Improving Network QoS Through Preference-Driven Web Caching



Yee Man Chan
Jonathan Womer
Jeff MacKie-Mason
Sugih Jamin
University of Michigan and Beyond

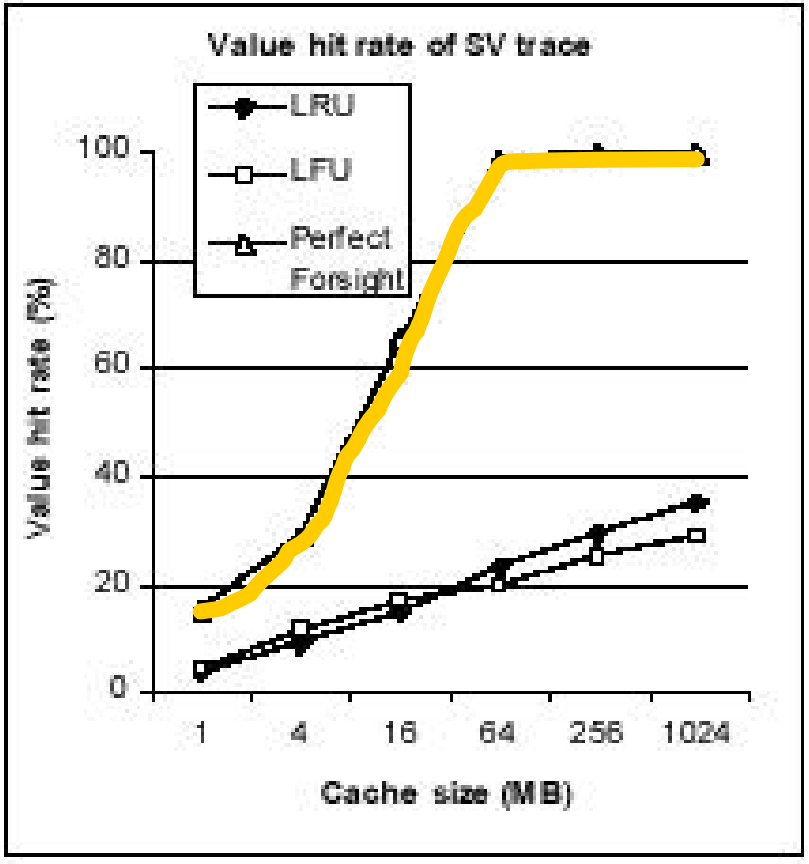
The Problem

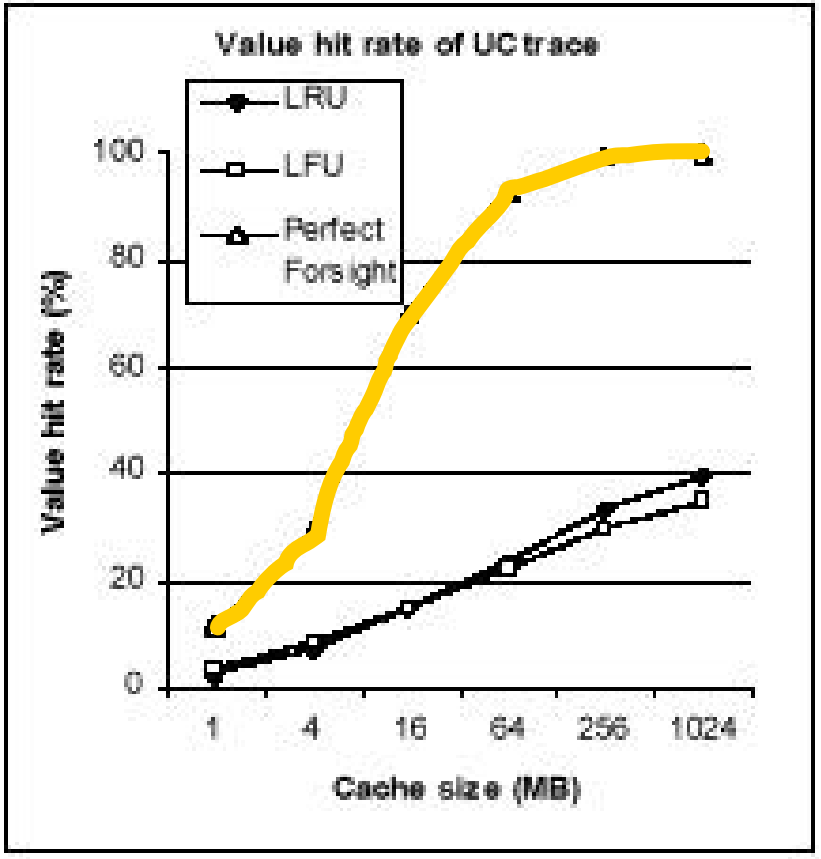
- Rapid growth of convergent **multi-service** networks
- Highly variable, **unreliable**, stochastic service quality
- Increasing demand for multiple qualities-of-service (**QoS**)

Our Idea: QoS via distributed file services

- QoS for network transport is one (much-studied!) possibility
- **Dynamic distributed file storage** to reduce transport loads is another
- How well might a market-based file caching system work?
- What are the **incentives problems**?

Potential for market-based Web caching





© 1999 Chan, Womer, MacKie-Mason, Jamin

Implementation Approach

- Users send **messages** to cache "manager"
  - requests, usage and forecast reports, bids
- Cache behaviors function of messages:
  - **Replacement**
  - **Service**
- Design Service and Replacement
  - to get useful messages (**revelation problem**)
  - to deliver most valued objects (**value-sensitivity**)

© 1999 Chan, Womer, MacKie-Mason, Jamin

Implementation: Market for disk space

- Auction off **write access** (give universal read access)
- One-shot 20 min. clearing, agents bid to control weights on document placement in cache
  - **incentive compatibility possible** with, e.g., 2nd price auction
- NOTE: Provides **deterministic** QoS

Implementation: Market for Objects (A-swLFU)

- Simple, **easy-to-implement** extension to standard LFU caching
  - weight frequencies by server-declared value per byte delivered
  - paying for priority of **keeping object in the cache**
- Takes advantage of strong statistical regularities in Web request streams
- NOTE: Provides **stochastic** QoS

Data: Disk space market

- 3 actual L3 Web cache request streams from NLANR project
- Palo Alto, Silicon Valley and University of Illinois (UIUC), 15-28 Aug 98

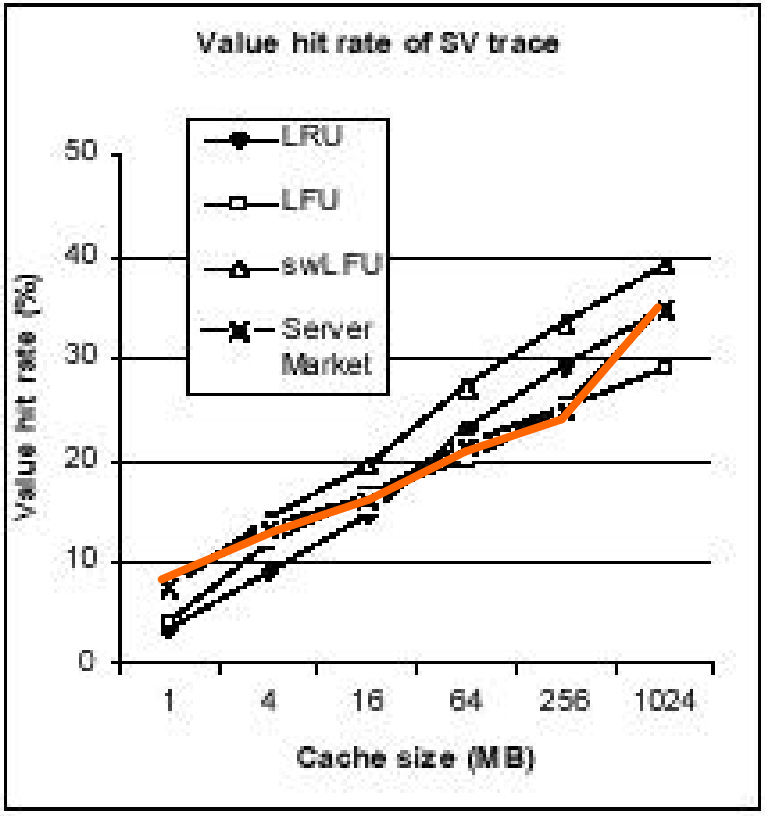
	Palo Alto	Silicon Valley	Univ Ill. Urbana-Cha
# servers	114,381	124,698	105,710
# URLs	3,412,105	3,744,274	2,884,598
# requests	7,011,622	7,897,659	5,568,112
Bytes requested	131,665,275,664	161,620,444,331	127,346,723,989
Value requested	264,025,996,908,451	319,466,493,484,667	264,106,487,028,562
Infinite Cache			
size (bytes)	60,037,623,775	66,976,225,688	51,825,514,504
hit rate	51.3364	52.5901	48.1943
byte HR	54.4013	58.5596	59.3036
value HR	48.5422	57.467	56.5745

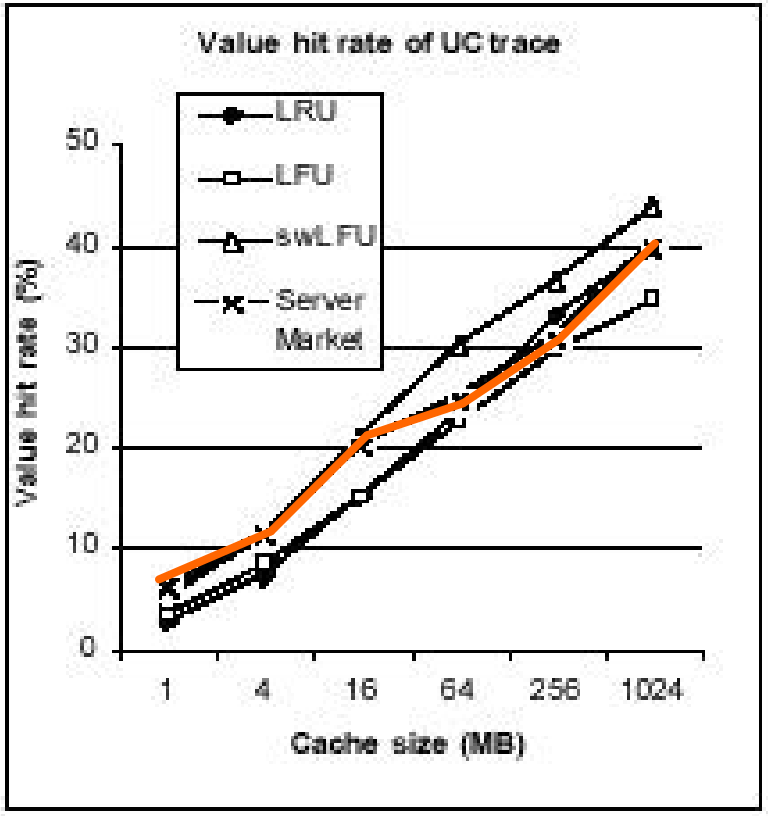
Data: Object market

- 6 Web cache request streams from NLANR, March 1999
- Example: Champaign-Urbana

	Univ Ill. Urbana-Champaign	
# requests	**26,024,662**	
# servers	**250,484**	
# documents	**7,615,462**	
bytes requested	**397,548,684,913**	
mean # requests	**3.4**	
Infinite Cache		
unique bytes	**150,119,984,279**	**(150 GB)**
max. hit rate	**70.7%**	
max. byte HR	**64.4%**	

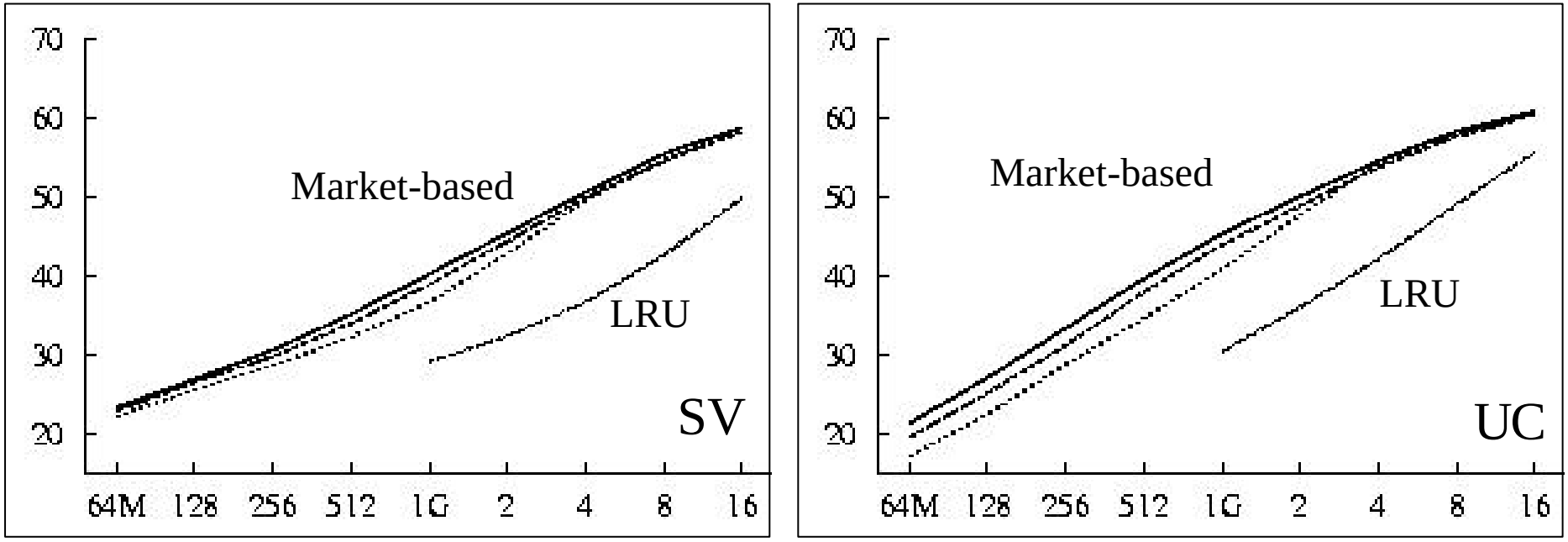
Performance of disk space market





For small cache sizes, can match A-swLFU algorithm of Kelly et al., with incentive compatibility

Performance of object market



Value hit rate as function of cache size for three market-based variants and LRU for comparison.

Performance summary

- Both markets provide **variable QoS**: higher valued file access obtains higher hit rate
- Also **higher aggregate user value** (only for small caches in disk space market)

Lessons Learned

- **Preference revelation:** Easier for disk space market
  - but forecasting problem horrible
  - and write access not easily correlated with value
- **Exponential distribution** of object requests **overwhelms value sensitivity** unless value range is about 10^5
- Big gains in single cache unlikely. Market for multiple networked caches? (Cf. Akamai)

References

- http://www-personal.umich.edu/~jmm/

- Chan, Womer, Jamin and MacKie-Mason, "The Case for Market-based Push Caching." *Proceedings of the Second International Conference on Telecommunications and Electronic Commerce,* Nashville, TN, November 1999.
- Kelly, Chan, Jamin and MacKie-Mason, "Biased Replacement Policies for Web Caches: Differential Quality-of-Service and Aggregate User Value," Proceedings of the Fourth International Web Caching Workshop, San Diego, CA, March 31-April 2, 1999 (refereed).
- Chan, Womer, MacKie-Mason and Jamin, "One Size Doesn't Fit All: Improving Network QoS Through Preference-driven Web Caching." Presented at Second Berlin Internet Economics Workshop, May 28-29, 1999.
- Kelley, Jamin and MacKie-Mason, "Variable QoS from Shared Web Caches: User-Centered Design and Value-Sensitive Replacement." Draft, June 1999. In submission to *IEEE Infocom 2000.*