Preserving Peer Replicas by Rate-Limited Sampled Voting

Petros Maniatis, Mema Roussopoulos,
TJ Giuli, David S. H. Rosenthal,
Mary Baker, Yanto Muliadi

Stanford University

Digital Preservation of Academic Materials

- Academic publishing is moving from paper to electronic media
  - Instead of purchasing paper copies, libraries rent access to on-line digital materials
- Librarians are scared with good reason
  - Access depends on the fate of the publisher
  - Time is unkind to bits after decades
  - Plenty of enemies (ideologies, governments, corporations)
- **Goal: Preserve access for local patrons, for a very long time**

The Preservation Problem

- Preserve access to a population of "document" replicas
  - "Document" may be a year-long run of a journal
- Requirements
  - Very low-cost hardware, operation and administration
  - No central control
  - Respect for access controls
  - A long-term horizon
- Must anticipate and degrade gracefully with
  - Undetected bit rot
  - Sustained attacks

Protocol Threats

- Assume conventional platform/social attacks
- Mitigate further damage through protocol
- Top adversary goal: **Stealth Modification**
  - Modify replicas to contain adversary's version
  - Hard to reinstate original content after large proportion of replicas are modified
- Other goals
  - Denial of service
  - System slowdown
  - Content theft

The LOCKSS Solution

- Peer-to-peer auditing and repair system for replicated documents / no file sharing
- A peer periodically audits its own replica, by calling an opinion poll
- When a peer suspects an attack, it raises an alarm for a human operator
  - Correlated failures
  - IP address spoofing
  - System slowdown
- New iteration of a deployed system

Sampled Opinion Poll

- Each peer holds
  - *reference list* of peers it has discovered
  - *friends list* of peers it knows externally
- Periodically (faster than rate of bit rot)
  - Take a sample of the reference list
  - Invite them to send a hash of their replica
- Compare votes with local copy
  - Overwhelming agreement (>70%) ☞ Sleep blissfully
  - Overwhelming disagreement (<30%) ☞ Repair
  - Too close to call ☞ Raise an alarm
- To repair, the peer gets the copy of somebody who disagreed and then reevaluates the same votes

Reference List Update

- Take out voters in the poll
  - So that the next poll is based on different group
- Replenish with some "strangers" and some "friends"
  - Strangers: Accepted nominees proposed by voters
  - Friends: From the friends list
  - The measure of favoring friends is called *churn factor*

LOCKSS Defenses

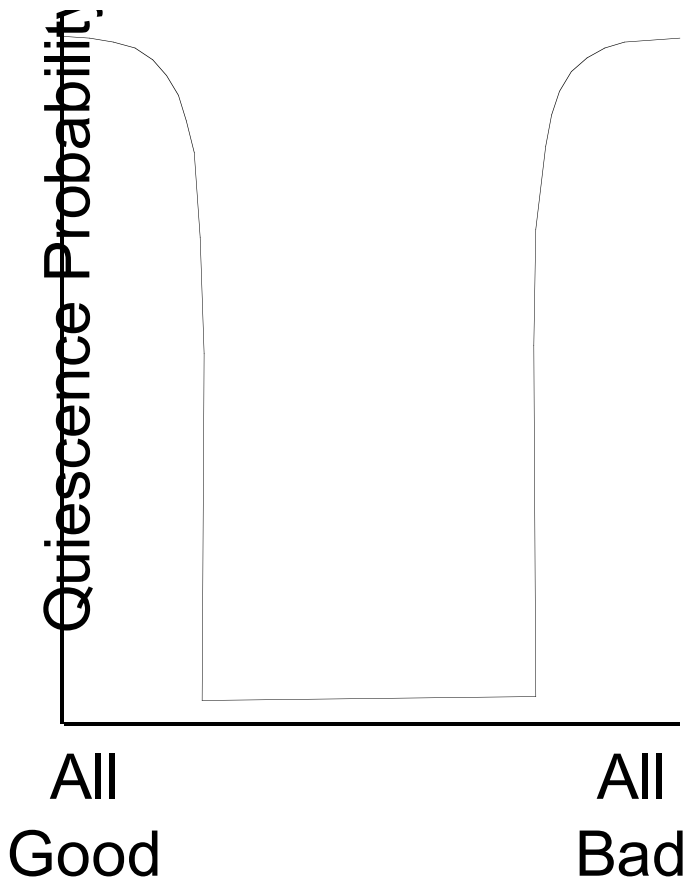
- Limit the rate of operation
- Bimodal system behavior
- Churn friends into reference list

Limit the rate of operation

- Peers determine their rate of operation autonomously
  - Adversary must wait for the next poll to attack through the protocol
- No operational path is faster than others
  - Artificially inflate "cost" of cheap operations
  - No attack can occur faster than normal ops

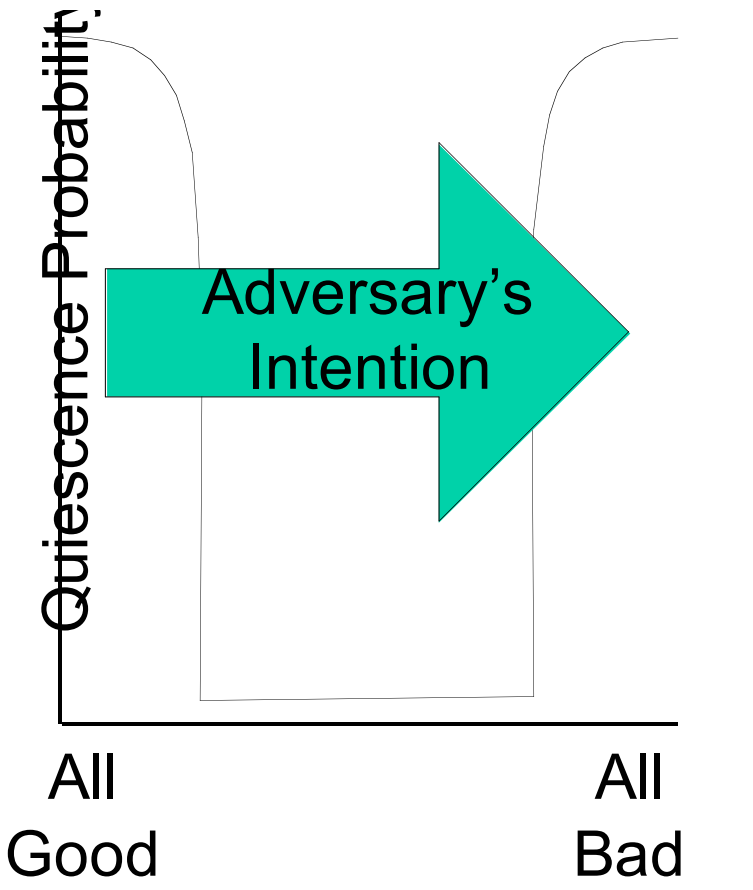
Bimodal System Behavior

- When most replicas are the same, no alarms
- In between, many alarms
- To get from mostly correct to mostly wrong replicas, system must pass through "moat" of alarming states



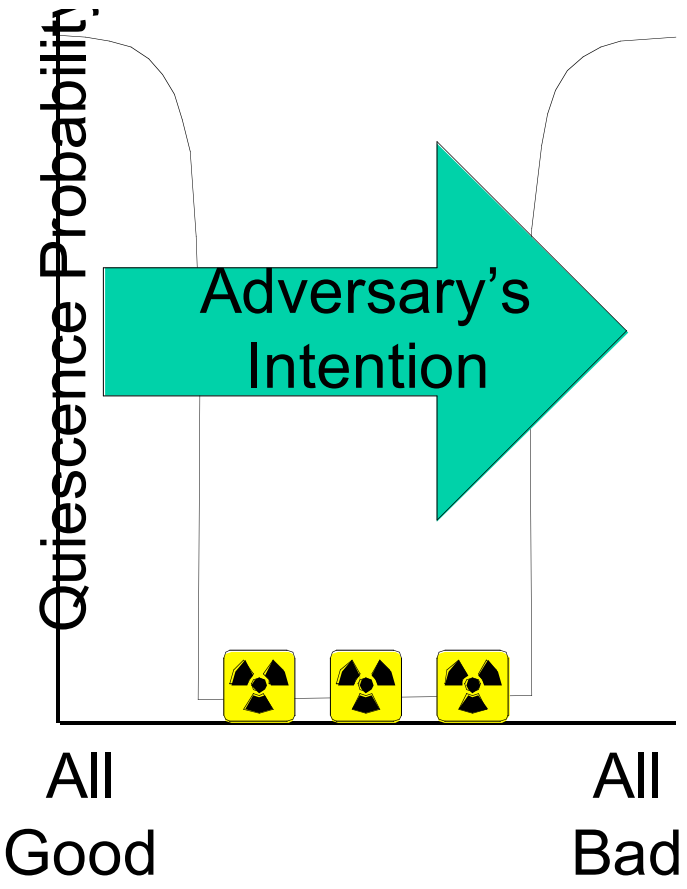
Bimodal System Behavior

- When most replicas are the same, no alarms
- In between, many alarms
- To get from mostly correct to mostly wrong replicas, system must pass through "moat" of alarming states



Bimodal System Behavior

- When most replicas are the same, no alarms
- In between, many alarms
- To get from mostly correct to mostly wrong replicas, system must pass through "moat" of alarming states



Churn Friends into Reference List

- Churn adjusts the bias in the reference list
- High churn favors friends
  - Reduces the effects of Sybil attacks
  - But offers easy targets for focused attack
- Low churn favors strangers
  - It offers Sybil attacks free reign
    - Bad peers nominate bad; good peers nominate some bad
  - Makes focused attack harder, since adversary can predict less of the poll sample
- Goal: strike a balance

Evaluation Methodology

- Model a very powerful, realistic adversary
- Identify major goals of adversary attacks
- Devise and implement rational strategies
- Measure the impact of each strategy
  - locally (on library patrons)
  - globally (on document survival)

Adversary Model

- Unconstrained resources
  - Purchased (cheap) or spoofed (cheaper) identities
  - Limitless computational power
- Initially takes over proportion of peer population
  - Through platform exploits, bribery, broken kneecaps
  - Taken over peer is *subverted* for the duration
- Perfect coordination
  - Instantaneous communication with and control of minions
  - Load balancing of attack effort
  - Flawless content preservation
- Perfect knowledge of peers' operational parameters

Stealth Modification Strategy

- Goal is to replace good replicas with bad replicas without being detected
- First he *lurks*
  - Objective: Increase his foothold in unsubverted peers' reference lists
  - Follow the protocol, but always nominate bad peers
- Then he *attacks*
  - Objective: Convince unsubverted peer its replica is damaged
  - Jump in, supplying a "repair" copy of the bad replica
  - When adversary has too few minions in a poll, he votes with correct replica to avoid an alarm
  - The greater his foothold in unsubverted peer's reference list, the less likely the alarm

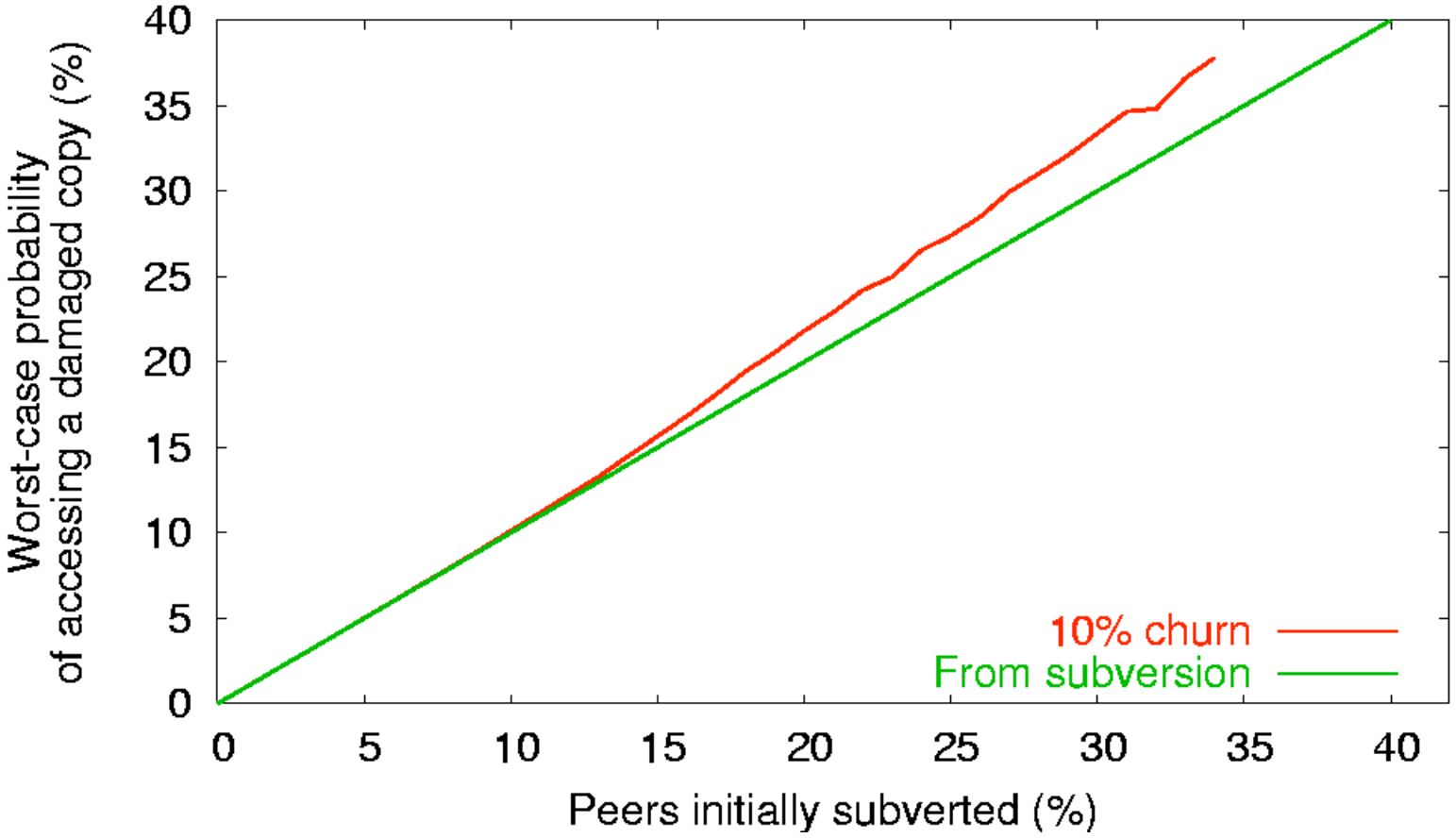
Evaluation

- We use Narses, an application-layer protocol simulator
- Scenarios
  - 1000 original peers, in clusters of friends
  - Initially, 0 – 40% are subverted
  - Lurk for up to 20 years
  - Attack for up to 10 more years
  - Report worst-cases over ~200 runs per data point (recent results)

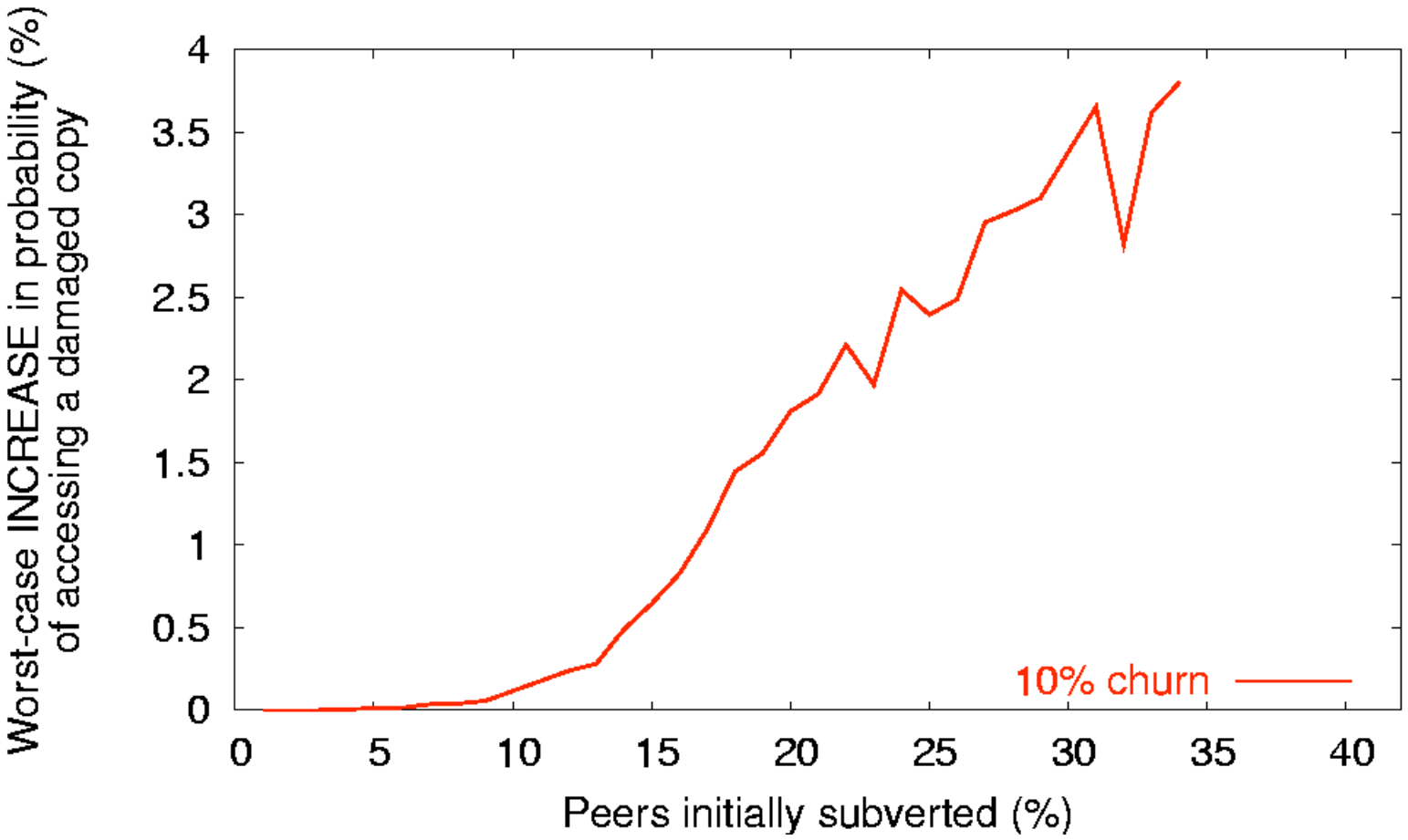
Metrics

- Metrics
  - What's the probability that an access reaches a bad replica
  - What's the probability that the document is damaged irrecoverably
- **How big is the effect of the worst protocol attack on top of the effect of the initial subversion?**

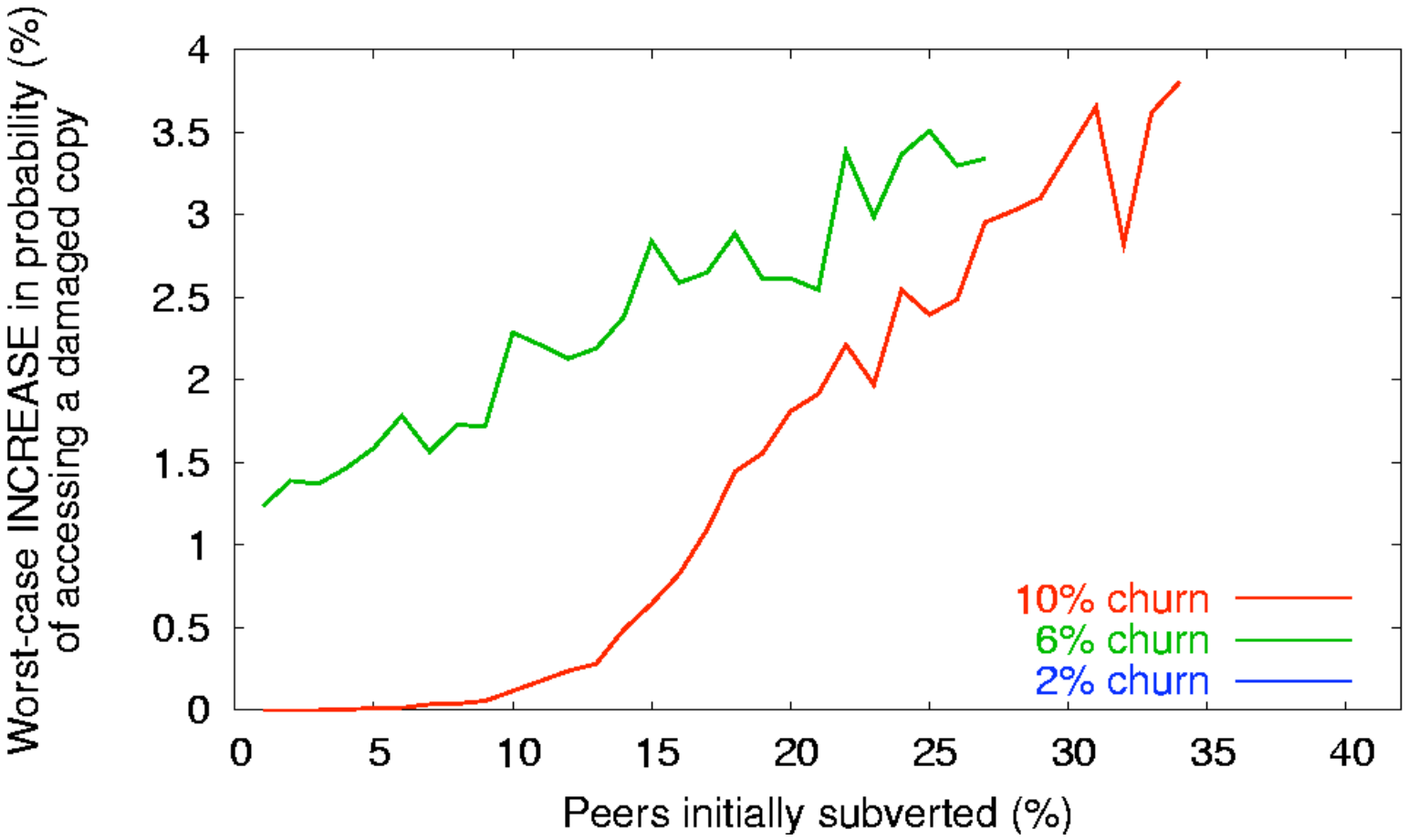
Probability of Accessing Bad Replica



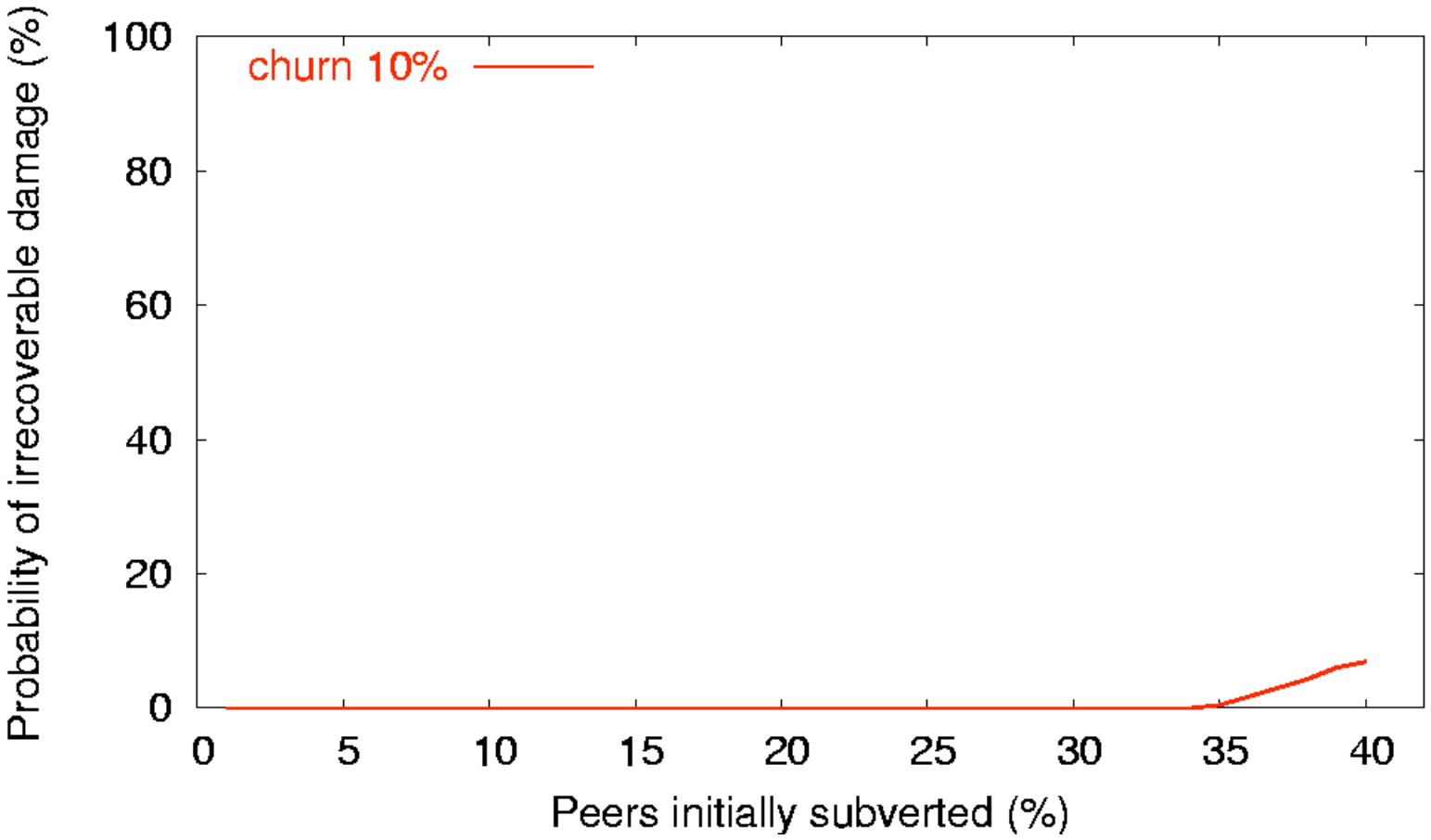
Probability of Accessing Bad Replica (Incremental)



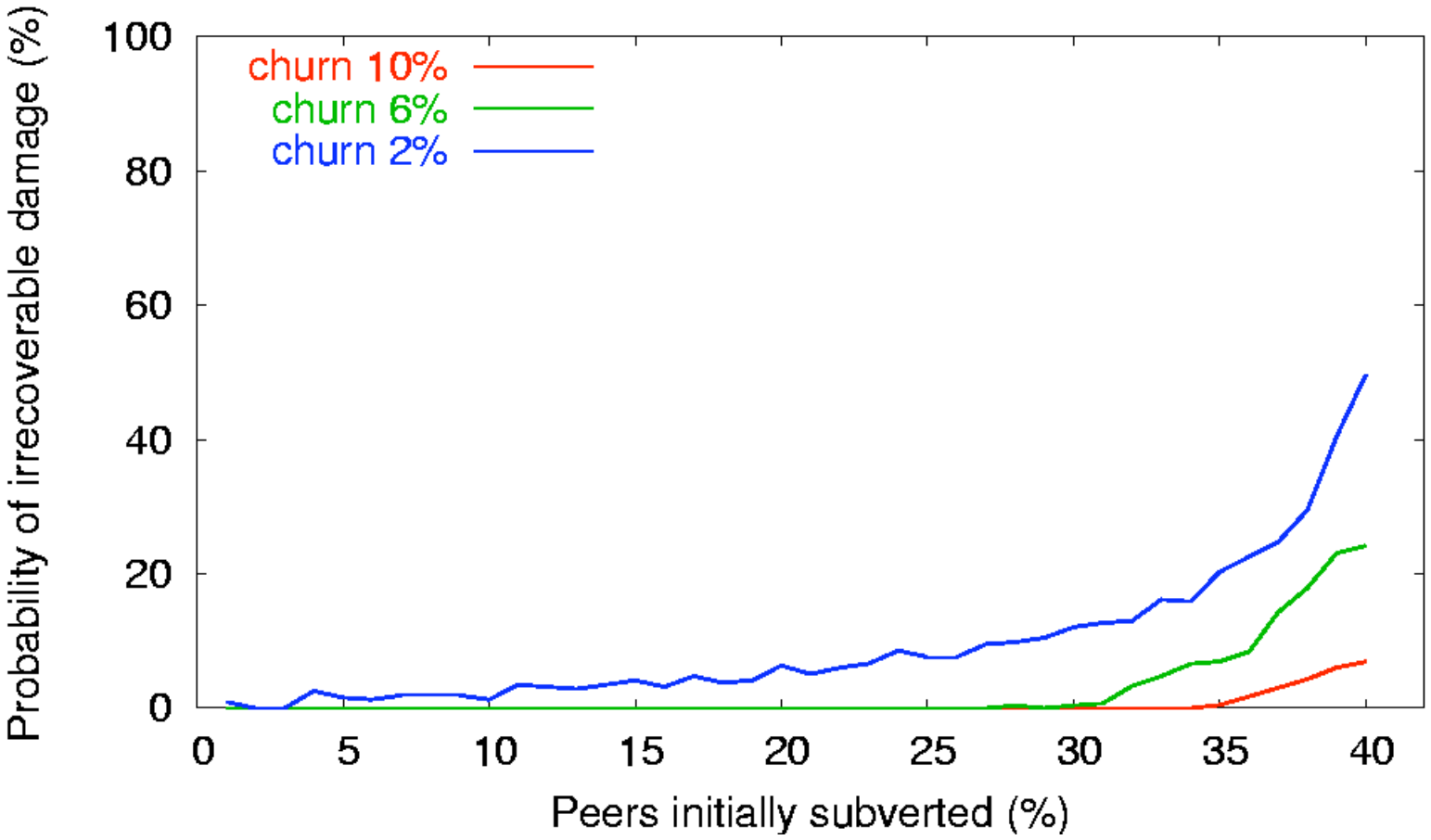
Probability of Accessing Bad Replica (Incremental)



Probability of Irrecoverable Damage



Probability of Irrecoverable Damage

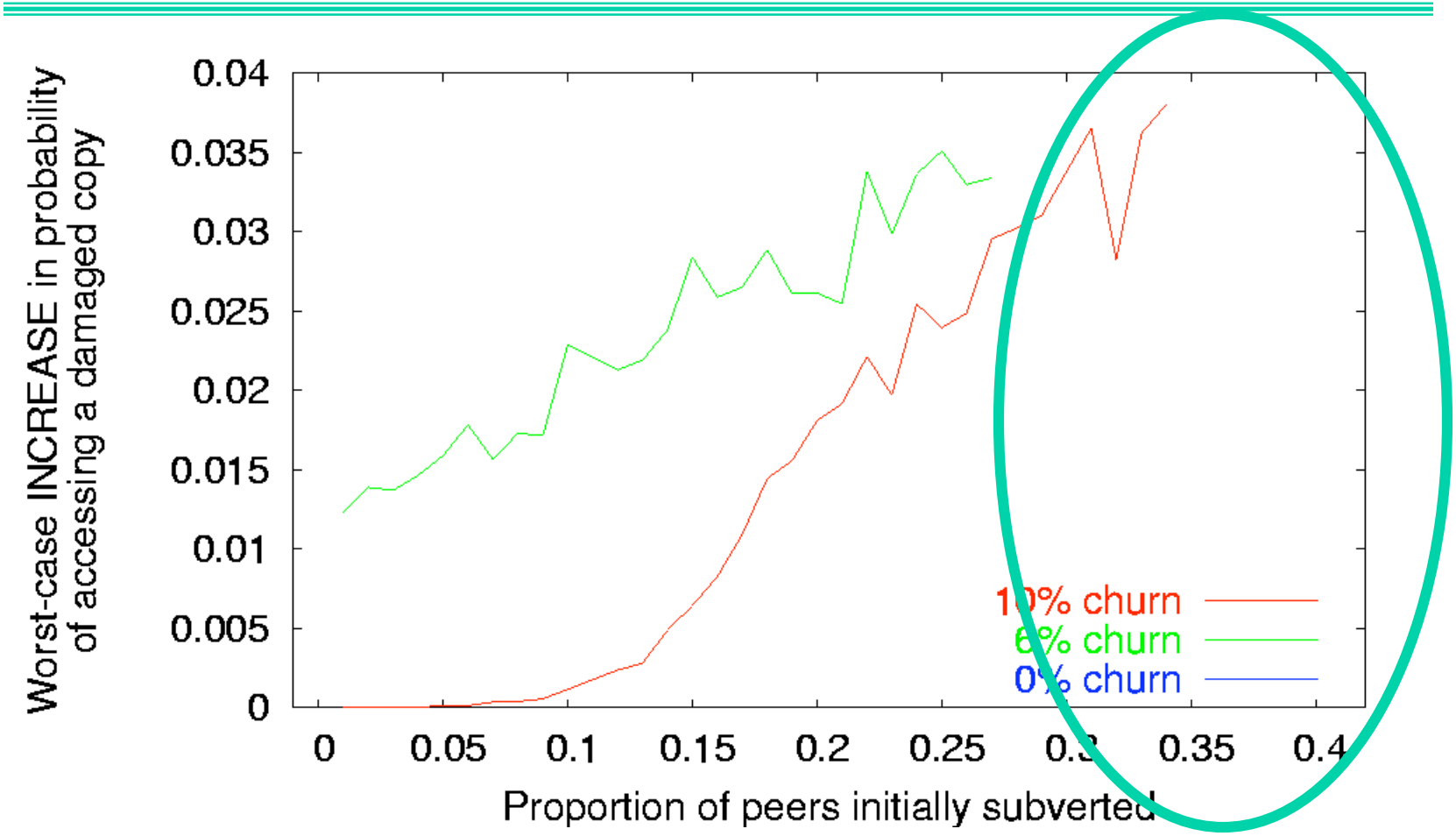


Conclusions

- P2P digital preservation system
  - Rate limiting, bimodal intrusion detection, churn
- Results
  - Resistant to attacks for low subversions
  - Degrades gracefully for greater subversions
- Status
  - Promising results for other attacks (DDoS)
  - Beta running, new protocol to be deployed in 2004
  - Simulator, current and new protocol on sourceforge

http://www.lockss.org/

Global Viewpoint – Document Survival



Popular Yet Flawed Alternatives

- Use super-fabulous RAID
  - Can be complementary, but alone cannot ensure survivability when failures do occur (e.g., human error)
- Encrypt or sign to ensure integrity / cryptography for strong identities
  - Preserving public keys just as hard a problem
- Boost efficiency with erasure codes etc.
  - Storage space is not an issue and all replicas must be whole

Discovery

- Participants in a poll nominate peers for inclusion in the caller's reference list
  - They can pick nominees however they want
- Caller invites some nominees into the poll
  - Equal number from each nominator
  - Enough to keep the reference list populated
  - Nominees vote in same way as original invitees
- Nominee's vote is **only** used to determine acceptance
  - Nominee is accepted only if it votes with "correct" result
  - Nominee vote is **not** counted in vote tabulation

Scaling Considerations

- Large libraries have about 5000 electronic titles, most with short history
- Median title/year size is 300 MBytes, max is 2.5 GBytes
- For storage and polls, each title/year costs no more than $5/year
- Rack of ~10 PCs could cover large collections