

DECADE Survey

draft-ietf-decade-survey-01

Editors: Richard Alimi, Akbar Rahman, Y. Richard Yang

Other Authors: Zhihui Lu, Pang Tao, Juan Carlos Zuniga

Survey Overview

- In-network storage used in many contexts
 - One common use is to increase efficiency of content distribution
- Existing systems have been useful in their own contexts
 - Systems' capabilities reflect their specific contexts
- Survey evaluates in context of DECADE
 - DECADE is targeted for P2P but may also support other applications
- Feedback on addition topics to include is appreciated!

Survey Outline

■ Classification methodology

- ❑ Storage system components

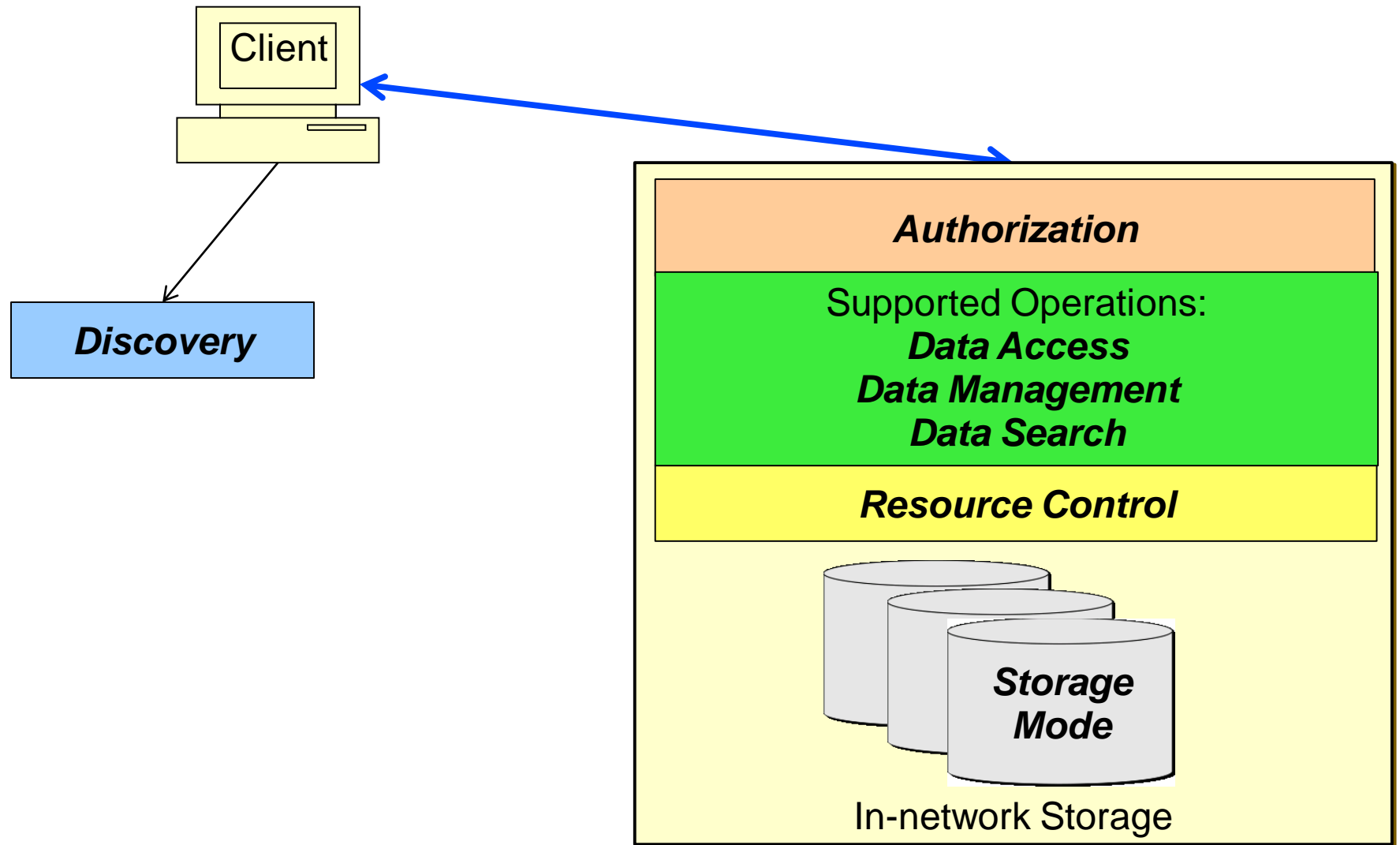
■ In-Network storage systems

- ❑ Sampling of key existing and experimental systems
- ❑ Applicability to DECADE
- ❑ Analysis of components
- ❑ Overall observations

■ Storage access and related protocols

- ❑ Sampling of key existing and experimental protocols
- ❑ Analysis of components
- ❑ Overall observations

In-network Storage System Components

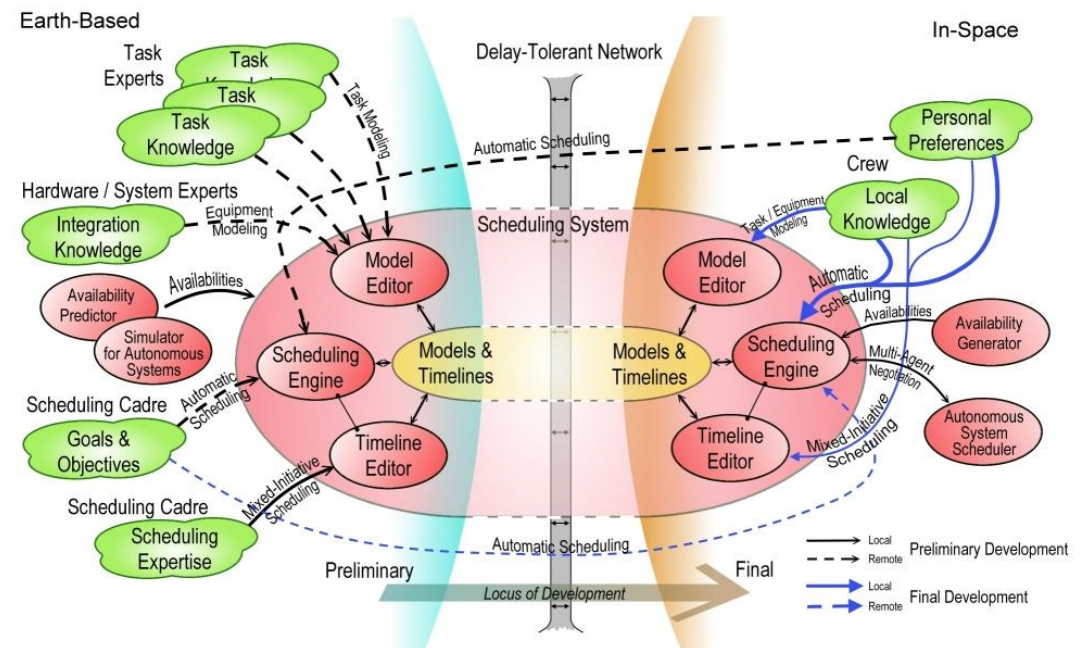


In-Network Storage Systems

(only showing new items added to survey since last IETF)

Delay-Tolerant Network

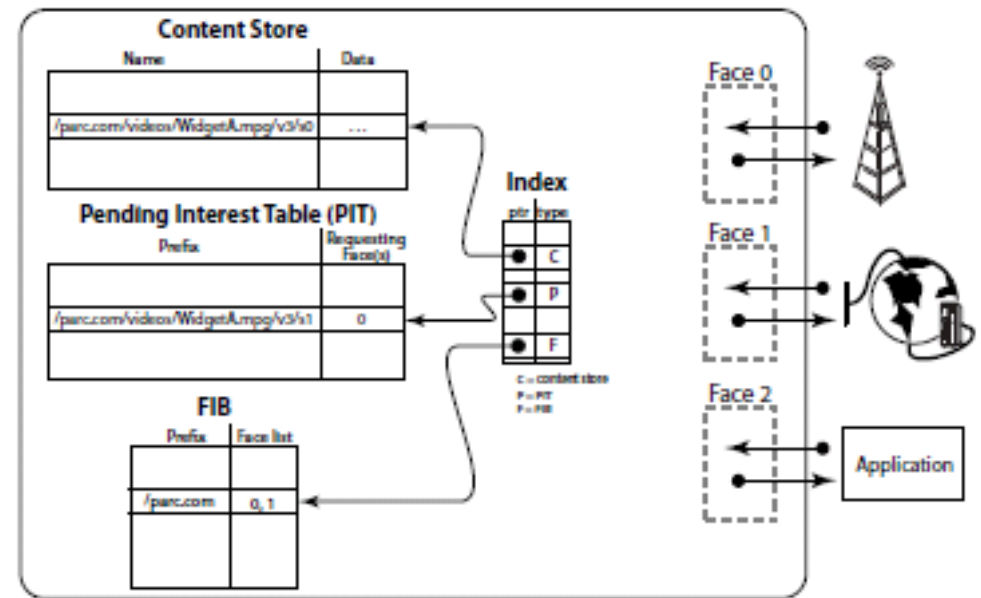
- Originally designed for “Interplanetary Internet” and then adapted for sensor-networks and other high delay environments
- Store-and-forward overlay layer called “Bundle Layer”, defined between transport and application layers
- Focus on in-network storage to overcome long network delays



Discovery	URI is the basis of addressing, and subsequent DTN routing
Authorization	Public-restricted or private
Data Access	Users implicitly cause content to be stored by starting a transaction
Data Mgmt	Via Time To Live parameter associated with DTN transaction
Data Search	Not provided
Resource Ctrl	Not provided
Storage Mode	DTN Bundle Layer segments data for subsequent routing /storage

Named Data Networking

- NDN is a specific project following Information-Centric Networking approach
- A research initiative that proposes to replace IP addresses with “content names”
- NDN routers may store a copy of a data item that it has routed to service new requests for same content name
- Proponents argue that capacity of network can be extended if network storage has a more central role



Discovery	Content names are the basis of addressing and discovery
Authorization	Public-unrestricted, Public-restricted or private
Data Access	Users implicitly cause content to be stored by requesting data
Data Mgmt	Via Time To Live parameter associated with NDN transaction
Data Search	Interest packet will identify data name. Routers will forward using name based forwarding tables
Resource Ctrl	Not provided (but being researched)
Storage Mode	Named data packets

Photo Sharing

- Online photo storage service for end users
- Typical end user interface is through an HTTP browser
- Photos are stored as files which can be organized into meta-structures (e.g. albums, galleries)
- Focused on long term storage instead of short term caching

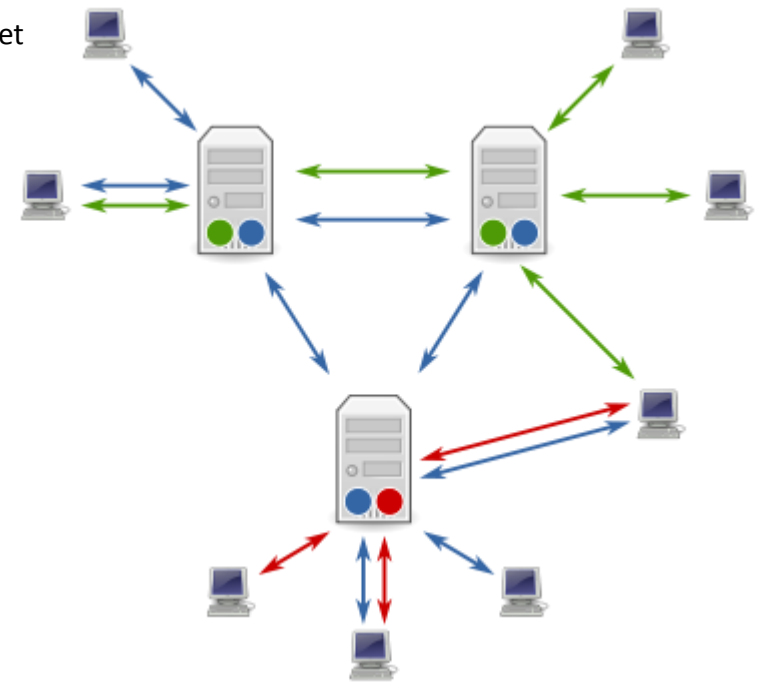
Kodak Gallery



<i>Discovery</i>	Manual (via DNS lookup of well-known hostname)
<i>Authorization</i>	Public-unrestricted or private
<i>Data Access</i>	Read, write
<i>Data Mgmt</i>	Delete
<i>Data Search</i>	User can search for photo tags
<i>Resource Ctrl</i>	Not provided
<i>Storage Mode</i>	File-based (organized into albums, etc.)

Usenet

- Distributed Internet based discussion (message) system arranged as a set of “newsgroups”
- Usenet messages are distributed and stored among a large conglomeration of servers
- Messages are copied from server to server until all servers have all messages



Discovery	Manual (via DNS lookup of well-known hostname)
Authorization	Public-unrestricted or private (to members of that newsgroup)
Data Access	Read, write
Data Mgmt	Limited ability to delete
Data Search	User can manually search through newsgroups by subject
Resource Ctrl	Not provided
Storage Mode	Messages are stored as files organized into newsgroups

Observations

- Majority of the surveyed systems were designed for client-server architecture
 - Exceptions are a few of the newer technologies(e.g. BranchCache and P2P Cache) which do support a P2P mode
- Many of the surveyed systems were designed for caching rather than long term storage
 - DECADE should investigate both modes of storage and the various trade offs involved
- Majority of the authorization models of the surveyed systems do not support a decoupling of the resource owner and user
 - DECADE may need to evolve authorization model to support this decoupling

Storage and Other Related Protocols

(only showing new items added to survey since last IETF)

HTTP



- Key protocol for the World Wide Web
 - And therefore used by many web based services
- Stateless client-server protocol
 - Follows RESTful model
- Often associated with downloading content (GET) from web servers, but also supports uploading (PUT/POST) of content

<i>Discovery</i>	Manual (IP address or via DNS lookup of well-known hostname)
<i>Authorization</i>	Public-unrestricted, public-restricted and private
<i>Data Access</i>	Basic read/write operations
<i>Data Mgmt</i>	Not provided
<i>Data Search</i>	Not provided
<i>Resource Ctrl</i>	Not provided
<i>Storage Mode</i>	File-based (which map to URI path hierarchy)

Observations (1/2)

- All of the surveyed protocols were designed for client-server architectures
 - However, it is possible that some of the protocols could be adapted to work in a P2P architecture
- Several popular in-network storage systems use HTTP as their key protocol even though HTTP is not classically considered a storage protocol
- Majority of the surveyed protocols do not support:
 - Low latency access (e.g. for live streaming)
 - Resource control (for users to manage access by other peers)

Observations (2/2)

- Most of the surveyed protocols do support:
 - ❑ User ability to read/write content
 - ❑ Access control
 - ❑ Error indication
 - ❑ Ability to traverse firewalls and NATs

Conclusions

- There are many successful in-network storage systems and protocols, but they may have been designed for uses cases different from those defined for DECADE
 - Most surveyed systems and protocols were designed for client-server architectures and not P2P
 - However, important lessons (observations) can be learned from the surveyed systems and applied to DECADE design

Comments and questions?

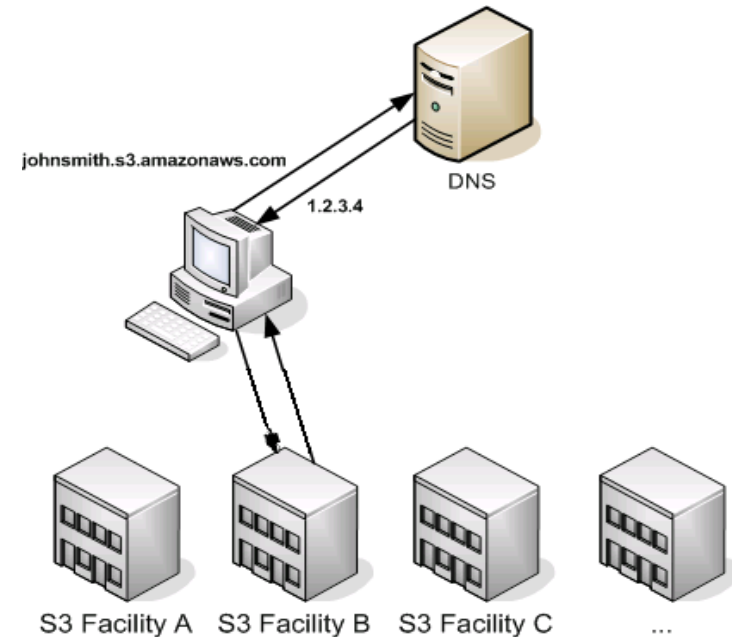
Backup

(Surveyed items already discussed in last IETF)

In-Network Storage Systems

Amazon S3

- Online storage service for end users
- Storage organized into buckets containing data objects
- Popular backend storage for other applications
- Related services
 - Windows Azure Blob service



Discovery	Manual (via DNS lookup of well-known hostname)
Authorization	Typically not provided
Data Access	Read, write
Data Mgmt	Delete
Data Search	User may enumerate bucket contents to find desired file
Resource Ctrl	Not provided
Storage Mode	Object-based (organized into buckets)

BranchCache

- Caches and shares content within branch office

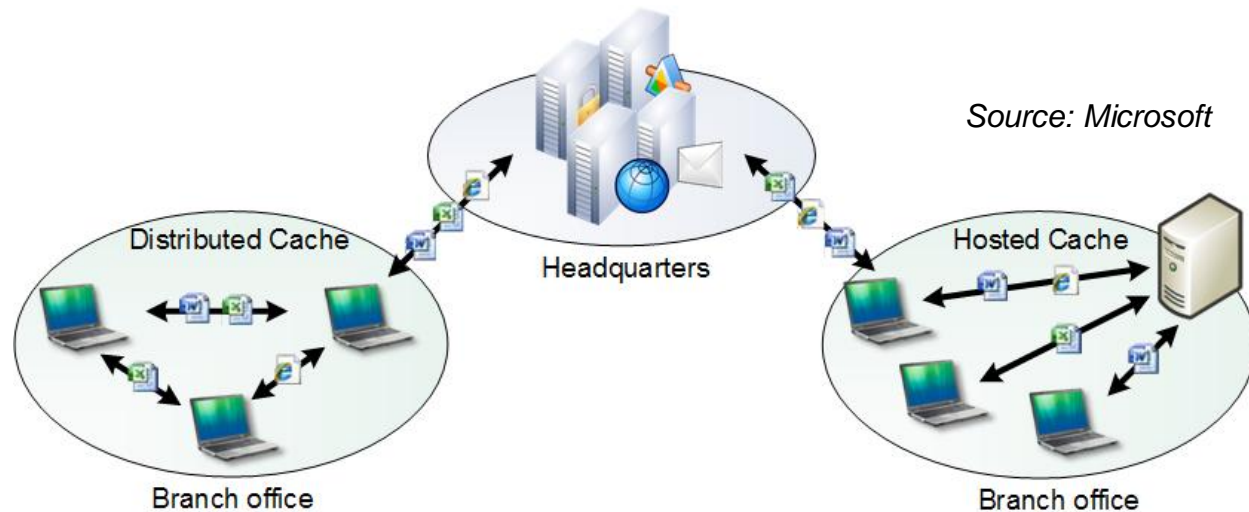
- ☐ Reduce WAN link utilization
- ☐ Improve application responsiveness

- Transparent to end-user

- ☐ Instrument networking stack

- Hosted Cache and Distributed modes

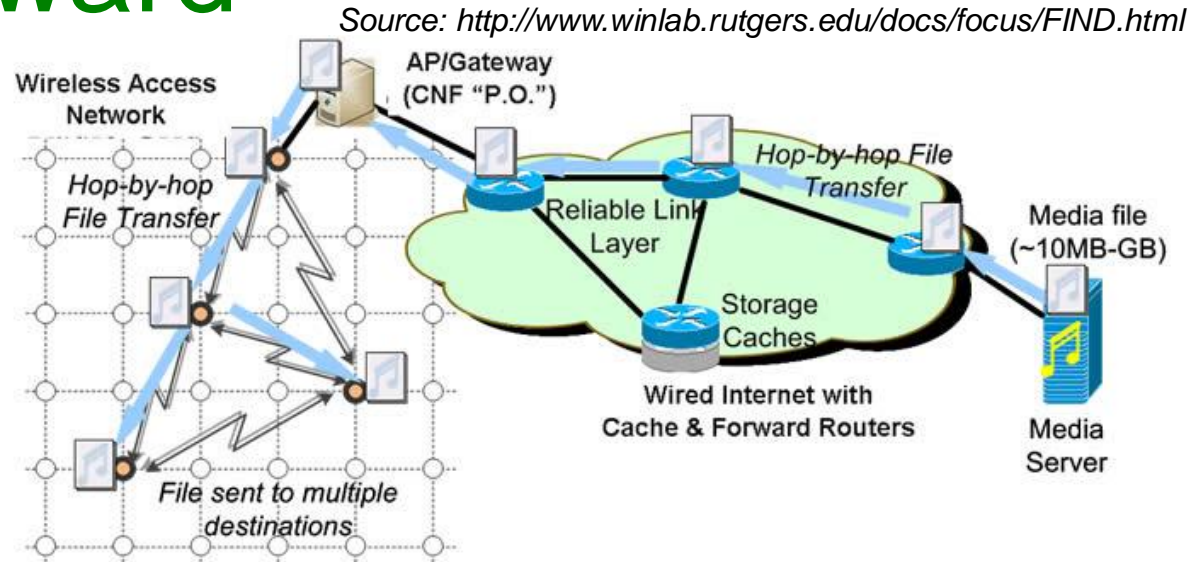
- Maintains end-to-end security



Discovery	Distributed: multicast Hosted: provisioning or manual
Authorization	Keys derived from content server; data decryptable by auth'd clients
Data Access	Read/write (transparent to client) Write is according to caching policy
Data Mgmt	Not provided to end user
Data Search	Not provided to end user
Resource Ctrl	Hosted: admin-controlled policy Distributed: backoff and throttling
Storage Mode	Object-based

Cache-and-Forward Architecture

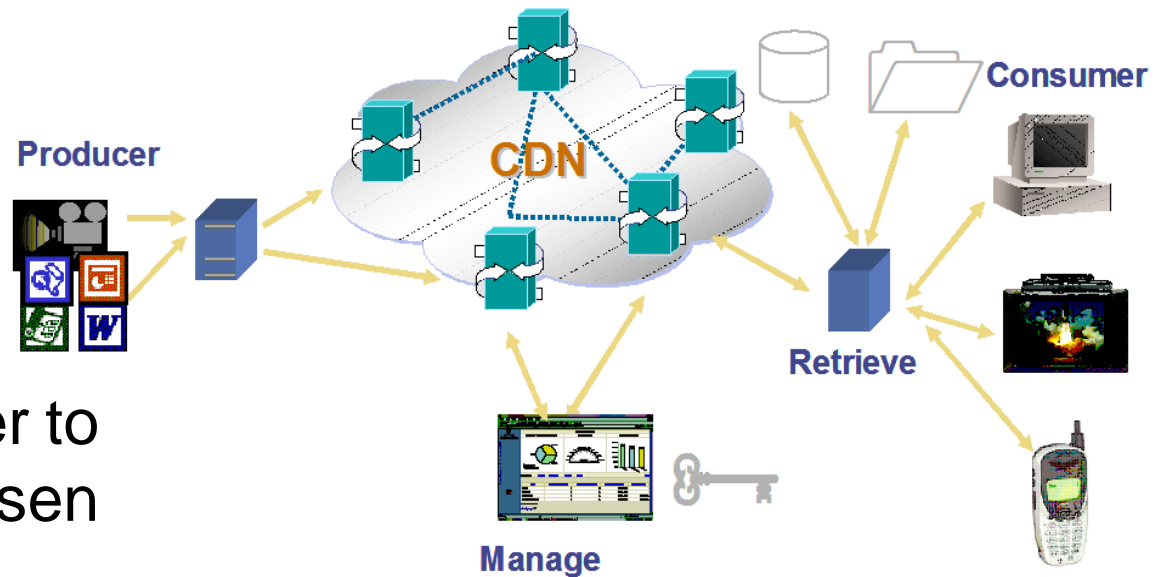
- Proposal for content delivery in future Internet
- Storage placed at some nodes within network
 - ❑ At or nearby routers
- Store-and-forward
 - ❑ Disconnected mobile users
 - ❑ In-network caching
- Focus on large data files



Discovery	Lookup cache-&-forward node via location-independent content ID
Authorization	Not provided
Data Access	Read/write (transparent to client) Write according to caching policy
Data Mgmt	Not provided
Data Search	Not provided
Resource Ctrl	Not provided
Storage Mode	Object-based (with objects representing individual files)

CDNs

- Distribute content to cache/edge servers closer to users; direct users to chosen servers
- Content owner has management frontend
- Typically have extensive infrastructure
 - Distribution amongst CDN nodes, cache management, request routing, etc

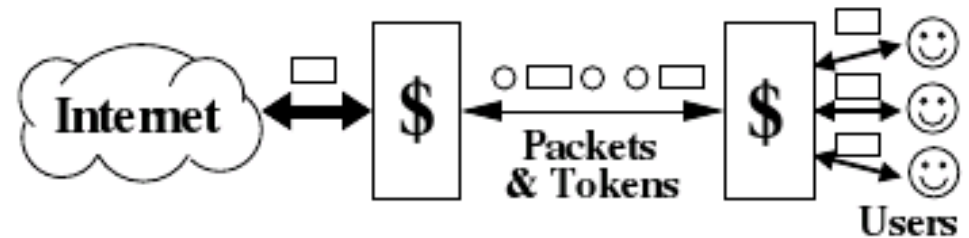


Discovery	DNS or other redirection
Authorization	Typically not provided
Data Access	Read-only for clients Writable for content provider
Data Mgmt	Only to content provider
Data Search	Only to content provider
Resource Ctrl	Not provided
Storage Mode	File-based

Network Traffic Redundancy Elimination (RE)

Source: N. Spring, D. Wetherall. "A protocol-independent technique for eliminating redundant network traffic", SIGCOMM 2000.

- Identify and remove repeated content in network transfers



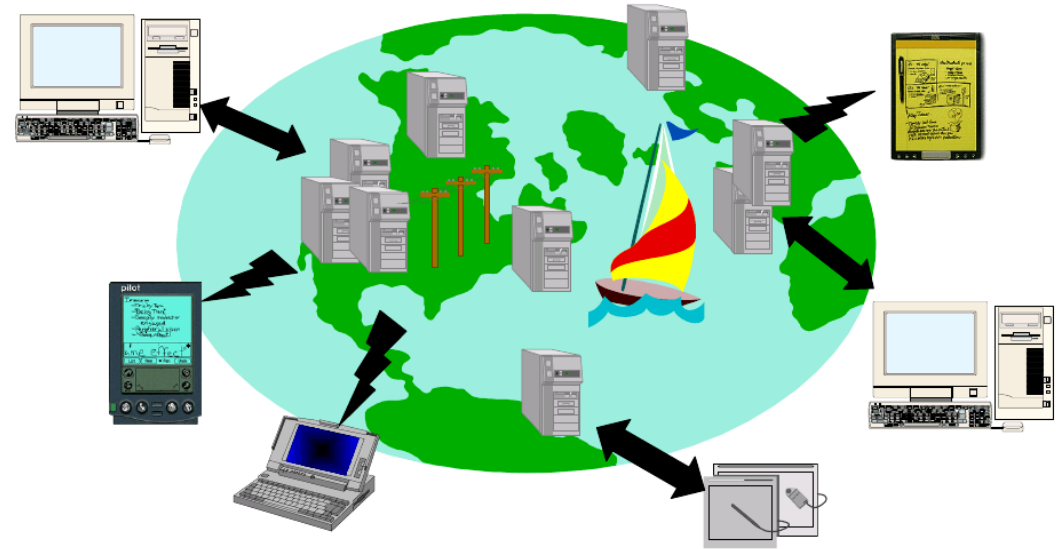
■ Packet-level RE

- ❑ Forwarding elements equipped with storage
- ❑ Cache data from forwarded packets
- ❑ Upstream routers can replace previously-forwarded data with fingerprint

Discovery	Not necessary; implemented entirely within network elements
Authorization	Preserves endpoint control
Data Access	Read/write (transparent to user)
Data Mgmt	Not provided
Data Search	Not provided
Resource Ctrl	Content provider still moderates packet sending rate
Storage Mode	Object-based (with objects being data from transferred packets)

OceanStore

- Research storage system from UC Berkeley
- Aim is to provide globally-distributed storage
- Multiple storage providers pool resources together
- Focus on
 - ☐ Resiliency
 - ☐ Self-organization
 - ☐ Self-maintenance

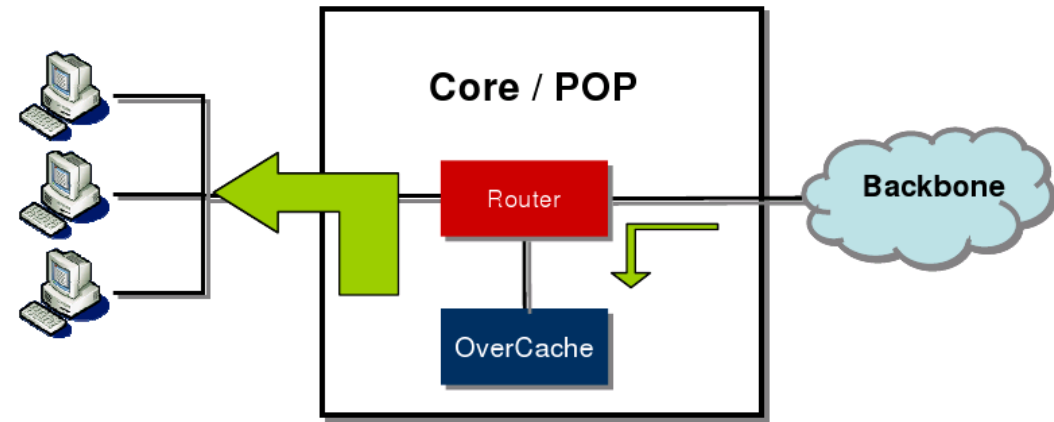


Discovery	Manual (via DNS lookup of well-known hostname)
Authorization	Provided (specifics unclear from published paper)
Data Access	Read, write
Data Mgmt	Allows update of existing objects; multiple versions may be retained
Data Search	Not provided
Resource Ctrl	Not provided
Storage Mode	Object-based (though, NFS and HTTP interfaces built on top of it)

P2P Cache Transparent

Source: http://www.oversi.com/images/stories/white_paper_july.pdf

- Cache P2P content and serve locally
- Implements P2P application protocols to avoid changes to P2P clients
- Uses DPI to avoid explicit discovery by P2P clients
 - Acts as intermediary in session with remote peer



Discovery	DPI (transparent to client)
Authorization	Not provided
Data Access	Read/write (transparent to client) according to caching/ISP policy
Data Mgmt	Not provided
Data Search	Not provided
Resource Ctrl	Not provided
Storage Mode	Object-based (chunks of content stored)

P2P Cache

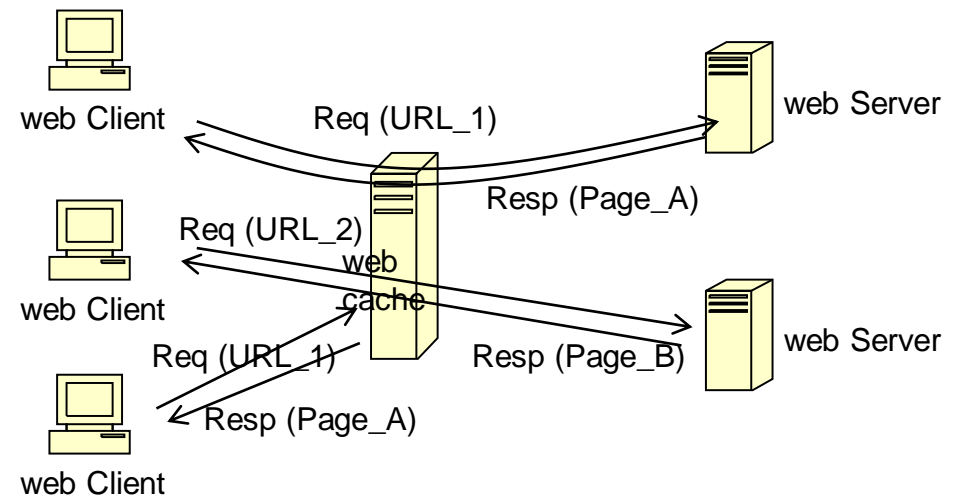
Non-Transparent

- Cache frequently-used P2P content and serve locally
- Implements P2P application protocols to avoid changes to P2P clients
- Explicitly peers with a client

<i>Discovery</i>	Normal discovery in P2P overlay (tracker, DHT, PEX, etc.)
<i>Authorization</i>	Not provided
<i>Data Access</i>	Read/write Write according to caching policy
<i>Data Mgmt</i>	Not provided
<i>Data Search</i>	Not provided
<i>Resource Ctrl</i>	Not provided
<i>Storage Mode</i>	Object-based (chunks of content stored)

Web Cache

- Cache web content and serve locally
 - HTML pages, images, etc
- Server indicates cachability, clients indicate if cached response is acceptable
- HPTP: Extension for P2P
 - Proposes to share P2P content using HTTP
 - Aims to use existing web caches



Discovery	Manual configuration, DNS, or transparent (DPI)
Authorization	Not provided
Data Access	Read/write according to caching/ISP policy
Data Mgmt	Not provided
Data Search	Not provided
Resource Ctrl	Not provided
Storage Mode	Object-based (keyed by HTTP request fields)

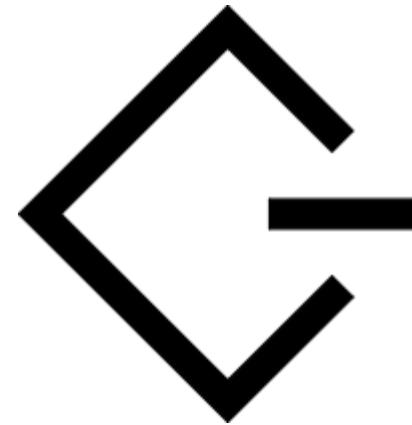
Storage and Other Related Protocols

iSCSI

■ SCSI objectives

- ❑ Enable communication with storage devices
- ❑ Initiator sends commands to target (device)
- ❑ Block-based access
 - No filesystem

■ iSCSI enables commands to be sent over TCP



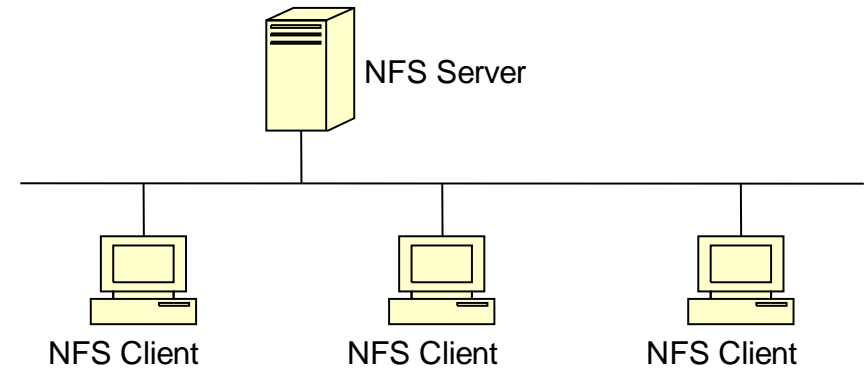
<i>Discovery</i>	Manual or iSNS
<i>Authorization</i>	CHAP for initiators and targets; not provided at block-level
<i>Data Access</i>	Read and write at specific position (LBA offset) into storage device
<i>Data Mgmt</i>	Not directly provided; may be implemented via read/write
<i>Data Search</i>	Not directly provided; may be implemented via read
<i>Resource Ctrl</i>	Not provided
<i>Storage Mode</i>	Block-based

NFSv4, NFSv4.1

- Allow client to access network storage in manner similar to local storage

- Major features

- ☐ Authentication mechanisms
- ☐ Delegation to clients
- ☐ Locking
- ☐ Split metadata and data (pNFS)
- ☐ Access control supports ACLs and modes
- ☐ Named attributes



<i>Discovery</i>	Manual (IP address or via DNS lookup of well-known hostname)
<i>Authorization</i>	User-based; processes using ACL
<i>Data Access</i>	Traditional FS operations (e.g., open/close, read/write, remove)
<i>Data Mgmt</i>	Traditional FS operations (rename, link, getattr/setattr)
<i>Data Search</i>	Enumerate directory to find desiredfile (e.g. readdir, lookup)
<i>Resource Ctrl</i>	User-based storage quota
<i>Storage Mode</i>	File-based

OAuth

■ NOT a storage protocol

- ❑ Included here due to its authentication model

■ “client” vs. “resource owner”

- ❑ OAuth separates them
- ❑ Resource owner can provide limited access to a client

■ Features of credentials

- ❑ Expiration time
- ❑ Allow revocation by owner

<i>Discovery</i>	N/A
<i>Authorization</i>	Client creates delegation request; approved by resource owner
<i>Data Access</i>	N/A
<i>Data Mgmt</i>	N/A
<i>Data Search</i>	N/A
<i>Resource Ctrl</i>	N/A
<i>Storage Mode</i>	N/A

WebDAV

- Distributed authoring for web resources

- ☐ And various other uses

- Major features

- ☐ Properties, Locking

- Extensions

- ☐ Versioning (RFC3253)

- ☐ SEARCH (RFC5323)

- ☐ ACL (RFC3744)

- ☐ Tickets for authorization (draft-ito-dav-ticket-00)

<i>Discovery</i>	Manual (IP address or via DNS lookup of well-known hostname)
<i>Authorization</i>	User-based; permissions include read, write, etc
<i>Data Access</i>	Traditional filesystem operations (e.g., read, write); no update
<i>Data Mgmt</i>	Traditional filesystem operations (e.g., move, delete)
<i>Data Search</i>	Enumeration, or list by user-supplied criteria
<i>Resource Ctrl</i>	User- or collection-based storage quota
<i>Storage Mode</i>	File-based (organized by collections)