International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

The InterPlanetary File System: A Technical Survey of Architecture, Performance, Security, and Applications

Azizgul Azizhussaini

Department of Computer Science, Hemchandracharya North Gujarat University
Email: *azizhussaini1193[at]gmail.com*

Abstract: ***The InterPlanetary File System (IPFS) emerges as the most researched decentralized storage protocol within the academic literature landscape. It offers a transition from location-based addressing to content-based networking. This survey compiles peer-reviewed works from venues such as ACM SIGCOMM, IEEE INFOCOM, USENIX Security, and NDSS, covering the protocol's design, experimental performances, security issues, and application potential. Ultimately, we find that IPFS is performance feasible for most applications (99% of DHT lookups complete in under 1.7s, it operates in 27 countries and across 2700+ autonomous systems), but security breaches abound (for example, Sybil attacks on the Kademlia DHT can eclipse content in an estimate of $4 in cloud computing (CVE-2023-26248)). Such risks align with a tension between decentralization and centralization as 95% of the content is served by the top-50 peers. This survey aids researchers and practitioners in navigating the IPFS state-of-the-art and highlights gaps for open challenges in the arenas of DHT PUT response times and Bitswap privacy threats; thus, future work relating to decentralized web infrastructure is derived.***

Keywords: InterPlanetary File System, IPFS, Distributed Hash Table, Kademlia, Content-Addressable Storage, Peer-to-Peer Networks, Decentralized Web, libp2p, Bitswap

1. Introduction

In recent years, web consolidation has become increasingly prevalent where a majority of web traffic comes from a handful of companies and even micro-webs elect to run on large, established cloud infrastructures [1]. The financial implications of outages are massive, too; in 2013, Amazon's eCommerce site was reported to lose over $66,000 per minute when down. As such, in light of centralization, the "Decentralized Web" era has emerged with a plethora of technologies that attempt to champion user control via open-source, community driven software deployments.

The InterPlanetary File System (IPFS) is the largest and most well-known decentralized web today [2]. IPFS is an open-source, content-addressable peer-to-peer network for distributed storage and delivery with millions of content retrievals per day and dozens of third-party applications relying on it. IPFS deploys a unique protocol stack in a standardized fashion that brings together disparate sources including BitTorrent (swarming), Kademlia (distributed hash tables), Git (Merkle DAGs), and Self-Certifying File Systems.

IPFS is more than an academic curiosity. IPFS is the technical foundation for many decentralized solutions including NFT markets like OpenSea [3] as well as blockchain-oriented storage and censorship-resistant content delivery. IPFS has been used to evade censorship during country-wide internet blackouts due to its location-independent means of content access.

This survey contributes a comprehensive technical assessment of IPFS through a multimodal approach to peer reviewed literature. We structure our findings across six major themes:

1) Protocol Architecture: Detailed examination of content addressing, Kademlia DHT, Bitswap, and the libp2p networking stack
2) Performance Characteristics: Empirical measurements of latency, throughput, and scalability from major studies
3) Security Analysis: Vulnerability assessments including eclipse attacks, Sybil attacks, and privacy concerns
4) Comparative Positioning: Analysis against BitTorrent, traditional CDNs, and other decentralized storage systems
5) Application Domains: Healthcare, scientific data sharing, blockchain integration, and IoT
6) Open Challenges: Identified limitations and future research directions

2. Background and Foundational Concepts

2.1 From Location-Based to Content-Based Addressing

Where traditional web architecture is concerned, it implements location-based addressing via URLs in which people access content based upon where it is and not based upon the characteristics that make up the content. Therefore, much is incorrect with such developments beneath that access fragile links (if something moves, it can no longer be accessed), fails to authenticate access (people cannot authenticate that they got what they needed), and creates single points of failure (all access to all content is through the server hosting it and if that goes down, everyone is out of luck).

But much of Information Centric Networking (ICN) research supports content-based addressing, accessing content based upon a cryptographic hash of the content itself. For example, if a user requests information based upon a hash and any node has the information and delivers it, the user can cryptographically access what they wanted and authenticate it

was what they needed. This has been supported by ICN literature regarding NDN [4] and NetInf architecture.

2.2 Distributed Hash Tables and Kademlia

Distributed Hash Tables (DHTs) provide a decentralized lookup service mapping keys to values across participating nodes. The Kademlia protocol, introduced by Maymounkov and Mazières [5], forms the routing foundation for IPFS. Key properties of Kademlia include:

- XOR Distance Metric: Distance between node IDs is computed as $d(x,y) = x \oplus y$, interpreted as an unsigned integer • k-Buckets: Each node maintains a routing table of $k$buckets, where bucket i contains nodes with distance $2^i \leq d < 2_{i+1}$
- $O(\log n)$ Lookup Complexity: Queries converge to the target in logarithmic time relative to network size
- Preference for Long-Lived Peers: The protocol preferentially retains stable nodes to resist churn

The XOR metric satisfies the triangle inequality ($d(x,z) \leq d(x,y) + d(y,z)$), enabling efficient routing proofs. IPFS extends standard Kademlia with a replication parameter $k = 20$, meaning content is stored on the 20 closest peers to its content identifier.

2.3 Merkle Directed Acyclic Graphs

IPFS organizes content as Merkle DAGs (Directed Acyclic Graphs), where each node is identified by the cryptographic hash of its contents, including hashes of child nodes. This structure, familiar from Git version control, provides:

- Content Integrity: Any modification changes the root hash
- Deduplication: Identical subtrees share the same hash
- Incremental Verification: Subtrees can be verified independently

The InterPlanetary Linked Data (IPLD) layer [6] provides a unified data model for content-addressed data structures, enabling interoperability between different systems using content addressing.

3. Protocol Architecture

The IPFS architecture comprises seven conceptual layers: Identities, Network, Routing, Exchange, Objects, Files, and Naming. We examine each component in detail based on protocol specifications and academic analyses.

3.1 Content Identifiers (CIDs)

Content Identifiers (CIDs) are self-describing contentad dressed identifiers that combine multiple encoding schemes. A CIDv1 has the structure:

<multibase><version><multicodec><multihash>

- Multibase: Encoding specification (e.g., base32, base58btc)
- Version: CID version identifier (0 or 1)
- Multicodec: Content format type (e.g., dag-pb, raw) Multihash: Self-describing hash (algorithm + length + digest)

CIDv0, the original format, uses base58btc-encoded SHA256 multihashes and is recognizable by the "Qm" prefix (46 characters). CIDv1 provides forward compatibility for different hash functions and encoding formats [7].

The multihash format enables cryptographic agility—the system can transition to new hash functions (e.g., from SHA256 to SHA-3) without breaking existing content references.

As of 2022, SHA-256 remains the default, producing 32-byte hashes regardless of content size.

3.2 Kademlia DHT Implementation

IPFS implements the Kademlia DHT with several modifications detailed in the libp2p specifications [8]. The DHT stores three types of records:

1) Provider Records: Map CIDs to peer IDs of content providers
2) Peer Records: Map peer IDs to multiaddresses (network addresses)
3) IPNS Records: Map IPNS keys to content paths

The lookup algorithm proceeds iteratively:

Algorithm 1: IPFS DHT Lookup
1) Initialize query-queue Q with k closest peers to target X
2) Initialize queried set $P_q = \emptyset$
3) While closest 3 unqueried peers not exhausted:
   a) Select peer p closest to X from Q
   b) Query p: "Who are the k closest peers to X?"
   c) Add p to P_q
   d) Add returned peers to Q (sorted by distance)
4) Return k closest successfully queried peers

The IPFS network maintains separate WAN (public) and LAN (local) DHTs. WAN DHT nodes operate in client mode (query-only) or server mode (respond to queries) based on whether they are publicly dialable [9].

Routing Table Structure: Each peer maintains up to 256 k-buckets (capped at 15 in practice), each storing up to $k = 20$ peers. The routing table refresh occurs every 10 minutes, where for each bucket, a random address in that bucket's range is looked up to discover new peers.

3.3 Bitswap Protocol

Bitswap is IPFS's block exchange protocol for sending and receiving content-addressed blocks [10]. Unlike requestresponse protocols, Bitswap is message-based, with messages containing wantlists and/or blocks.

The protocol has evolved through multiple versions:

- v1.0.0: Basic wantlist and block exchange
- v1.1.0: Added CID version support
- v1.2.0: Introduced HAVE/DONT_HAVE responses for content routing

A Bitswap message contains:

```
message Message { message Wantlist { enum WantType
  { Block = 0; Have = 1; } message Entry {
   bytes block = 1;  // CID
     int32 priority = 2;       // Normalized priority bool
     cancel = 3;    // Revokes entry WantType wantType
     = 4; bool sendDontHave = 5;
   } repeated Entry entries = 1; bool full = 2; //
   Complete vs. delta
  }
  optional Wantlist wantlist = 1; repeated bytes blocks =
  2;
  repeated BlockPresence blockPresences = 3;
}
```

Bitswap sessions optimize retrieval of related blocks (e.g., file chunks). When fetching a Merkle DAG, the session broadcasts WANT-HAVE for the root CID, then progressively requests child blocks from responsive peers [10]. This enables Bitswap to retrieve blocks in as few as 2 RTTs when directly connected to a provider.

3.4 libp2p Networking Stack

libp2p provides IPFS's modular networking foundation, supporting transport-agnostic communication. Key components include:

Transport Protocols: TCP, QUIC, WebSocket, WebRTC, with QUIC providing native encryption and multiplexing.

Connection Security: TLS 1.3 or Noise protocol for encrypted channels. libp2p does not use the CA system, instead deriving trust from peer public keys.

Stream Multiplexing: Multiple logical streams over a single connection using Yamux or mplex protocols. This reduces NAT traversal overhead and connection establishment costs.

NAT Traversal: libp2p implements ICE-like protocols for hole-punching, plus relay protocols (circuit-relay) when direct connection is impossible. The AutoNAT protocol helps peers discover their public reachability status.

Peer Discovery: Multiple mechanisms including DHTbased routing, mDNS for local networks, and bootstrap peers for initial connectivity.

3.5 IPNS: Mutable Naming

The InterPlanetary Name System (IPNS) provides mutable pointers to immutable content [11]. An IPNS name is derived from a public key hash, and records are signed by the corresponding private key:

```
IPNS_name = Hash(public_key)
IPNS_record = Sign(private_key, { value:
      "/ipfs/<CID>", sequence: N, validity:
      timestamp,
      TTL: duration
})
```

IPNS records are published to the DHT with configurable republication intervals (default: 4 hours) and cache lifetimes (default: 24 hours). The sequence number prevents replay attacks, and validity timestamps enable expiration.

4. Empirical Performance Analysis

This section synthesizes findings from major measurement studies, particularly the seminal SIGCOMM 2022 paper by Trautwein et al. [1] and subsequent work.

4.1 Network Topology and Scale

Trautwein et al.'s crawler-based measurement campaign from a German server, running every 30 minutes, revealed:

- Network Size: 15,000–27,000 dialable DHT server peers with strong diurnal periodicity
- Geographic Distribution: Presence in 152 countries across 2,700+ autonomous systems
- Infrastructure Distribution: Significant presence outside major cloud providers

The study observed that only approximately 2.5% of peers remain online for more than 24 hours [12], indicating substantial churn that impacts routing table stability.

4.2 DHT Lookup Performance

DHT GET operations (content retrieval) demonstrate acceptable latency:

Table 1: DHT GET Latency Percentiles [1]

Percentile	Latency
50th (Median)	< 500ms
90th	< 1.2s
99th	< 1.7s

Critically, Trautwein et al. found no correlation between object size and retrieval latency (Pearson coefficient: 0.13), indicating that delays are dominated by size-agnostic DHT lookups rather than content download time.

4.3 DHT PUT Performance Crisis

DHT PUT operations (content publication) exhibit severe performance issues. Trautwein et al.'s follow-up INFOCOM 2024 study [13] measured baseline PUT latencies:

Table 2: DHT PUT Latency (Baseline) [13]

Percentile	Latency
50th (Median)	6.3s
90th	20.6s
95th	49.8s

The DHT walk phase accounts for over 90% of total PUT latency. Their proposed "Optimistic Provide" mechanism achieved sub-second median latency (0.51s) with speed-up factors of 11.3× to 46.7× while reducing network overhead by over 40%.

4.4 Retrieval Stretch Factor

To compare IPFS retrieval against traditional HTTPS, Trautwein et al. measured the "cost of decentralization" as the ratio of IPFS retrieval time to estimated HTTPS retrieval time:

- Median stretch factor: < 4× compared to HTTPS
- 80th percentile: stretch < 2×
- Gateway caching significantly improves performance, with only 15.6% of bytes requiring uncached retrieval

4.5 Content Accessibility and Centralization

Shi et al.'s SIGMETRICS 2024 study [3] analyzed over 4 million CIDs and revealed concerning centralization patterns:

- Content Distribution: Approximately 50% of stored files are NFT-related
- Provider Concentration: Top-50 peers (mostly cloudhosted) serve 95% of content
- Replication: Only 2.71% of data files show replication across more than 5 nodes
- Large File Performance: Retrieval throughput degrades for large files due to chunk CID resolution overhead

4.6 Video Streaming Performance

Wu et al.'s WWW 2023 study [14] tested 28,000+ videos and found:

- Median RTT for content retrieval: 60ms
- At 8 Mbps bandwidth (median for US/EU users): 90% of videos experience stalls
- At 25 Mbps (over-provisioned): 50% still experience stalls
- Traditional ABR algorithms fail because multi-source retrieval prevents accurate throughput estimation

Their proposed "Telescope" algorithm, which accounts for IPFS's multi-provider characteristics, reduced stalls by 95% compared to traditional ABR.

4.7 Private Network Performance

Lajam and Helmy [15] evaluated IPFS in private networks and found performance significantly worse than FTP, with local I/O disk operations dominating latency. Paradoxically, reading performance degraded as data became more popular in the network due to increased coordination overhead.

5. Security Analysis

Academic security analysis has identified fundamental vulnerabilities in IPFS's architecture, with multiple attack demonstrations published at top security venues.

5.1 Eclipse Attacks on DHT

Prünster, Marsalek, and Zefferer [16] demonstrated practical node eclipse attacks at USENIX Security 2022. The attack exploits:

1) Failure to favor long-lived peers in routing table updates
2) Lack of protection for lower routing table buckets
3) Ability to quickly evict honest peers via pregenerated Sybil IDs

The attack used 29TB of pregenerated Peer IDs to game libp2p's reputation system. Results showed arbitrary IPFS nodes can be eclipsed (isolated from the network) with moderate effort. This vulnerability was assigned CVE-2020-10937 and led to mitigations in go-ipfs versions 0.5–0.7. Mitigations deployed include:

- IP diversity requirements in routing tables (max 3 peers per /16 subnet)
- Constant scoring for relay nodes to prevent reputation inflation
- Preservation of active connections during idle periods

5.2 Content Censorship via DHT

Sridhar et al. [17] demonstrated content-targeted censorship at NDSS 2024. The attack:

1) Generates $e \geq k$ Sybil identities closer to target CID than honest peers
2) Positions Sybils to receive all provider records and resolution queries
3) Drops provider records and ignores queries for target CID

Attack Cost: Approximately $4 on AWS (single t3.x large instance with 4 vCPUs) makes arbitrary content undiscoverable within seconds to 48 hours.

Detection and Mitigation: The authors proposed statistical peer ID density estimation and region-based queries, achieving 99.6% detection rate and 100% mitigation of detected attacks. This vulnerability was assigned CVE-202326248.

5.3 Persistent Sybil Vulnerabilities

Cholez and Ignat [18] demonstrated at ARES 2024 that IPFS continues lacking basic defense mechanisms against localized Sybil attacks. Their attack from a single computer achieved control over arbitrary DHT entries using only 20 strategically placed Sybil IDs (brute- forced in approximately 90 minutes on an 8-core desktop).

Their 2025 follow-up work proposed SR-DHT-Store [19], a region-based query approach that addresses attacks achieving approximately 80% denial rates against previous mitigations.

5.4 Privacy Vulnerabilities

Balduf et al. [20] demonstrated through 15-month passive monitoring at IEEE ICDCS 2022 that Bitswap's broadcast design fundamentally leaks user interests to all connected peers.

Their monitoring identified:

- User content requests and access patterns
- Public IPFS/HTTP gateways including hidden node IDs
- Correlation of requests across sessions

Daniel and Tschorsch [21] proposed privacy-enhanced protocols using Bloom filters for interest concealment, though these trade client privacy for provider privacy.

5.5 Network-Level Attacks

Matter and Tran [22] simulated BGP routing attacks against 3,000 CIDs using real topology data. A single malicious AS can censor 75% of content for 57% of requesters by hijacking only 62 prefixes to achieve 70% of full attack effectiveness.

5.6 Malicious Content Propagation

Patsakis and Casino [23] proposed the RIGA algorithm for malware distribution exploiting IPFS's anonymity, content permanence, and gateway architecture. Research on sensitive file leaks [24] identified over 2,000 files with sensitive information including API keys and SSH private keys, with 60% of patched GitHub credential files remaining accessible on IPFS.

6. Comparative Analysis

6.1 IPFS versus BitTorrent

Daniel and Tschorsch's comprehensive IEEE Communications Surveys & Tutorials comparison [7] identified key architectural differences:

Table 3: IPFS vs. BitTorrent Comparison

Aspect	IPFS	BitTorrent
Namespace	Global	Pertorrent
Dedup.	Crossfile	Withintorrent
DHT Size	~25K	>1M *nodes*
Discovery	DHT + Bitswap	DHT + Trackers

The SIGCOMM 2022 study [1] measured stretch factors typically under 4× compared to HTTPS, with stretch under 2× for 80% of content retrievals from central European vantage points. Gateway caching significantly improves performance, but this reliance introduces centralisation pressures that contradict IPFS design goals.

6.2 IPFS versus Traditional CDNs

Merlec and In's systematic comparison [25] rated systems across multiple dimensions:

Table 4: Decentralized Storage Comparison

System	Spd	Lat	Scl	Red
IPFS	Mod	Mod	Mod	High
BitTorrent	High	Low	High	High
Filecoin	High	Low	High	High
Storj	High	Low	High	High
Arweave	Mod	Mod	High	V.High
Swarm	Mod	Mod	Mod	High

6.3. IPFS versus Filecoin

A comprehensive analysis of 35 studies [26] revealed:

- IPFS: Simpler integration, low-risk real-time scenarios, average latency ~210 ms
- Filecoin: Proof-of-Replication (PoRep) and Proofof-Spacetime (PoST) provide higher data availability (~99.9%), but they also come with higher latency and operational overhead.

Filecoin uses financial incentives to address IPFS's lack of persistence guarantees: storage providers stake FIL collateral and risk being slashed for storage failures. Emerging best practices include hybrid architectures that combine Filecoin data assurance with IPFS accessibility.

7. Application Domains

7.1. Healthcare Data Management

A significant application area is healthcare, where many systems combine blockchain technology with IPFS to store medical records in a safe and scalable manner.

Using ciphertext-policy attribute-based encryption (CPABE) for fine-grained access control, Sun et al. [27] proposed an attribute-based encryption scheme for electronic medical records in IPFS. The plan overcomes blockchain scalability constraints by storing encrypted data in IPFS while preserving hash references on Ethereum.

Using Hyperledger Fabric and IPFS storage, Chenthara et al. [28] created a "Healthchain" with role-based access control smart contracts (PCHDMAC-SC). Their assessment showed efficacy in terms of interoperability, scalability, privacy, and data security.

A scalable blockchain model that demonstrated ransomware resilience through data redundancy across consortium IPFS nodes was presented by Jayabalan and Jeyanthi [29].

7.2 Sharing Scientific Data

As FAIR-compliant (Findable, Accessible, Interoperable, Reusable) persistent identifiers [30], IPFS CIDs have been positioned. The content-addressed nature ensures:

- Findability: CIDs serve as globally unique identifiers
- Accessibility: Content retrievable from any provider
- Integrity: Hash verification ensures unmodified data
- Reusability: Immutable references enable citation

COVID-19 open research datasets hosted on IPFS/Filecoin for improved accessibility and persistence are among the notable deployments.

7.3 Blockchain and NFT Integration

IPFS is widely used by OpenSea and other NFT marketplaces to store metadata permanently. About 50% of IPFS content is related to NFT, according to the SIGMETRICS 2024 study [3]. IPFS CIDs are stored in smart contracts as unchangeable references to off-chain media, and pinning infrastructure is provided by Pinata and related services. Among the difficulties are:

- Pinning providers determine the accessibility of content.
- There is no on-chain confirmation of IPFS's continuous availability.
- Without additional infrastructure, metadata permanence is not guaranteed.

7.4 Edge Computing and IoT

A study that was published in Scientific Reports [31] combined IPFS storage, sharding, and DPoS consensus to achieve a maximum throughput of about 11.094 ms at 500 TPS for limited IoT networks.

M2M distributed protocols using IPFS on Raspberry Pi devices [32] proved feasible for PubSub communications and edge synchronisation with private swarms.

7.5 Digital Preservation

Alam's "InterPlanetary Archival Record Object" (IPARO) [33] implemented immutable linked lists of records and suggested decentralised version tracking using IPFS and IPNS primitives. By dividing WARC response records for deduplication, the InterPlanetary Wayback (ipwb) project offers distributed archive replay.

8. Open Challenges and Future Directions

8.1. Performance Obstacles

DHT PUT Latency: DHT write performance is still the most important constraint. Although this particular problem is addressed by Optimistic Provide [13], more widespread implementation and standardisation are required.

Chunk CID resolution reduces throughput for large files. Prefetching techniques and parallel resolution are two possible solutions.

Traditional ABR algorithms are complicated by multisource retrieval in video streaming. Telescope [14] and other IPFS-aware streaming solutions show promise but need wider adoption.

8.2 Persistence Guarantees

Best-effort storage is offered by IPFS, but availability is not guaranteed. Garbage collection eventually eliminates unpopular unpinned content, with only 2.71% of files exhibiting significant replication. Among the research avenues are:

- Mechanisms of economic incentive other than Filecoin
- Systems of reputation for storage providers
- Predictive pinning of content according to access patterns

8.3 Privacy improvements

Bitswap's broadcast mechanism inherently leaks user interest. Areas for future extension include:

- Private Information Retrieval (PIR) protocols [34]
- Onion routing to request content
- Interest masking through bloom filters

8.4 Security improvements

The DHT remains an attack vector, many vulnerabilities are constantly discovered. Areas for future extension include:

- Proof-of-Work for peer id generation
- Social trust networks for peer authentication
- Quantum-resistant primitives in content addressing

8.5 Decentralization improvements

Decentralization is an important factor, however, centralized solutions don't help: cloud-hosted nodes offer most content, gateway reliance creates single points of failure, and high churn means many people are served from a small group of stable infrastructure providers. Thus, decentralized solutions would be appropriate if:

- Incentives were tied to geo-distributed usage
- Client-side caching was more effective
- Browser-natives IPFS support was more mature

8.6 Future research avenues

Important areas of research were outlined by Protocol Labs:

- Federated Learning via IPFS: preliminary results of accuracy convergence to centralized FL within 1‰
- Mobile Support: Pub/sub on portable devices, energy consumption study
- 5G/6G Support: Device-to-device solutions for unloading bandwidth

9. Conclusion

This survey has provided a comprehensive technical study of the InterPlanetary File System from peer-reviewed academic studies. The IPFS is the largest content addressed storage network operating today spanning 2,700+ autonomous systems and millions of daily retrievals. Research findings include:

1) Performance: DHT lookups are completed in under 1.7s at the 99th percentile, however, PUT operations suffer from extreme latency (6.3s median baseline) and require tuning
2) Security: DHT vulnerabilities exist to allow content censorship for ~$4 and while recent updates have addressed CVE-2020-10937 and CVE-2023-26248, they remain active
3) Centralization: 95% of content comes from the top-50 peers, clearly demonstrating a reliance on cloud-hosted nodes despite the goal of decentralization
4) Use Cases: IPFS works well in healthcare, scientific publishing, NFTs and IoT which commonly link access with blockchain technology and persistence with Filecoin.

The biggest gap within the literature today is between IPFS and centralization since it prides itself on decentralization but continues to centralize. Researchers should continue exploring the DHT for security clearance as this is where the most relevant developments are found while tuning performance to specific workloads may have the most meaningful impact. Practitioners should be aware that today, the best option is to use IPFS for accessibility and then Filecoin for persistence.

As the decentralized web develops, the IPFS is a crucial layer of underpinnings for infrastructure that supports such decisions for content dissemination, data sovereignty and resilience across the internet for the foreseeable future.

Acknowledgments
This survey relies on research from Protocol Labs, institutions of higher learning and researchers globally and the open source community surrounding IPFS. We thank ProbeLab for providing measurement infrastructure, and security researchers for responsible disclosure efforts.

References

[1] D. Trautwein, A. Raman, G. Tyson, I. Castro, W. Scott, M. Schubotz, B. Gipp, and Y. Psaras, "Design and Evaluation of IPFS:
[2] A Storage Layer for the Decentralized Web," in *Proc. ACM SIGCOMM*, Amsterdam, Netherlands, Aug. 2022, pp. 739–752.
[3] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," *arXiv preprint arXiv:1407.3561*, 2014.
[4] L. Shi, et al., "A Closer Look into IPFS: Accessibility, Content, and Performance," *Proc. ACM on Measurement and Analysis of Computing Systems*, vol. 8, no. 2, Article 20, 2024. [4] L. Zhang, et al., "Named Data Networking," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 3, pp. 66–73, 2014.
[5] P. Maymounkov and D. Mazières, "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric," in *Proc. IPTPS*, LNCS vol. 2429, 2002, pp. 53–65.
[6] H. Sanjuan, S. Haad, and A. Pgte, "Merkle-CRDTs: MerkleDAGs meet CRDTs," *arXiv preprint arXiv:2004.00107*, 2020.
[7] E. Daniel and F. Tschorsch, "IPFS and Friends: A Qualitative Comparison of Next Generation Peer-to-Peer Data Networks," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 31–52, 2022.
[8] libp2p Project, "Technical Specifications for the libp2p Networking Stack," GitHub Repository, 2024. [Online]. Available: https://github.com/libp2p/specs
[9] IPFS Documentation, "Distributed Hash Tables (DHT)," 2024. [Online]. Available: https://docs.ipfs.tech/concepts/dht/
[10] A. de la Rocha, D. Dias, and Y. Psaras, "Accelerating Content Routing with Bitswap: A Multi-Path File Transfer Protocol in IPFS and Filecoin," *Protocol Labs Technical Report*, 2021.
[11] N. Fotiou, V. A. Siris, and G. C. Polyzos, "Enabling Self-Verifiable Mutable Content Items in IPFS," *arXiv preprint arXiv:2105.08395*, 2021.
[12] E. Daniel and F. Tschorsch, "Passively Measuring IPFS Churn and Network Size," in *Proc. IEEE ICDCSW*, 2022, pp. 60–65.
[13] D. Trautwein, et al., "IPFS in the Fast Lane: Accelerating Record Storage with Optimistic Provide," in *Proc. IEEE INFOCOM*, 2024, pp. 1920–1929.
[14] Z. Wu, C. R. Yang, S. Vargas, and A. Balasubramanian, "Is IPFS Ready for Decentralized Video Streaming?" in *Proc. ACM Web Conference*, 2023.
[15] O. A. Lajam and T. A. Helmy, "Performance Evaluation of IPFS in Private Networks," in *Proc. DSDE*, 2021, pp. 77–84.
[16] B. Prünster, A. Marsalek, and T. Zefferer, "Total Eclipse of the Heart – Disrupting the InterPlanetary File System," in *Proc. USENIX Security*, Boston, MA, Aug. 2022, pp. 3735–3752.
[17] S. Sridhar, O. Ascigil, N. Keizer, F. Genon, S. Pierre, Y. Psaras, E. Rivière, and M. Król, "Content Censorship in the InterPlanetary File System," in *Proc. NDSS*, San Diego, CA, Feb. 2024.
[18] T. Cholez and C. Ignat, "Sybil Attack Strikes Again: Denying Content Access in IPFS with a Single Computer," in *Proc. ARES*, Vienna, Austria, Jul. 2024.
[19] T. Cholez, F. Netto, and C. Ignat, "Active Sybil Attack and Efficient Defense Strategy in IPFS DHT," *arXiv preprint arXiv:2505.01139*, May 2025.
[20] L. Balduf, S. Henningsen, M. Florian, S. Rust, and B. Scheuermann, "Monitoring Data Requests in Decentralized Data Storage Systems: A Case Study of IPFS," in *Proc. IEEE ICDCS*, 2022, pp. 658–668.
[21] E. Daniel and F. Tschorsch, "Privacy-Enhanced Content Discovery for Bitswap," in *Proc. IFIP Networking*, 2023.
[22] L. Matter and M. Tran, "Network-level Censorship Attacks in the InterPlanetary File System," *arXiv preprint*, 2025.
[23] C. Patsakis and F. Casino, "Hydras and IPFS: A Decentralised Playground for Malware," *International Journal of Information Security*, 2019.
[24] L. Balduf, et al., "Secrets are Forever: Characterizing Sensitive File Leaks on IPFS," in *Proc. IFIP Networking*, 2024.
[25] M. M. Merlec and H. P. In, "Blockchain-Based Decentralized Storage Systems for Sustainable Data Self-Sovereignty: A Comparative Study," *MDPI Sustainability*, vol. 16, Article 7671, 2024.
[26] R. Garcia, et al., "Comparative Security and Performance Evaluation of IPFS and Filecoin for Off-chain Blockchain Storage," *Systematic Literature Review*, 2024.
[27] J. Sun, X. Yao, S. Wang, and Y. Wu, "Blockchain-based Secure Storage and Access Scheme for Electronic Medical Records in IPFS," *IEEE Access*, vol. 8, pp. 59389–59401, 2020.
[28] S. Chenthara, K. Ahmed, H. Wang, F. Whittaker, and Z. Chen, "Healthchain: A Novel Framework on Privacy Preservation of Electronic Health Records Using Blockchain Technology," *PLOS ONE*, vol. 15, no. 12, e0243043, 2020.
[29] J. Jayabalan and N. Jeyanthi, "Scalable Blockchain Model Using Off-chain IPFS Storage for Healthcare Data Security and Privacy," *Journal of Parallel and Distributed Computing*, vol. 164, pp. 152–167, 2022.
[30] M. Vukolov, "IPFS for Scientific Data: FAIR Principles and Persistent Identifiers," in *Proc. European Open Science Cloud Symposium*, 2021.
[31] M. Ahmad, et al., "Performance Enhancement in Blockchain Based IoT Data Sharing Using Lightweight Consensus Algorithm," *Scientific Reports (Nature)*, 2024.
[32] G. D. Putra, et al., "Decentralized P2P Broker for M2M and IoT Applications," *MDPI Proceedings*, vol. 54, no. 1, Article 24, 2020.
[33] S. Alam, "Decentralized Web Archiving and Replay via InterPlanetary Archival Record Object (IPARO)," in *Proc. iPRES*, 2023.
[34] M. Mazmudar, S. Veitch, and R. Mahdavi, "Peer2PIR: Private Queries for IPFS," to appear in *Proc. IEEE S&P*, 2025.