DatashareNetwork: A Decentralized Privacy-Preserving Search Engine for Investigative Journalists

Kasra EdalatNejad (EPFL)
Wouter Lueks (EPFL), Julien Martin, Soline Ledésert (ICIJ)
Anne Lhôte (ICIJ), Bruno Thomas (ICIJ), Laurent Girod (EPFL), Carmela Troncoso (EPFL)
Journalists

The International Consortium of Investigative Journalists is a global network of 267 investigative journalists in 100 countries who collaborate on in-depth investigative stories.
The Panama Papers: Exposing the Rogue Offshore Finance Industry

A giant leak of more than 11.5 million financial and legal records exposes a system that enables crime, corruption and wrongdoing, hidden by secretive offshore companies.
11.5 Million documents (Centralized)

**The Panama Papers: Exposing the Rogue Offshore Finance Industry**

A giant leak of more than 11.5 million financial and legal records exposes a system that enables crime, corruption and wrongdoing, hidden by secretive offshore companies.
11.5 Million documents (Centralized)

Journalists from over 80 countries
Digitalization
The LuxLeaks documents contain information about H4201030M.pdf, which includes details about LuxCo, its participation in H3G Holdings, and the fiscal value of LuxCo's contributions. The text mentions organizations such as Fairfax Financial, mercapital ftsr_page.pdf, and H4202032M.pdf, along with various locations and entities involved in the LuxLeaks scandal.
Datashare
Datashare

Local

Central
DatashareNetwork

Local

Journalist ↔ Journalist

Central
Survey: sharing

Are you willing to share your documents?

- Yes: 60%
- No: 40%
- Only with screening talk: 0%
1. Privacy-preserving search
1. Privacy-preserving search
2. Screening talk
Out of band collaboration/sharing

Find

Contact

No retrieval
Survey: threat model

Journalists  ICIJ  Third party
Survey: threat model

Journalists

ICIJ

Third party
Survey: threat model

Journalists

ICIJ

Third party
Survey: threat model

Journalists

ICIJ

Third party
Threat model
Coerce or compromise

Hacking

Violence

Subpoena
Malicious for protection
Setting

267 1000 journalists
Setting

1000 journalists

1000 documents per journalist

Corpus

Owner A

Owner B

Owner C
Privacy
Privacy: querier

I’m searching for: “mickey mouse scandal”

Protect:
- The query
- The querier’s identity
Privacy: owner

I’m searching for: “mickey mouse scandal”

Protect:
• The corpus
Privacy: screening

Protect:
- The content of conversation
- The existence of conversation
Build a system
Challenge: asynchrony
Communication

Querier

Communication server

Owner B

Owner C

33
Challenge: find a match
Document search

Matching status

Protect:
• The query
• The corpus
• The result
Document search

Querier

Owner

Query

Responses

Multiset private set intersection

Matching status
Private Set Intersection (PSI)
Private Set Intersection (PSI)

Existing

"mickey mouse"

mouse
Mickey

Querier

PSI

Owner

duck
Donald
Minnie
Mickey

{Mickey}
Multiset Private Set Intersection (MS-PSI)

[∅, {Mickey, mouse}, {Mickey}]

Alice

MS-PSI

Bob

mouse

Mickey

Donald duck

mouse

Mickey Minnie

Donald

Minnie
Multiset Private Set Intersection (MS-PSI)

Querier

Owner

Mickey mouse

[∅, {Mickey, mouse}, {Mickey}]
Integrating

Querier

Communication server

Owner B

Owner C
Integrating

1. Query Q

Querier

Communication server

Owner B

Owner C
Integrating

1. Query \( Q \)
2. Retrieve query

Querier

Communication server

Owner B

Owner C
Integrating

1. Query

2. Retrieve query

3. Respond

Querier → Communication server → Owners B, C
Integrating

1. Query Q

2. Retrieve query

3. Respond

Querier → Communication server → Owner

Owner B → Communication server

Owner C → Communication server
Integrating

1. Query Q
2. Retrieve query
3. Respond
4. Retrieve responses
Challenge: enable screening talk
Anonymous messaging
Privacy: layers

MS-PSI  Document search  Messaging  DatashareNetwork
Privacy: layers

MS-PSI  Document search  Messaging  DatashareNetwork

Same privacy as PSI*
Privacy: layers

MS-PSI

Document search

Messaging

DatashareNetwork

Same privacy as PSI*

Only searched keywords
Privacy: layers

MS-PSI
Same privacy as PSI*

Document search
Only searched keywords

Messaging
Unobservable

DatashareNetwork
Privacy: layers

MS-PSI
Same privacy as PSI*

Document search
Only searched keywords

Messaging
Unobservable

DatashareNetwork
Achieves privacy goals
Privacy: a new (stronger) setting
DatashareNetwork
DatashareNetwork

Requirement analysis

Querier

Communication server

Owner

Owner
DatashareNetwork

Requirement analysis

Multiset private set intersection

Communication server
DatashareNetwork

Requirement analysis

Multiset private set intersection

Communication server

Owner

Querier

Document search
DatashareNetwork

Requirement analysis

Multiset private set intersection

Anonymous screening

Communication server

Document search
DatashareNetwork

Requirement analysis

Multiset private set intersection

Anonymous screening

Communication server

Document search

End-to-end system design and privacy analysis
DatashareNetwork

Requirement analysis

Multiset private set intersection

Anonymous screening

Document search

End-to-end system design and privacy analysis

Formal proofs
Thank you for your attention!