



Interpol 2nd I-SOP Webinar

Analysis of IPFS (InterPlanetary File System)

Jihun Son



Korea University
School of Cybersecurity



Korea University
Digital Forensic Research Center

Jihun Son(Hunjison)

Education

- [~2020] B.S. Law, Korea National Policy Univ.
- [2022~] M.S. Information Security, Korea Univ.

Experience

- [2020~] Investigator, Seoul Metropolitan Police Agency
- [2022~] Digital Forensic Research Center, Korea Univ.

Publication

- Messengers / Wallets / Decentralized storages

Forensic analysis of instant messengers: Decrypt Signal, Wickr, and Threema J Son, YW Kim, DB Oh, K Kim Forensic Science International: Digital Investigation 40, 301347	16	2022
Forensic analysis of MetaMask cryptocurrency wallet artifacts J Son, J Park J. Digit. Forensics 16, 151-165	3	2022
IF-DSS: A forensic investigation framework for decentralized storage services J Son, G Kim, H Jung, J Bang, J Park Forensic Science International: Digital Investigation 46, 301611		2023

Today's topic !



Technical something..

- [2021] 3 reports on bug bounty program of Naver
- [2022] Hackingcamp, Codegate, Codeengn conference

Online content



Table of Contents

- ❑ Introduction
- ❑ What is IPFS?
- ❑ Forensic Analysis of IPFS
 - Remote-side Investigation Case
 - Local-side Investigation Case
- ❑ IF-DSS: Forensic Investigation Framework

Keywords



InterPlanetary File System



Forensic Framework



Phishing

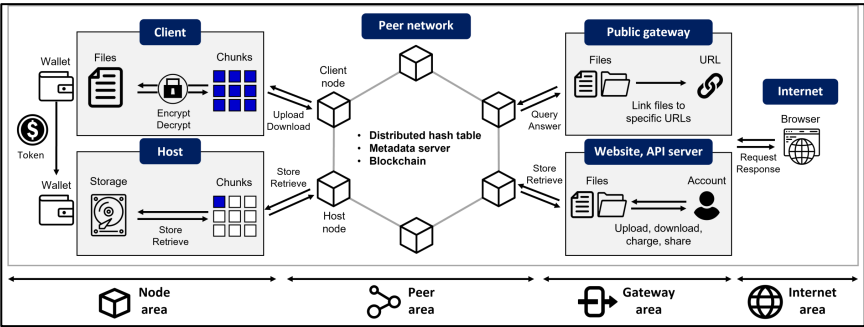
Copyright Infringement

Introduction

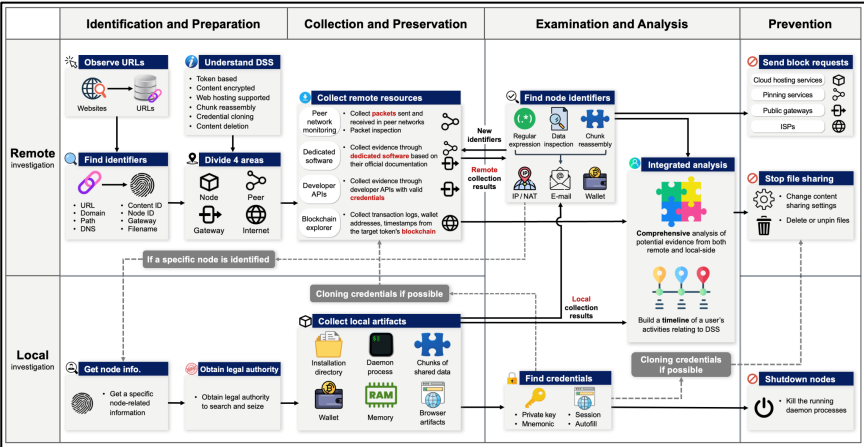
IPFS(InterPlanetary File System)

- **IPFS is popular, but it has been abused**
 - Filecoin has a market capitalization of over 2 billion, which is 31st in the overall coin market
 - Over 300,000 phishing URLs are reported by major vendors each month
 - This number is growing rapidly
 - It can be abused illegally because of its privacy and censorship resistance
 - Anna's Archive(part of **Z-Library**) has uploaded 6 million books on IPFS network
- **Forensic Investigation of IPFS is challenging**
 - All data uploaded to DSS are distributed across the node, peer, gateway, and Internet areas
 - Service provider possess limited user data and often reluctant to cooperate with law enforcement agencies
 - Existing forensic frameworks for cloud storage and P2P have limitations for DSS

Summary

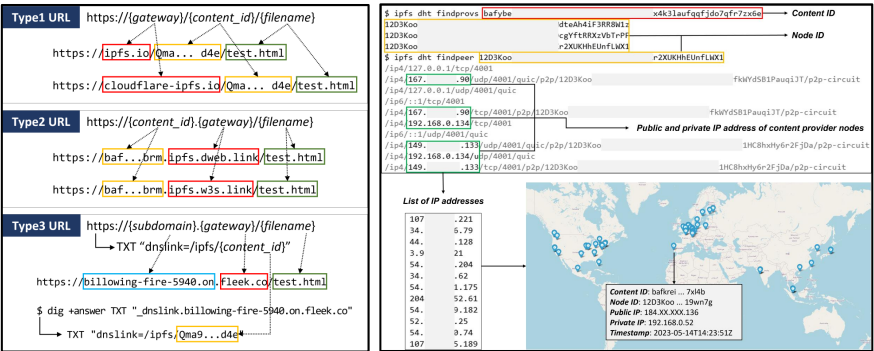


An overview: How IPFS works

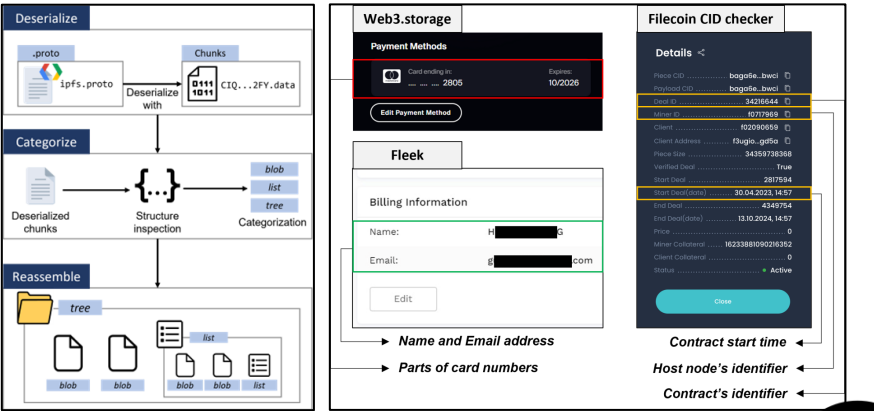


Forensic Investigation Framework

Case 1 : Phishing URLs Remote-side



Case 2 : IPFS host node Local-side



Tools Dataset Manuals

IF-DSS: forensic Investigation Framework of Decentralized Storage Services



Congratulations! This project has been accepted to DFRWS APAC 2023 and Forensic Science International: Digital Investigation

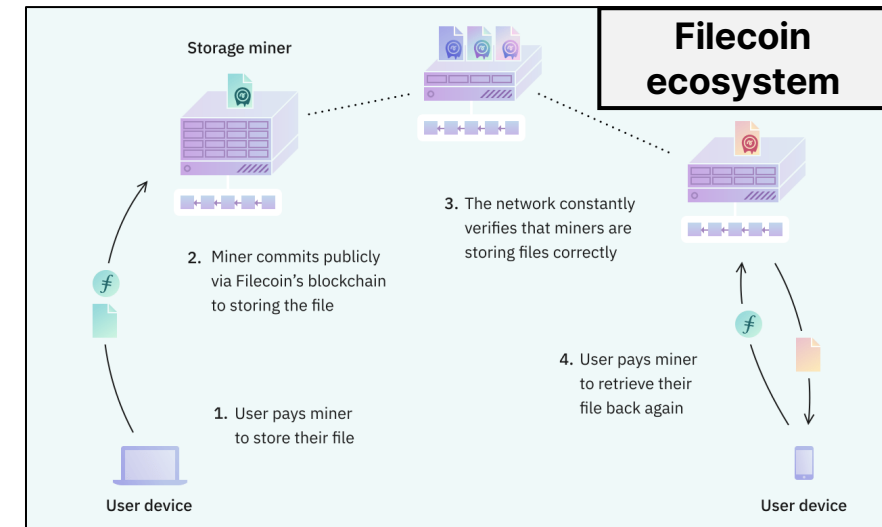
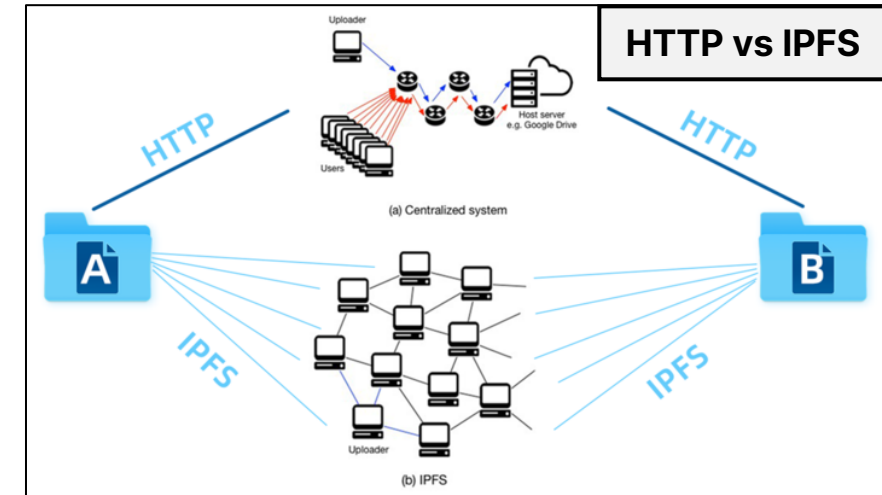
We propose a novel forensic investigation framework for DSS, named IF-DSS, as depicted in Figure 2. The framework comprises detailed steps that are categorized into remote and local investigations, depending on the location where potential digital evidence may be stored. Building upon the traditional digital forensic framework, our IF-DSS framework incorporates the necessary steps for effectively responding to DSS, including (1) identification and preparation, (2) collection and preservation, (3) examination and analysis, and (4) prevention.

What is IPFS?

What is IPFS?

IPFS and Filecoin

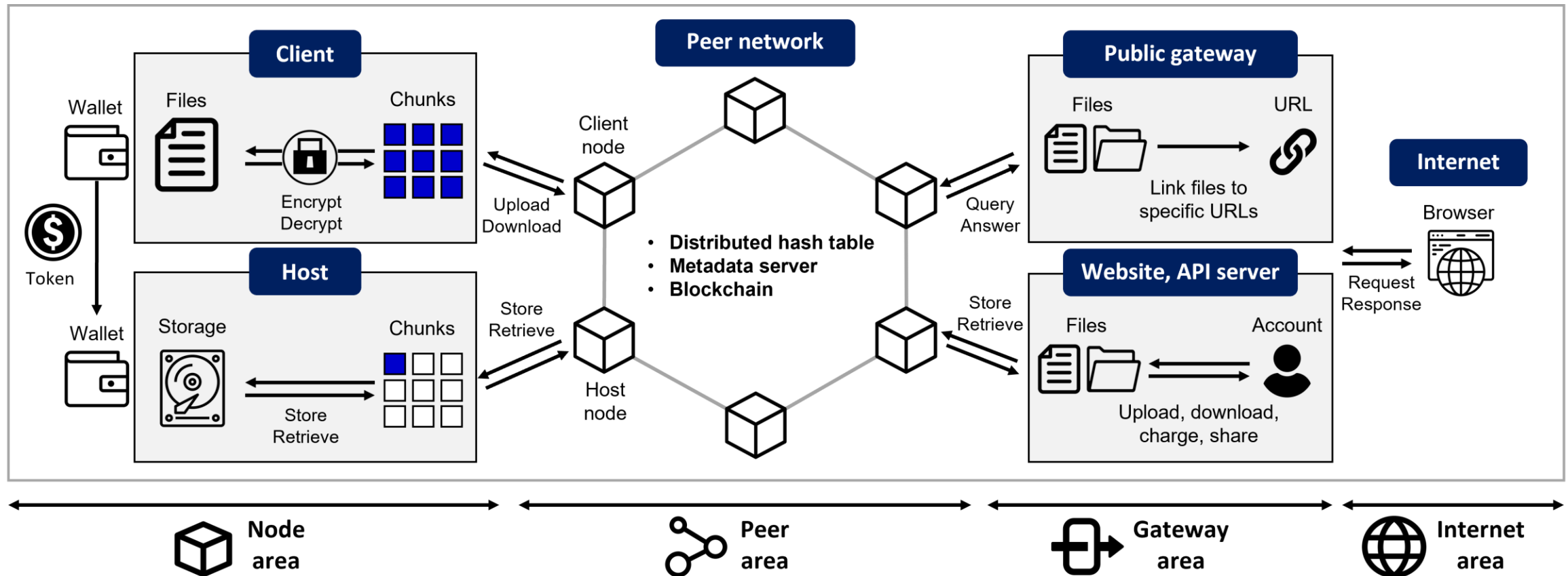
- **IPFS**  IPFS
 - Peer-to-peer, **Decentralized** data storage!
 - **Faster** than normal HTTP
 - No single failure (Resilience)
- **Filecoin**  Filecoin
 - Data storage service based on IPFS
 - Filecoin is ERC-20 token
 - Incentive mechanism for IPFS host nodes



What is IPFS?

An overview: How IPFS and Filecoin works

- It consists of 4 areas

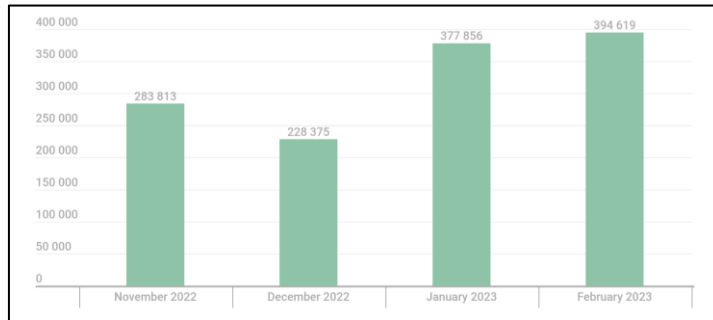


What is IPFS?

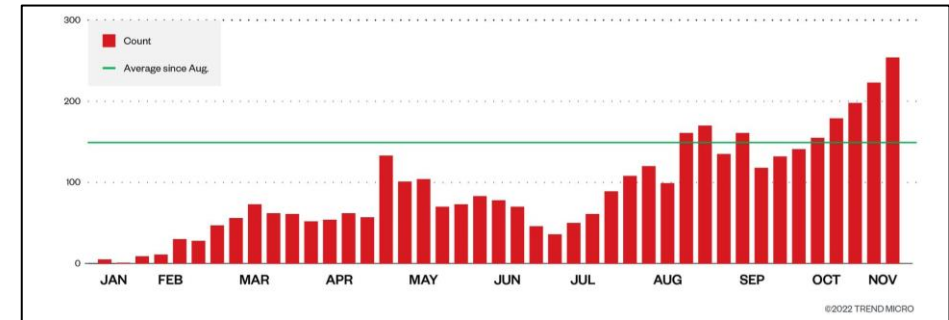
Potential Crime Scenarios

■ Phishing

- IPFS provides free web-hosting domains, SSL certification, and no content-deletion



← Kaspersky
Trend Micro →



■ Copyright Infringement

- Anna's archive(part of **Z-Library**) have uploaded about 6 million books on IPFS

Anna's Archive

The largest truly open library in human history. ★ Mirrors Sci-Hub, LibGen, Z-Lib, and more.

25,116,839 books, 99,425,860 papers — preserved forever. Learn more...

Recent downloads: ioteur: Dieu se tait. Le diable murmure • End Game • Comprendre le monde - 6e éd. • Sous vide grillé: the

- Option #7: [Libgen.rs Fiction](#) (click "GET" at the top)
- Option #8: [Libgen.li](#) (also click "GET" at the top)
- Option #9: [IPFS Gateway #1](#) (you might need to try multiple times with IPFS)
- Option #10: [IPFS Gateway #2](#)
- Option #11: [IPFS Gateway #3](#)
- Option #12: [Bulk torrent downloads](#) (experts only)

Forensic Analysis of IPFS

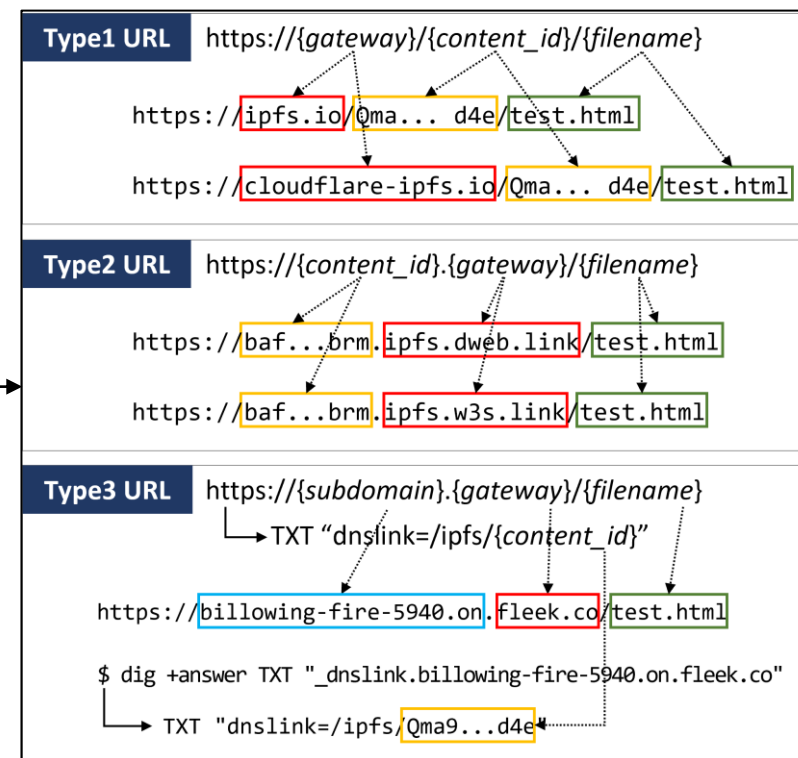
Remote and Local Investigation

1) Remote Investigation

Investigating phishing URLs hosted on the IPFS network

- We identified 3,662 IPFS-related URLs out of 61,528 Phishing URLs (from *Phishtank*)
- There are 3 types of IPFS URLs
 - Content ID(CID), Gateway name, Filename can be identified
 - Type3 URL has its CID in the TXT record of the DNS

phish_id	url
8146175	https://bafybeictu25z5ncxaqkhrotbm57aplkm3g2aexbbckk76gg2juyzp64ju.ipfs.dweb.link/adminspaze.html
8146140	https://bafkreigdqockwil32afzmsayaszse6way4sftsaikhfohufxwel2mdp6yi.ipfs.dweb.link/#noc@protocol.ai
8145974	https://ipfs.io/ipfs/bafybeigr6gnras7lkwdn74px2pdfhnnlo46xxggrenmpjefjtvixsi6jvu/#aaaa@example.jp
8145854	https://cloudflare-ipfs.com/ipfs/bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu
8145853	https://gateway.pinata.cloud/ipfs/bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu
8145852	https://bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu.ipfs.infura-ipfs.io
8145851	https://gateway.ipfs.io/ipfs/bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu
8145850	https://ipfs.io/ipfs/bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu
8145849	https://bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu.ipfs.cf-ipfs.com
8145848	https://bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu.ipfs.nftstorage.link
8145846	https://nftstorage.link/ipfs/QmTVfgwb4L9pJcqiTbSxsykQqdng7daqcfqNYwCR3doZkQ
8145847	https://bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu.ipfs.dweb.link/
8145845	https://infura-ipfs.io/ipfs/QmTVfgwb4L9pJcqiTbSxsykQqdng7daqcfqNYwCR3doZkQ
8145844	https://ipfs.io/ipfs/QmTVfgwb4L9pJcqiTbSxsykQqdng7daqcfqNYwCR3doZkQ
8145843	https://nftstorage.link/ipfs/Qmc8tihErcRPih7EomwWsC4a8UMUTof1HTqRBuATmiG3aM
8145842	https://bafybeicmthcrr5cbjgisyb6gqu5pnytwk4yf2cyu2rkbhfmormsmfwf4qu.ipfs.cf-ipfs.com/
8145841	https://cf-ipfs.com/ipfs/QmTVfgwb4L9pJcqiTbSxsykQqdng7daqcfqNYwCR3doZkQ
8145840	https://infura-ipfs.io/ipfs/Qmc8tihErcRPih7EomwWsC4a8UMUTof1HTqRBuATmiG3aM
8145839	https://bafybeignab6rgfrvqahuqzo65shoe2h5azjrcjtfekkrm2xwnclw3mwgvhy.ipfs.cf-ipfs.com/
8145838	https://cf-ipfs.com/ipfs/Qmc8tihErcRPih7EomwWsC4a8UMUTof1HTqRBuATmiG3aM



1) Remote Investigation

■ Collecting remote resources

■ Dedicated software(*Kubo*)

- Command "**findprovs**": CID → node ID
- Command "**findpeer**": node ID → IP address
- IP-based locations could be found

■ Another methods

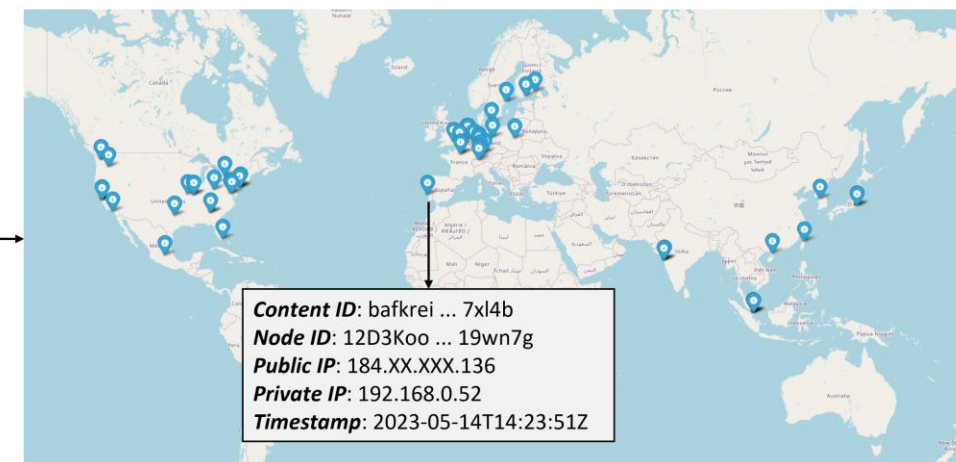
- Peer network monitoring (from peer area)
- Developer API (from gateway area)
- Blockchain explorer (from blockchain area)

```
$ ipfs dht findprovs bafybe x4k3laufqqfjdo7qfr7zx6e → Content ID
12D3Koo dteAh4iF3RR8W1z → Node ID
12D3Koo cgYftRRXzVbTrPF
12D3Koo r2XUKHhEUnfLWX1
$ ipfs dht findpeer 12D3Koo r2XUKHhEUnfLWX1
/ip4/127.0.0.1/tcp/4001
/ip4/167. .90/udp/4001/quic/p2p/12D3Koo fkWYdSB1PauqiJT/p2p-circuit
/ip4/127.0.0.1/udp/4001/quic
/ip6:::1/tcp/4001
/ip4/167. .90/tcp/4001/p2p/12D3Koo fkWYdSB1PauqiJT/p2p-circuit
/ip4/192.168.0.134/tcp/4001
/ip6:::1/udp/4001/quic
/ip4/149. .133/udp/4001/quic/p2p/12D3Koo 1HC8hxHy6r2FjDa/p2p-circuit
/ip4/192.168.0.134/udp/4001/quic
/ip4/149. .133/tcp/4001/p2p/12D3Koo 1HC8hxHy6r2FjDa/p2p-circuit
```

Public and private IP address of content provider nodes

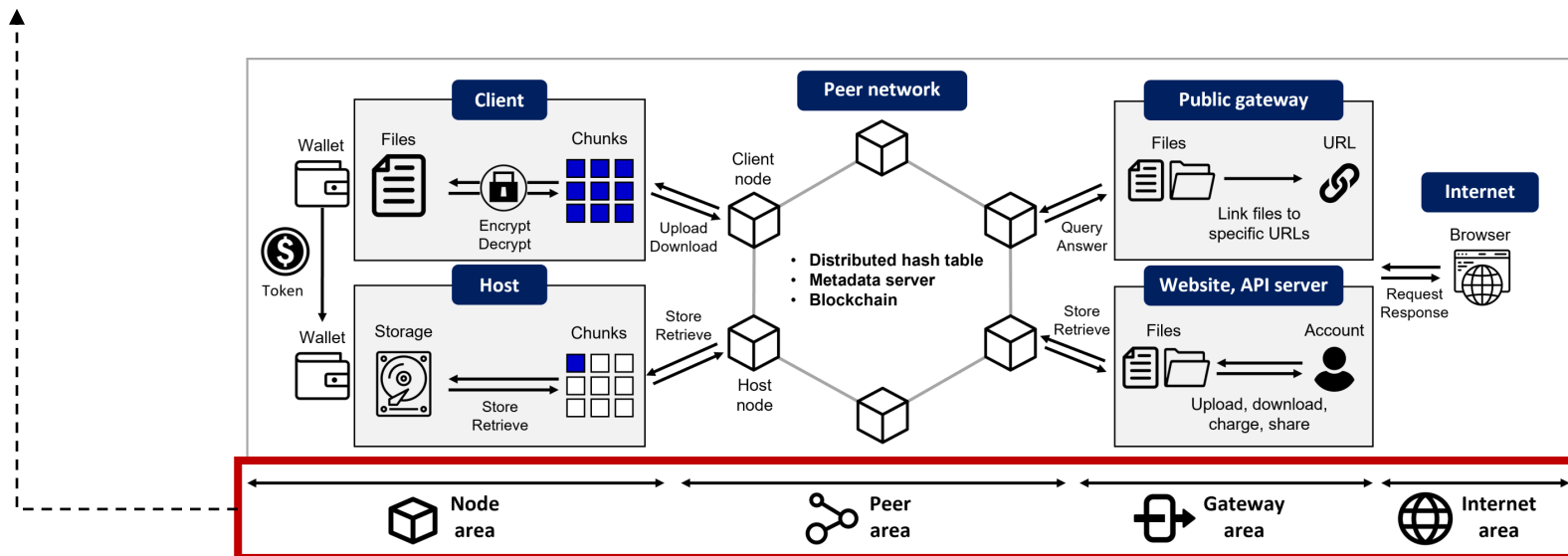
List of IP addresses

107	.221
34.	6.79
44.	.128
3.9	21
54.	.204
34.	.62
54.	1.175
204	52.61
54.	9.182
52.	.25
54.	0.74
107	5.189



1) Remote Investigation

- Preventing further file distribution
 - Sending content-blocking requests to ...
 - [Internet area] ISP: No response
 - [Gateway area] ipfs.io(IPFS's official gateway): Blocked in a day
 - [Peer area] Pinata(pinning services): No response
 - [Node area] AWS(cloud hosting service): Blocked in a day



2) Local Investigation

■ Case study

- We created an IPFS node that hosted a *distributed-Wikipedia-mirror*, and assumed it is illegal content

■ Collecting local artifacts

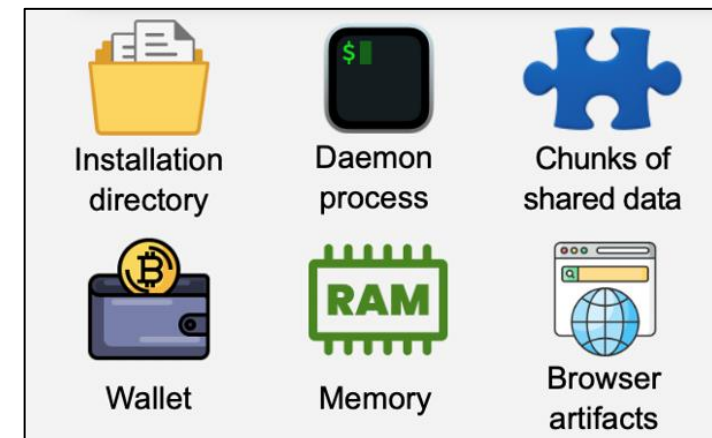
- Installation directory of Kubo(IPFS's dedicated software)
 - File chunks, Program logs ...
- Chrome artifacts
 - History database: Visiting records
 - Credentials: Website cookies
- Process memory dumps



Distributed Wikipedia Mirror Project [↗](#)

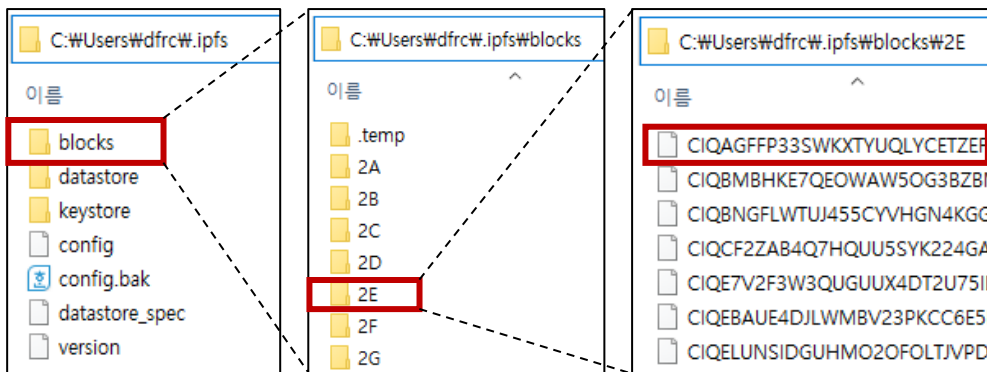
Putting Wikipedia Snapshots on IPFS and working towards making it fully read-write.

OS	Path
Windows	/Users/<Username>/ipfs
Linux	/home/<Username>/ipfs



2) Local Investigation

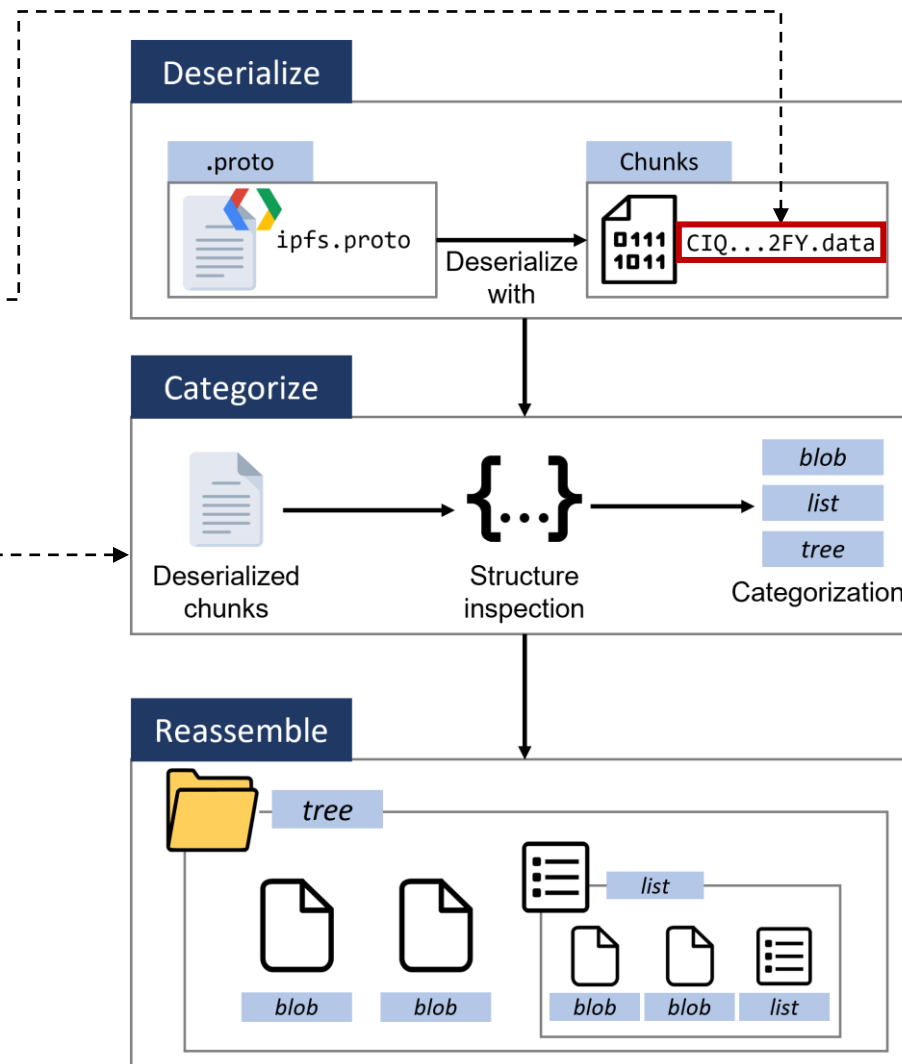
■ Reassembling IPFS file chunks



```
// BLOB
{"data": 0x123456 ...}

// LIST
{"data": ["BLOB", "LIST", "BLOB"],
"links":{"hash":Qm..., "size": 1234}}

// TREE
{"data": ["BLOB", "LIST", "BLOB"],
"links":{"hash":Qm..., "size": 1234, "name": "README.md"}}
```



2) Local Investigation

■ Cloning Credentials

- IPFS Installation directory is copied to another laptop
- Chrome's cookies are copied to another laptop

■ Collecting Remote resources

- Accessing through the cloned credentials, we obtained an email address and a parts of registered card numbers from Web3.storage and Fleek websites
- Uploaded files can be searched on Filecoin explorer, which retrieve miner IDs and timestamps

Web3.storage

Payment Methods

Card ending in: ... 2805 Expires: 10/2026

Edit Payment Method

Fleek

Billing Information

Name: H [redacted] G

Email: g [redacted] .com

Edit

Filecoin CID checker

Details

Piece CID b9a6e...bwci

Payload CID b9a6e...bwci

Deal ID 34216644

Miner ID f0717969

Client f02090659

Client Address f3ugio...gd5a

Piece Size 34359738368

Verified Deal True

Start Deal 2817594

Start Deal(date) 30.04.2023, 14:57

End Deal 4349754

End Deal(date) 13.10.2024, 14:57

Price 0

Miner Collateral 16233881090216352

Client Collateral 0

Status Active

Close

→ Name and Email address

→ Parts of card numbers

← Contract start time

← Host node's identifier

← Contract's identifier


Dataset / Tools / Manuals

■ Github Repository (<https://github.com/hunjison/IF-DSS>)

hunjison Update README.md	
dataset	First commit to this repo
img	First commit to this repo
output	First commit to this repo
src	First commit to this repo
LICENSE	Initial commit
README.md	Update README.md
main.py	First commit to this repo
requirements.txt	First commit to this repo

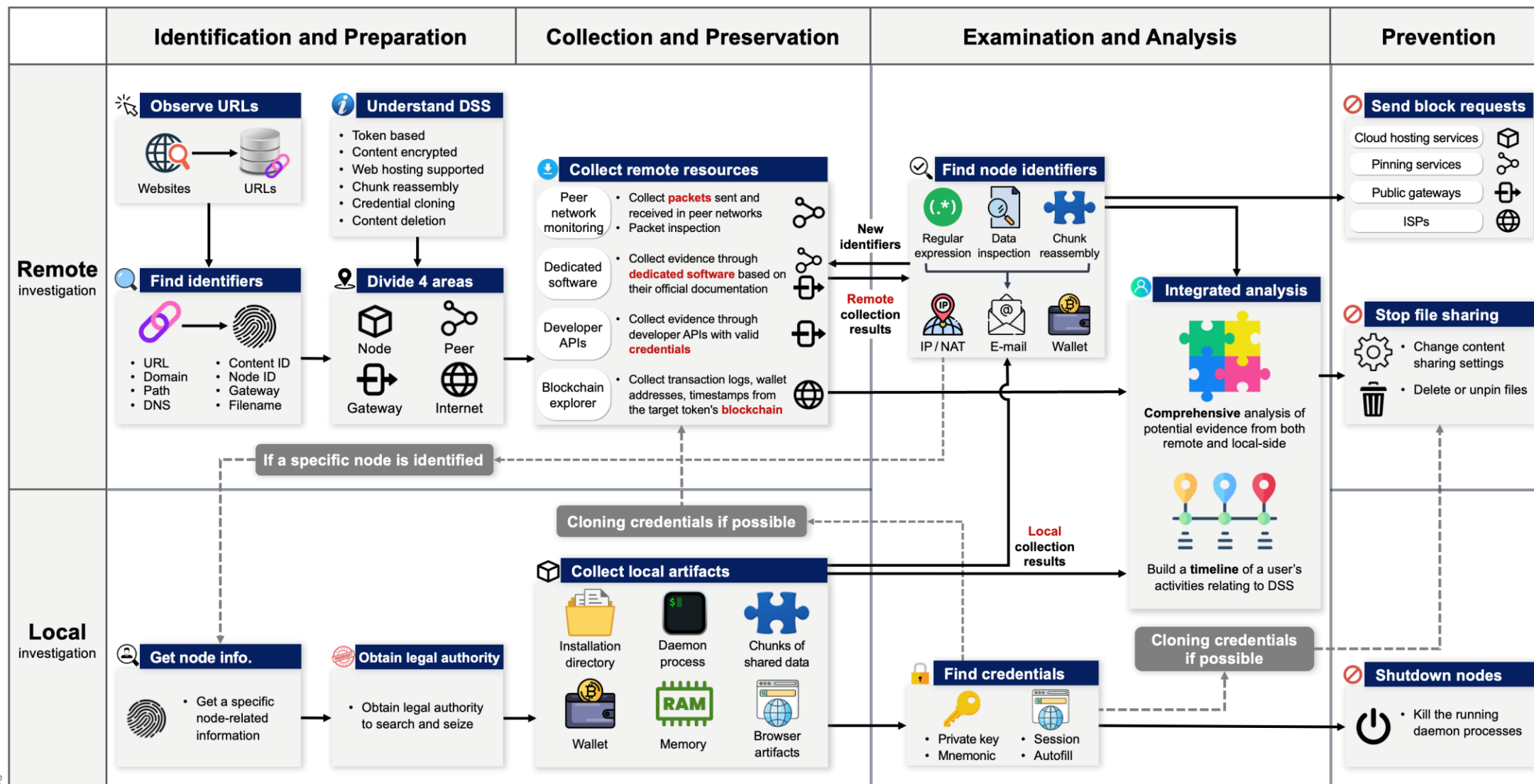
Dataset	
.temp	CIQPX5BA6HOG7FDEPQ7KA74QR2EN2M6Z6CXLCP
2A	CIQPRLFORIE3JY4I4YQJW7HP3BMRMNWTYBDCRQRY
2B	CIQPONYFKYCSOBU5NHGNEI4H7D5A22SPUVILB2MFJ5GMSJXNNGBV2PI.data
2D	CIQPIWSBLA54TLABIYJR2TKJXOS7M4LJZ5TX64XV3KM2DR4YKJY3UA.data
2E	CIQPB26ZEULRXCAUG3SOT5XKIOH6EOQZNZLLA7CARNJFX3TX6BJA72Y.data
2F	CIQPAU57E625A2BFWB5HNI3CL3UTRIXFCDVILTP35JDXDDDCGMPV6I.data
2G	CIQP5FWA2UMEL3PPXQM36ITKEUKZ32NKRUQIC5RHTDGDGCDV6XQ2QBVA.data
	CIQOUN3F2W2ATMFJXOMKM2J6UV6W4BOF34KDOV5IJ4PHVKYPL6PVCQ.data

1.4 Prevention	Manuals
We have prepared information that may be helpful for prevention.	
<ul style="list-style-type: none"> IPFS-based Services <ul style="list-style-type: none"> IPFS Pinning service : A service that pins a specific CID to prevent it from being garbage collected and deleted. ex) Pinata, Fleek, Filebase, Web3Storage, ntf.storage, Estuary IPFS Gateway service : A service that connects a specific CID to a corporate-owned gateway. ex) Pinata, Infura, Filebase, Fleek, Cloudflare Cloud IP Range <ul style="list-style-type: none"> AWS IP Range: AWS provides its own IP address ranges in JSON format. AZURE IP Range: AZURE provides its own IP range in JSON format. GCP IP Range: GCP does not publicly disclose IP ranges, but it is possible to check IP ranges through DNS. <pre># \$ cat gcp_ip_range.sh for LINE in \$(dig txt _cloud-netblocks.googleusercontent.com +short tr " " "\n" grep include); do do dig txt \${LINE} + short done tr " " "\n" grep ip4 cut -f 2 -d : sort -n</pre> Cloudflare IP Range: Cloudflare provides its own IP ranges in a txt format. 	

Tools
<pre>\$ python main.py</pre>  <p>Usage: main.py [OPTIONS] COMMAND [ARGS]...</p> <p>IF-DSS: forensic Investigation Framework for Decentralized Storage System Code examples for case studies related to IPFS with Filecoin</p> <p>Options:</p> <pre>--help Show this message and exit.</pre> <p>Commands:</p> <pre>imap Command to perform parsing parse Command to perform parsing reassemble Command to perform reassembly track Command to perform tracking trackdns Command to perform DNS operation</pre>

IF-DSS: Forensic Investigation Framework

A Forensic Investigation Framework for Decentralized Storage Services



A Forensic Investigation Framework for Decentralized Storage Services

- **Decentralized Storage Services**
 - IPFS, Filecoin based, BTFS(BitTorrent File System), ...

Comparison table between seven well-known DSSs in consideration of six forensics-related features.

Features	IPFS	Filecoin based	BTFS	Internet Computer	Storj	Sia	Arweave based
Cryptocurrency	N/A	Filecoin (FIL)	BitTorrent (BTT)	Internet Computer (ICP)	Storj (STORJ)	SiaCoin (SC)	Arweave (AR)
Data encryption	No	No	No	No	Yes	Yes	No
Web hosting	Yes	Depends	Yes	Yes	Yes	No	Depends
Chunk reassembly	Yes	No	Yes	No	No	No	No
Credential cloning	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Content deletion	No	No	No	Yes	Yes	Yes	No

*N/A means there is no cryptocurrency based.

A Forensic Investigation Framework for Decentralized Storage Services

- This paper provides a detailed explanation of each step of the IF-DSS framework



THANK YOU

