An information-centric overlay network architecture for content distribution and mobility support

Ph.D. Presentation

Konstantinos Katsaros

Advisor: Prof. George C. Polyzos

Athens University of Economics and Business

Department of Informatics

Motivation

- Internet model: end-to-end principle
  - Need to resolve a specific end-host to retrieve data
  - Location coupled with identity
- Internet use: information-centric
  - “Anyone” that can provide the required data is fine
  - E.g. P2P, CDNs, etc.
    - P2P file sharing traffic accounted for 53% of Internet traffic in 2008/2009 [1]
- Internet use: mobile end hosts
  - Fixed identity, changing location
- Model mismatch …

[1] IPOQUE. (2009) Internet Study 2008/2009. [Online]. Available: http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009

Motivation (Cont.)

- Arbitrary overlay content delivery structures
  - Created at the edge of the network
  - Ignoring:
    - Network topology
    - Data location
    - Data popularity
- Inefficient use of network resources
  - E.g., up to 95% percent of inter-domain traffic could be avoided in BitTorrent [1]
  - P2P traffic throttling attempts, network neutrality issues
- Mobility not inherently supported
  - "Add-on" solutions e.g., Mobile IP, SIP
  - *Walled gardens…*

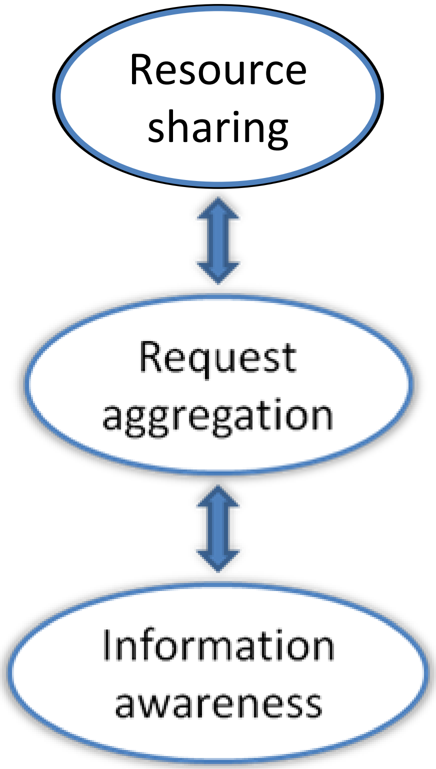
[1] R. Cuevas, N. Laoutaris, X. Yang, G. Siganos, and P. Rodriguez, "Deep Diving into BitTorrent Locality," CoRR, vol. abs/0907.3874, 2009.

Outline

- Motivation
- Objectives
- Architecture overview
- Thesis overview
- Multicast
- Multicast & Caching
- *Hierarchical Pastry*
- *Mobility support*
- Related Work
- Summary & Conclusions
- Future Work

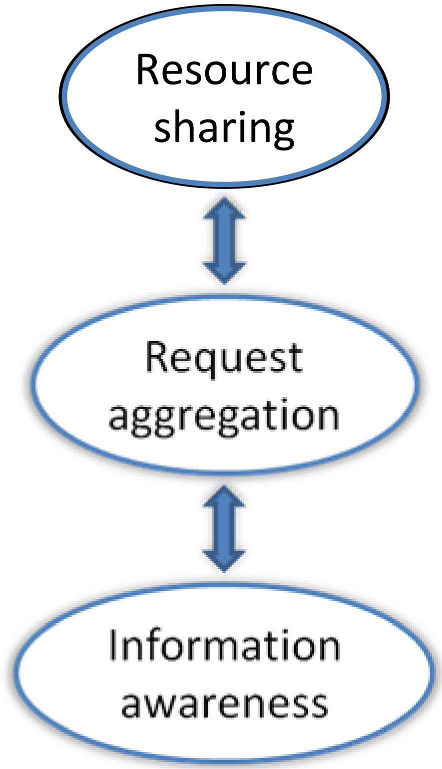
Objectives

- **Efficient** network **resource utilization**
  - Resource sharing mechanisms
  - E.g., multicast, caching
- **Scalability**
  - Unlimited size of the information domain
- **Usage model simplification**
  - End hosts not engaging in translating what to where
- Enhancing end user **quality of experience**
- Inherent **mobility support**



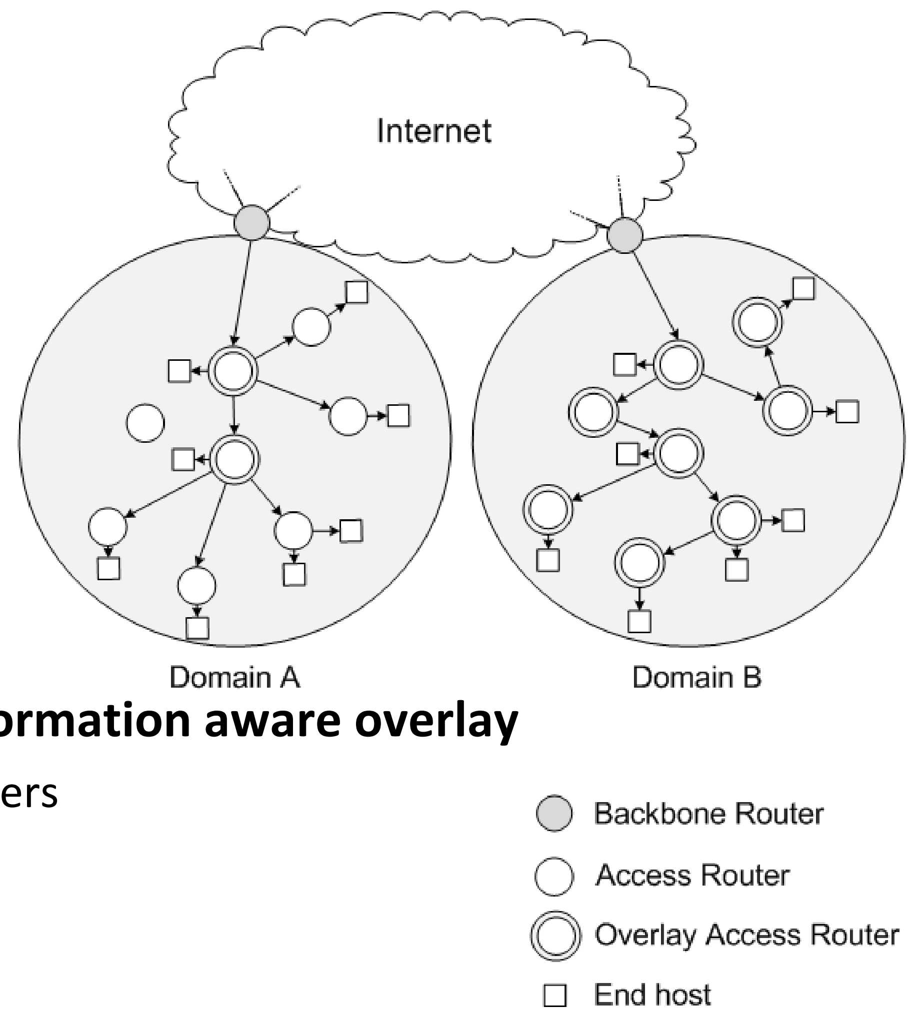
Objectives (Cont.)

- **Facilitating deployment**
  - *Clean-slate* requires replacing existing functionality
  - Available network layer solutions (e.g., IP Multicast)
    - Practically not available
    - Not easy to deploy gradually
    - Difficult group management
  - Content Delivery Networks
    - Overlay deployment
    - Proprietary, ad-hoc solutions
    - Cumbersome configuration/maintenance



Architecture overview

- **Overlay Access Routers (OARs)**
  - Deployed inside access networks
  - Gradual deployment is feasible
- Providing **overlay multicast**
  - Based on Scribe over Pastry DHT
    - Scalable
    - Adaptive to physical topology
- Providing **caching**
  - Multiple cache locations
  - Close to end-hosts
- Proxy-ing end host **access to the information aware overlay**
  - Based on (Pastry/Scribe) flat identifiers
  - Request aggregation
- Supporting **mobility**



Thesis overview

- **Multicast**
  - *Router Assisted Overlay Multicast (RAOM)*
  - Deploying multicast functionality in an overlay fashion
- **Multicast & Caching**
  - *MultiCache*
  - Enabling caching of data delivered by multicast trees
- **Adapting to the inter-network structure**
  - *H-Pastry*
  - Canonical version of Pastry
- **Mobility Support**
  - *Overlay Multicast Assisted Mobility (OMAM)*
  - Revisiting multicast assisted mobility

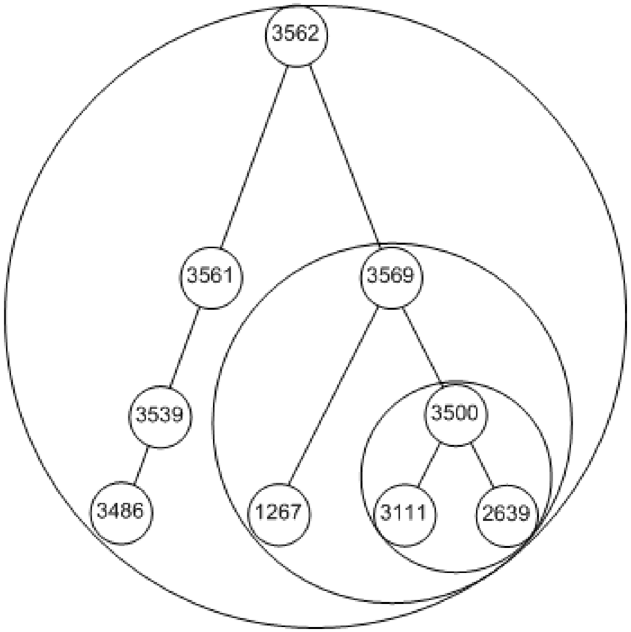
Overlay routing/forwarding substrate

Pastry

- Prefix Routing
  - Exponentially decreasing distance in ID space
  - Exponentially increasing length of overlay (hops)
- **Short routes** property
  - Proximity metrics
  - Reduced stretch
- **Route convergence** property
  - The average distance traveled by each of two messages before their routes converge is approximately equal to the distance between their respective source.

Scribe

- Fully distributed Publish/Subscribe multicast scheme
- Reverse path routing
  - Forwarding state established with JOIN messages towards RV

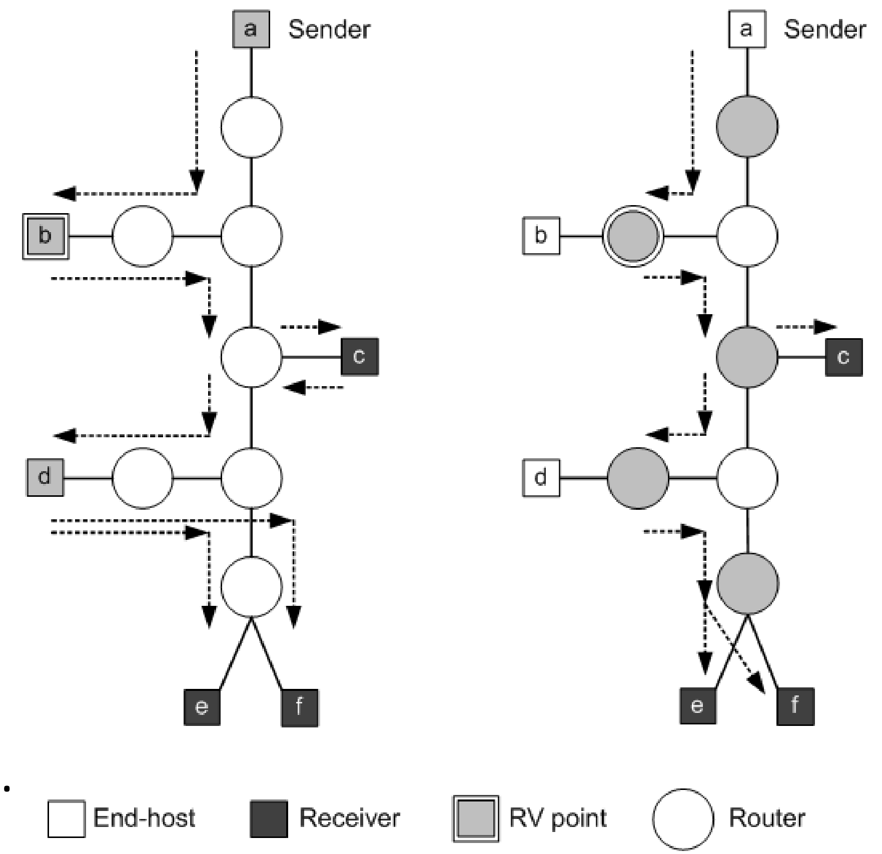


Router Assisted Overlay Multicast (RAOM)

MULTICAST

Router Assisted Overlay Multicast (RAOM)

- Avoiding reliance on end-hosts
- A forwarding end-node may become a bottleneck
  - ADSL uplinks
  - E.g., node *d*
- Multiple downloads by neighboring end-nodes
  - E.g., nodes *e* and *f*
- Increased churn
  - Application-specific participation …
- How much do we gain…?



RAOM Trees properties

Content delivery paths & traffic load

- ***Path stretch***

$$= \frac{\text{\# IP hops in the overlay}}{\text{\# IP hops of shortest path tree}}$$

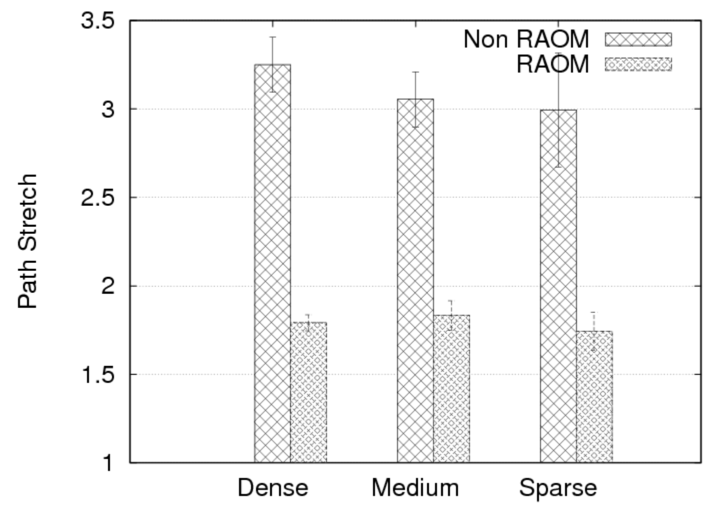
  - Over all trees and receivers
  - 40-45% decrease

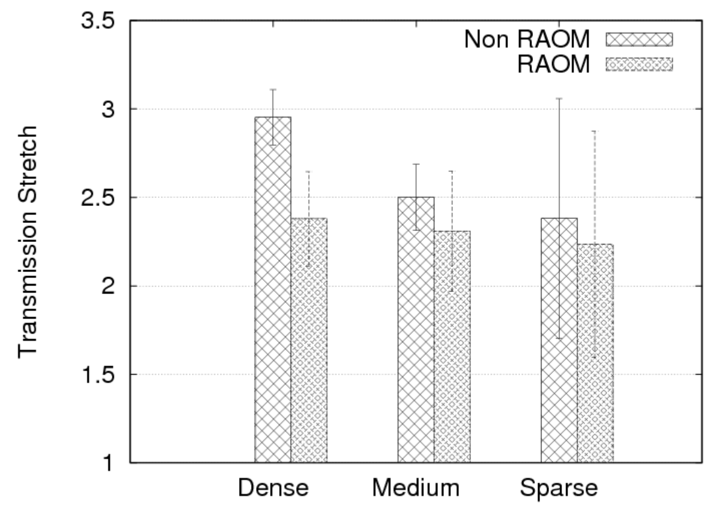
- ***Transmission stretch***

$$= \frac{\text{\# hop-by-hop transmissions}}{\text{IP multicast \# transmissions}}$$

  - Over all trees
  - 7-19% decrease

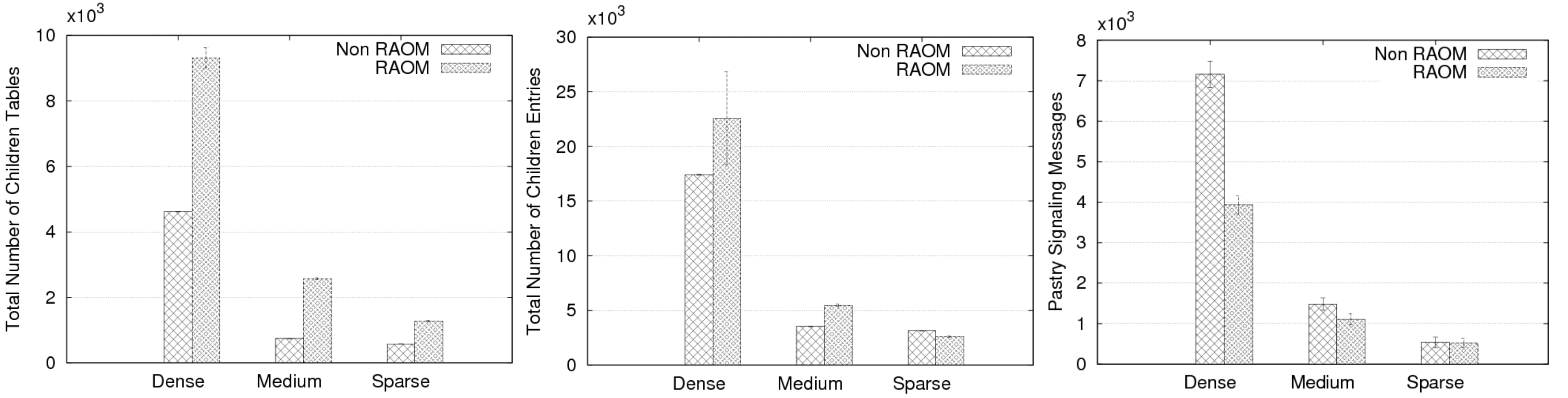
#Routers	*#End Hosts*		
1400	4000 (Dense)	1000 (Medium)	500 (Sparse)





RAOM Trees properties

Control plane overhead

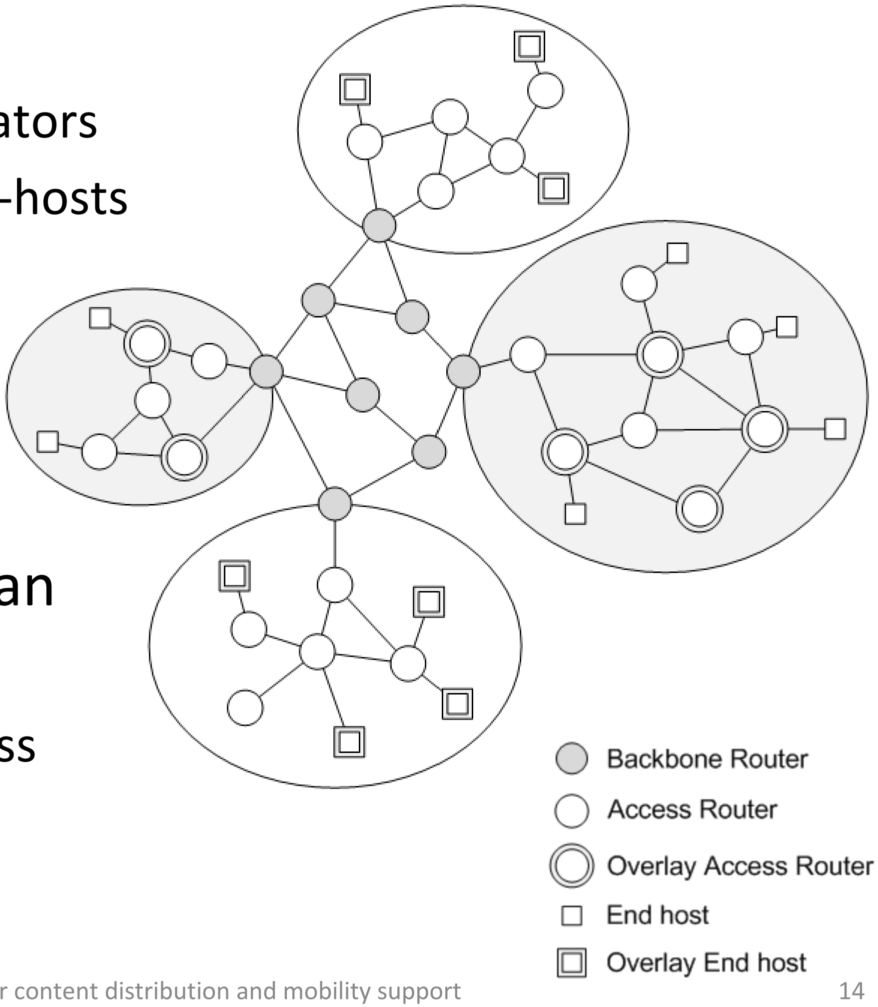


- *Node stress*
  - Forwarding load
    - Including proxy entries
  - Number of children tables
    - Number of groups served
  - Number of children entries
    - Number of nodes served
  - Overhead increase
    - Proxy routers aggregate overhead from multiple end-nodes
- *DHT signaling*
  - 4-45% decrease
  - Exactly due to the fewer overlay nodes

#Routers	*#End Hosts*		
1400	4000 (Dense)	1000 (Medium)	500 (Sparse)

Incremental deployment

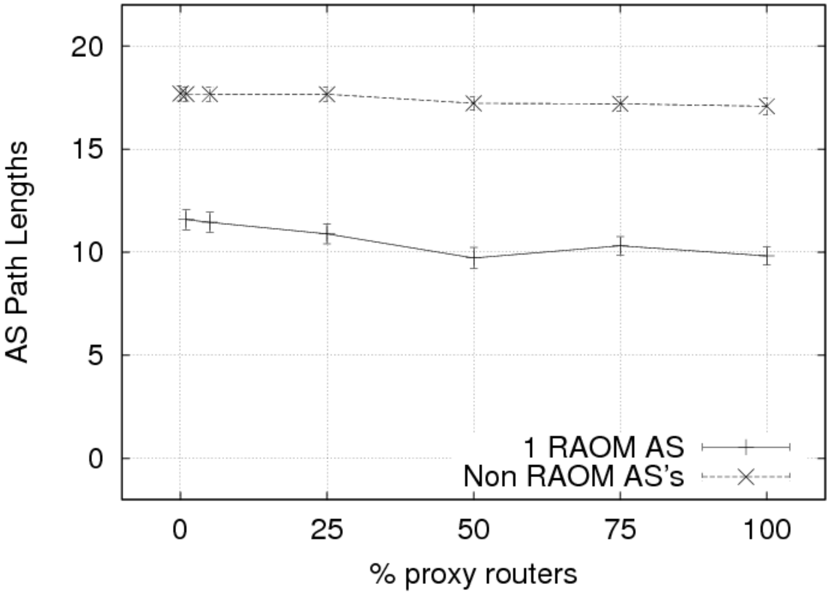
- Progressive transition to a RAOM-enabled inter-network
- **Inter-domain** dimension
  - Gradual adoption by network operators
  - Non-adopting ISPs: reliance on end-hosts
- **Intra-domain** dimension
  - Gradual deployment inside access networks
    - Progressive deployment of OARs
- Investigating performance from an ISP point of view:
  - How would this evolutionary process affect per ISP performance?

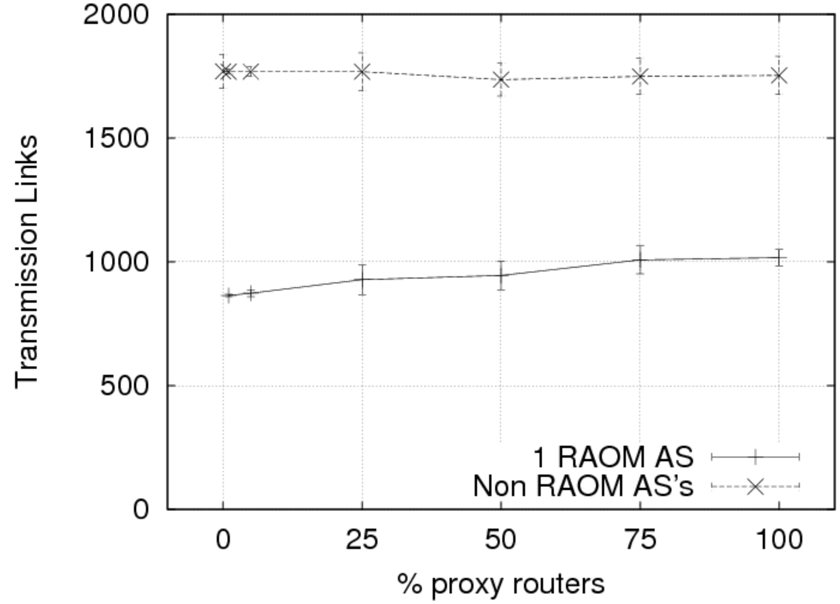


Incremental deployment

RAOM initialization

- **Q:** Are there performance related **incentives** for operators to first adopt RAOM?
- **A: Yes**. Even if there is only 1 RAOM enabled AS it will have significant advantages in terms of path length and transmission stress.



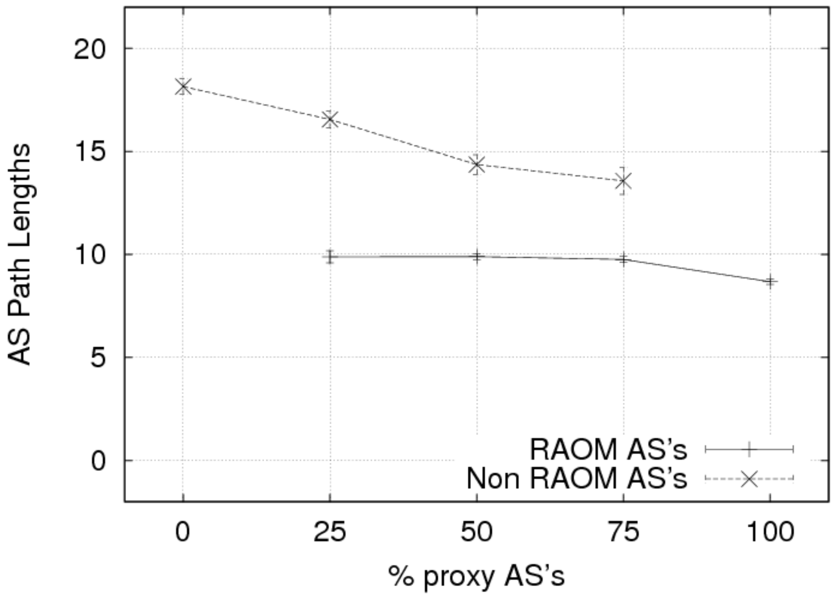


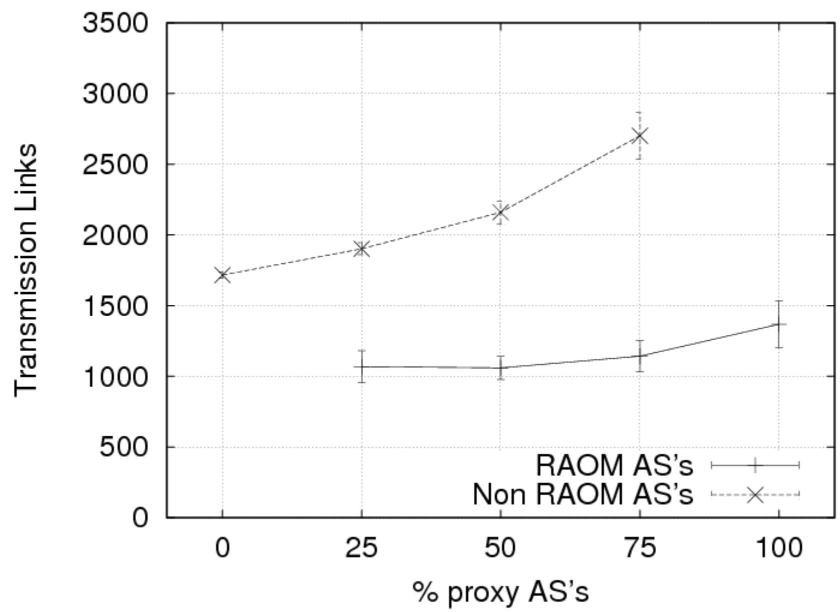
- 1 RAOM enabled AS vs. Non RAOM enabled AS's
  - Path length: ↓ 41%
  - Transmission Load: ↓ 44%

Incremental Deployment

Interdomain incremental deployment
(50% Intradomain deployment density)

- **Q:** How would progressive adoption affect performance?
- **A:** After initial deployment, RAOM supporting network operators have a consistent advantage over non RAOM supporting ones.



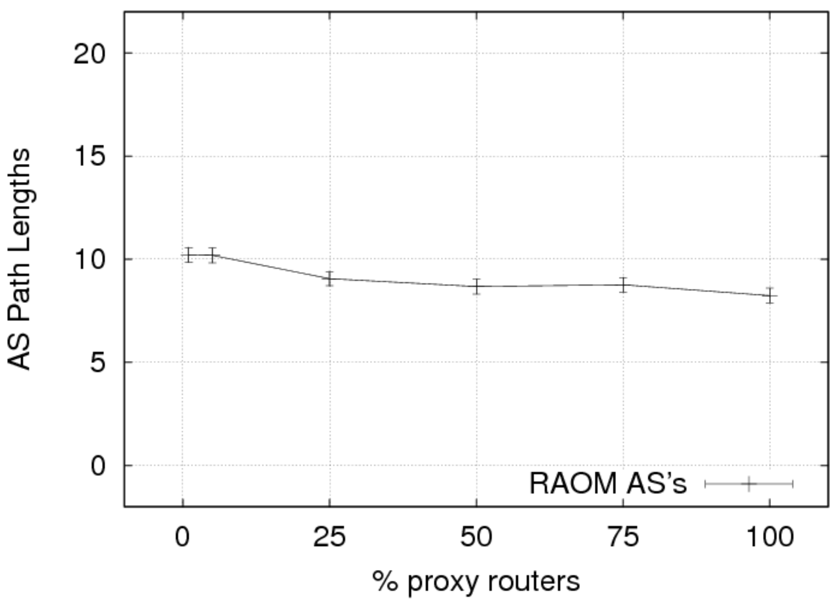


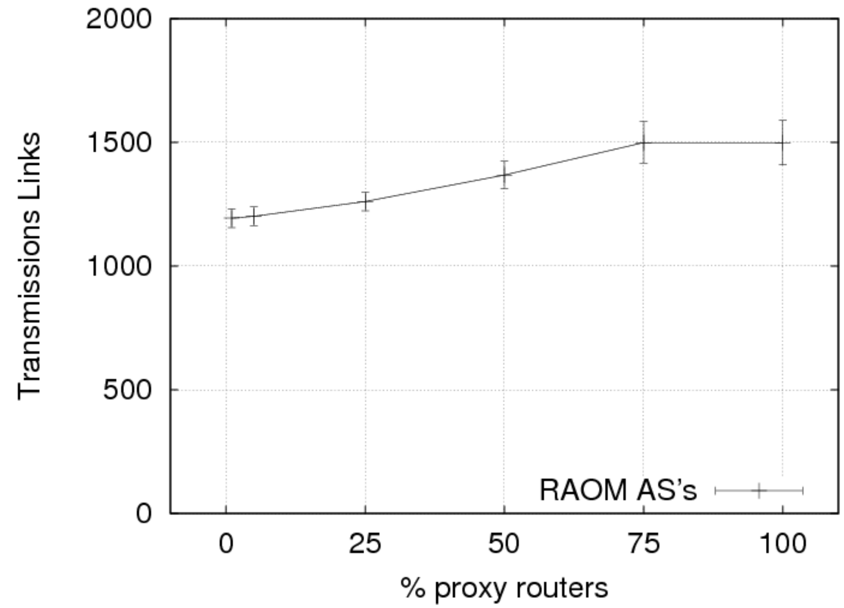
- RAOM enabled AS's vs. Non RAOM enabled AS's
  - Path length: ↓ 28-40%
  - Transmission Load: ↓ 43-57%

Incremental Deployment

Intradomain incremental deployment
(100% Interdomain deployment density)

- **Q:** Would increased OAR density improve performance?
- **A:** Sparse deployments enough to reap the benefits





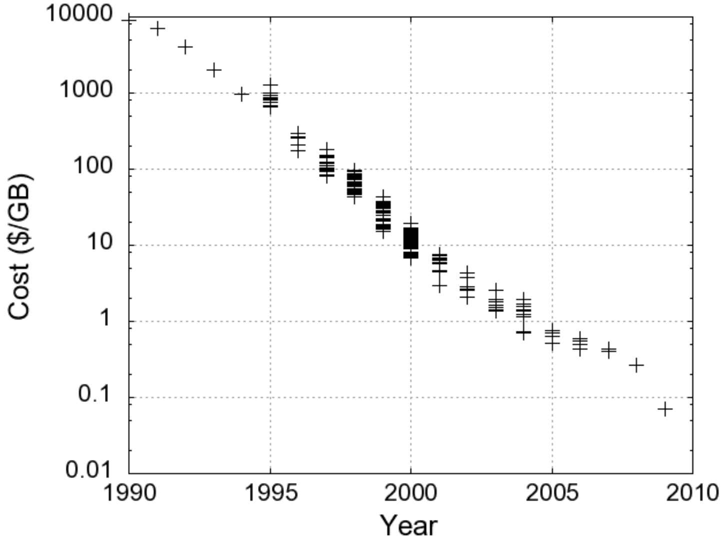
- Gradual deployment of OARs
  - Path length: marginal improvement with greater density
  - Transmission load: denser deployments incur higher branching

MultiCache

MULTICAST & CACHING

Caching

- Taking advantage of **low storage costs**
- **Target**: exchange traffic with storage
- Addressing asynchronous requests
- **Benefits**
    - ***User:***
      Fast delivery
    - ***Network operator:***
      Localization of traffic
    - ***Content provider:***
      Reduced content provision load

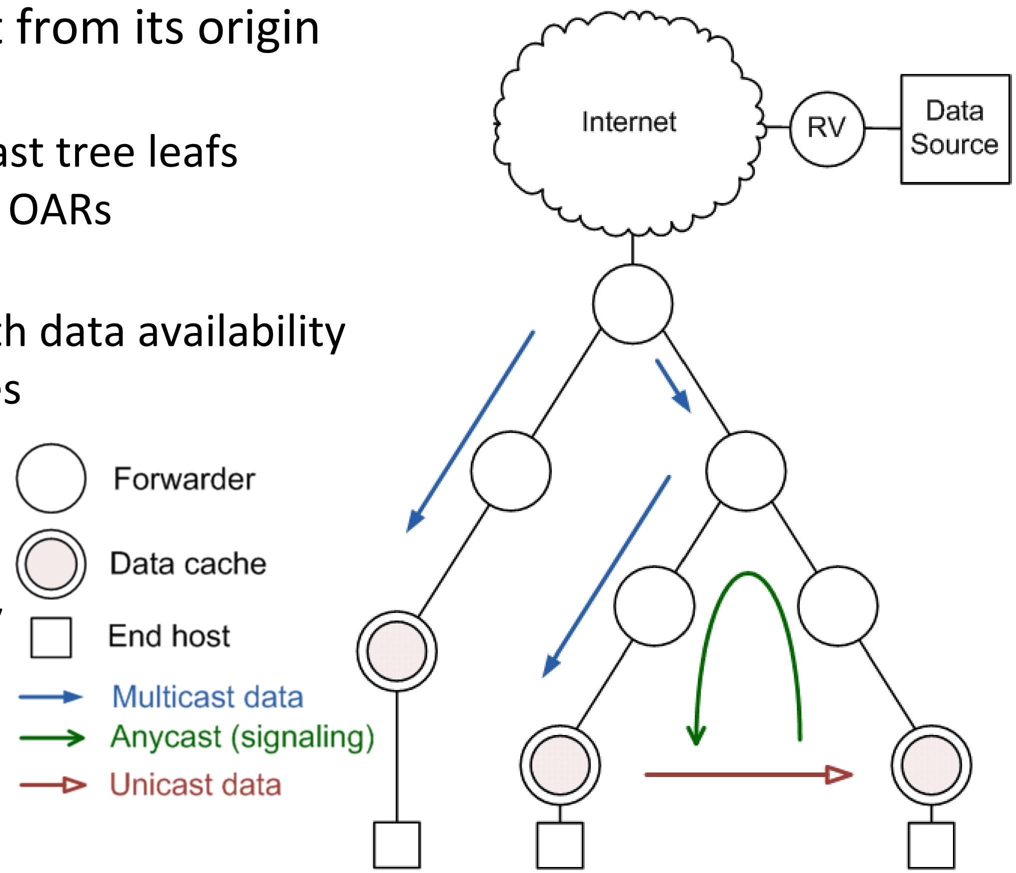


Storage cost evolution [1]

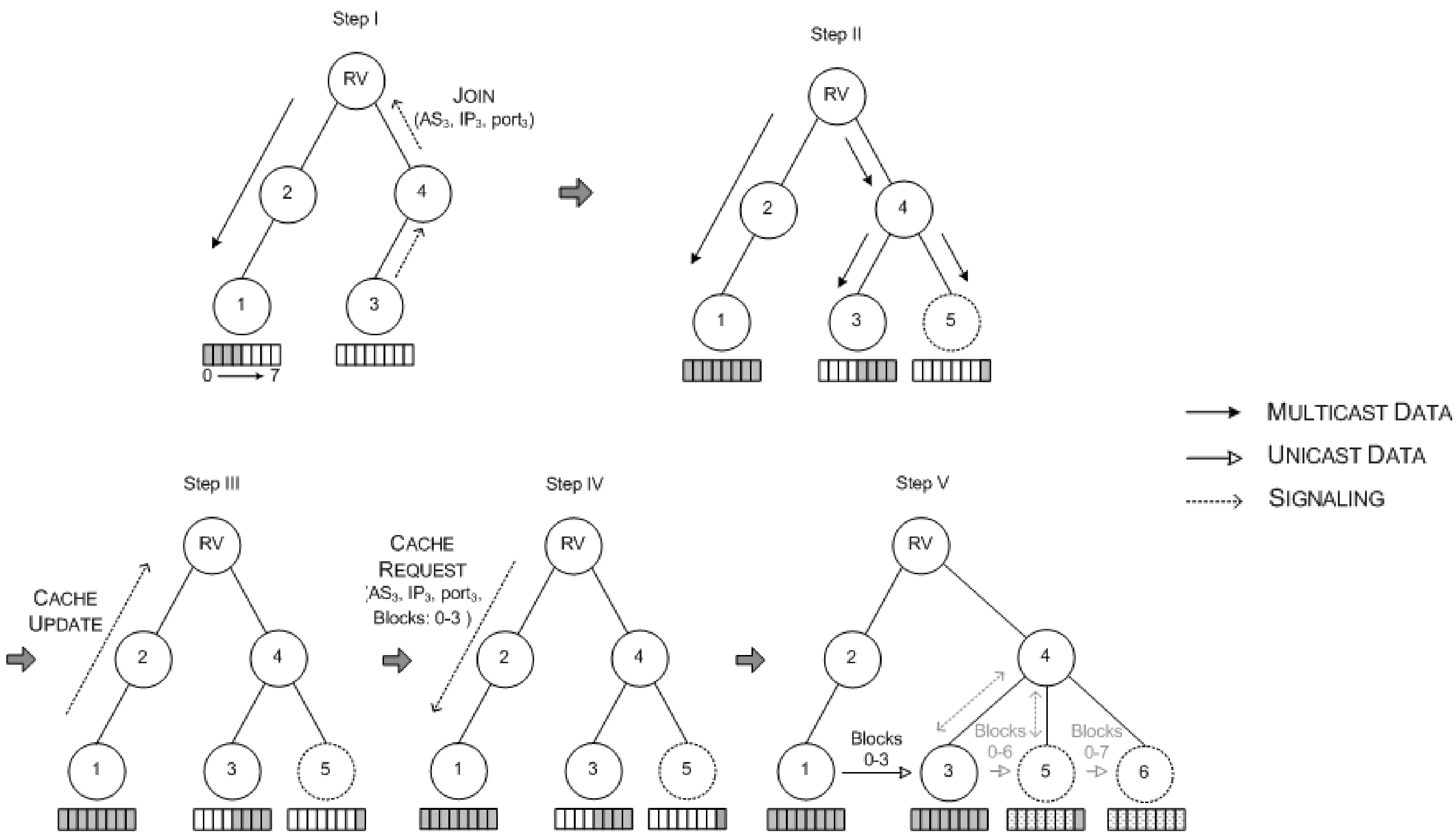
[1] Historical notes about the cost of hard drive storage space, http://ns1758.ca/winch/winchest.html

Functionality Overview

- Focusing on **content distribution**
- Overlay **multicast** brings content from its origin
- **Caching**
  - Data @ proxy OARs, i.e., multicast tree leafs
  - Forwarding state @ Forwarding OARs
- **Anycasting** cache requests
  - Correlating forwarding state with data availability
  - **Localizing traffic** inside sub-trees
    - Route Convergence property
- **Unicasting** cached data
  - Avoiding stretched paths …
  - Creating additional cached copy
- **Content fragmentation**
  - Piece level
    - Parallelizing transfers
    - Enabling partial caching
  - Block level
    - Facilitating cache provision…



MultiCache: example



Cache replacement: policies

- Common policies
  - Least Recently Used (LRU)
    - Evicting the LRU fragment of the LRU file
- MultiCache specific policies
  - Most Recently Used - Intra Domain (MRU-Intra)
    - Evicting the fragment most recently delivered to an OAR *k* of the same domain
    - Increased probability of the fragment not evicted by OAR *k*
  - Most Frequently Used - Intra Domain (MFU-Intra)
    - Evicting the fragment most frequently delivered to OARs of the same domain
    - Increased probability for an alternative caching location to exist
  - **Enforced on fragments**
    - Fragments not associated with files
    - No control signaling and state overhead

Performance evaluation

Workload and metrics

- Simulation based evaluation
- GT-ITM topologies
- **BitTorrent-like workload**
  - Mandelbrot-Zipf file popularity [1]
  - Exponential decay arrival process for file requests [2]
  - Fixed *file* arrival rate [2]
  - Trace sampled file sizes
- **Metrics**
  - Cache hit ratio (CHR) (%)
  - Intra-domain cache hit ratio (CHR - Intra) (%)
  - Distance to block source (hop count)
  - Egress Interdomain Traffic (MB)
  - Intradomain Link Stress
  - Download Time (sec)

[1] M. Hefeeda and O. Saleh, "Traffic modeling and proportional partial caching for peer-to-peer systems," *IEEE/ACM Transactions on Networking*, vol. 16, no. 6, pp. 1447–1460, 2008.

[2] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, "A performance study of BitTorrent-like peer-to-peer systems," *IEEE JSAC*, vol. 25, no. 1, pp. 155–169, 2007.

Performance evaluation

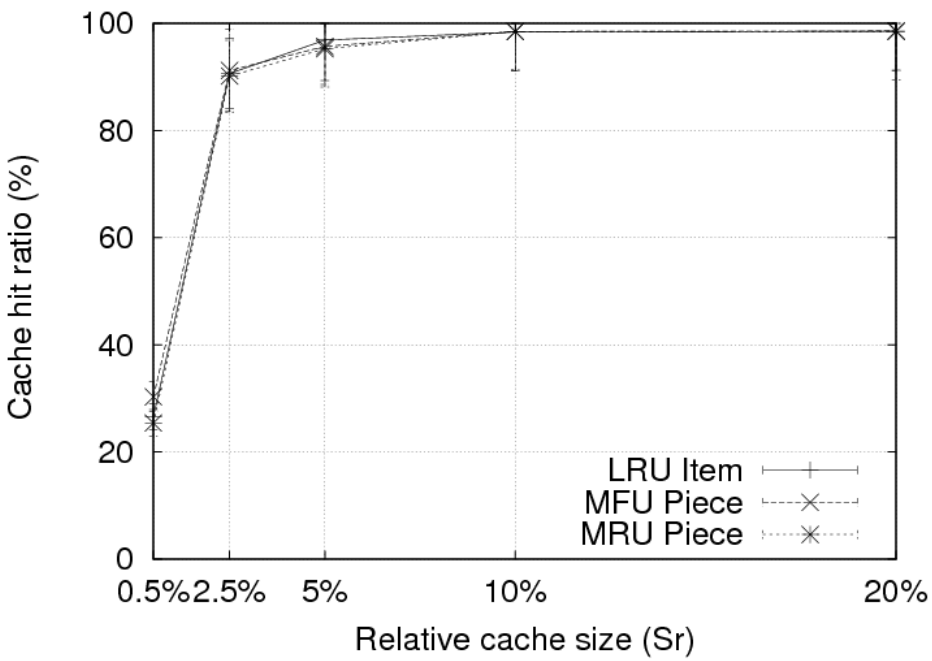
Important parameters

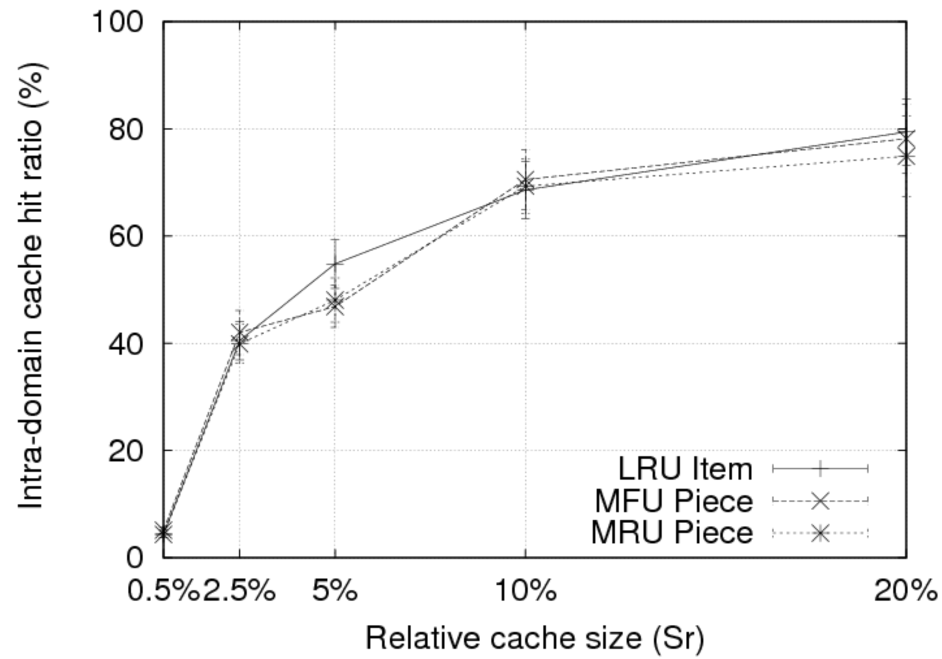
- **Relative Cache Size**, S_r
  - Percentage of "infinite cache size "
    I.e., minimum cache size to avoid replacements [1]
- **Deployment density**, $d \in (0,1]$
  - Fraction of access routers enhanced with overlay functionality
- **Host dispersion**
  - Affects traffic locality
    - Widely dispersed requests in favor of non localized traffic
  - Expressed with factor $l \in [0,1]$
    - $l = 0$, all nodes uniformly dispersed throughout the AS's
    - $l = 1$, all nodes inside a single AS

[1] L. Fan, P. Cao, J. Almeida, and A. Z. Broder, "Summary cache: a scalable wide-area web cache sharing protocol," *IEEE/ACM Transactions on Networking*, vol. 8, no. 3, pp. 281-293, 2000.

Results: cache size &replacement policies

l = 0, d = 0.25, 8 MB Piece, LRU vs. MFU–Intra vs. MRU–Intra

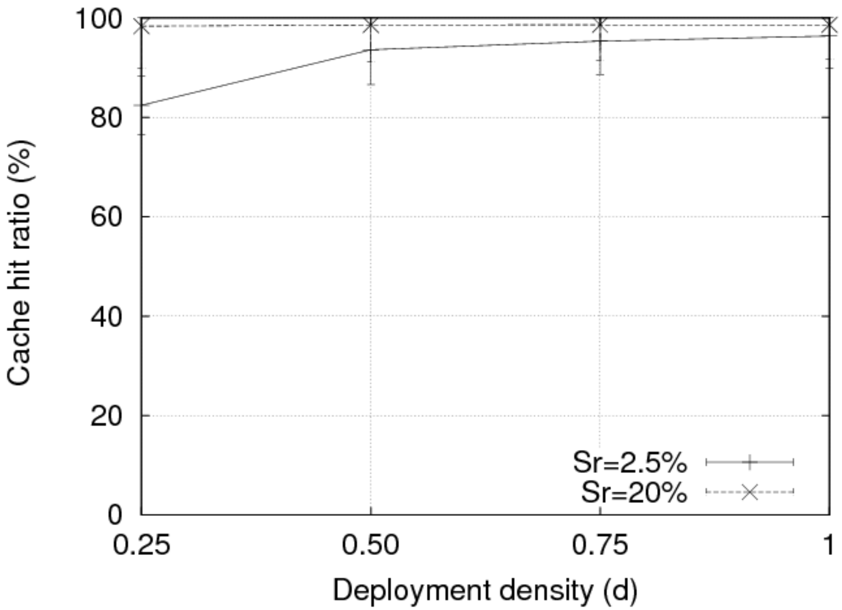


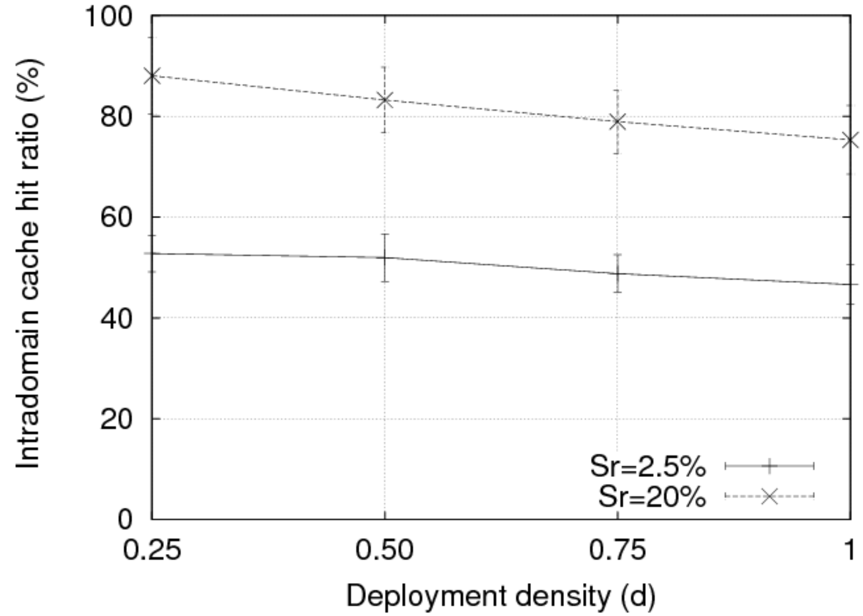


- High cache hit ratios
  - Taking advantage of cache multiplicity
  - Reducing overlay multicast
- Localizing traffic
  - High CHR–Intra values for $S_r \geq 2.5\%$
- Same performance with simpler, file-oblivious MFU/MRU-Intra policies

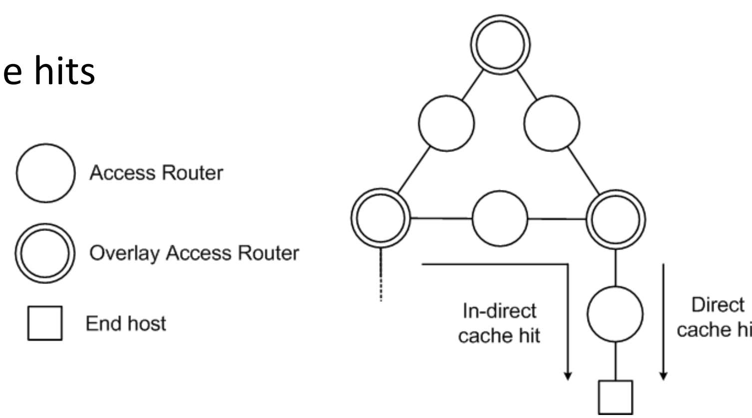
Results: incremental deployment

l = 0.5, 8 MB Piece, MFU-Intra



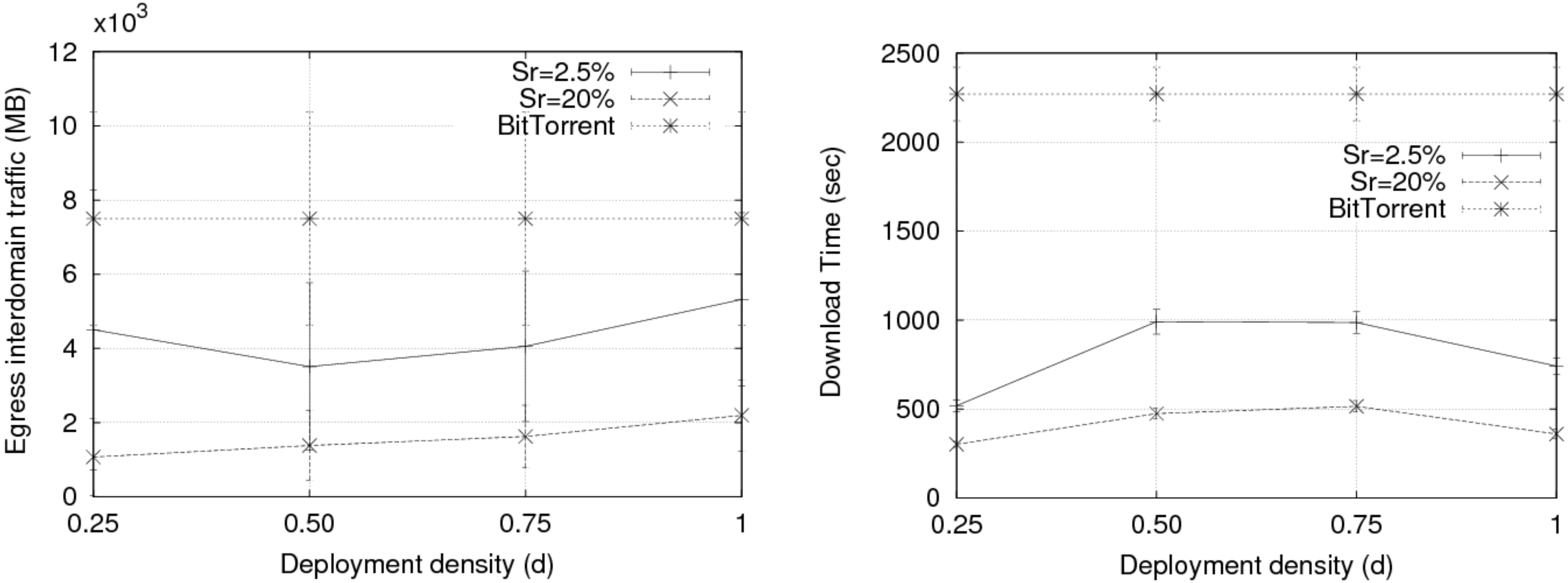


- Sparse deployments
  - Increased request aggregation, more *direct* cache hits
  - Low cache sizes trigger multicast
- Dense deployments
  - Increased total caching space
  - Denser trees, increased branching
  - Increasing interdomain cache provision



Results: MultiCache Vs. BitTorrent

l = 0.5, 8 MB Piece, MFU-Intra



- Egress inter-domain traffic reduced by 53,19% (S_r=2.5%, d=0.5)
  - Similar trend for intra-domain traffic
- Download time reduced by 56,35% (S_r=2.5%, d=0.5)

– Impact of caching

– No search for data

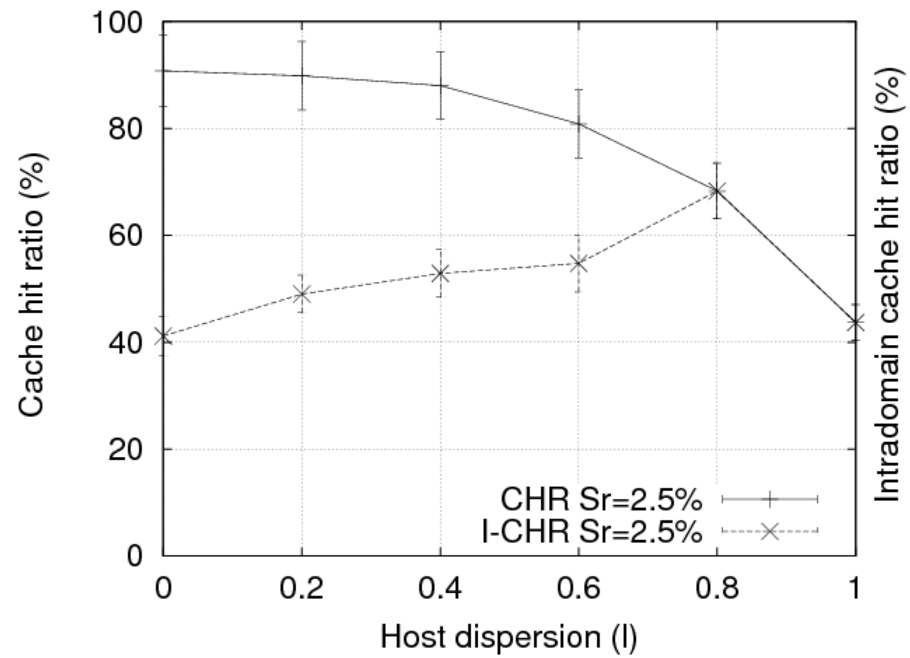
– No uplink bottlenecks

→ Sparse deployments of large size caches preferable

Results: host dispersion

d = 0.25, 8 MB Piece, MFU-Intra

- Lower dispersion imposes greater stress on caches
  - Reducing CHR
  - Multicast gradually takes over
- Taking advantage of localized request patterns
  - Increasing CHR-Intra
  - Up to the cache size limit



Scalability analysis (I)

- **Overlay routing**
  - Pastry DHT: $O(\log N)$
- **Multicast forwarding state size**
  - Subject to **overlay size** and the imposed **workload**
- Overlay size: $N_o = dN$
  - Assuming ubiquitous adoption
- Overlay shortest paths:

$$L_i = \binom{\log_{2^b} N_o}{i} p^i (1-p)^{\log_{2^b} N_o - i} \quad , \quad p = \frac{1}{2^b}$$ [1]

d	Intradomain deployment density
N	# access routers
C	Number of end host requests
2^b	Pastry numbering base

- Request aggregation
  - Average number of requests at i overlay hops : $C_i = min[CL_i, N_oL_i]$

[1] Loguinov, D., Kumar, A., Rai, V., and Ganesh, S. 2003. Graph-theoretic analysis of structured peer-to-peer systems: routing distances and fault resilience. In Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications (Karlsruhe, Germany, August 25 - 29, 2003). SIGCOMM '03. ACM, New York, NY, 395-406.

Scalability analysis (II)

- Multicast tree branch merging
  - *MaxMerge*: all merging branches stem from the same node
  - *MinMerge*: branches merge in couples
- Measuring the average number of forwarding links per tree level (J_i)

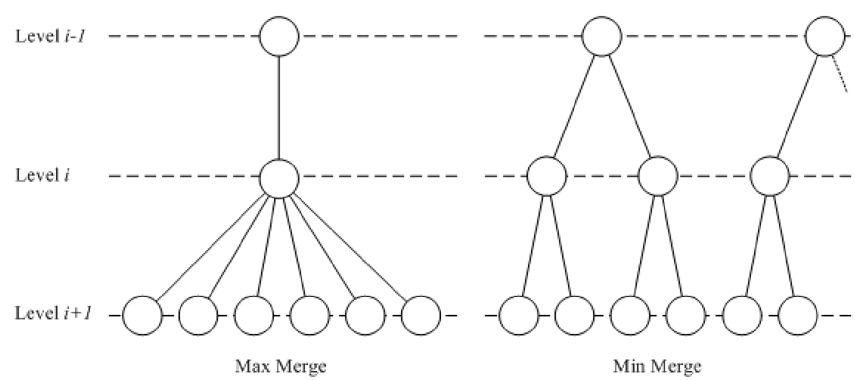
$$J_i = F_i(1 - \min[\frac{C}{N_o}, 1]) + C_{i+1}$$

$$F_i = \begin{cases} (1 - S_{i+1})J_{i+1} + 1 & , \ MaxMerge \text{ case} \\ (1 - S_{i+1})J_{i+1} + \frac{S_{i+1}J_{i+1}}{2} & , \ MinMerge \text{ case} \end{cases}$$

where: $S_i = 1 - (1 - \frac{1}{N_o \cdot L_i})^{J_i}$

- Forwarding state uniformly distributed

$$K_i^j = \begin{cases} \frac{J_i^j}{J_{i-1}^j} & , i > 0 \\ J_i^j & , i = 0 \end{cases}$$



Scalability analysis (III)

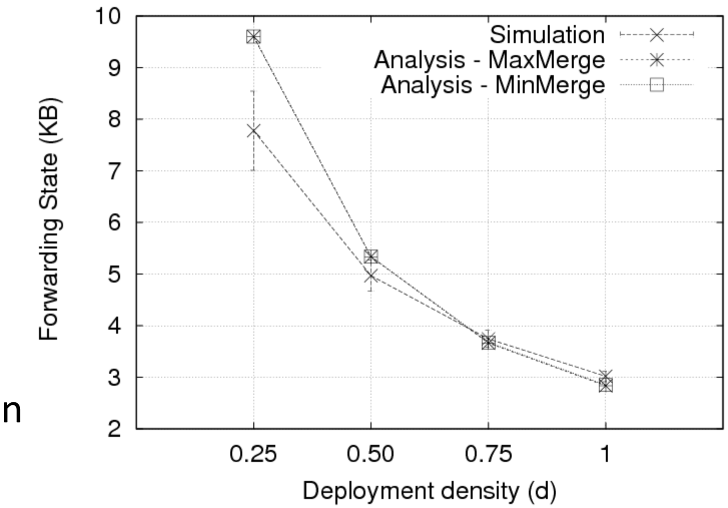
- Workload
  - M items, Mandelbrot-Zipf popularity

$$P(j) = \frac{K}{(j+\rho)^\alpha}, j \in [1, M] \quad , \quad K = 1/(\sum_1^M (i+\rho)^\alpha)$$

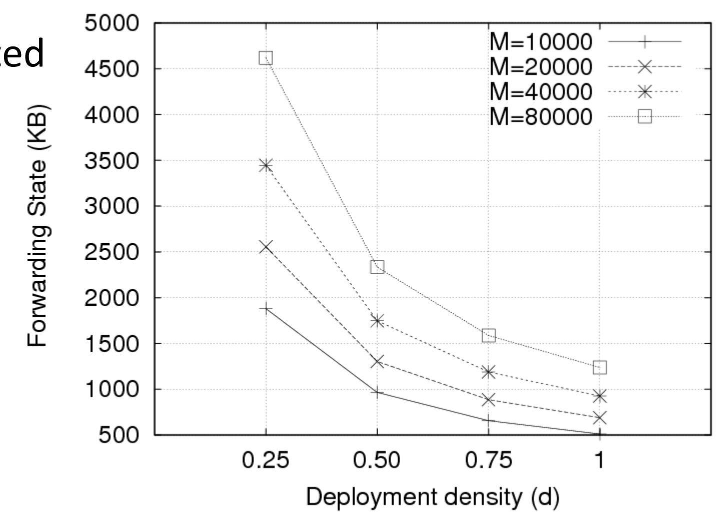
  - Average number of end- host requests per item k *normalized* to the size (U) of the end host population

$$C^k = \frac{P(k)}{P(0)} \cdot U$$

  - Each item of size (S) fragmented to pieces of size (P)
  - The total workload (across all trees) equally distributed to participating OARs
- Memory requirements
  - Forwarding link memory footprint: 256 bit
    - 128 bit (Pastry ID) + 32 bit (IP)
      + 32 (AS Number) + 64 bit (*blockfield*)
  - **Less than 5MB (< 150K entries)** of forwarding state per OAR for the concurrent support of **3.25×10^6 multicast trees** providing content to **10^4 users**



(N= 2400, U= 1500, S=P, M = 500)



(N= 15000, U= 10000, S=656MB, P= 16MB)

Related work (I)

- Publish-Subscribe Internet Routing Paradigm (PSIRP) [SIGCOMM ’09]
  - Clean slate approach
  - Based on flat identifiers
  - Source routing using Bloom filters
  - ➢ Deployment concerns
  - ➢ Source routing not suitable for mobility
- Content Centric Networking (CCN) [CoNEXT ’09]
  - Flooding content requests
  - Reverse path state establishment
  - Hierarchical organization of information
  - IP compatible
  - ➢ Scalability concerns

Related work (II)

- Data Oriented Network Architecture (DONA) [SIGCOMM ’07]
  - *Routing on names*, flat identifiers
  - Name resolution allowing anycast, application-transparent caching
  - IP routing
  - ➢ Not focusing on:
    - Resource sharing
    - Mobility
- SplitStream [SOSP ’03]
  - Employing *regular* Scribe forests for streaming applications
    - Content stripes delivered through multiple trees
  - Targeting at disjoint delivery paths
  - ➢ End-host uplink bottlenecks not removed
  - ➢ Tree-reconfiguration mechanisms violate locality properties
  - ➢ No caching

Summary & Conclusions

- Proposed an overlay architecture for content distribution & mobility support
  - Based on the joint operation of multicast and caching
  - Head-to-head comparison against BitTorrent
  - ✓ Substantial reduction of inter-domain traffic (↓ 53%)
  - ✓ Substantial improvement of download times (↓ 56%)
- Investigated incremental deployment
  - Both at the inter-domain and the intra-domain level
  - ✓ Consistent benefits for network operators
  - ✓ Sparse deployments of high storage capacity OARs preferable
- Designed the Canonical version of Pastry , H-Pastry (*not shown*)
  - Considering both proximity metrics & routing preferences
  - ✓ Substantial reduction of stretch (↓ 60%), while constraining traffic within AS boundaries
- Proposed overlay multicast assisted mobility (*not shown*)
  - Localizing routing updates
  - ✓ Substantial reduction of packet loss during handoff (↓ 47%)

Future work

- Gaining finer control over inter-domain service provision
  - MultiCache over H-Pastry
  - Expressing routing policies with H-Pastry
- Broadening application space
  - E.g., streaming
- Considering alternative deployment scenarios
  - E.g., MultiCache-enabled set-top-boxes
- Mobility support
  - Location management
  - Interactive applications
  - Mobility prediction
  - Impact of caching
  - OMAM Vs. HMIPv6 (RFC 4140)

Publications - I

[Information-centric networking]

1. K. Katsaros, G. Xylomenos, and G.C. Polyzos, "On the incremental deployment of overlay information centric networks," Proc. ICT Future Network and Mobile Summit, Florence, Italy, June 2010.
2. A. Zahemsky, D. Lagutin, B. Gajic, C. Reason, D. Trossen, C. E. Rothenberg and K. Katsaros, “Experimentally-driven research in Publish/Subscribe Information-centric Inter-Networking,” Proc. TridentCom, Berlin, Germany, May 2010.
3. K. Katsaros, G. Xylomenos, and G.C. Polyzos, “A hybrid overlay multicast and caching scheme for information-centric networking,” Proc. 13th IEEE Global Internet Symposium, San Diego, CA, USA, March 2010.
4. K. Katsaros, G. Xylomenos, and G.C. Polyzos, “MultiCache: an incrementally deployable overlay architecture for information-centric networking,” Proc. IEEE INFOCOM, WiP, San Diego, CA, USA, March 2010.
5. K. Katsaros, V.P. Kemerlis, Ch. Stais, G. Xylomenos, "A BitTorrent Module for the OMNeT++ Simulator," Proc. 17th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), London, UK, September 2009.
6. K. Katsaros, N. Bartsotas and G. Xylomenos, "Router assisted overlay multicast," Proc. 5th Euro-NGI Conference on Next Generation Internet Networks (NGI 2009), Aveiro, Portugal, July 2009.
7. K. Katsaros, N. Fotiou, G.C. Polyzos, G. Xylomenos, "Overlay Multicast Assisted Mobility for Future Publish/Subscribe Networks," Proc. ICT Mobile Summit, Santander, Spain, June 2009.
8. K. Katsaros, N. Fotiou, G.C. Polyzos, G. Xylomenos, "Supporting Mobile Streaming Services in Future Publish/Subscribe Networks," Proc. Wireless Telecommunications Symposium (WTS 2009), Prague, Czech Republic, April 2009.
9. G. Xylomenos, K. Katsaros and V.P. Kemerlis, "Peer Assisted Content Distribution over Router Assisted Overlay Multicast," 1st Euro-NF workshop on Future Internet Architecture - New Trends in Service and Networking Architectures, Paris, France, November 2008.

Publications - II

[Cognitive radio]

10. K.V. Katsaros, P.A. Frangoudis, G.C. Polyzos and G. Karlsson, "Design Challenges of an Open Spectrum Access Scheme," Proc. First IEEE International Workshop on Cognitive Radio and Networks (CRNETS 2008 - In conjunction with IEEE PIMRC 2008), Cannes, French Riviera, France, September 2008.
11. V.G. Douros, P.A. Frangoudis, K. Katsaros and G.C. Polyzos, "Power Control in WLANs for Optimization of Social Fairness," Proc. 12th Pan-Hellenic Conference on Informatics (PCI 2008), Samos, Greece, August 2008.

[Mobile grid]

12. K. Katsaros and G.C. Polyzos, "Evaluation of scheduling policies in a Mobile Grid architecture," Proc. International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2008), Edinburgh, UK, June 2008.
13. K. Katsaros and G.C. Polyzos, "Mobility-aware Grid Computing," *The Encyclopedia of Information Science and Technology*, Second Edition, Mehdi Khosrow-Pour, ed., IGI Group, 2008.
14. K. Katsaros and G.C. Polyzos, "Towards the Realization of a Mobile Grid," Proc. 3rd ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT), Student Workshop, New York, USA, December 2007.
15. K. Katsaros and G.C. Polyzos, "Optimizing Operation of a Hierarchical Campus-wide Mobile Grid," Proc. 18th IEEE Annual International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Athens, Greece, September 2007.
16. K. Katsaros and G.C. Polyzos, "Optimizing Operation of a Hierarchical Campus-wide Mobile Grid for Intermittent Wireless Connectivity," Proc. 15th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN 2007), Princeton, NJ, USA, June 2007.

Publications - III

[Multicast]

17. G.C. Polyzos and K. Katsaros, “Multicast,” *The Encyclopedia of Electrical and Electronics Engineering*, J.G. Webster, ed., John Wiley & Sons, 2007.

[Multicast in 3G Networks/MBMS]

18. G. Xylomenos, K. Katsaros, and V. Tsakanikas, "Support of Multiple Content Variants in the Multimedia Broadcast/Multicast Service," *International Journal of Communication Systems*, Wiley (to appear).
19. G. Xylomenos and K. Katsaros, “A Multiple Content Variant Extension of the Multimedia Broadcast/Multicast Service,” Proc. IST Mobile Summit, Mykonos, Greece, June 2006.
20. G. Xylomenos and K. Katsaros, “Supporting Multiple Content Variants in the Multimedia Broadcast/Multicast Service”, Proc. World Telecommunications Congress (WTC), Budapest, Hungary, May 2006.
21. G. Xylomenos, K. Katsaros and G.C. Polyzos, “Analysis of a Multiple Content Variant Extension of the Multimedia Broadcast/Multicast Service”, Proc. European Wireless, Athens, Greece, April 2006.

Thank you

Konstantinos Katsaros

Mobile Multimedia Laboratory
Department of Computer Science
Athens University of Economics and Business
ntinos@aueb.gr