CSC2231: Cooperative Caching

http://www.cs.toronto.edu/~stefan/courses/csc2231/05au

Stefan Saroiu

Department of Computer Science

University of Toronto

Administrivia

- **No lectures next week both Monday and Thursday**
- **Report reviews due on Thursday at noon**
- **Long lecture on Thursday in two weeks**
- **Most remaining lectures will have two papers assigned**
  - Read both; submit reviews to both

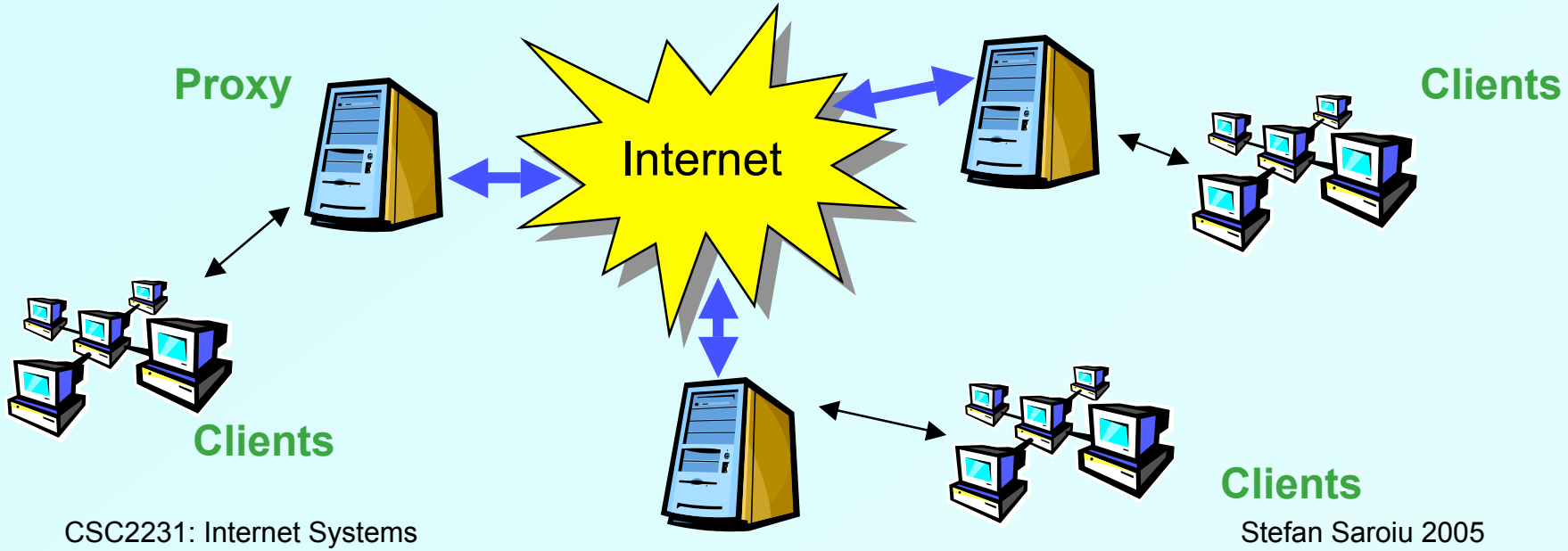
All slides today taken from a talk given by Alec Wolman

Cache Effectiveness

- **Previous work has shown that hit rate increases with population size**
- **A single proxy cache has practical limits**
  - Load, network topology, organizational constraints
- **One technique to scale the client population is to have proxy caches cooperate**

Cooperative Web Proxy Caching

- **Sharing and/or coordination of cache state among multiple Web proxy cache nodes**
- **Effectiveness of proxy cooperation depends on:**
  - ♦ **Inter-proxy communication distance**
  - ♦ **Size of client population served**
  - ♦ **Proxy utilization and load balance**



Cooperative Web Caching

- **How much benefit does cooperative caching provide in the Web environment?**

Previous Research

- **Cooperative proxy caching was a popular research topic:**

  [e.g. Chankhunthod et al. 96, Zhang et al. 97, Fan et al. 98, Krishnan et al. 98, Menaud et al. 98, Tewari et al. 98, Touch 98, Karger et al. 99 ...]
- **Focus was on highly scalable algorithms**
- **Some seek to scale to the entire Web**

Challenges

- **No real understanding of document sharing across diverse organizations**
- **Little analytic or empirical evaluation of these algorithms using realistic workloads for large-scale client populations**

- **Problem:**
  - Evaluating cooperative proxy caching requires multiple simultaneous traces of Web proxies, across a diverse set of organizations

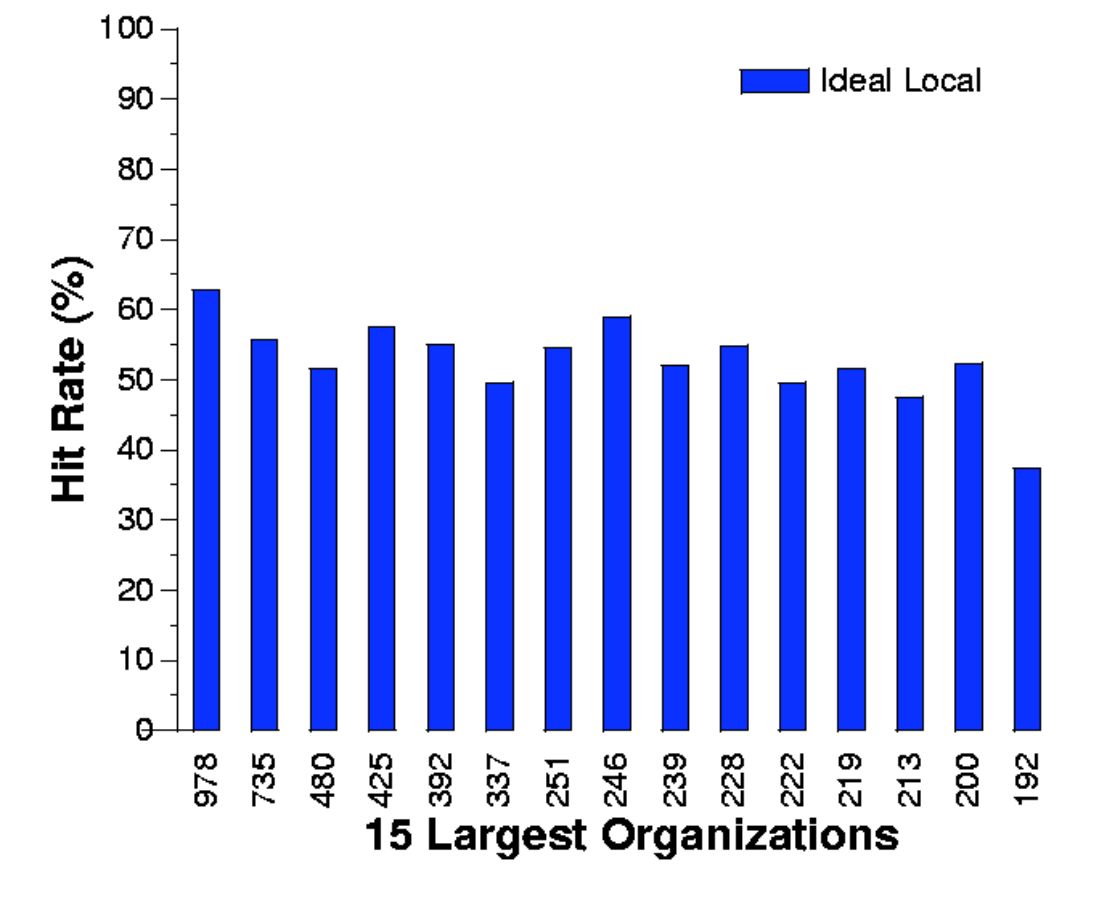
Cooperation Across Organizations

- **By considering each UW organization as an independent "company" with its own clients and its own proxy, we can empirically evaluate cooperative caching across diverse client populations**

- **Place a proxy cache in front of each organization. What is the benefit of cooperative caching among these 170 proxies?**

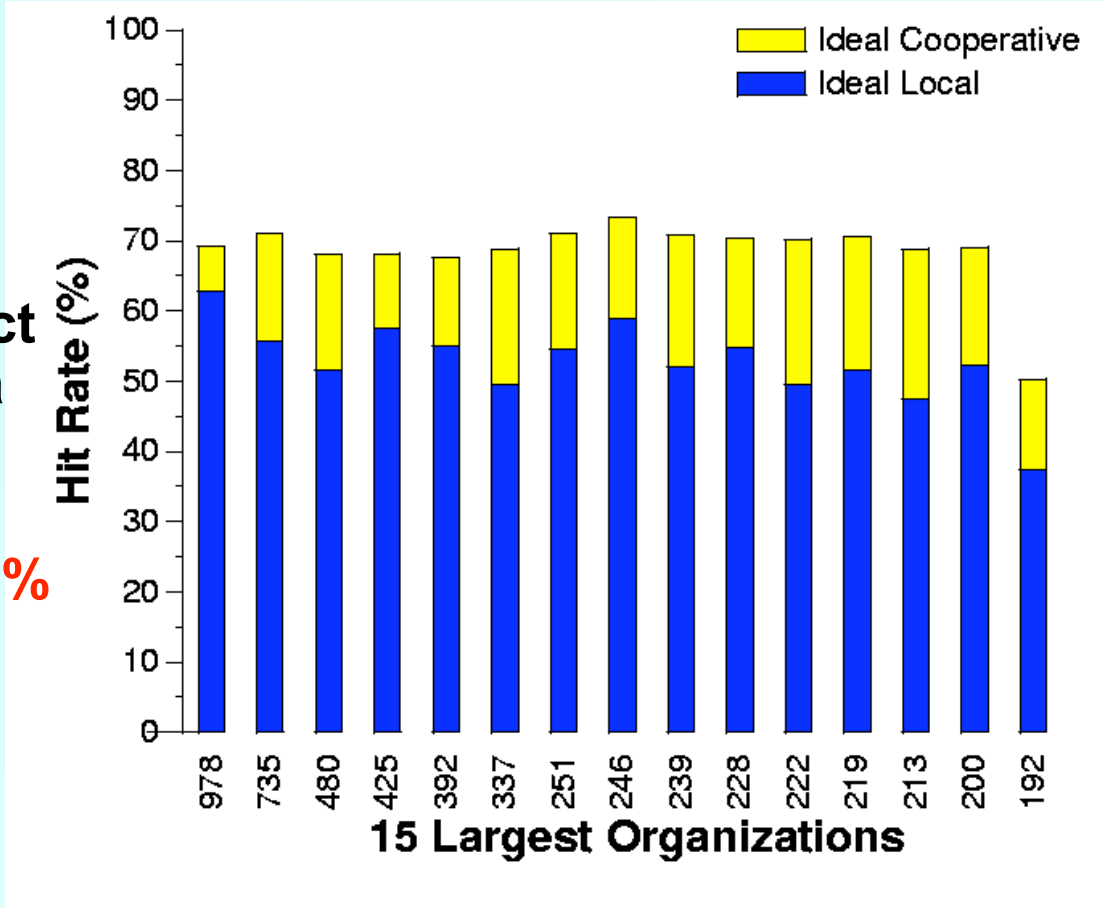
Ideal Hit Rates for UW proxies

- **Ideal hit rate - infinite storage, ignore cacheability, expirations**
- **Average ideal local hit rate: 43%**



Ideal Hit Rates for UW proxies

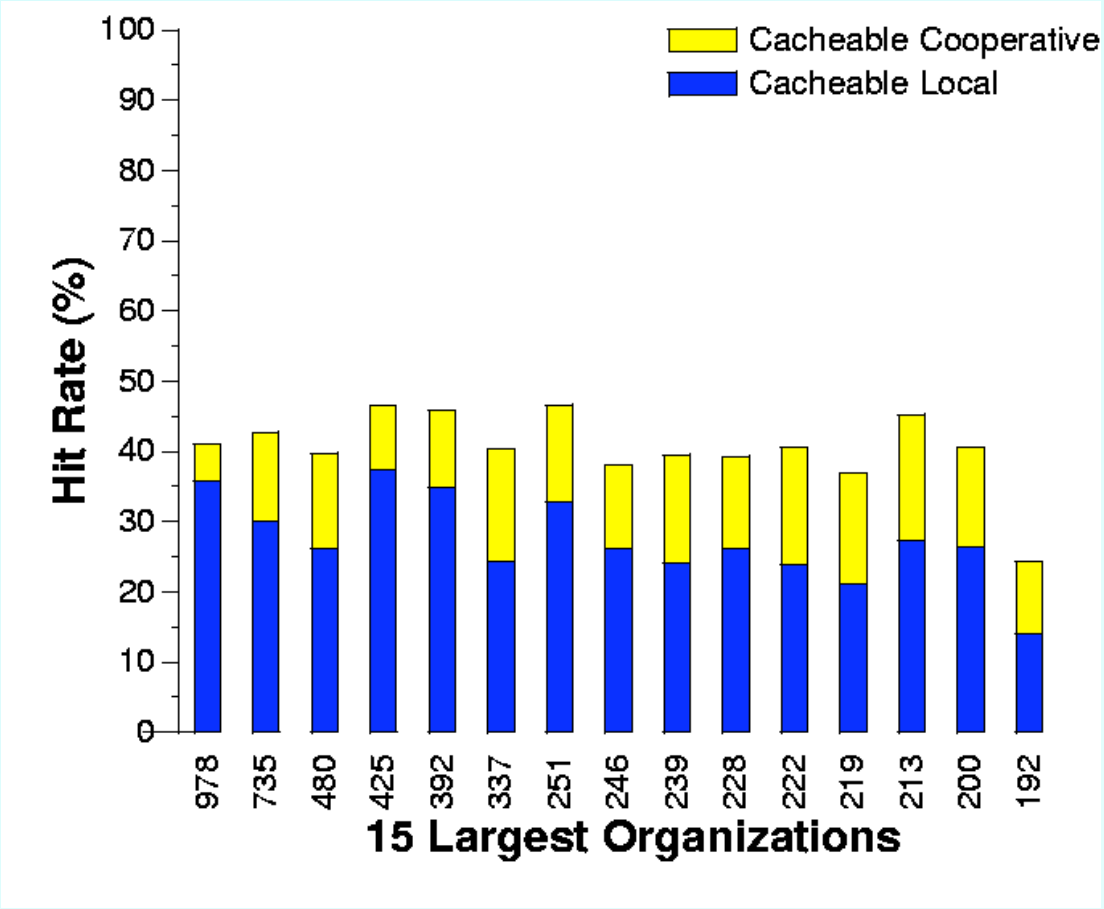
- **Ideal hit rate - infinite storage, ignore cacheability, expirations**
- **Average ideal local hit rate: 43%**
- **Explore benefits of perfect cooperation rather than a particular algorithm**
- **Average ideal hit rate increases from 43% to 69% with cooperative caching**



Cacheable Hit Rates for UW proxies

- **Cacheable hit rate - same as ideal, but doesn't ignore cacheability**
- **Cacheable hit rates are much lower than ideal (average is 20%)**

- **Average cacheable hit rate increases from 20% to 41% with (perfect) cooperative caching**

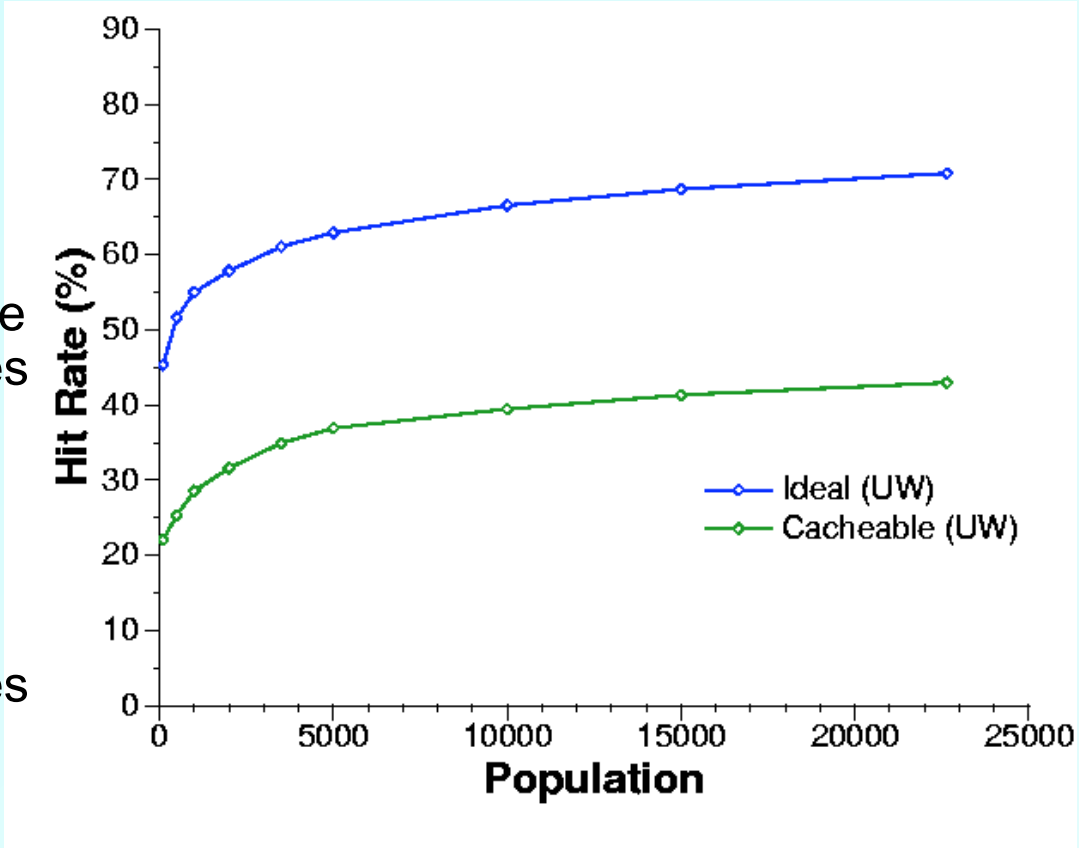


Scaling Cooperative Caching

- **Organizations of this size can benefit significantly from cooperative caching**
- **We don't need cooperative caching to handle the entire UW population size**
  - A single proxy (or small cluster) can handle this entire population!
  - No technical reason to use cooperative caching for this environment
  - In the real world, decisions of proxy placement are often political or geographical
- **How effective is cooperative caching at scales where a single cache will not work?**

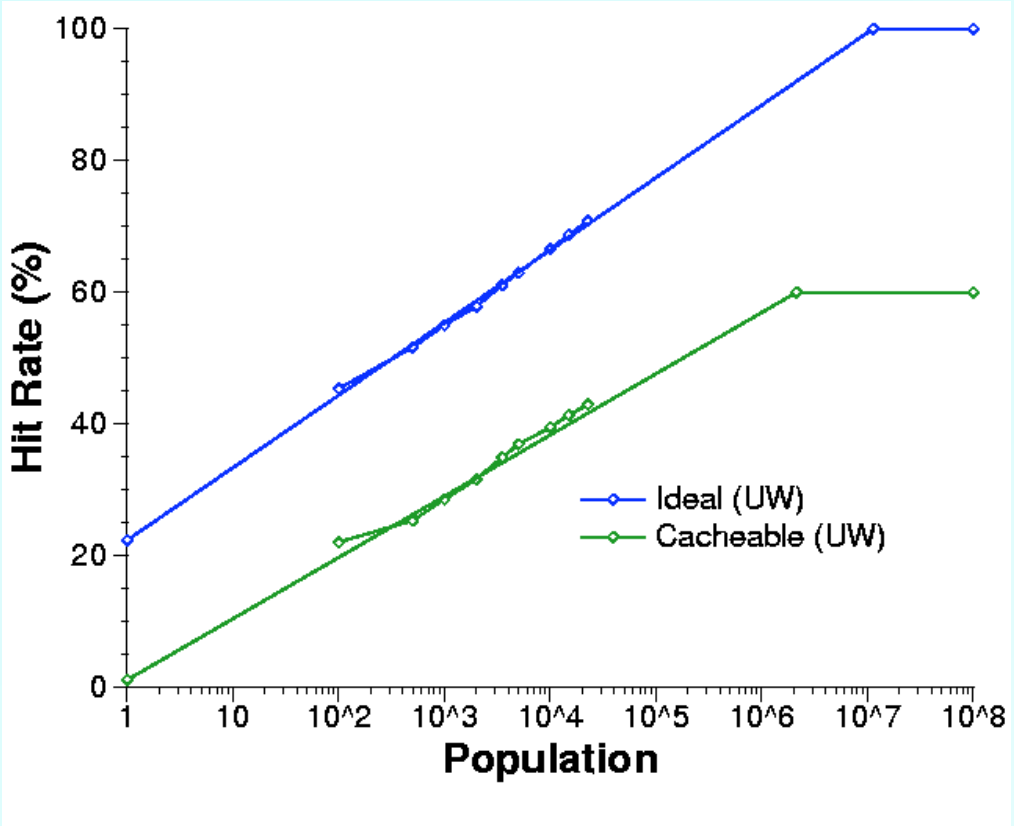
Hit Rate vs. Client Population

- **Curves similar to other studies**
  - [Duska97, Breslau98]
- **Small organizations**
  - Significant increase in hit rate as client population increases
  - The reason why cooperative caching is effective for UW
- **Large organizations**
  - Marginal increase in hit rate as client population increases



Extrapolation to Larger Client Populations

- **Use least squares fit to create a linear extrapolation of hit rates**
- **Hit rate increases logarithmically with client population - to increase hit rate by 10%:**
  - Need 8 UWs (ideal)
  - Need 11 UWs (cacheable)



Question

- **What is the benefit of cooperative caching among large organizations?**

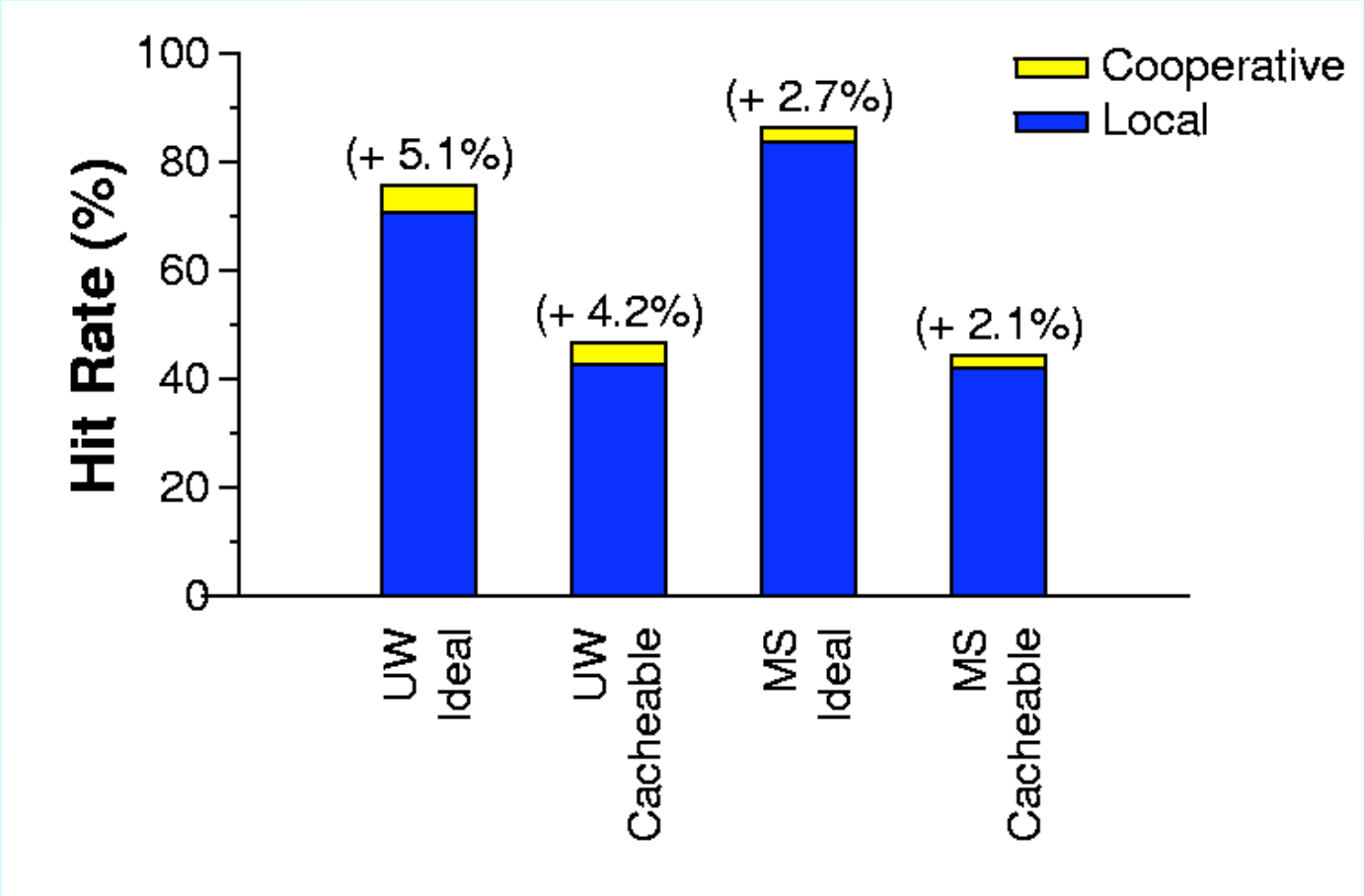
UW & Microsoft Cooperation

- **What if we ran a wire across Lake Washington, to connect UW & Microsoft?**
- **We collected a Microsoft proxy trace during same time period as the UW trace**
  - Combined population is ~80K clients
  - Increases the UW population by 3.6x
  - Increases the MS population by 1.4x

UW & Microsoft Traces

Trace	UW	MS
Duration	**7 days**	**6.25 days**
HTTP objects	**18.4 million**	**15.3 million**
HTTP requests	**82.8 million**	**107.7 million**
Avg. requests/sec	**137**	**199**
Total Bytes	**677 GB**	**N/A**
Server	**244,211**	**360,586**
Clients	**22,984**	**60,233**
Population	**~50,000**	**~40,000**

UW & MS Cooperative Caching



- **Is this worth it?**

Conclusions

- **A negative result: without significant workload changes, designing highly-scalable cooperative proxy-cache schemes is unnecessary**
  - Largest benefit is achieved with small populations (up to 2K-5K clients)
  - Limited benefit of cooperation when we combined the UW & Microsoft populations
  - Document cacheability is a severe limitation with current workloads

Discussion

- **What movies should Blockbuster store given that space is at a premium?**

Discussion

- **What movies should Blockbuster store given that space is at a premium?**
  - BB is like a movie cache
- **If the Blockbuster near your house doesn't carry the movie you want, should you try a different Blockbuster?**

Discussion

- **What movies should Blockbuster store given that space is at a premium?**
  - BB is like a movie cache
- **If the Blockbuster near your house doesn't carry the movie you want, should you try a different Blockbuster?**
  - It depends how big your local BB store is
  - Probably not…