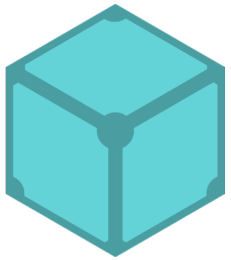
Stony Brook University

Secrets are forever:
Characterizing sensitive file leaks on IPFS

Zhengyu Wu, Brian Kondracki, Nick Nikiforakis, Aruna Balasubramanian

IFIP/IEEE Networking 2024

FAR BEYOND



IPFS

Operational since 2015

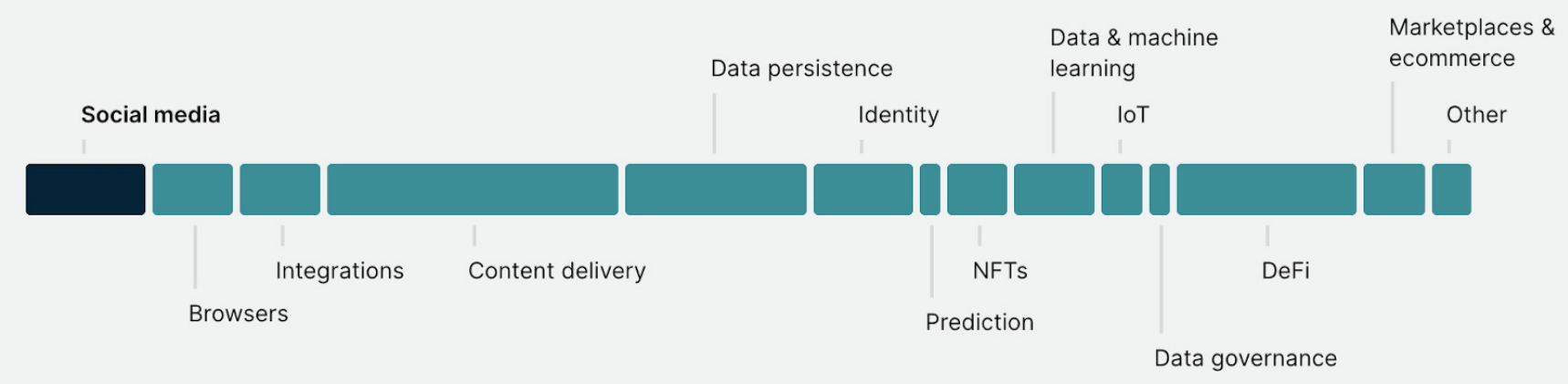
- Number of Requests: over 1 **billion** requests (weekly)
- Volume of Traffic: **Hundreds of TBs**
- Unique weekly users: **Tens of Millions**

Stony Brook University

IPFS

Operational since 2015

- Number of Requests: over 1 **billion** requests (weekly)
- Volume of Traffic: **Hundreds of TBs**
- Unique weekly users: **Tens of Millions**



FAR BEYOND

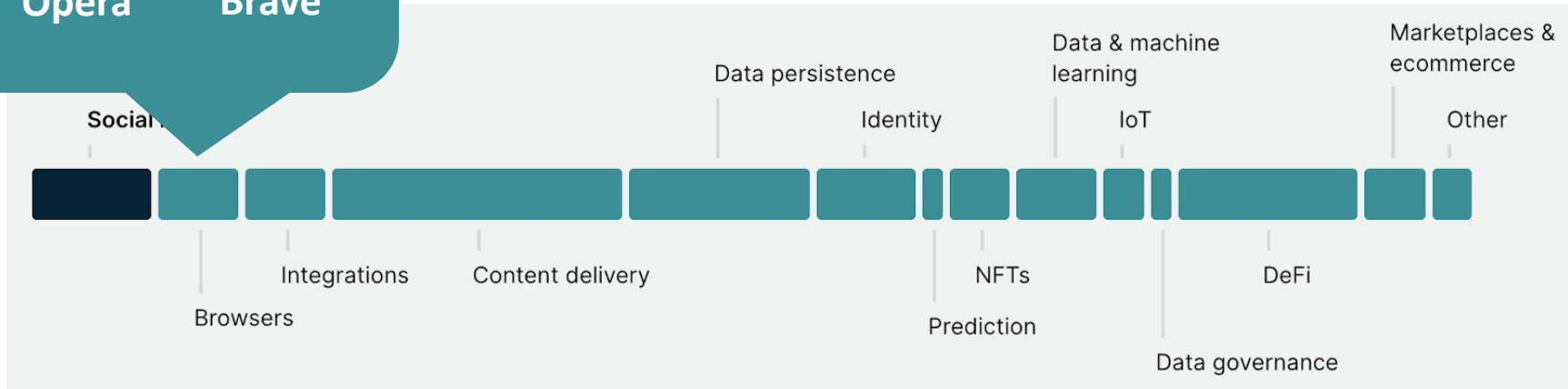
Stony Brook University

IPFS

Operational since 2015

- Number of Requests: over 1 **billion** requests (weekly)
- **Hundreds of TBs**
- s: **Tens of Millions**

Chromium Opera Brave



FAR BEYOND

Stony Brook University

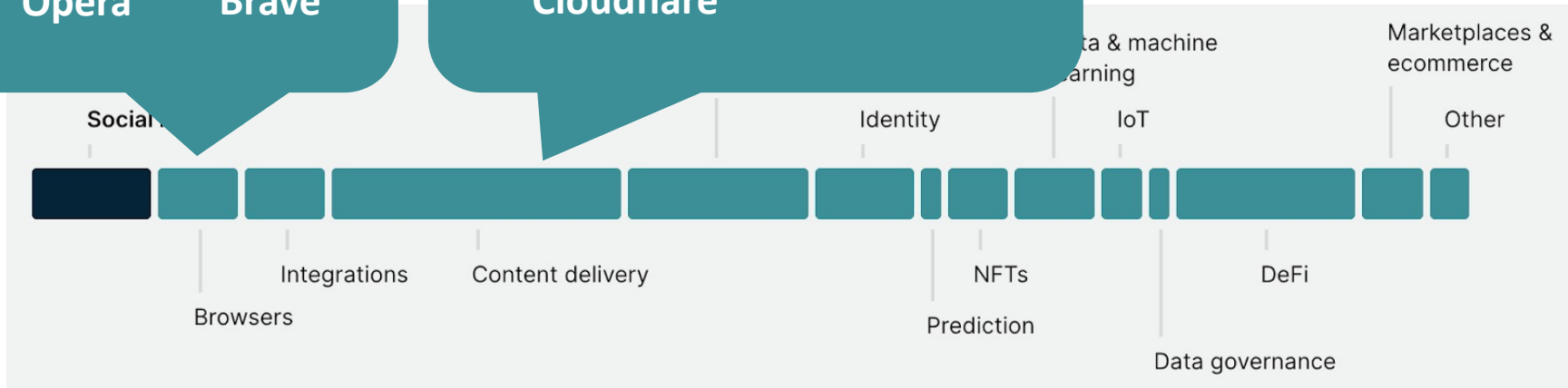
IPFS

Operational since 2015

- Number of Requests: over 1 **billion** requests (weekly)

Chromium Opera Brave

Cloudflare and more ...



FAR BEYOND

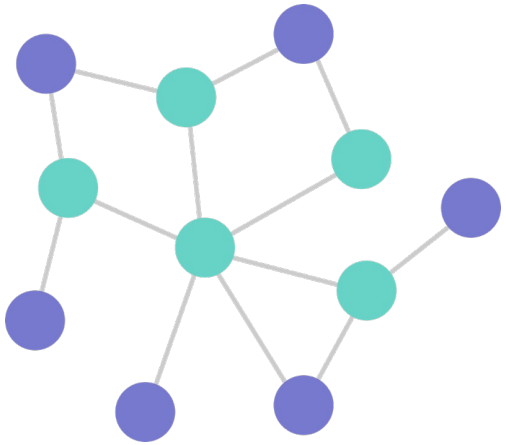
Stony Brook University

IPFS workflow

IPFS workflow

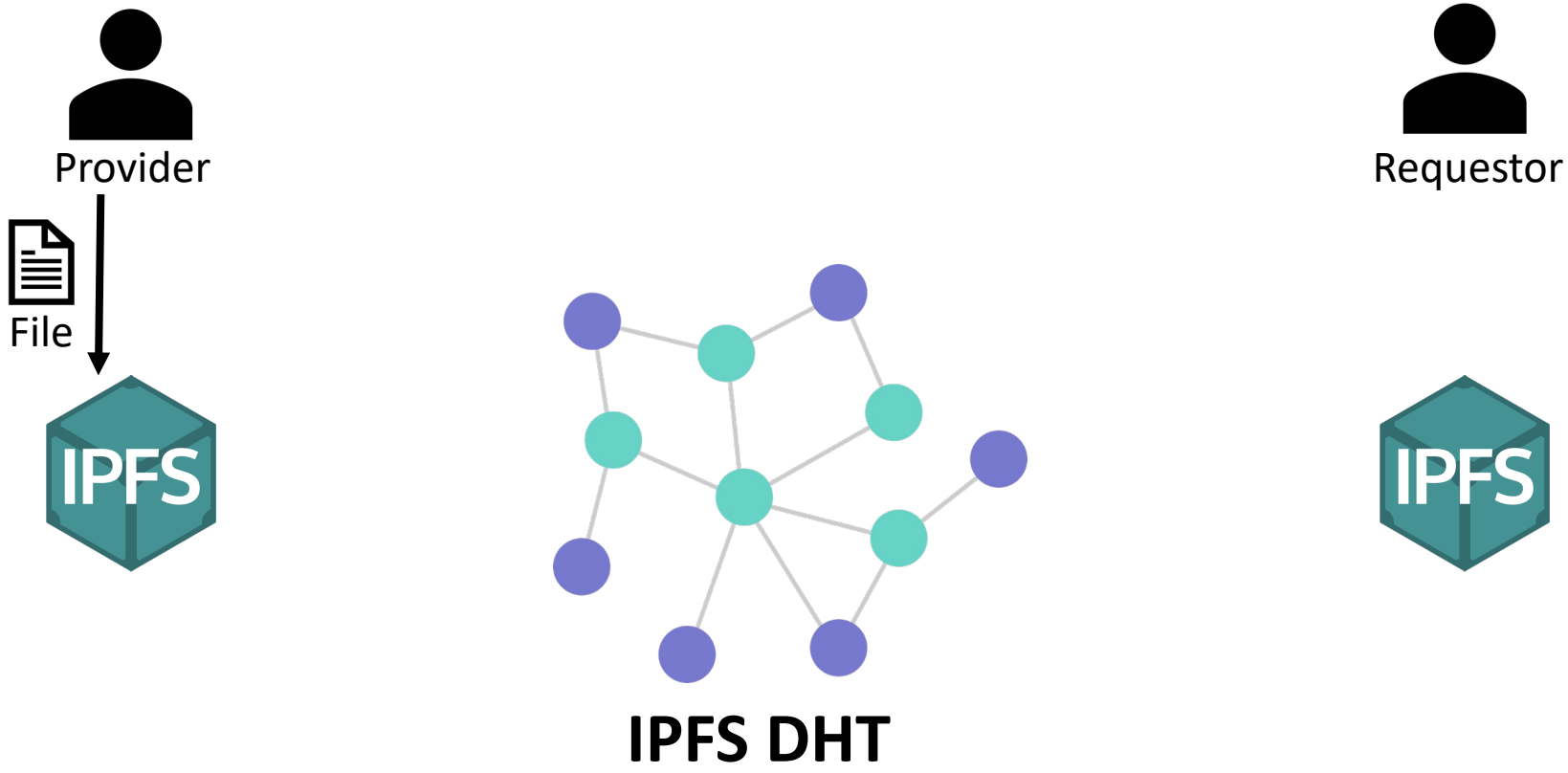
Provider

Requestor

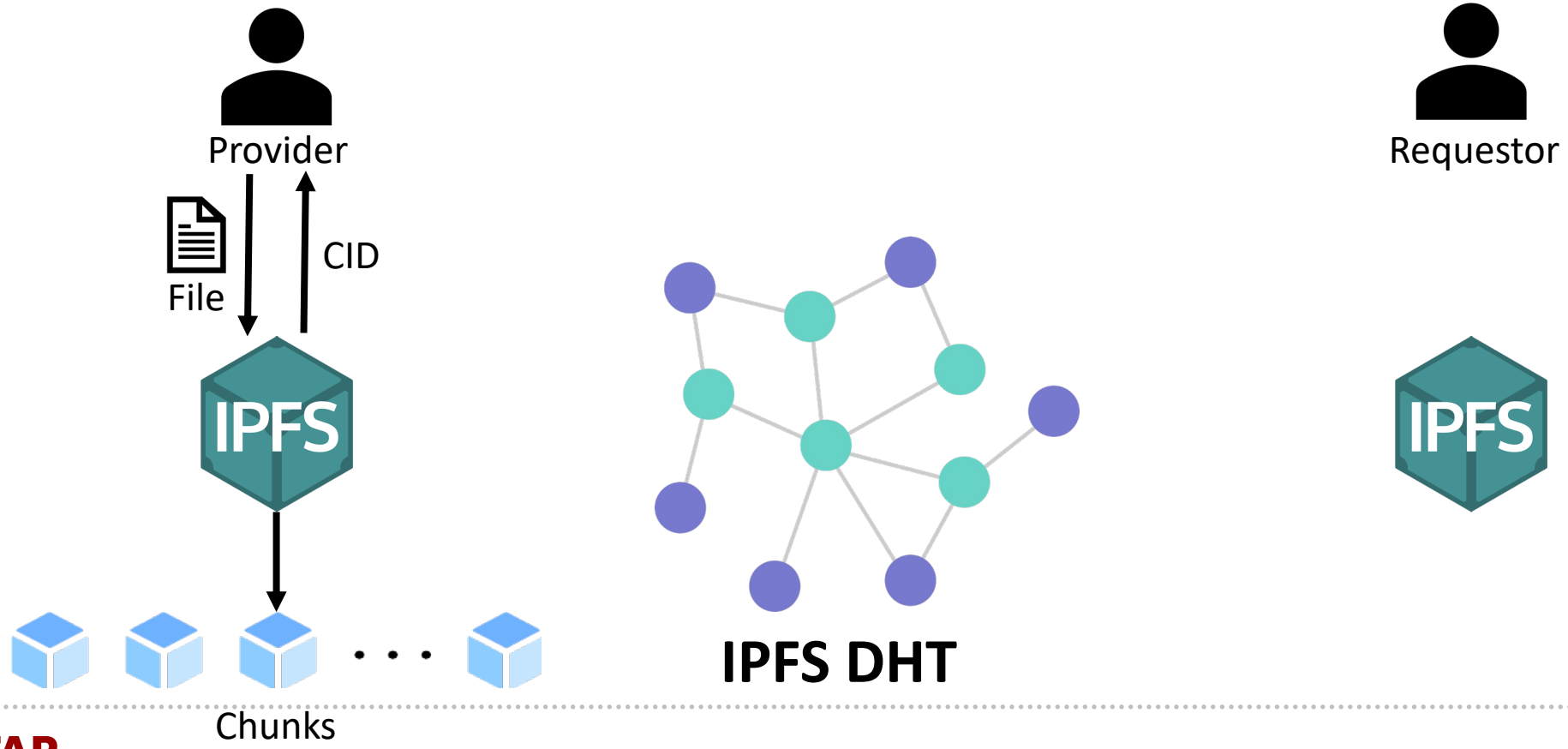


IPFS DHT

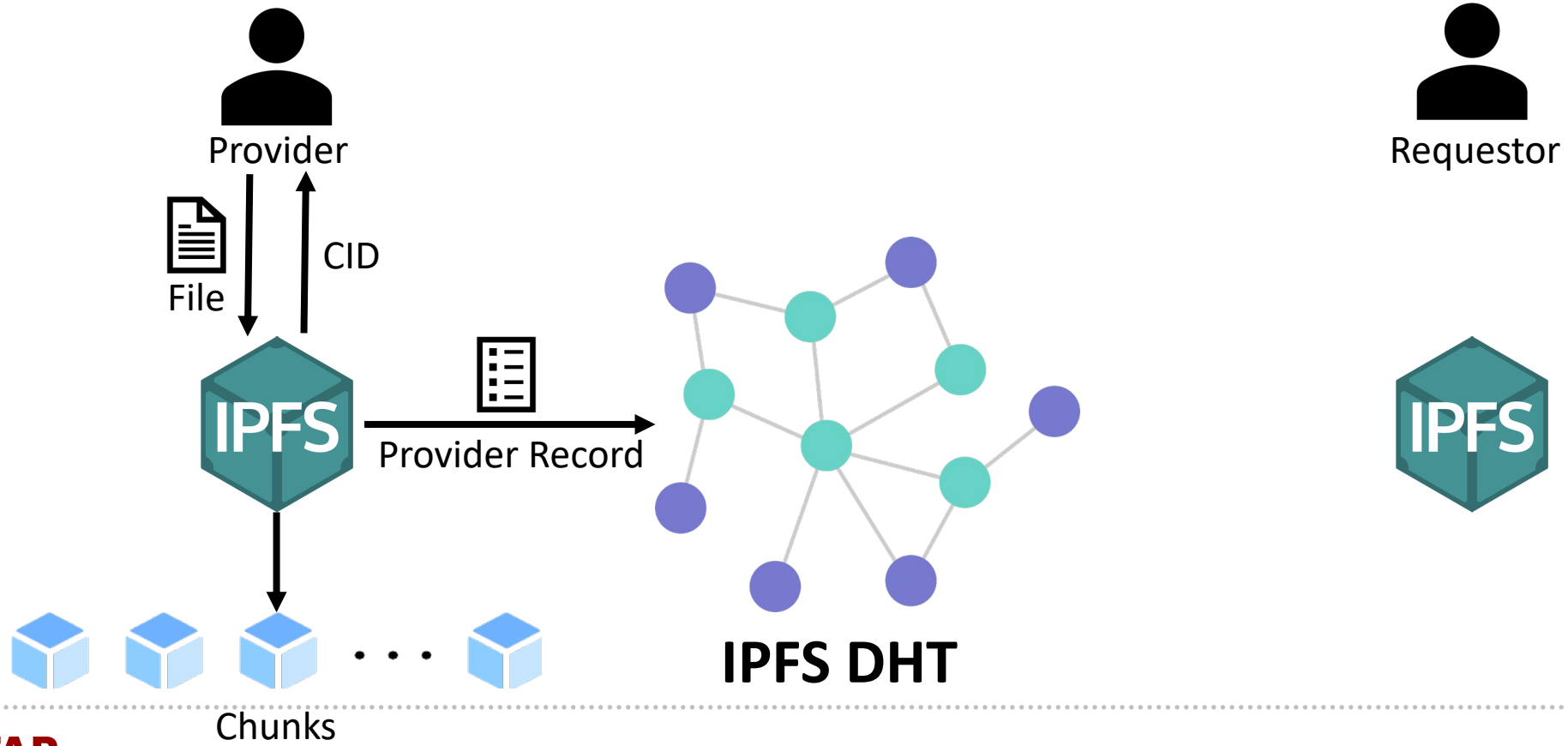
IPFS workflow



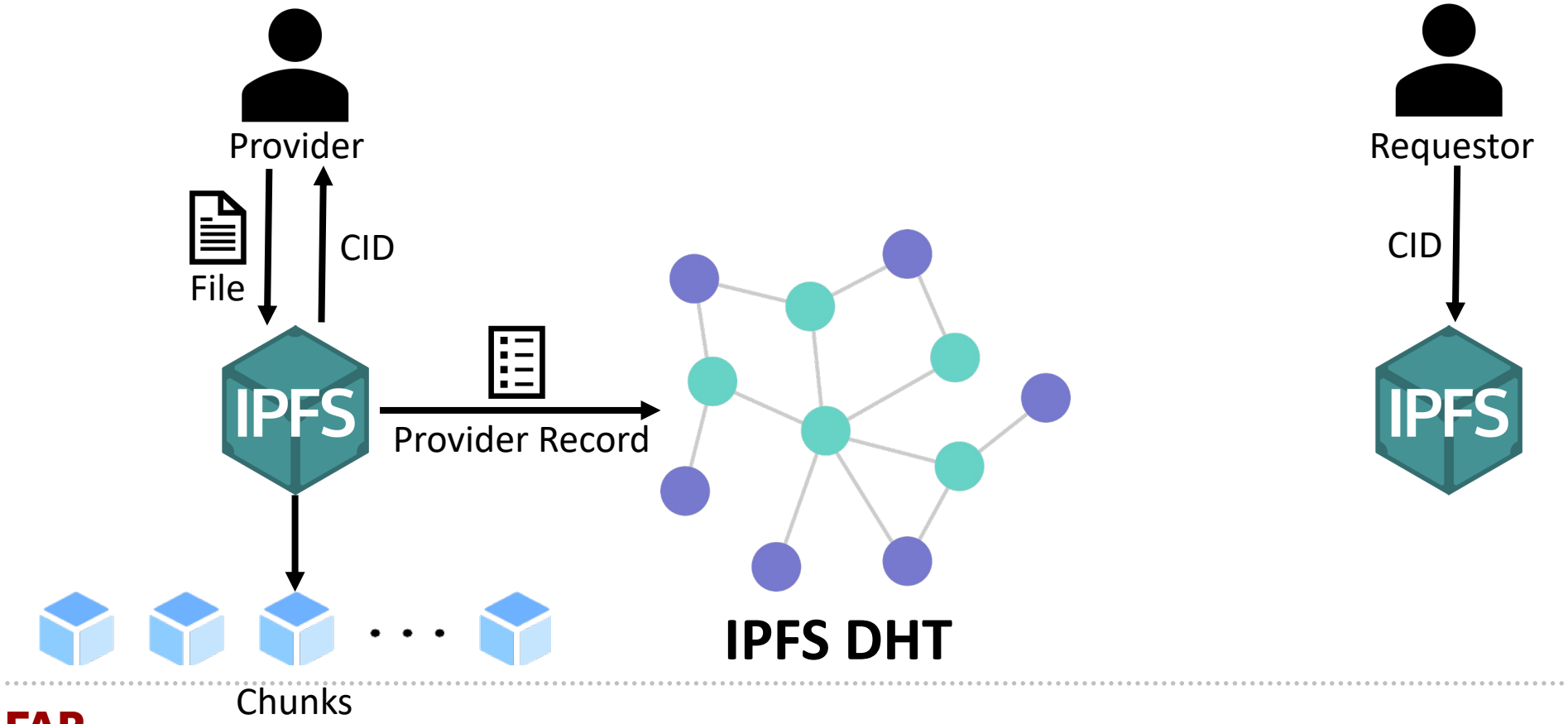
IPFS workflow



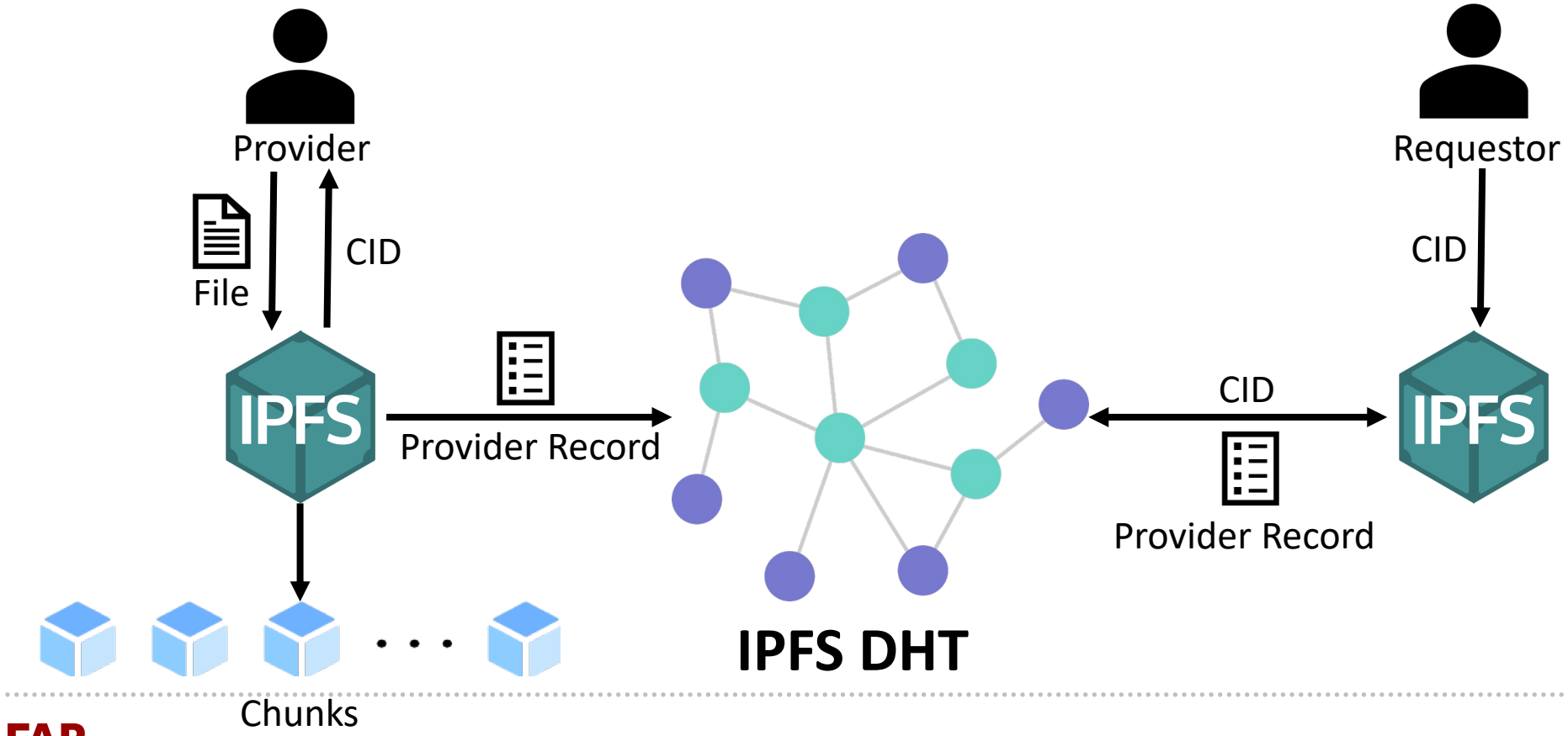
IPFS workflow



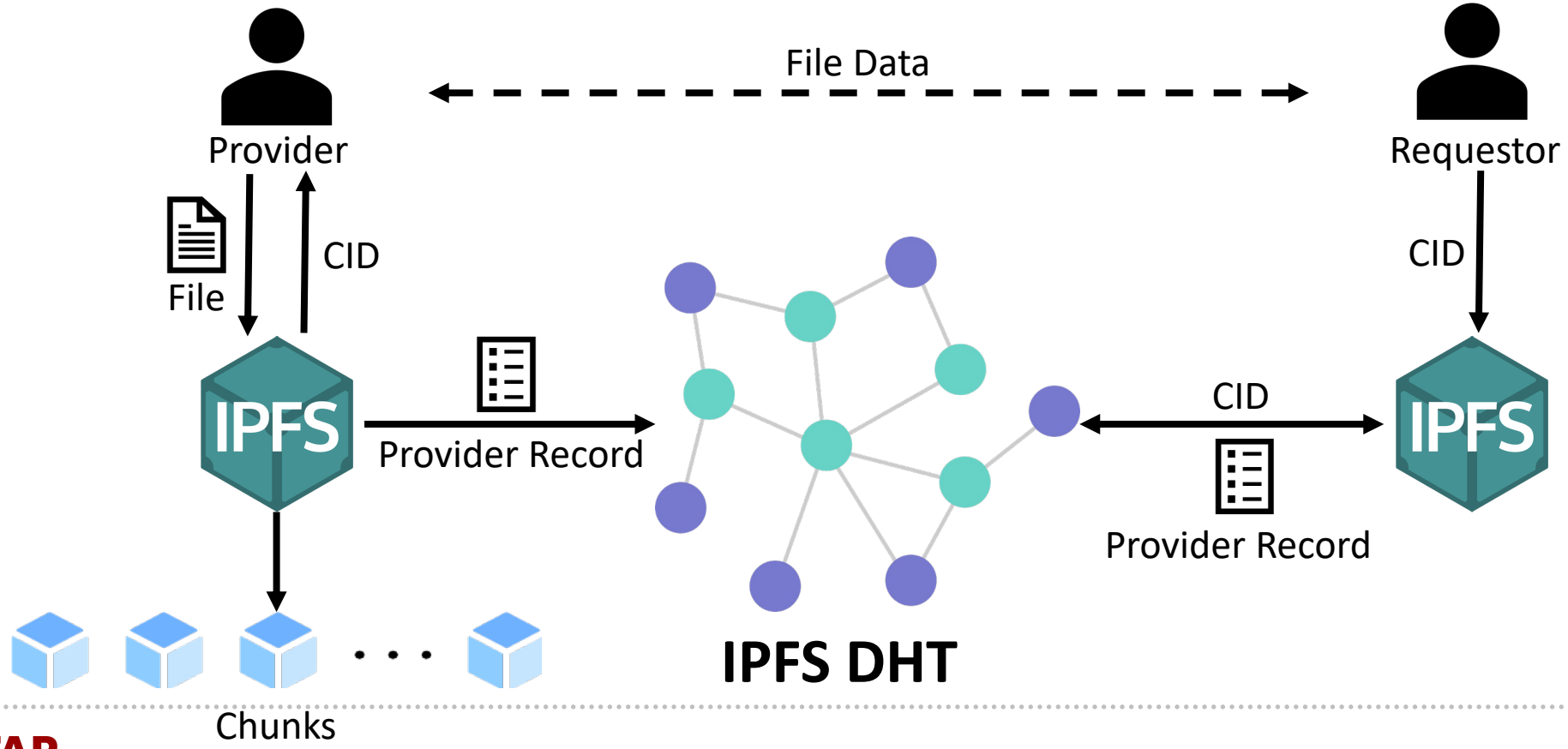
IPFS workflow



IPFS workflow



IPFS workflow



Stony Brook University

How is IPFS different from other storage systems?

How is IPFS different from other storage systems?

- IPFS is a public peer-to-peer network
  - Files shared on IPFS are all ***public***. Anyone can retrieve the file using the CID

How is IPFS different from other storage systems?

- IPFS is a public peer-to-peer network
  - Files shared on IPFS are all ***public***. Anyone can retrieve the file using the CID
- IPFS distribution nature
  - Any files can be cached if the file is being requested by other peers
  - Even if the original owner removed the file, the file is available

How is IPFS different from other storage systems?

- IPFS is a public peer-to-peer network
  - Files shared on IPFS are all ***public***. Anyone can retrieve the file using the CID
- IPFS distribution nature
  - Any files can be cached if the file is being requested by other peers
  - Even if the original owner removed the file, the file is available
- IPFS File Longevity
  - Files can remain in the network for a long time as long as a peer is hosting the file

Stony Brook University

Are users sharing sensitive files on the IPFS platform?

Web URL

OneDrive URL

IPFS URL

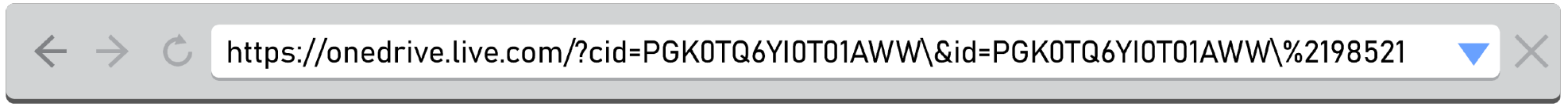
FAR BEYOND

Stony Brook University

Are users sharing sensitive files on the IPFS platform?

Web URL

OneDrive URL



IPFS URL

FAR BEYOND

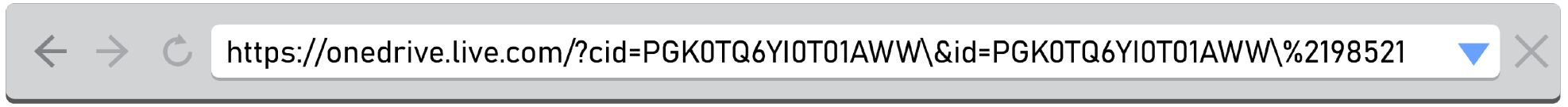
Stony Brook University

Are users sharing sensitive files on the IPFS platform?

Web URL

Public

OneDrive URL



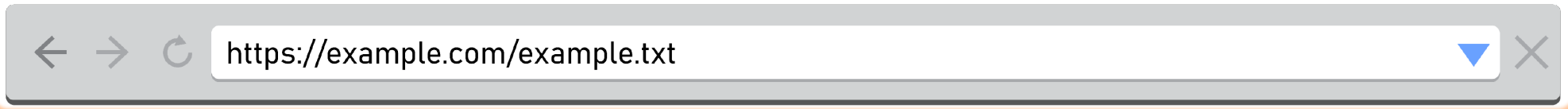
IPFS URL

FAR BEYOND

Stony Brook University

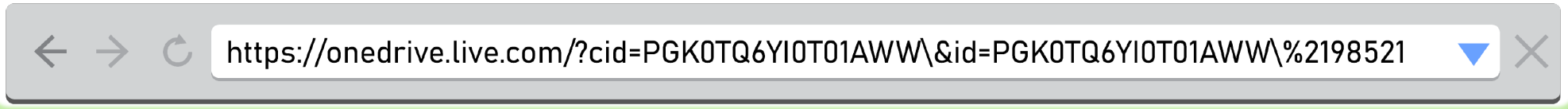
Are users sharing sensitive files on the IPFS platform?

Web URL



Public

OneDrive URL



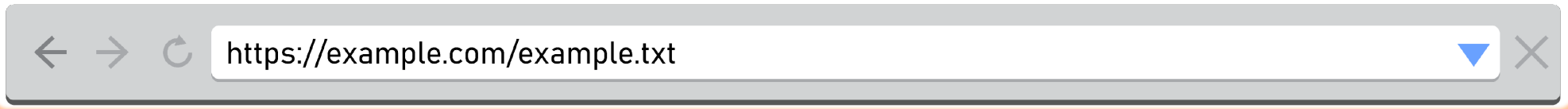
IPFS URL

FAR BEYOND

Stony Brook University

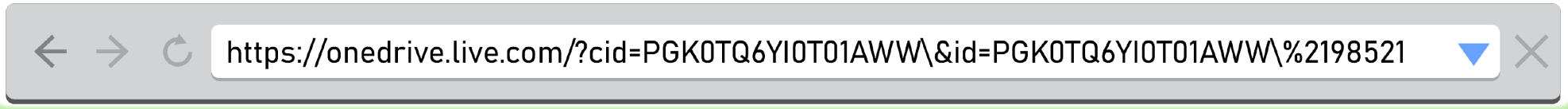
Are users sharing sensitive files on the IPFS platform?

Web URL



Public

OneDrive URL



Private

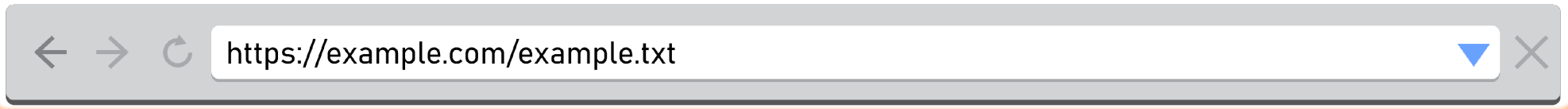
IPFS URL

FAR BEYOND

Stony Brook University

Are users sharing sensitive files on the IPFS platform?

Web URL



Public

OneDrive URL

Private

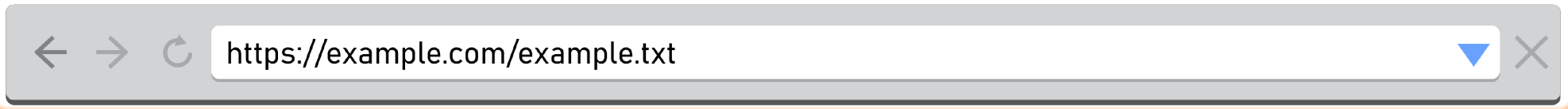
IPFS URL

FAR BEYOND

Stony Brook University

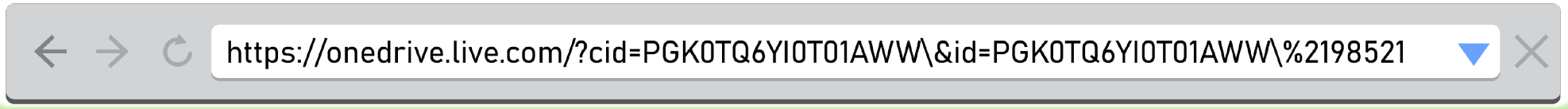
Are users sharing sensitive files on the IPFS platform?

Web URL



Public

OneDrive URL



Private

IPFS URL

Looks private But public

Sensitive file sharing study

- "How Bad Can It Git? Characterizing Secret Leakage in Public GitHub Repositories." – Meli, Michael,et al. NDSS'19.
- Findings:
  - Secret leakages of private keys and API keys are pervasive, affecting over 100k repositories
  - Thousands of new unique secrets are leaked every day

Sensitive file sharing study

- "How Bad Can It Git? Characterizing Secret Leakage in Public GitHub Repositories." – Meli, Michael,et al. NDSS'19.
- Findings:
  - Secret leakages of private keys and API keys are pervasive, affecting over 100k repositories
  - Thousands of new unique secrets are leaked every day

Our goal is to investigate sensitive leaks on IPFS

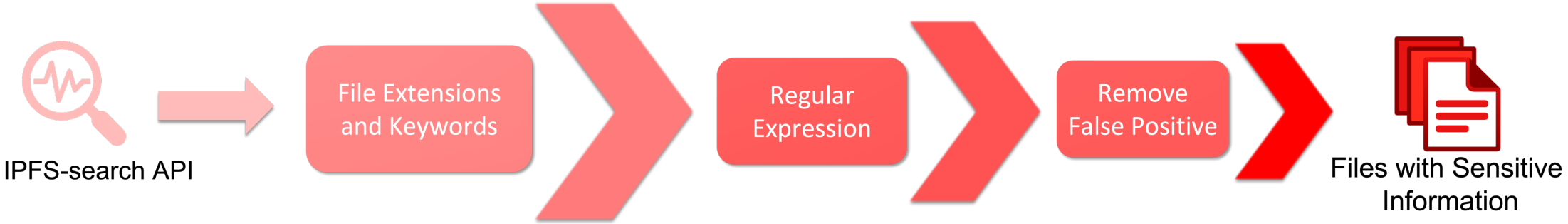
Challenges: Querying all files

- Files on IPFS ***do not*** have human-readable labels associated with them
  - Cannot target search directly on IPFS network
- Files on IPFS are distributed
  - Searching for files across all nodes is difficult
- Large volume of files are shared on IPFS
  - Study by Dennis[1] revealed ***billions*** of CID are shared on IPFS each day

[1] Trautwein, Dennis. "Hydra's Performance Contribution," January 9, 2023. https://github.com/probe-lab/network-measurements/blob/master/results/rfm21-hydras-performance-contribution.md.

Stony Brook University

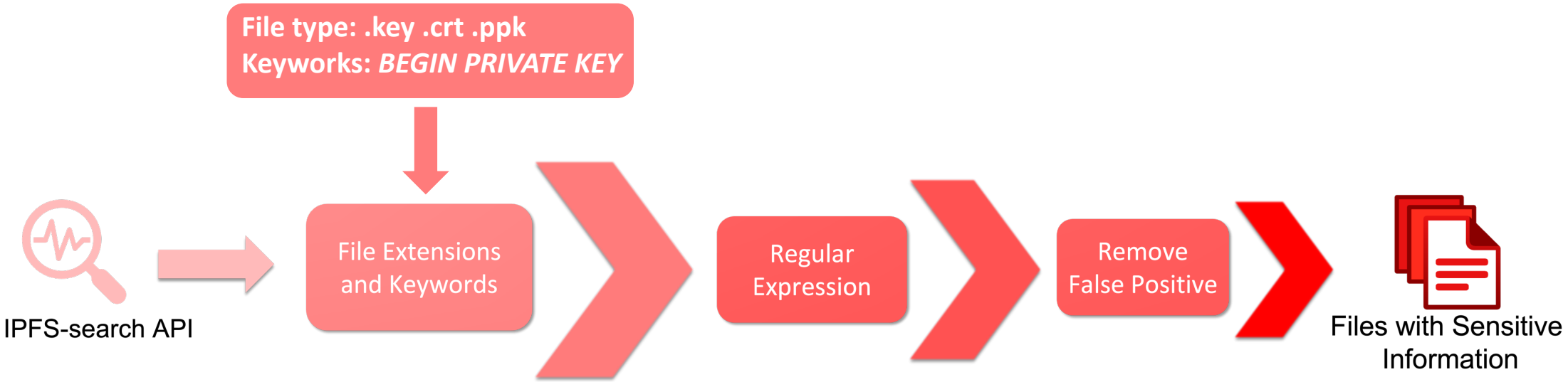
Multi-stage filter for sensitive files



FAR BEYOND

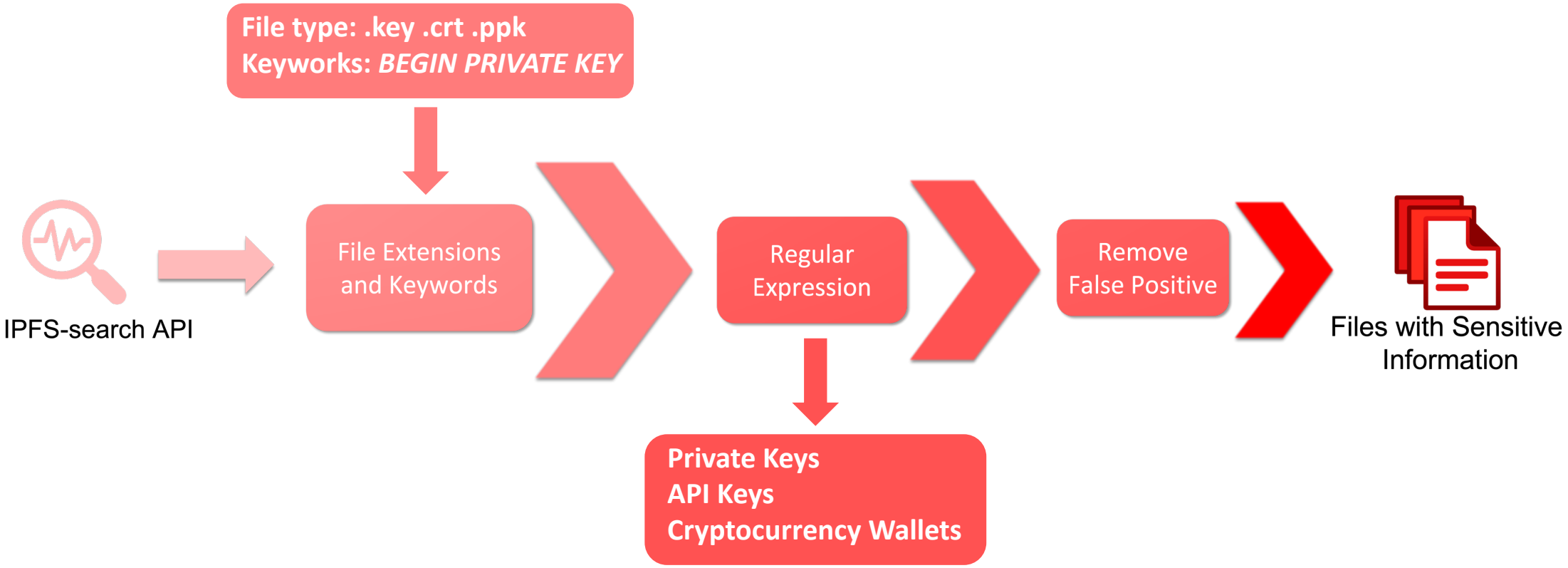
Stony Brook University

Multi-stage filter for sensitive files



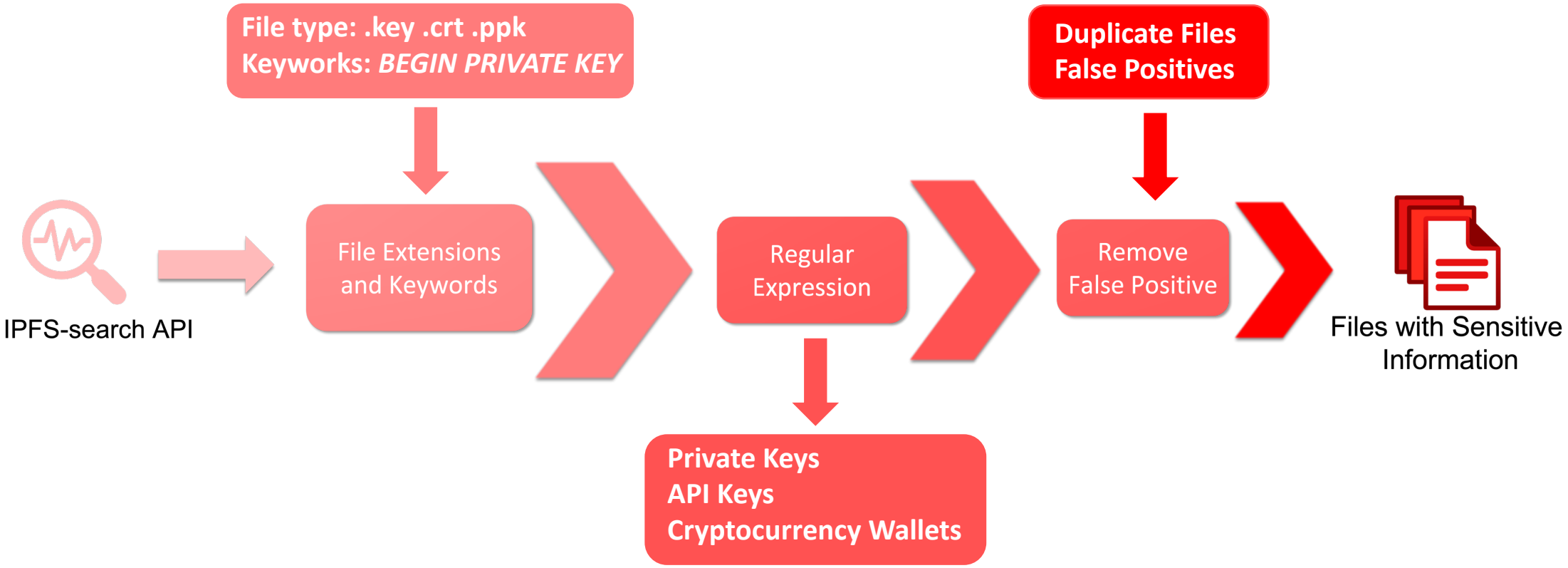
FAR BEYOND

Multi-stage filter for sensitive files



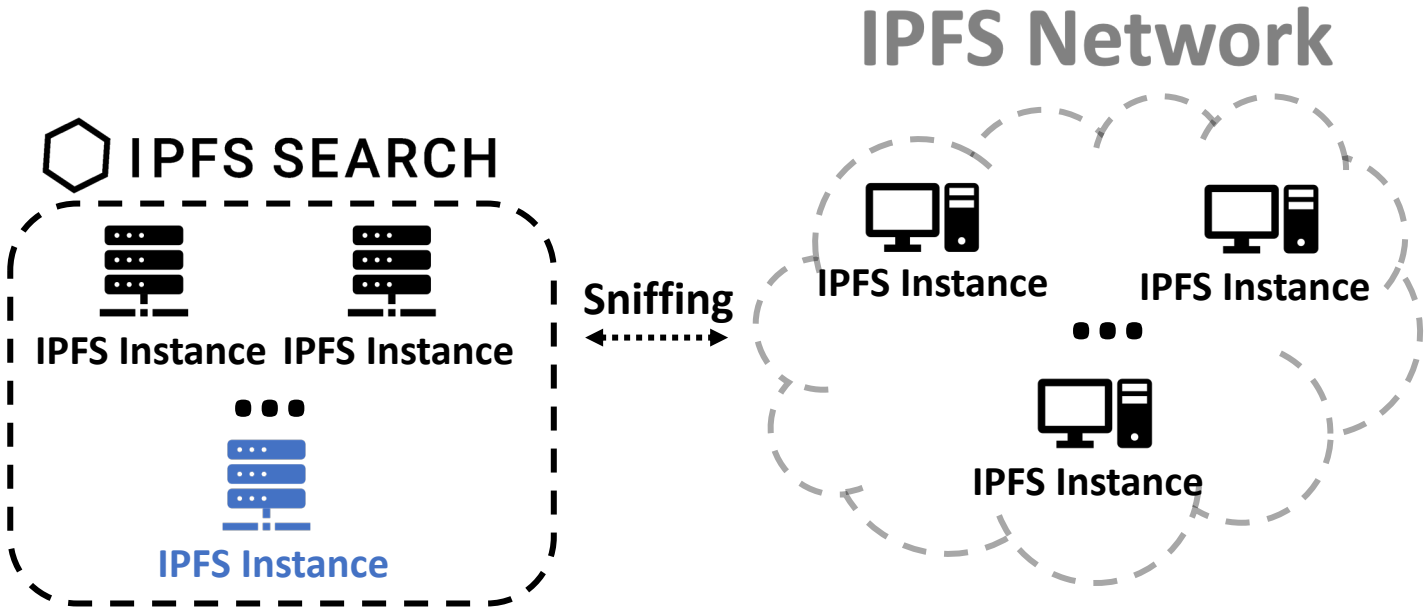
Stony Brook University

Multi-stage filter for sensitive files

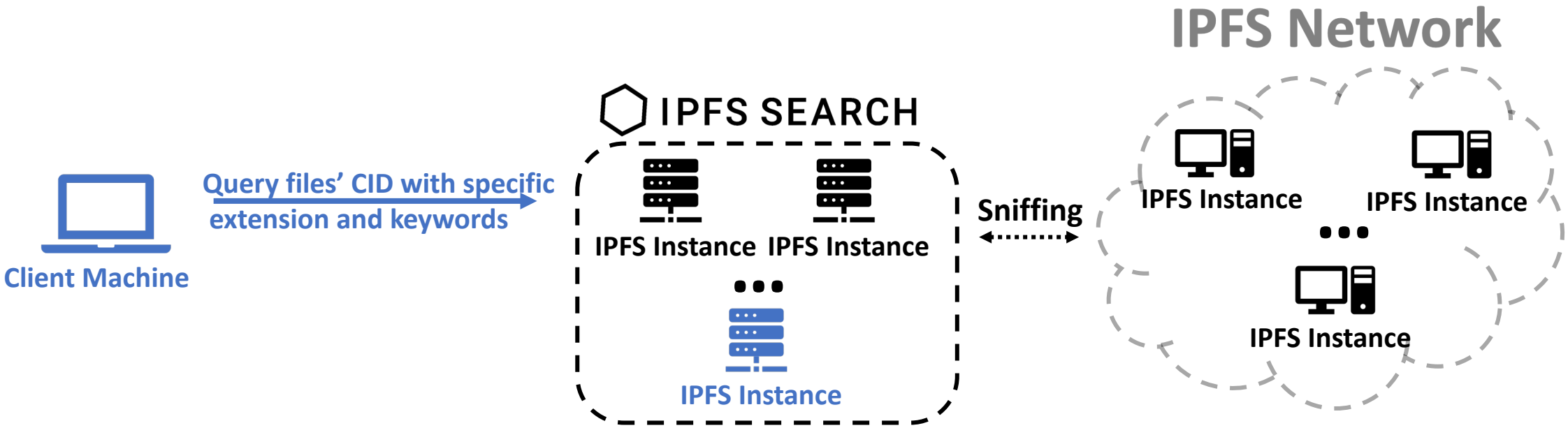


FAR BEYOND

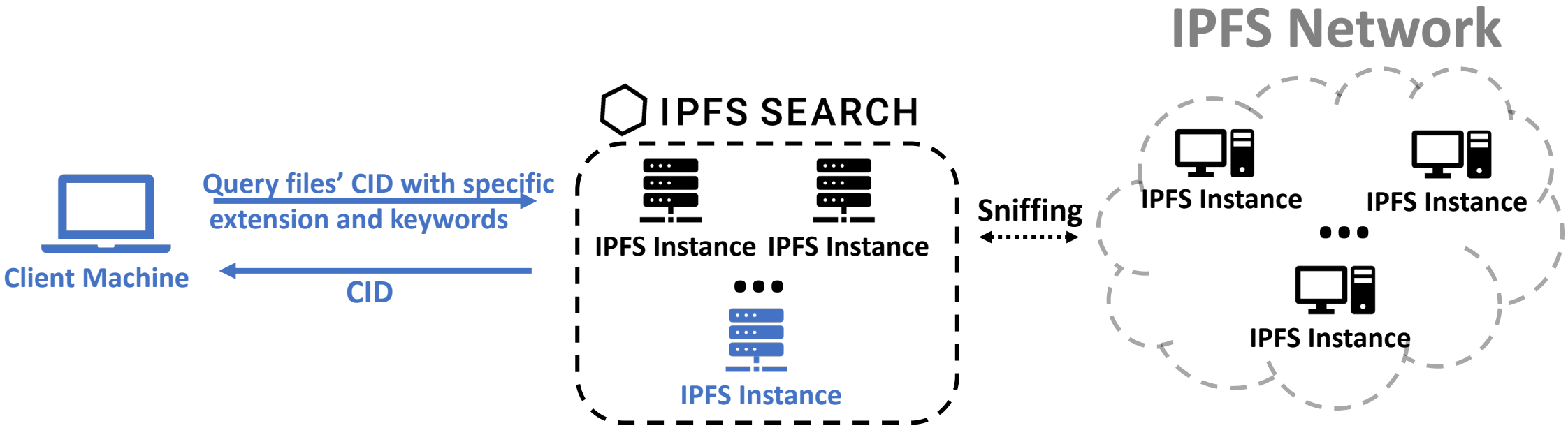
Measurement setup



Measurement setup

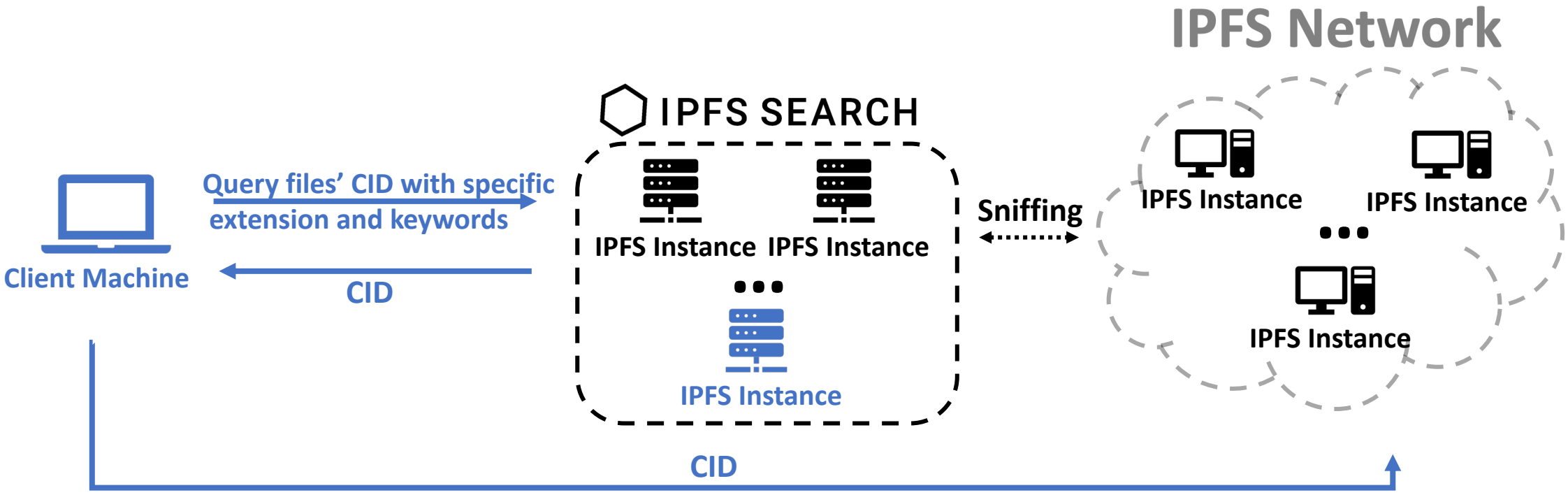


Measurement setup



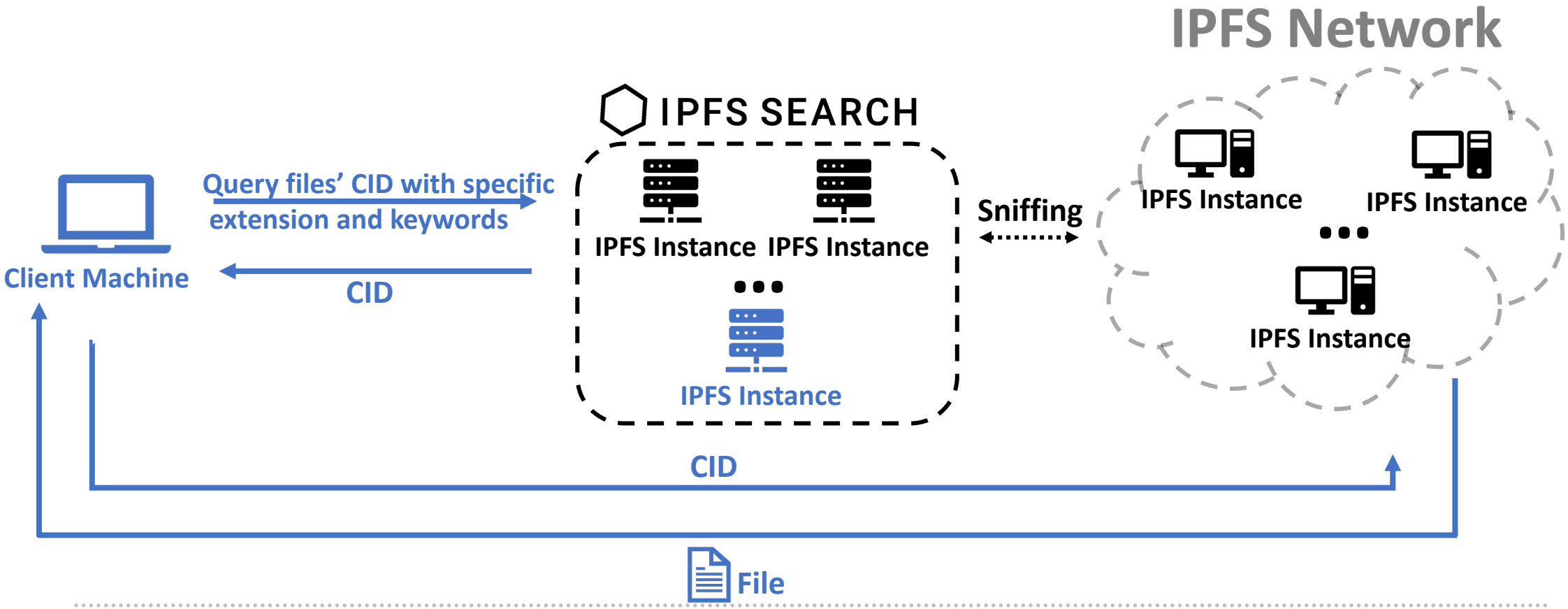
Stony Brook University

Measurement setup



FAR BEYOND

Measurement setup



Files collected

- Collect data from September 21, 2022, to November 30, 2022
- Total of 10,777 files were downloaded from the IPFS network
- Approximately **0.03%** of the total file index by IPFS-search during the same time period

Files collected

- Collect data from September 21, 2022, to November 30, 2022
- Total of 10,777 files were downloaded from the IPFS network
- Approximately **0.03%** of the total file index by IPFS-search during the same time period

File MIME Type	Count	File MIME Type	Count
gzip	4,353	json	902
plain/text	2,449	pgp-signature	396
zip	2,178	other (epub, html, ….)	526

Leaks in files

- We found ***236*** non-compressed files that contain sensitive leaks, out of 4,061 files
- We found ***1,788*** files that contained sensitive leaks, among the 6,716 compressed files

Leaks in files

- We found ***236*** non-compressed files that contain sensitive leaks, out of 4,061 files
- We found ***1,788*** files that contained sensitive leaks, among the 6,716 compressed files

Sensitive Match	Non-compressed File	Sensitive Match	Compressed File
RSA Private Key	77	General Private Key	124
SSH Private Key	22	EC Private Key	74
Google OAuthID	10	Google OAuthID	645
Google API	7	Google API	328

Stony Brook University

Leaks in files

- We found ***236*** non-compressed files that contain sensitive leaks, out of 4,061 files
- We found ***1,788*** files that contained sensitive leaks, among the 6,716 compressed files

We also conduct similar study by deploying our own search and find similar result

FAR BEYOND

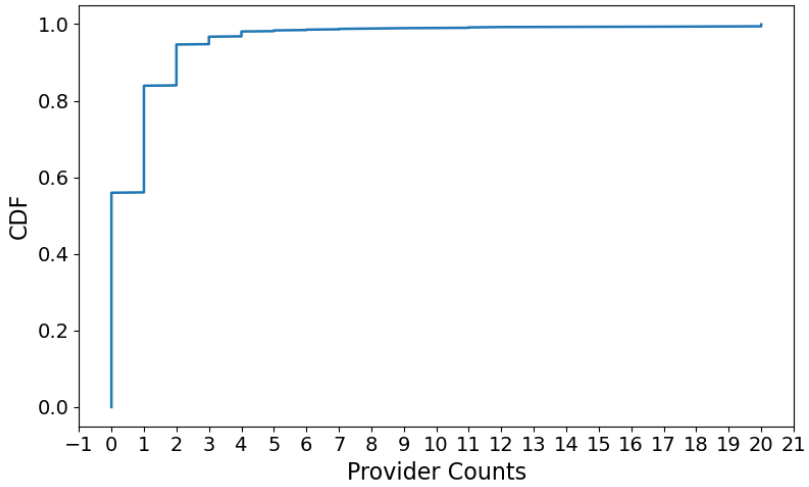
Stony Brook University

File longevity on IPFS

Stony Brook University

File longevity on IPFS

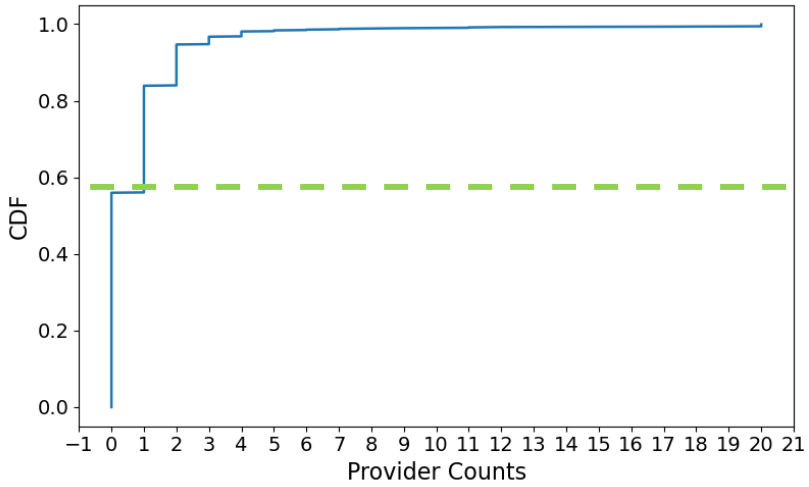
- 40% of the sensitive files are still available after 6-month



Stony Brook University

File longevity on IPFS

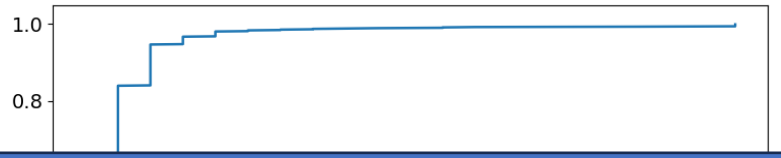
- 40% of the sensitive files are still available after 6-month



Stony Brook University

File longevity on IPFS

- 40% of the sensitive files are still available after 6-month



We also conduct study with files on GitHub, and found over 50% of the leaks are patched by owner. But the version that contains leaks are still available on IPFS

Conclusion

- **Sensitive Data on IPFS**: Our research uncovered that private keys and API tokens are openly shared and accessible via their CID on IPFS
- **Impact of Persistent Leaks**: Due to the distributed nature of IPFS, sensitive files can persist for a long time
- **Possible solution:** Encrypt CID requests during the announcement and retrieval phases

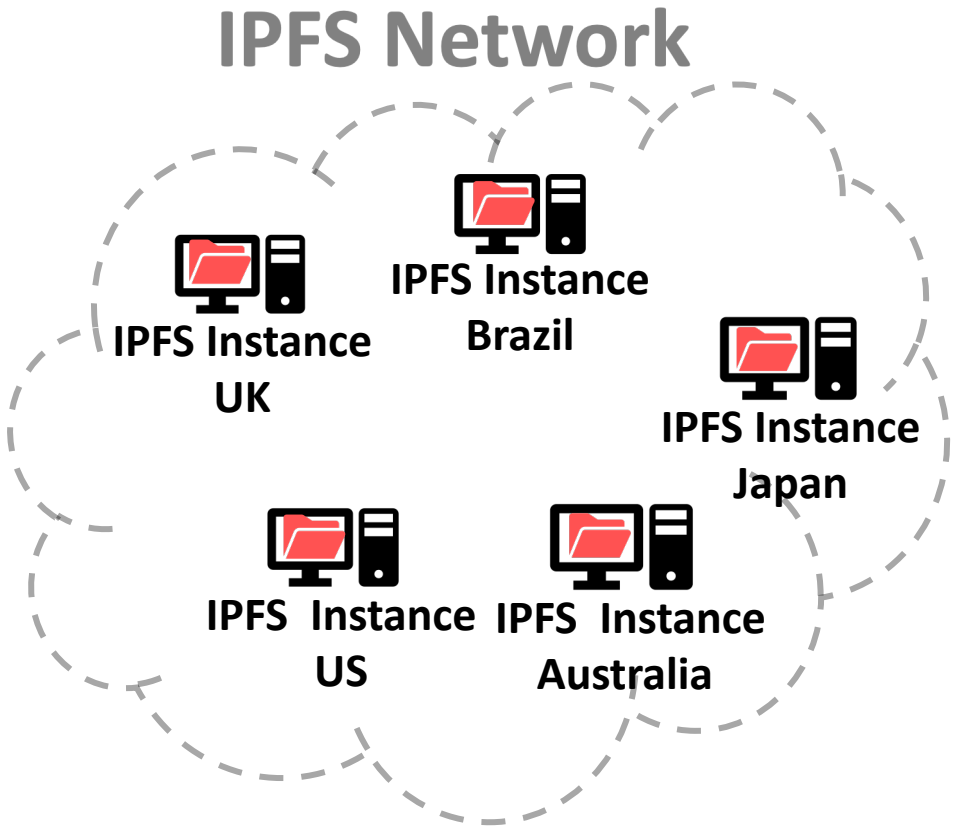
Stony Brook University

Q&A

FAR BEYOND

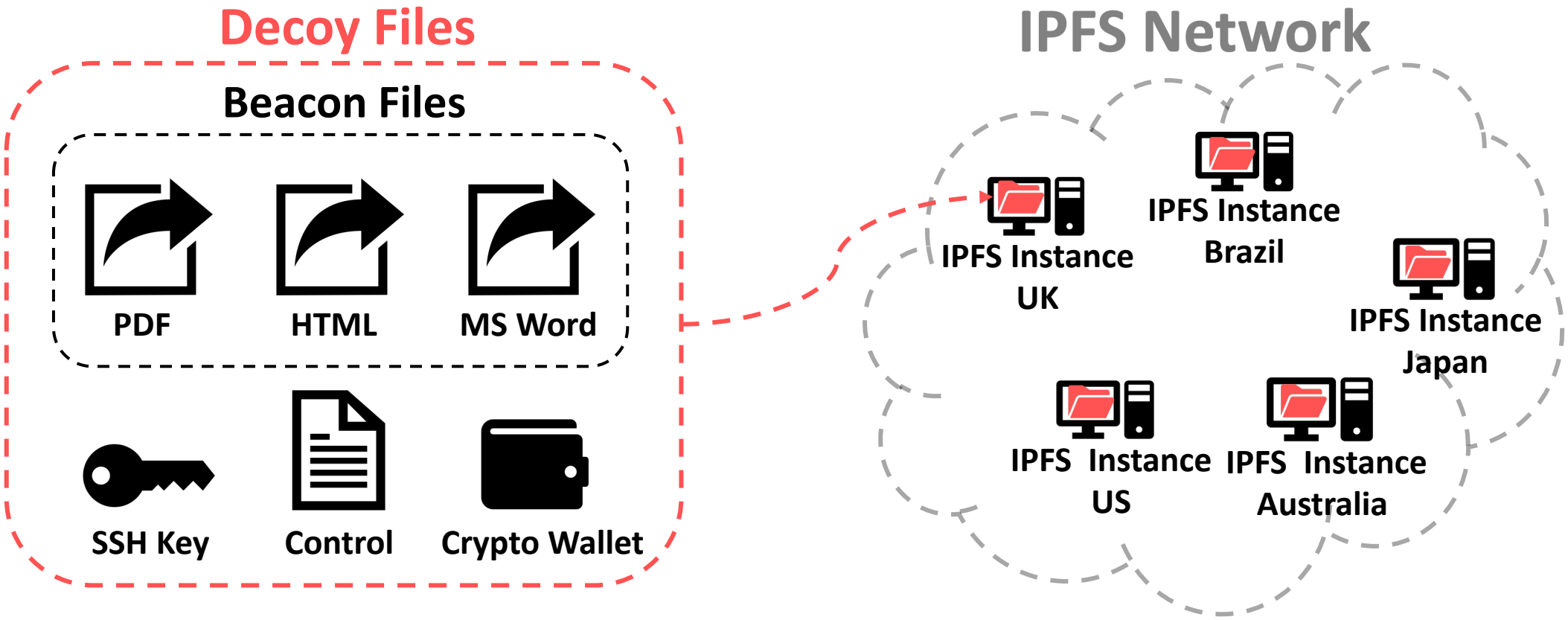
Honeypot Experiment

Honeypot Experiment



Stony Brook University

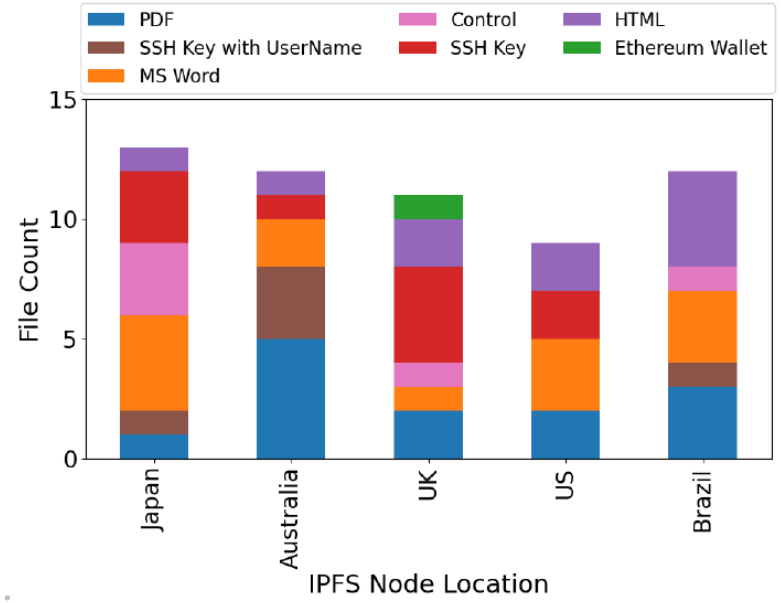
Honeypot Experiment



FAR BEYOND

Honeypot Result

- We observed different files are being download from our honeypot node
- We did not notice any attack being performed



Downloaded File Type From Different Location