TECHNISCHE UNIVERSITÄT DARMSTADT

GUARDIANS OF THE GALAXY:

CONTENT MODERATION IN THE INTERPLANETARY FILE SYSTEM

USENIX SECURITY '24

**Saidu Sokoto, Leonhard Balduf, Dennis Trautwein,
Yiluo Wei, Gareth Tyson, Ignacio Castro, Onur Ascigil,
George Pavlou, Maciej Korczynski, Björn Scheuermann, Michał Król**

TECHNISCHE UNIVERSITÄT DARMSTADT

MOTIVATION

When considering a P2P network…

…with easy access via HTTP from the outside…

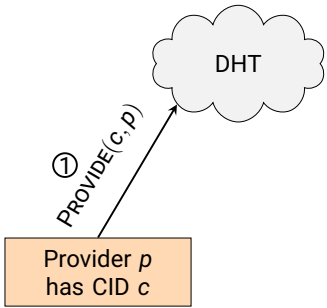
…**what could content moderation look like?**

OVERVIEW OF RESULTS

- We measure an existing solution to moderate problematic content in IPFS
- “Badbits” denylist list holds **400k+ entries**
- Majority of content **DMCA**-related or **phishing**, with stark growth recently
- Reaction times to **new** content can be slow
- Implementation challenging in decentralized setting

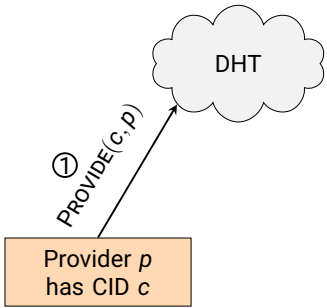
TECHNISCHE UNIVERSITÄT DARMSTADT

IPFS PRIMER

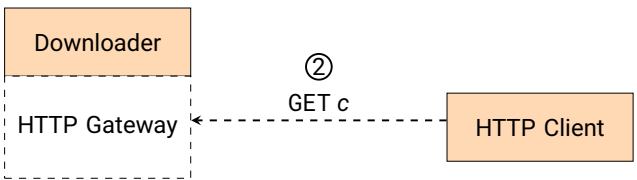


- Kademlia DHT

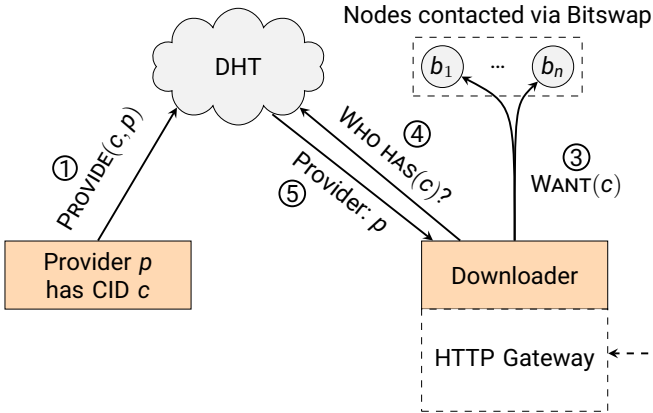
IPFS PRIMER



- Kademlia DHT
- Public gateway infrastructure
- HTTP ↔ IPFS translation



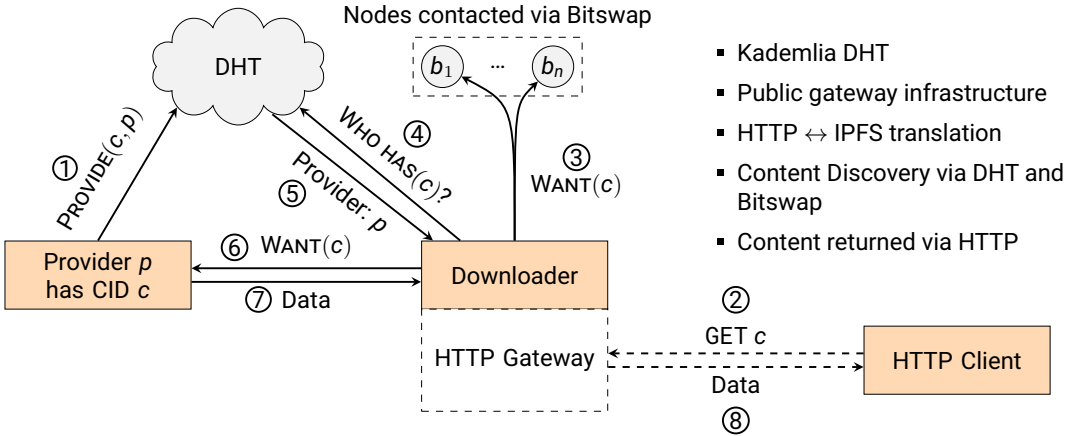
IPFS PRIMER



- Kademlia DHT
- Public gateway infrastructure
- HTTP ↔ IPFS translation
- Content Discovery via DHT and Bitswap

TECHNISCHE UNIVERSITÄT DARMSTADT

IPFS PRIMER



- Kademlia DHT
- Public gateway infrastructure
- HTTP ↔ IPFS translation
- Content Discovery via DHT and Bitswap
- Content returned via HTTP

HTTP ACCESS VIA PUBLIC GATEWAYS

(Public) Gateways translate HTTP↔IPFS

`https://ipfs.io/ipfs/bafybeigdy[…]tqy55fbzdi`

Gateway-as-a-Service providers offer chosen subdomains

`https://`**`web3-game`**`.example.com/ipfs/bafybeigdy[…]tqy55fbzdi`

TECHNISCHE UNIVERSITÄT DARMSTADT

HTTP ACCESS VIA PUBLIC GATEWAYS

(Public) Gateways translate HTTP↔IPFS

`https://ipfs.io/ipfs/bafybeigdy[…]tqy55fbzdi`

Gateway-as-a-Service providers offer chosen subdomains

`https://`**`microsoft-login`**`.example.com/ipfs/bafybeigdy[…]tqy55fbzdi`

CONTENT MODERATION?

- CIDs are opaque
- There is no central registry for all CIDs, *i.e.*, new content is difficult to find
- (Chosen-subdomain) public gateways make it easy to disseminate content using URLs
- Multiple URLs can point to the same content:

`https://ipfs.io/ipfs/bafybeigdy[…]tqy55fbzdi`

`https://cloudflare-ipfs.com/ipfs/bafybeigdy[…]tqy55fbzdi`

CONTENT MODERATION?

- CIDs are opaque
- There is no central registry for all CIDs, *i.e.*, new content is difficult to find
- (Chosen-subdomain) public gateways make it easy to disseminate content using URLs
- Multiple URLs can point to the same content:

`https://ipfs.io/ipfs/bafybeigdy[…]tqy55fbzdi`

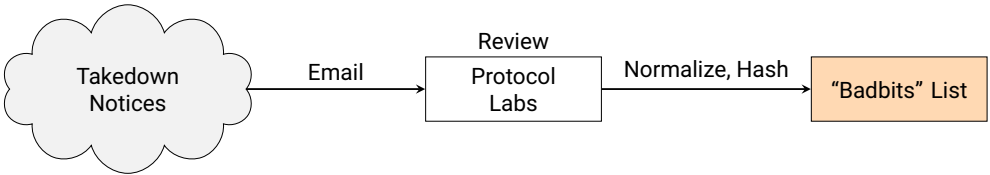
`https://cloudflare-ipfs.com/ipfs/bafybeigdy[…]tqy55fbzdi`

Traditional URL filtering **exists** but is **insufficient**.

TECHNISCHE UNIVERSITÄT DARMSTADT

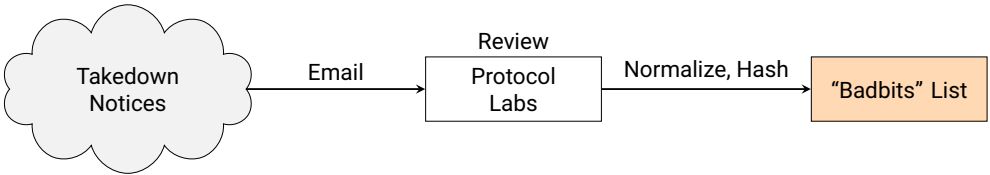
BADBITS LIST

Maintained by Protocol Labs, created in 2021, >400k entries (Jan ’24)



BADBITS LIST

Maintained by Protocol Labs, created in 2021, >400k entries (Jan ’24)

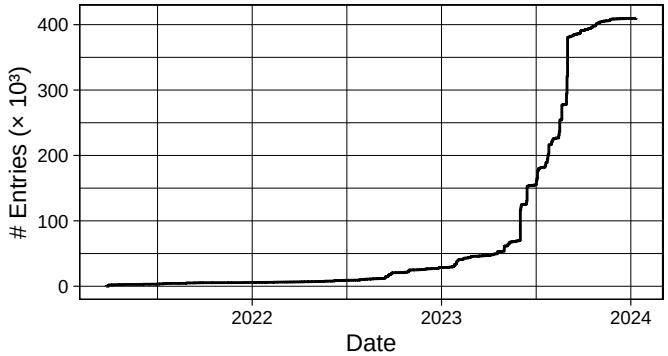


Support in recent client versions to check requested CIDs against the list:

TECHNISCHE UNIVERSITÄT DARMSTADT

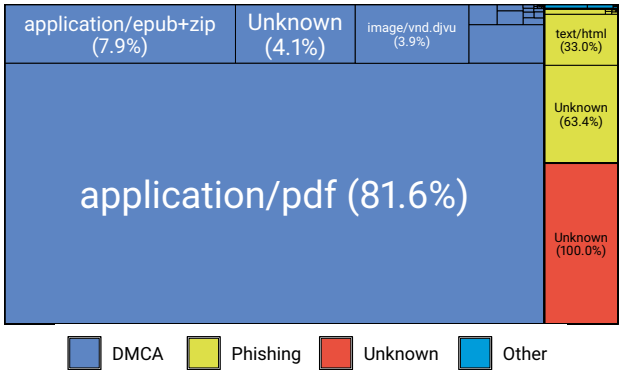
BADBITS LIST EVOLUTION

STARK GROWTH IN 2023



TECHNISCHE UNIVERSITÄT DARMSTADT

MOST ENTRIES: COPYRIGHT VIOLATIONS



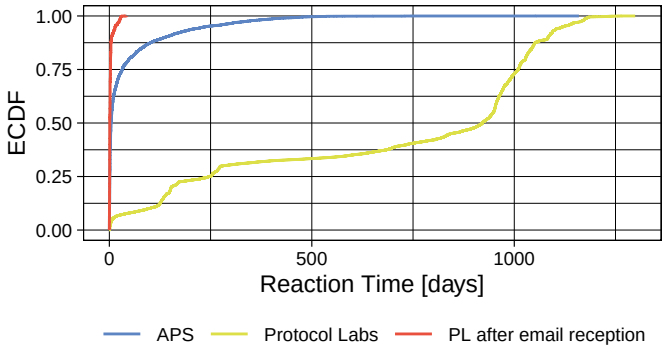
Most entries are DMCA-related

DMCA material mostly research papers

Phishing mostly HTML

TECHNISCHE UNIVERSITÄT DARMSTADT

HOW QUICKLY DO THE LISTS REACT?



Anti-Phishing Services relatively fast to react

PL (passively) very slow to react

PL (after email trigger) very quick

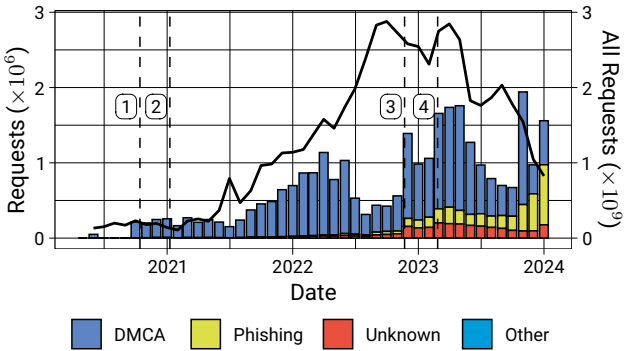
GROWING REQUEST ACTIVITY

UNIQUE PEER ID/CID COMBINATIONS PER MONTH

Copyrighted material is always popular

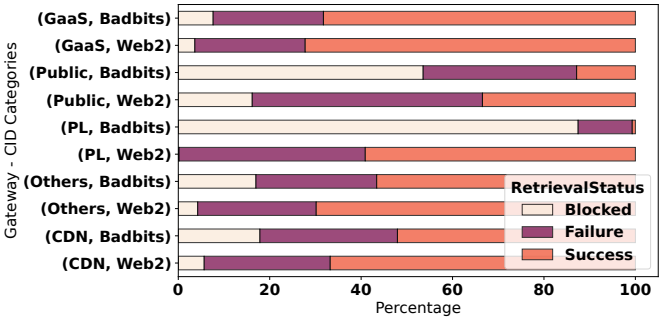
Recently: Much more phishing

Overall: <1 % of all requests



TECHNISCHE UNIVERSITÄT DARMSTADT

HOW DO GATEWAYS IMPLEMENT THE LISTS?



Support varies wildly by stakeholder

Web2.0 lists are generally not incorporated

RECOMMENDATIONS, FUTURE WORK

- Decentralized moderation solution
- Threat type annotations in the list, or different lists (DMCA, Malware, …)
- Improve Web2 filtering by teaching it to understand CIDs instead of using URLs
- Improve Web3 filtering by including Web2 threat feeds

CONCLUSION

- We measured an **existing content moderation** system in IPFS
- There is a **centralized** “badbits” denylist maintained by Protocol Labs based on email reports
- The list is **not annotated**, but most entries are DMCA-related
- Usage of the list is optional and **not widespread among clients**
- **Public gateways** are HTTP entry points which **often utilize the list**
- Listed content **broadly remains available** from the P2P network

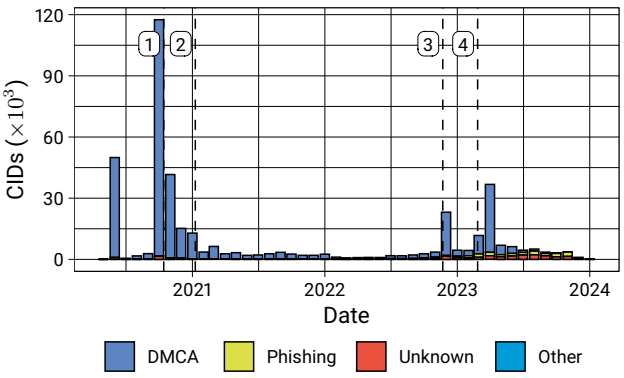
BACKUP SLIDES

Section 9

TECHNISCHE UNIVERSITÄT DARMSTADT

FIRST APPEARANCE IN REQUEST LOGS

CAN BE LINKED TO EXTERNAL EVENTS



① The Free Library project announces their operation on IPFS.

② The Nexus project announces release of latest index+dataset on IPFS.

③ Anna's Archive announces Z-Library on IPFS.

② The Nexus project announces web frontend "Standard Template Construct" on IPFS.