CS294-4
Peer-to-Peer systems

Introduction

August 25, 2003

John Kubiatowicz/Anthony Joseph

http://www.cs.berkeley.edu/~kubitron/courses/cs294-4-F03

What is Peer-to-Peer?

- P2P is a communications model in which each party has the same capabilities and either party can initiate a communication session.
  Whatis.com
- P2P is a class of applications that takes advantage of resources – storage, cycles, content, human presence – available at the edges of the internet.
  Clay Shirky, openp2p.com
- A type of network in which each workstation has equivalent capabilities and responsibilities.
  Webopedia.com
- A P2P computer network refers to any network that does not have fixed clients and servers, but a number of peer nodes that function as both clients and servers to other nodes on the network.
  Wikipedia.org

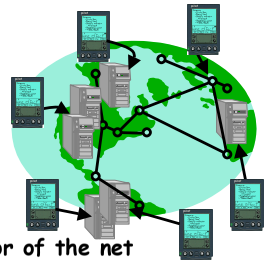
Is Peer-to-peer new?

- Certainly doesn't seem like it
  - What about Usenet? News groups first truly decentralized system
  - DNS? Handles huge number of clients
  - Basic IP? Vastly decentralized, many equivalent routers
- One view: P2P is a reverting to the old internet
  - Remember? (Perhaps you don't)
  - Once upon a time, all members on the internet were trusted.
  - Every machine had an IP address.
  - Every machine was a client and server.
  - Many machines were routers.
- What is new?
  - Scale: people are envisioning much larger scale
  - Security: Systems must deal with privacy and integrity
  - Anonymity: Protect identity and prevent censorship
  - (In)Stability: Deal with unstable components as the edges
    - » But, can systems designed this way be *more* stable?

Why the hype???

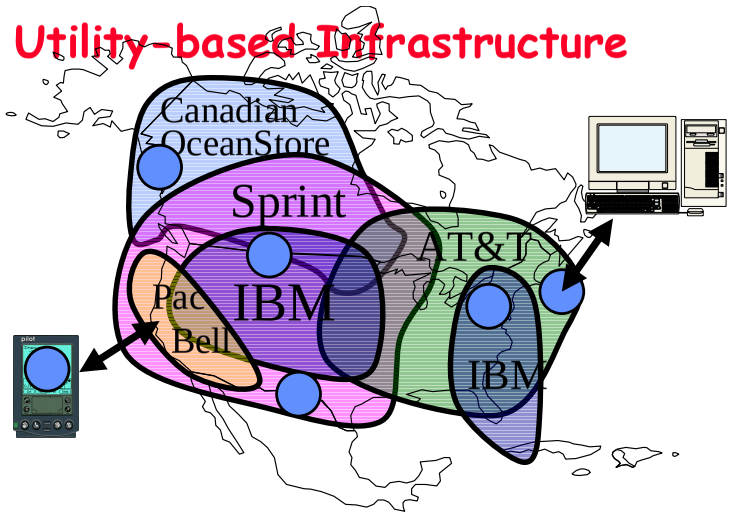
- File Sharing: Napster (+Gnutella, KaZaa, etc)
  - Is this peer-to-peer? Hard to say.
  - Suddenly people could contribute to active global network
    - » High coolness factor
  - Served a high-demand niche: online jukebox
- Anonymity/Privacy/Anarchy: FreeNet, Publis, etc
  - Libertarian dream of freedom from the man
    - » (ISPs? Other 3-letter agencies)
  - Extremely valid concern of Censorship/Privacy
  - In search of copyright violators, RIAA challenging rights to privacy
- Computing: The Grid
  - Scavenge the numerous free cycles of the world to do work
  - Seti@Home most visible version of this
- Management: Businesses
  - Suddenly Businesses have discovered extreme distributed computing
  - Does P2P mean "self-configuring" from equivalent resources?
  - Bound up in "Autonomic Computing Initiative"?

Who am I (Kubi)?
OceanStore Pundit



- Computing everywhere:
  - Desktop, Laptop, Palmtop
  - Cars, Cellphones
  - Shoes? Clothing? Walls?
- Connectivity everywhere:
  - Rapid growth of bandwidth in the interior of the net
  - Broadband to the home and office
  - Wireless technologies such as CMDA, Satelite, laser
- Where is persistent data????

Utility-based Infrastructure



- Data service provided by storage federation
- Cross-administrative domain
- Pay for Service

OceanStore:
Everyone's Data, One Big Utility

"The data is just out there"

- How many files in the OceanStore?
  - Assume 10^{10} people in world
  - Say 10,000 files/person (very conservative?)
  - So 10^{14} files in OceanStore!
  - If 1 gig files (ok, a stretch), get 1 mole of bytes!

Truly impressive number of elements…
… but small relative to physical constants

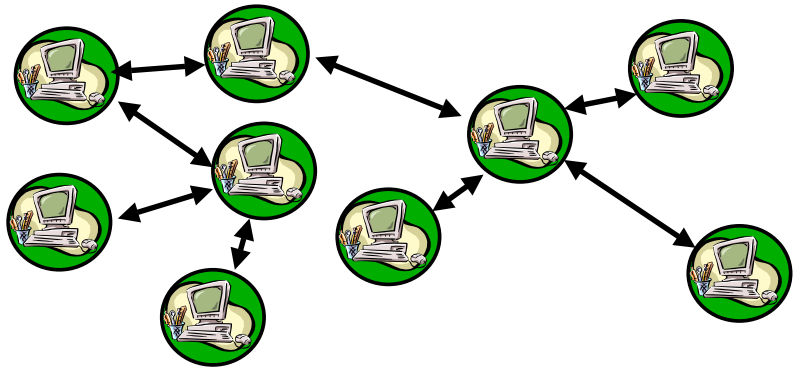
Aside: new results: 1.5 Exabytes/year (1.5×10^{18})

OceanStore Assumptions

- Untrusted Infrastructure:
  - The OceanStore is comprised of untrusted components
  - Individual hardware has finite lifetimes
  - All data encrypted within the infrastructure
- Responsible Party:
  - Some organization (*i.e. service provider)* guarantees that your data is consistent and durable
  - Not trusted with *content* of data, merely its *integrity*
- Mostly Well-Connected:
  - Data producers and consumers are connected to a high-bandwidth network most of the time
  - Exploit multicast for quicker consistency when possible
- Promiscuous Caching:
  - Data may be cached anywhere, anytime

Some Basic Questions about Peer-to-Peer

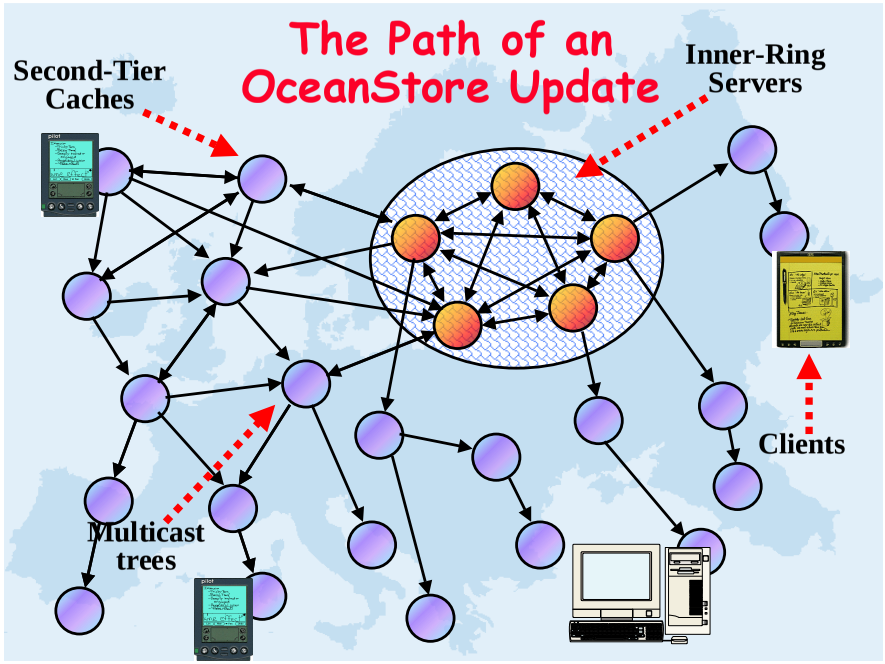
Does Full Symmetry Make Sense? "All nodes are created equal"



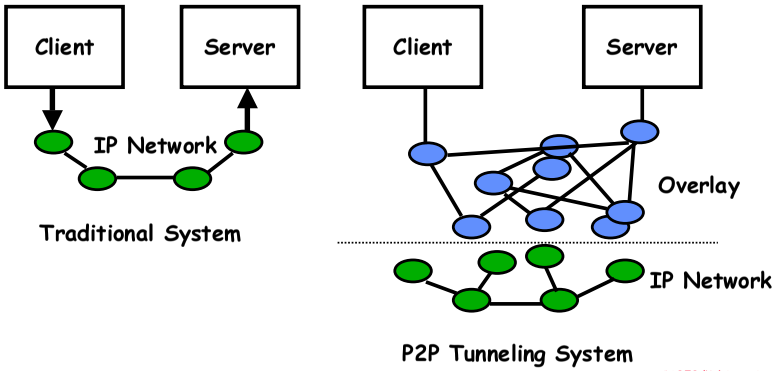
- Most distributed algorithms need points of serialization
  - UseNet may be one of few classes of algorithms that don't
- Nodes have distinguished capabilities
  - Better Connectivity, More memory, Better management, …

Possible advantages of Full Symmetry

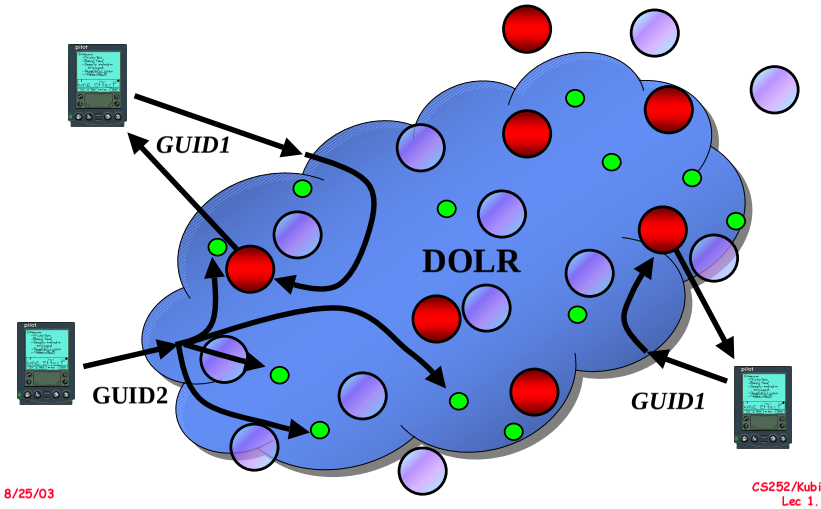
- We will call this "purist peer-to-peer"
- Anonymity/Deniability
  - When combined with cryptographic techniques, no distinguished nodes to go after
  - This is important property today!
- Algorithms Easier to Analyze
  - Differentiation always makes things harder to analyze
  - Almost all recent P2P papers use Purist Assumption
- How does (should?) Hierarchy and Equality interact??
  - This is a question we should answer by end of term



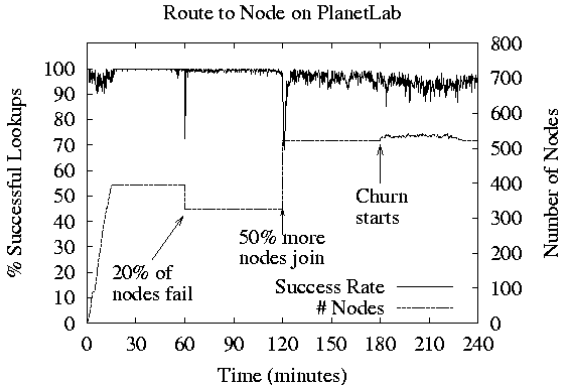
Can P2P Overlay Networking Buy you Something?



Enabling Technology: DOLR (Decentralized Object Location and Routing)



Planetlab DOLR Service

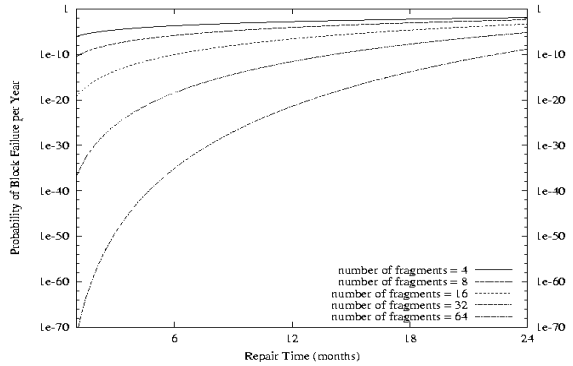


(May 2003: 1.5 TB over 4 hours)

Could experiment with Tapestry as a real service

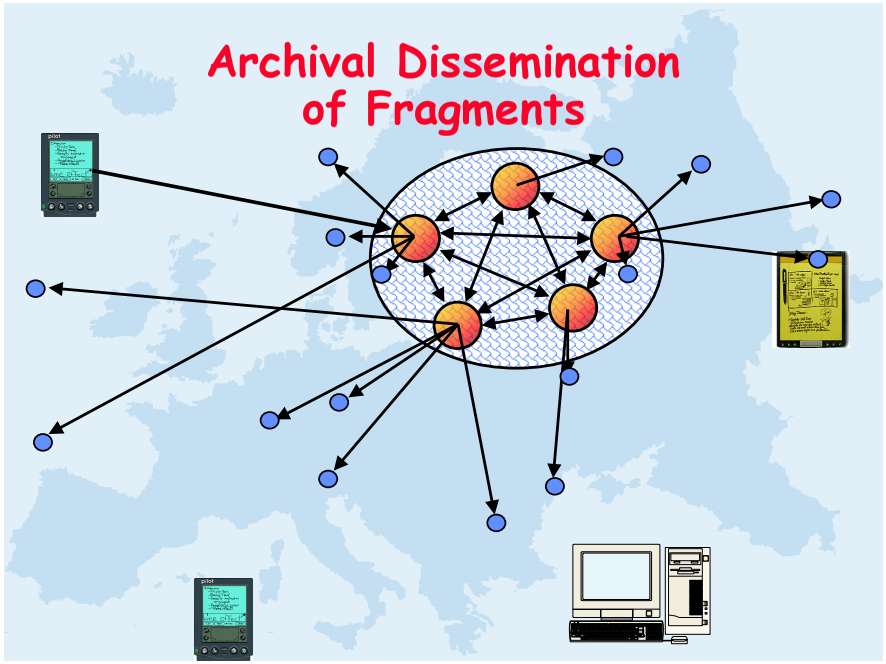
DOLR Model generalizes to many simultaneous apps

Can Replication be used more Effectively in Peer-to-Peer systems?

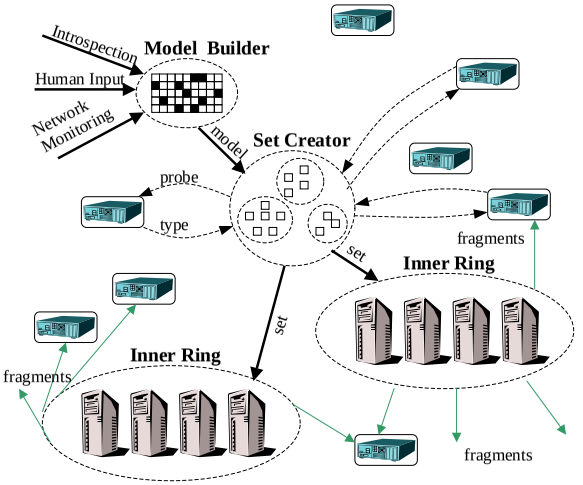


- Exploit law of large numbers for durability!
- 6 month repair, FBLPY:
  - Replication: 0.03
  - Fragmentation: 10^{-35}

Archival Dissemination of Fragments



The Dissemination Process: Achieving Failure Independence

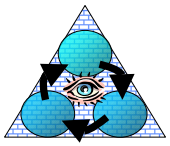


Peer-to-peer Goal: Stable, large-scale systems

- State of the art:
  - Chips: 10^8 transistors, 8 layers of metal
  - Internet: 10^9 hosts, terabytes of bisection bandwidth
  - Societies: 10^8 to 10^9 people, 6-degrees of separation
- Complexity is a liability!
  - More components ⇒ Higher failure rate
  - Chip verification > 50% of design team
  - Large societies unstable (especially when centralized)
  - Small, simple, perfect components combine to generate complex emergent behavior!
- Can complexity be a useful thing?
  - Redundancy and interaction can yield stable behavior
  - Better figure out new ways to design things…

Oceanstore Observation: Want Automatic Maintenance

- Can't possibly manage billions of servers by hand!
- System should automatically:
  - Adapt to failure
  - Exclude malicious elements
  - Repair itself
  - Incorporate new elements

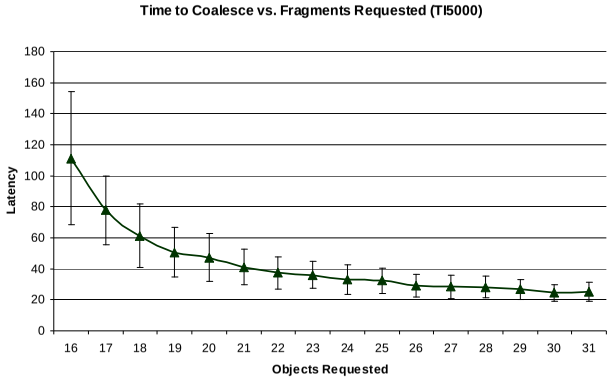


- System should preserve data over the long term (*accessible* for 1000 years):
  - Geographic distribution of information
  - New servers added from time to time
  - Old servers removed from time to time
  - Everything just works

The Thermodynamic Analogy

- Large Systems have a variety of *latent order*
  - Connections between elements
  - Mathematical structure (erasure coding, etc)
  - Equivalent/interchangeable elements
  - Distributions peaked about some desired behavior
- Permits "Stability through Statistics"
  - Exploit the behavior of aggregates (redundancy)
- Subject to Entropy
  - Servers fail, attacks happen, system changes
- Requires continuous repair
  - Apply energy (i.e. through servers) to reduce entropy
  - Introspection restores distributions

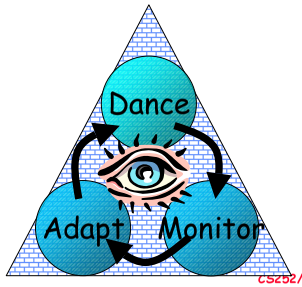
Statistical Advantage of Fragments



- Latency *and* standard deviation reduced:
  - Memory-less latency model
  - Rate ½ code with 32 total fragments

The Biological Inspiration

- Biological Systems are built from (extremely) faulty components, yet:
  - They operate with a variety of component failures ⇒ Redundancy of function and representation
  - They have stable behavior ⇒ Negative feedback
  - They are self-tuning ⇒ Optimization of common case
- Introspective (Autonomic) Computing:
  - Components for performing
  - Components for monitoring and model building
  - Components for continuous adaptation



ThermoSpective

- Many Redundant Components (Fault Tolerance)
- Continuous Repair (Entropy Reduction)

What does this really mean?

- Redundancy, Redundancy, Redundancy:
  - Many components that are roughly equivalent
  - System stabilized by consulting multiple elements
  - Voting/signature checking to exclude bad elements
  - Averaged behavior/Median behavior/First Arriving
- Passive Stabilization
  - Elements interact to self-correct each other
  - Constant resource shuffling
- Active Stabilization
  - Reevaluate and Restore good properties on wider scale
  - System-wide property validation
  - Negative feedback/chaotic attractor
- Observation and Monitoring
  - Aggregate external information to find hidden order
  - Use to tune functional behavior and recognize disfunctional behavior.

Problems?

- Most people don't know how to think about this
  - Requires new way of thinking
  - Some domains closer to thermodynamic realm than others: peer-to-peer networks fit well
- Stability?
  - Positive feedback/oscillation easy to get accidentally
- Cost?
  - Power, bandwidth, storage, ….
- Correctness?
  - System behavior achieved as aggregate behavior
  - Need to design around fixed point or chaotic attractor behavior (How does one think about this)?
  - Strong properties harder to guarantee
- Bad case could be quite bad!
  - Poorly designed ⇒Fragile to directed attacks
  - Redundancy below threshold ⇒ failure rate increases drastically

What about Game Theory?

- If the individual components are selfish can we:
  - Somehow get good aggregate behavior?
  - The search landscape for Game Theoretic and Thermodynamic systems is different!
- Are there aggregate mechanisms for enforcing correctness among selfish peers?

Course Information

- Time: Mon/Wed 10:30 – 12:00, 405 Soda Hall
- Faculty: John Kubiatowicz, Anthony Joseph
- Web: http://www.cs.berkeley.edu/~adj/cs294-4.f03

- Format: Two papers/day + occasional guest
- Course Prereqs: CS122 (networking), CS162 (OS)
- Breakdown (tentative):
  - Class presentation: 30% (two presentations each worth 15%)
  - Class project: 70%

- Projects: team-based and open ended.
  - Choose your own
  - Pick one from our projects page (still not totally decided)

What is a presentation?

- Presentations are in PowerPoint
- Target a 30minute presentation
  - That's about 15 slides
- You will be graded on several things:
  - Do you manage to present the high-level points of the paper?
  - Do you manage to get a good discussion going?
  - Do you relate this paper to other papers/topics in the course (CONTEXT!)
- For every paper find:
  - 3 most important points
  - 3 flaws with the assumptions/methodology/etc.

Why have Students Present?

- This is a lot of fun
- You get more viewpoints than just ours
- *We (Kubi and Anthony) get to learn something from you!!!!*
- I (Kubi) will consider this a great success if I learn something new from you this term.

List of Topics

- Study existing systems:
  - Gnutella, Freenet, Tapestry, RON, OceanStore, FarSite, Ivy, PAST, Pistache, …
- Study Fundamentals:
  - Byzantine Agreement, Logical Clocks, Reliable Group communication, Quorum systems, Game Theory, Authentication, Security..
- Study Behaviors
  - Measurements of existing systems
- Answer Questions:
  - Is peer-to-peer something new?
  - Can peer-to-peer philosophy offer something new?
  - Should we give up and do something else?