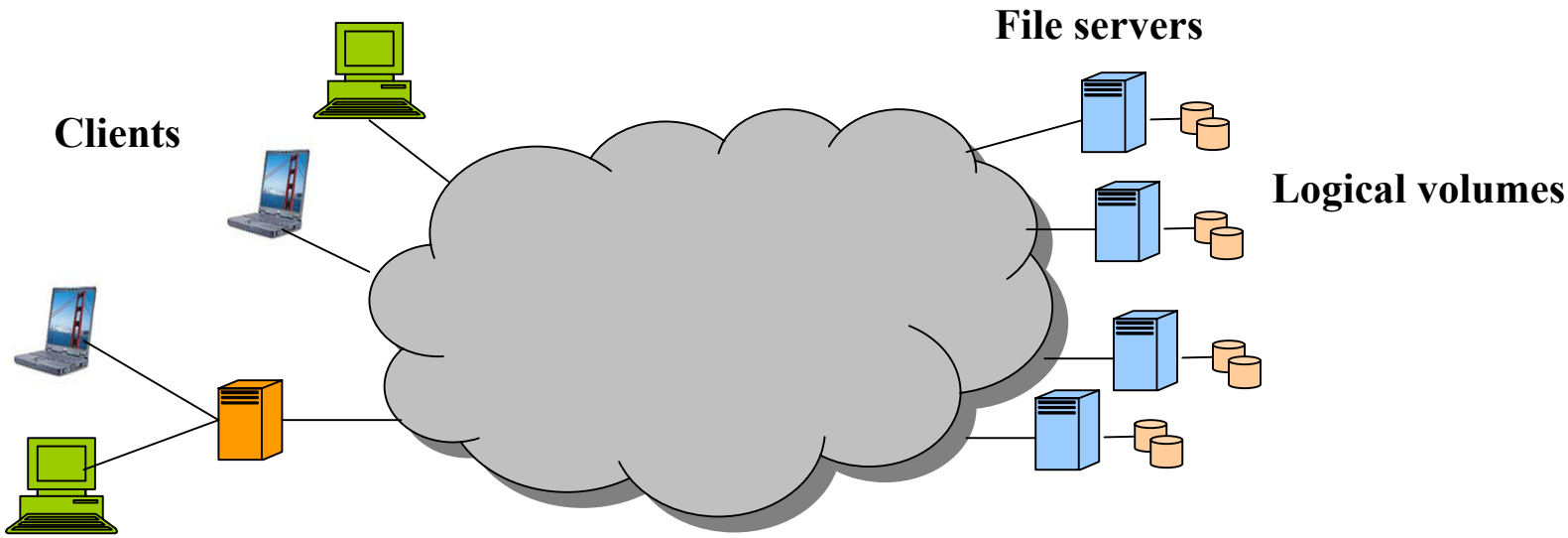
Locating Logical Volumes in Large-Scale Networks

Mallik Mahalingam, Christos Karamanolis, Magnus Karlsson and Zhichen Xu

HP Laboratories
Palo Alto, CA, U.S.A.
mmallik@hpl.hp.com

Motivation

- Storage resources are increasingly accessible over networks, intranets or even Internet



- Highly dynamic environment. Clients, files and volumes can all move to meet varying workload and infrastructure demands and conditions

Problem Definition

The location of a file is identified by a unique file ID containing a volume identifier

In a global and highly dynamic environment:

- **How can file servers find out what logical volumes they are in charge of?**
- **How can clients effectively find the file-server that serves a specific volume identifier?**
- **What infrastructure could be used to accomplish these goals, and how effective are they?**

Results and Contributions

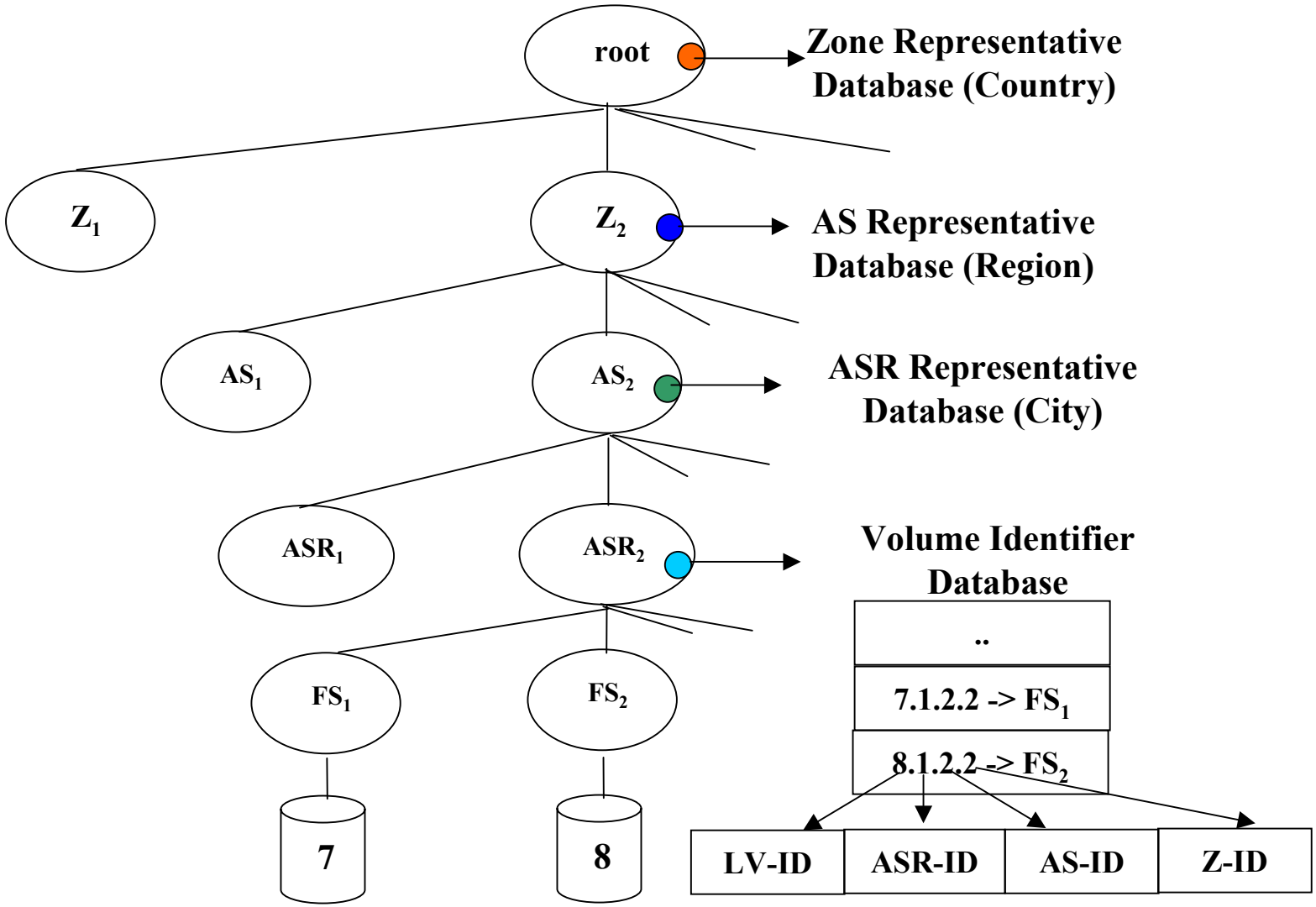
- Assignment of logical volumes to file servers
  - Based on DHCP
- Resolution of volume identifiers
  - Using DNS and prefix based routing principles

- We find that existing infrastructure such as DNS and DHCP is sufficient

Outline

- Logical volume identifier
- Logical volume assignment
- Logical volume discovery
  - DNS-approach
  - Prefix-based routing
- Simulation methodology
- Experimental results
- Conclusions

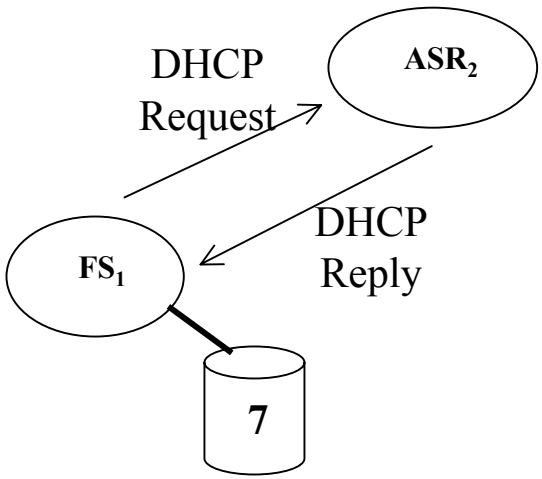
Logical Volume Identifier

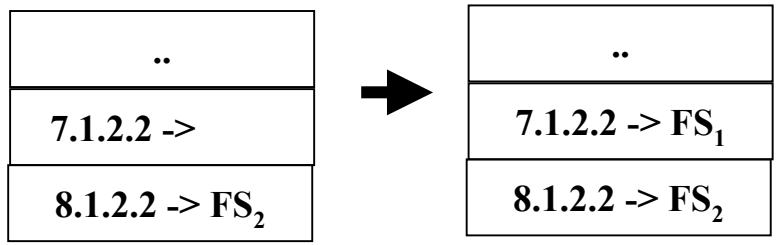


Volume Assignment

- Volumes are configured and managed via local administrative nodes (ASR representative)
- Volumes may be reassigned to different file servers due to fluctuating demand and/or changing system conditions
- Assignment of volumes to file servers is done using DHCP

Volume Assignment





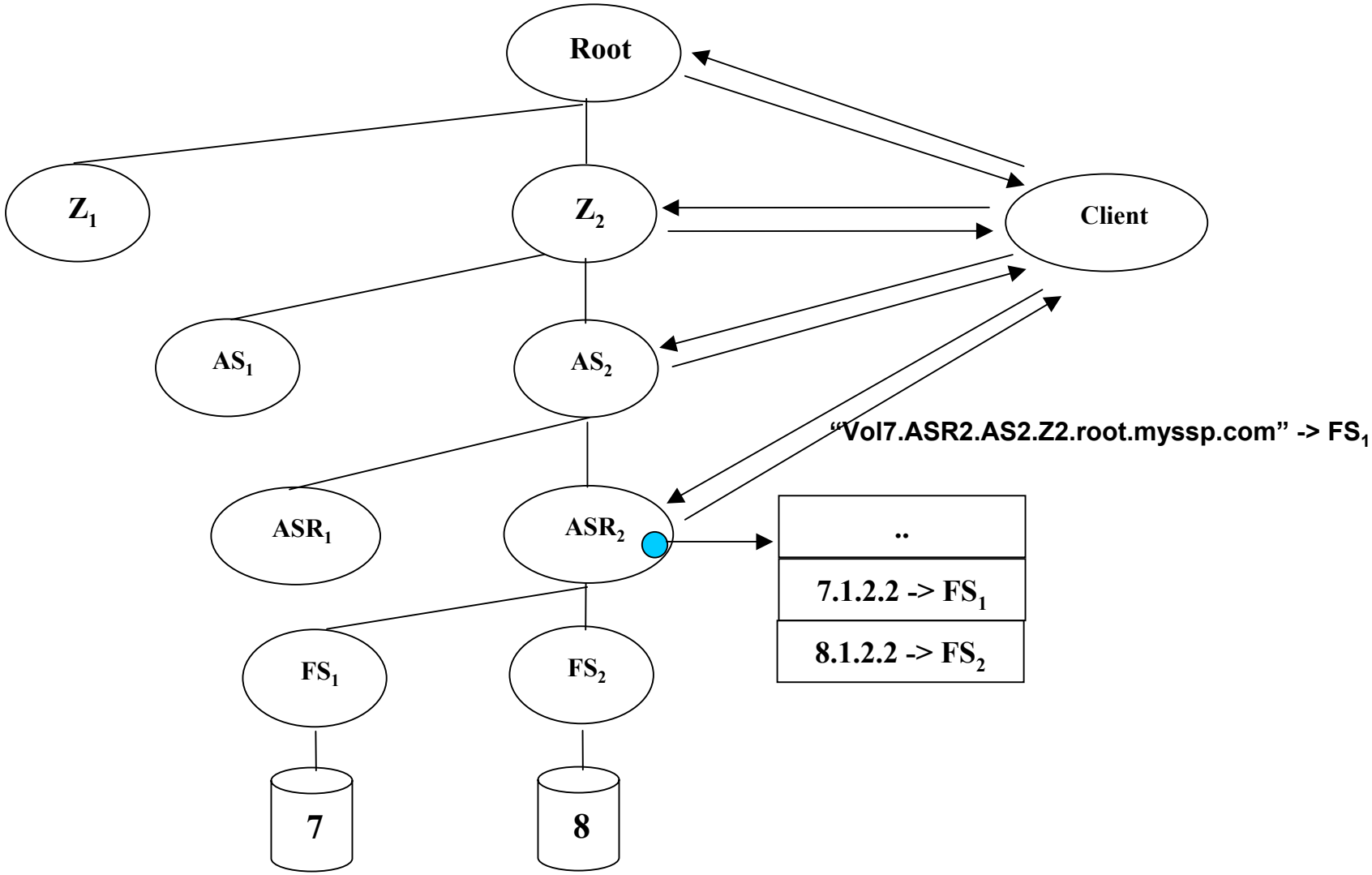
- If DHCP cannot be used:
  - File server is pre configured with logical volumes, or
  - File server is configured with an ASR address to contact for logical volume assignment
- Reassignment of logical volumes affects only the ASR data base

Volume ID Resolution

- In a distributed file system, file and directories can reside in different volumes served by different file servers
- When clients access a file, the server returns a file handle for that file
- The file handle contains a VID corresponding to the logical volume within which the file is physically stored
- Clients need to find the file server that “owns” the logical volume corresponding to that VID before accessing the file

Volume ID Resolution Using DNS

DNS resolution for "Vol7.ASR2.AS2.Z2.root.myssp.com"

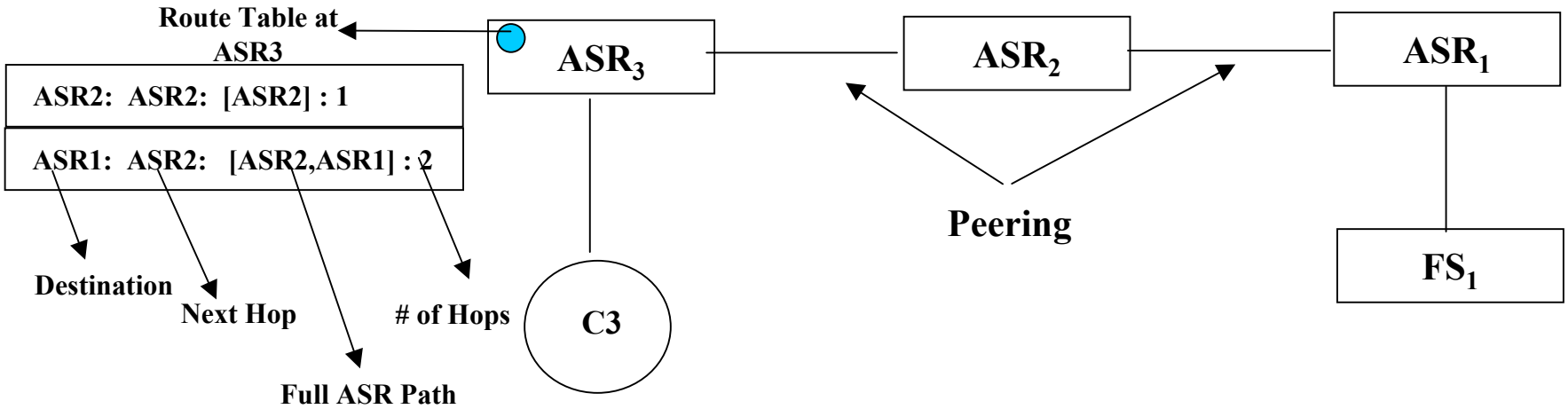


Volume ID Resolution Using DNS (2)

- **Advantage:** Uses existing DNS infrastructure
- **Disadvantage:** Might have to go through many network hops in order to resolve the VID

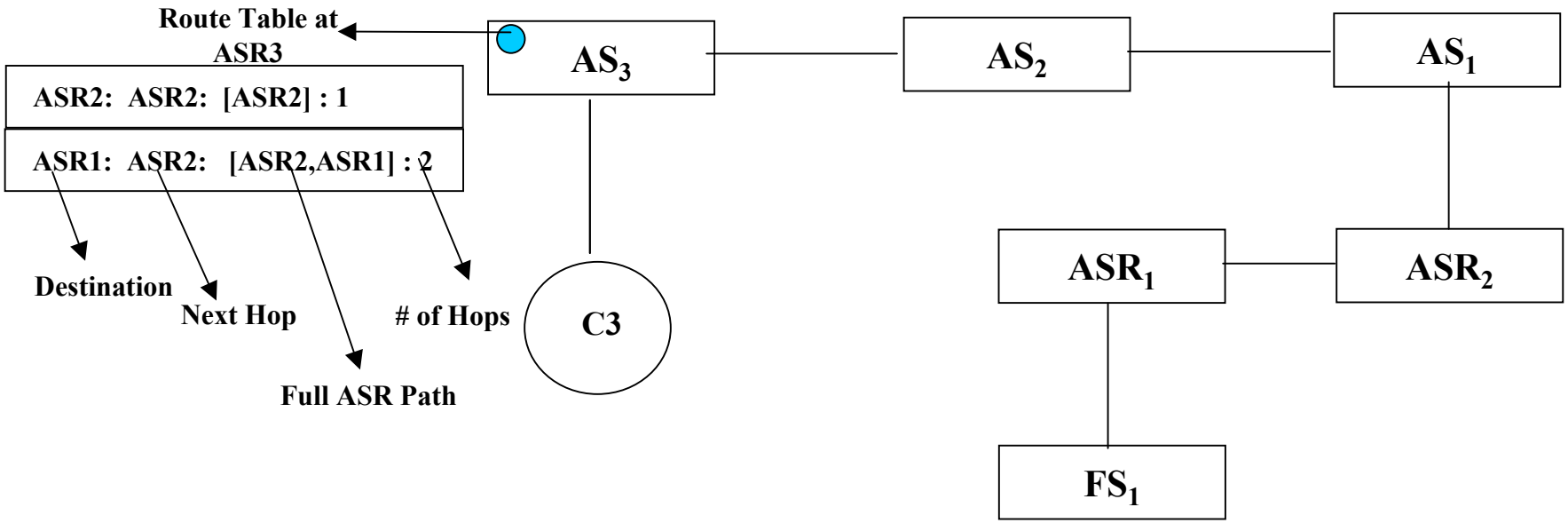
  Caching alleviates this problem when there is locality

Volume ID Resolution Using VIRP



- ASR-nodes could either be the administrative nodes or the clients themselves
- Could also be built into routers
- Each ASR knows about everybody else
- Drawback: ASR level routing might produce large routing tables

Volume ID Resolution Using VIRP+

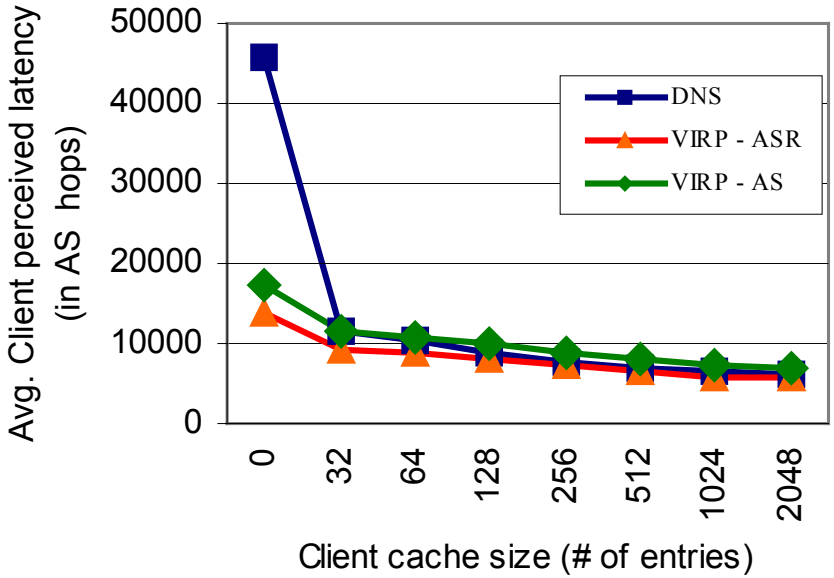


- Two-level routing
- ASR and AS nodes could either be administrative nodes or the clients themselves
- Advantage: Less routing information than previous approach
- Drawback: Usually leads to more network hops

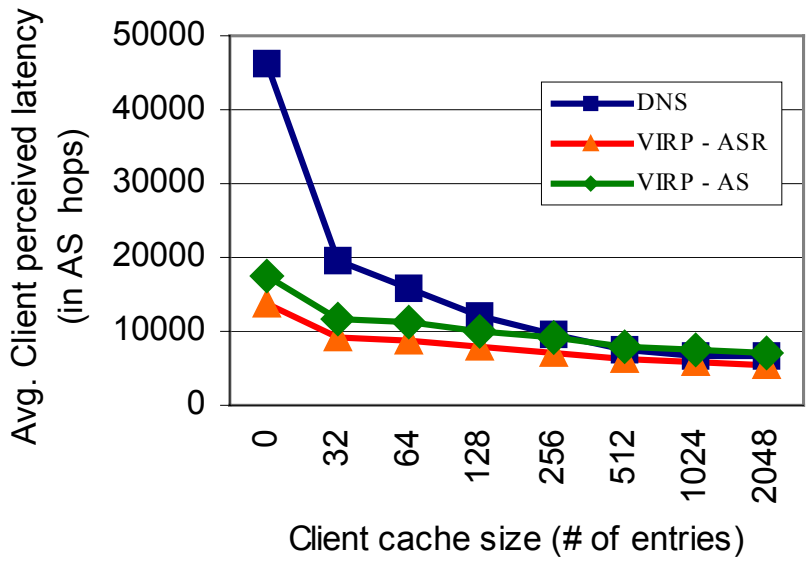
Simulation Methodology

- Access Patterns
  1. Worldcup 98 web-trace
  2. Uniformly random
- File Servers and Logical Volumes
  - 20,000 / 80,000
- Object distributions
  - Sequential / uniformly random
- Topology Distribution
  - Using real BGP-data
- Clients
  - Aggregated (using BGP) clients from Worldcup 98 web-trace

Worldcup 98, 20000 volumes



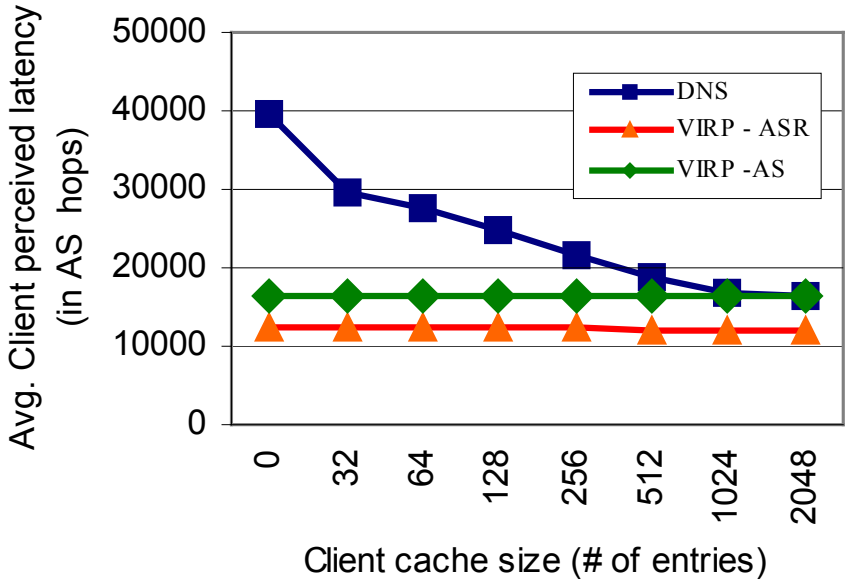
(a) Sequential object distribution



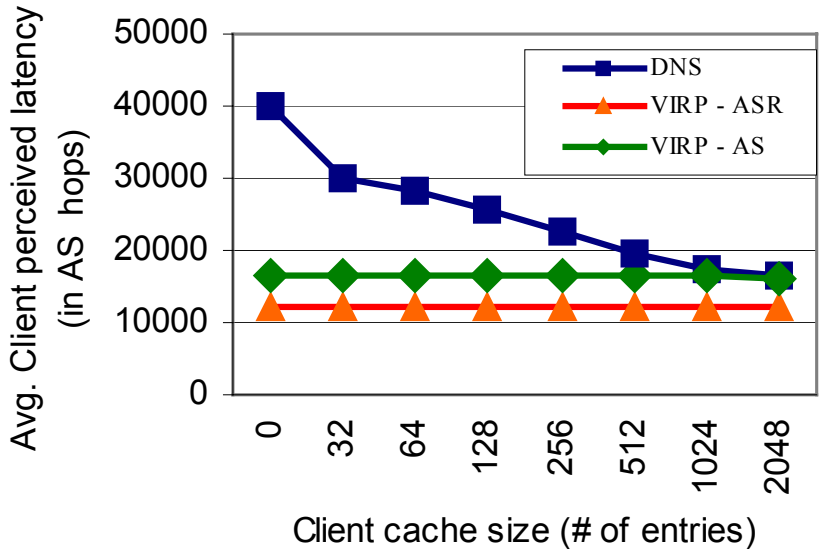
(b) Random object distribution

- **The DNS approach performs well when the locality is high**
- **As expected, VIRP-ASR performs the best**

Random accesses, 80000 volumes



(a) Sequential object distribution



(b) Random object distribution

- **The DNS-approach performs significantly worse for small cache sizes**
- **Still good for large cache sizes**

Conclusion

- VIRP with ASR-level aggregation outperforms other approaches
- For reasonable cache size, the DNS approach is comparable to the VIRP approach, even when locality is poor
- For reasonable cache sizes, the object distribution has little effect on client perceived latency
- The VIRP approach requires modification to existing infrastructure whereas DNS can be deployed using existing infrastructure