [**?**] in 2019
  ...just tolerate long loading times?
- open problem

So what now?

- Paper [**?**] in 2019
  ...just tolerate long loading times?
- open problem
- should we even use phone numbers?

I mean, how dangerous is it?

Hashes are secure

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Hashes are secure

Prefixes are known: hash reversal in milliseconds on consumer hardware

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Hashes are secure

Prefixes are known: hash reversal in milliseconds on consumer hardware

Problem: *Input entropy is low*

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Hashes are secure

Prefixes are known: hash reversal in milliseconds on consumer hardware

Problem: *Input entropy is low*

$\approx$ 37 bits

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Hashes are secure

Conclusion: Only PSI is actually secure!

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Hashes are secure

Prefixes are known: hash reversal in milliseconds on consumer hardware

Problem: *Input entropy is low*

$\approx$ 37 bits

Conclusion: Only PSI is actually secure!

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Enumerate Phone Number Space

reveal 100% of US signal users

reveal 10% of US whatsapp users

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Enumerate Phone Number Space

reveal 100% of US signal users

reveal 10% of US whatsapp users

	Whatsapp	Signal	Telegram
Rate limit	60K/d *	200K/d *	5K+100/d

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Telegram...

leaks information about non-registered users

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

I mean, how dangerous is it?

Telegram...

leaks information about non-registered users

Import Account: How many other users imported the same number

All the numbers are US: Large-scale Abuse of Contact Discovery in Mobile Messengers, *Hagen, Weinert, Sendner, Dmitrienko, Schneider*, NDSS 2021

What else?

Encrypted Backups

Messenger, but also in general

how to do?

Encrypted Backups

Messenger, but also in general

how to do?

passwords are no good

Encrypted Backups

Messenger, but also in general

how to do?

passwords are no good

hardware tokens get lost

Encrypted Backups

Messenger, but also in general

how to do?

passwords are no good

hardware tokens get lost

biometrics are still problematic

Encrypted Backups

Messenger, but also in general

how to do?

passwords are no good

hardware tokens get lost

biometrics are still problematic

Oblivious Pseudorandom Functions!

Encrypted Backups

Messenger, but also in general

how to do?

passwords are no good

hardware tokens get lost

biometrics are still problematic

Oblivious Pseudorandom Functions!

the hammer where every technical problem looks like an application

OPAQUE

Asymmetric Password-Authenticated Key Exchange Protocol

OPAQUE

Asymmetric Password-Authenticated Key Exchange Protocol

Server only stores encrypted backup

OPAQUE

Asymmetric Password-Authenticated Key Exchange Protocol

Server only stores encrypted backup

encryption key is derived from a password using OPRF

OPAQUE

Asymmetric Password-Authenticated Key Exchange Protocol

Server only stores encrypted backup

encryption key is derived from a password using OPRF

nice for all applications with passwords!

Conclusion

Contact Discovery is difficult

Real-life messengers mess up in many ways

Conclusion

Contact Discovery is difficult

Real-life messengers mess up in many ways

PSI is not an *academic* solution

Conclusion

Contact Discovery is difficult

Real-life messengers mess up in many ways

PSI is not an *academic* solution

Personal Opinion If you can't secure it, don't build it

Conclusion

Contact Discovery is difficult

Real-life messengers mess up in many ways

PSI is not an *academic* solution

Personal Opinion If you can't secure it, don't build it

also, use OPRFs for passwords!

Conclusion

Contact Discovery is difficult

Real-life messengers mess up in many ways

PSI is not an *academic* solution

Personal Opinion If you can't secure it, don't build it

also, use OPRFs for passwords!

If you want to chat, please do!
(I'll be around until the event ends)

Slide Template `https://github.com/sdesch/minimalist-beamer-latex`

Protocols for Private Messaging: Beyond TLS

Lena
mirren@ioc.exchange

April 15,2023

References