Summary on Blockchain Storage

Blockchain storage technologies have evolved from foundational peer-to-peer (P2P) systems like BitTorrent to sophisticated decentralized storage networks supporting immutability, security, and scalability. This evolution can be understood in several stages and key technologies.

1. Technological Precursors

Early P2P systems like BitTorrent allowed users to share files by downloading parts from multiple peers (seeders and leechers), managed via trackers or distributed hash tables (DHTs). Efficiency and fairness were ensured through strategies like rarest piece first and tit-for-tat.
Kademlia introduced a robust DHT mechanism using XOR-based distances and k-buckets for routing and locating peers efficiently, becoming a backbone for many decentralized systems.
Information-Centric Networking (ICN) proposed a shift from host-based to content-based communication, requiring new security models such as self-certifying names. These ensure data integrity by linking content to cryptographic hashes, making tampering detectable.

2. IPFS and Decentralized Web

The InterPlanetary File System (IPFS) builds on BitTorrent and Git principles to enable a content-addressed, distributed web. Files are split into blocks stored in a Merkle DAG, enabling block deduplication and verifiability. The Bitswap protocol handles block exchange, while Kademlia-based DHT enables peer discovery.

IPFS supports mutable pointers via IPNS and incentivizes storage through Filecoin, which creates a market of retrieval/storage miners. However, availability is not guaranteed unless managed by tools like IPFS Cluster for replication and redundancy.

3. Blockchain-Based Storage Networks

Swarm (Ethereum) uses Merkle trees and “areas of responsibility” for redundancy. Incentives include SWAP (payment tracking), SWEAR, and SWINDLE (compliance enforcement).

Hypercore Protocol allows sharing mutable, encrypted datasets using Hyperswarm DHT. Despite public keys being used to find data, discovery keys protect content confidentiality.
SAFE Network ensures privacy and anonymity via self-encryption and a dynamic, trust-based DHT structure. Only elders in a section can make network decisions.
Storj offers Amazon S3 compatibility and erasure coding for efficient retrieval. It employs satellite nodes to coordinate storage, uplinks for clients, and storage nodes to store actual data.
Arweave introduces a permanent, on-chain storage layer. Data is immutable and incentivized via a tit-for-tat-like protocol called Wildfire, with ranks influencing peer behavior and data persistence.

4. Use Cases and Considerations

These technologies address blockchain’s storage limitations by enabling off-chain storage, content authenticity, and decentralized redundancy. Use cases include blockchain scalability, IoT, healthcare, and agriculture. However, concerns remain regarding privacy, key management, and long-term availability.