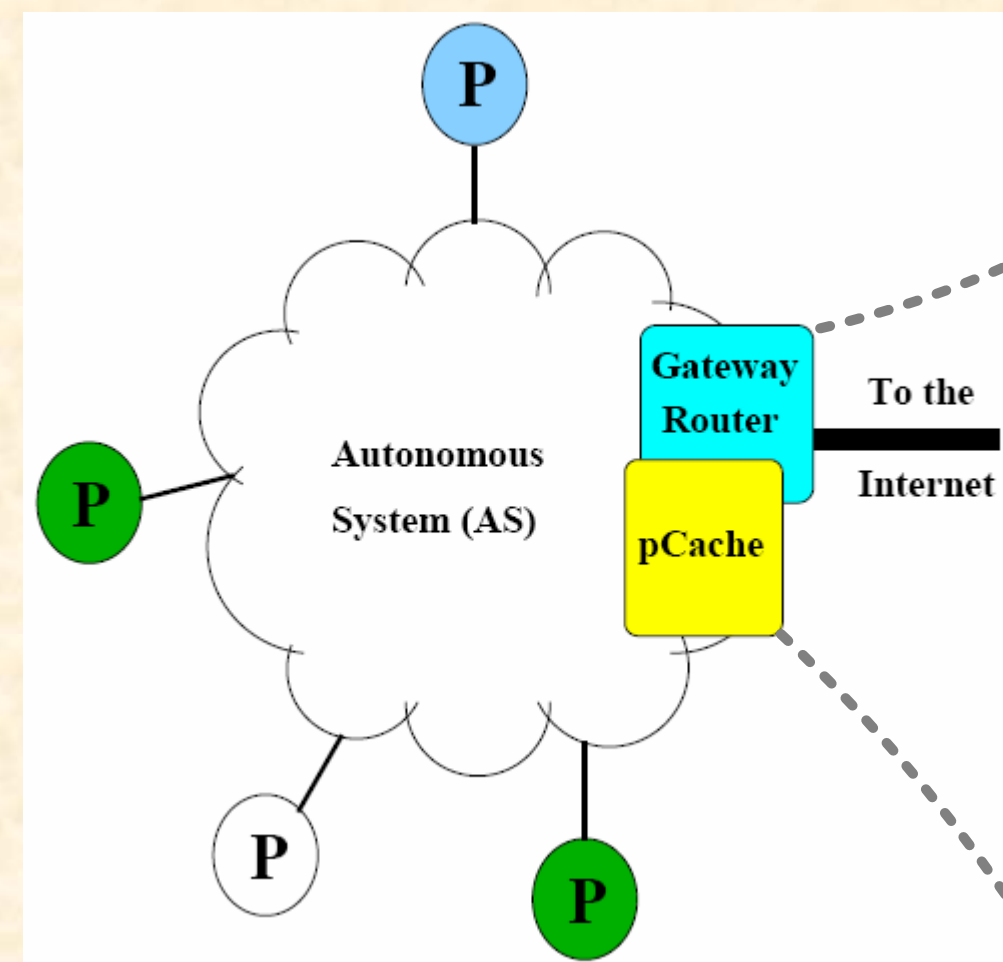
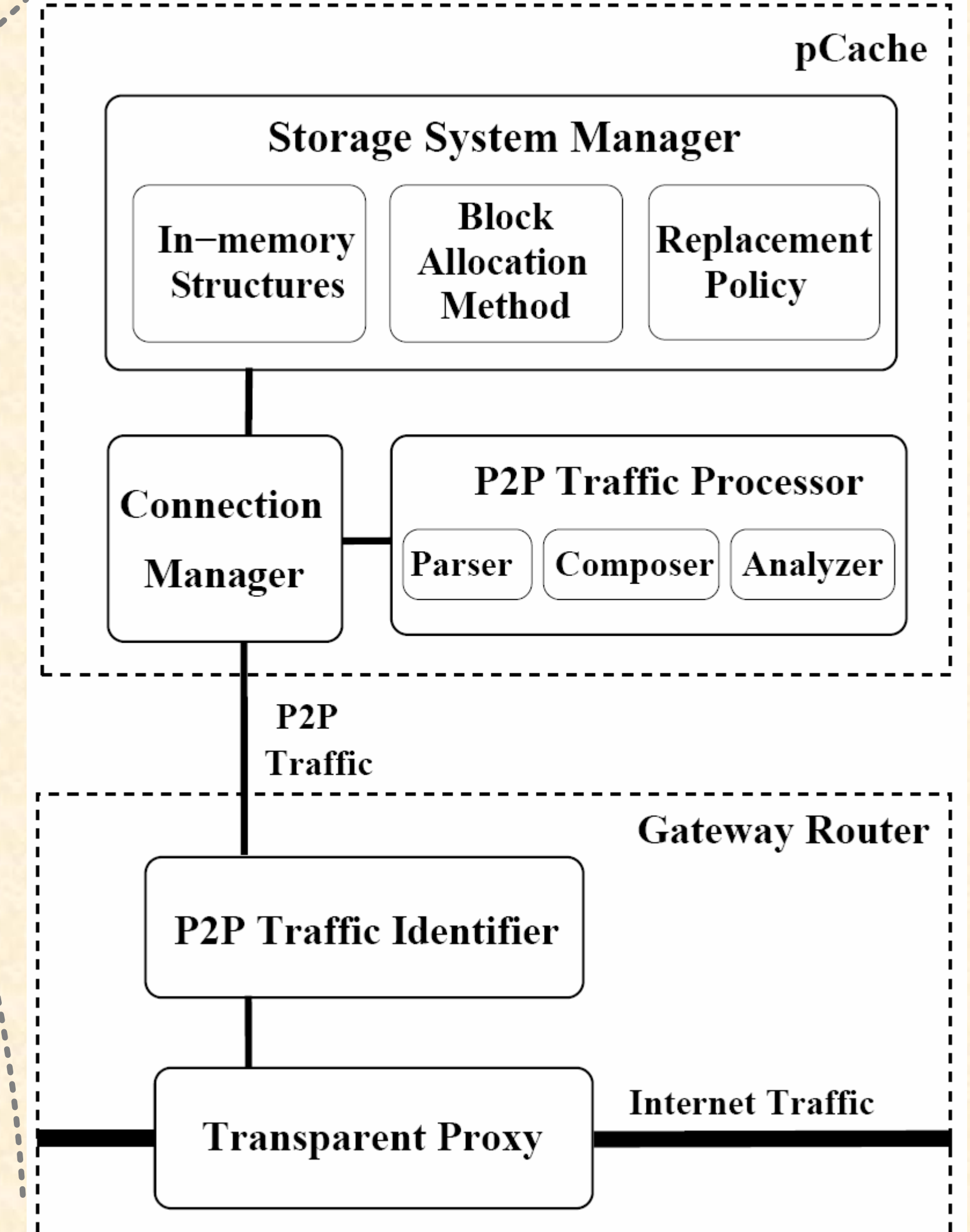


1. Motivation

- **P2P traffic is enormous**
 - Huge cost for ISPs, campuses, etc.
 - Increased load on backbone links
- **Previous works show benefits of P2P caching**
- **We designed an open-source P2P cache: pCache**
 - Works with different P2P systems
 - Fully transparent
 - Efficient storage system customized for P2P traffic

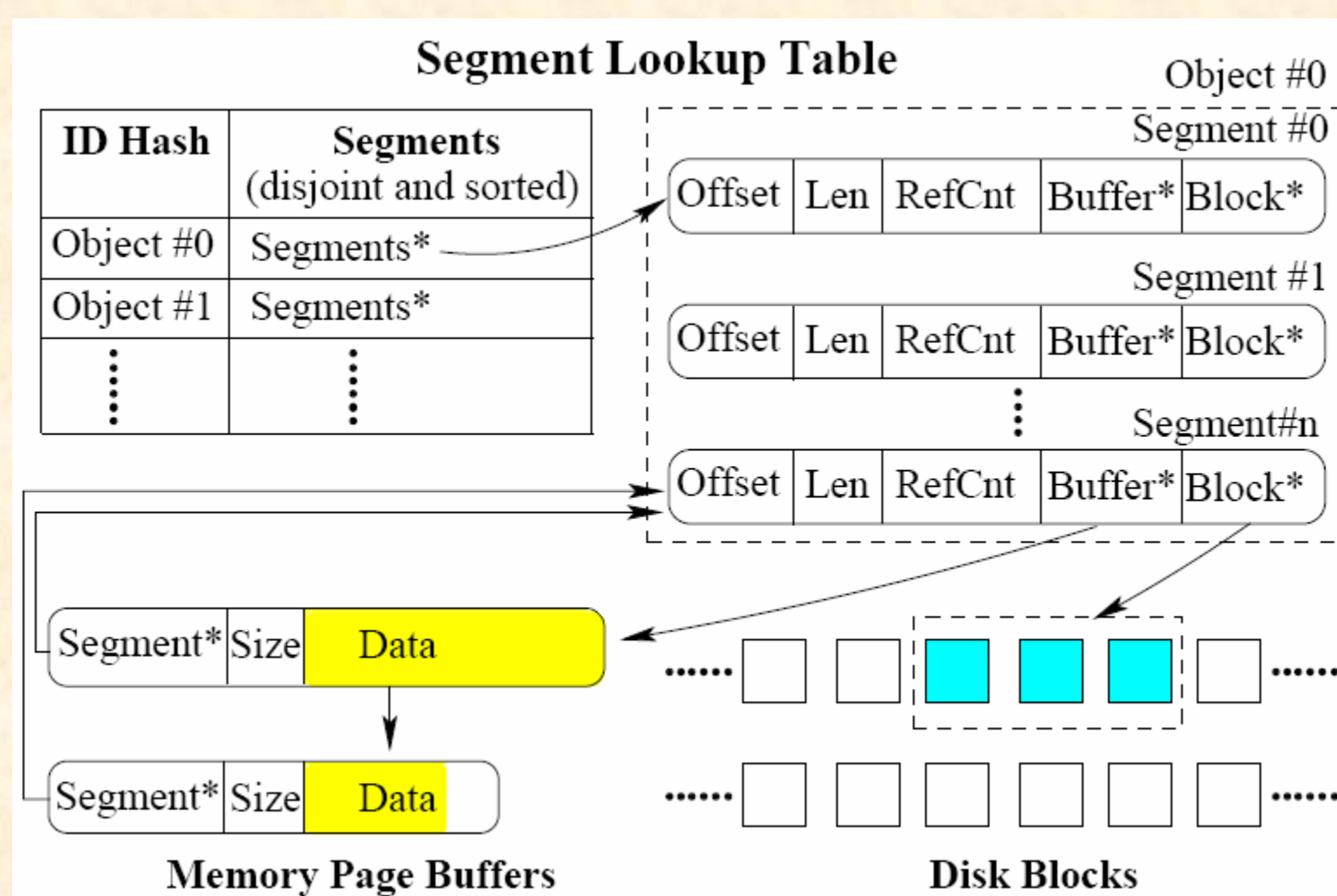


2. Overview



3. Storage System

- **Partial caching and serving of objects**
 - Requested segment sizes are highly variable
- **P2P-protocol-independence**
 - Support for cross-system caching
- **Different object replacement policies**
- **Minimize disk I/O operations**
 - Segment merging



4. Traffic Processor

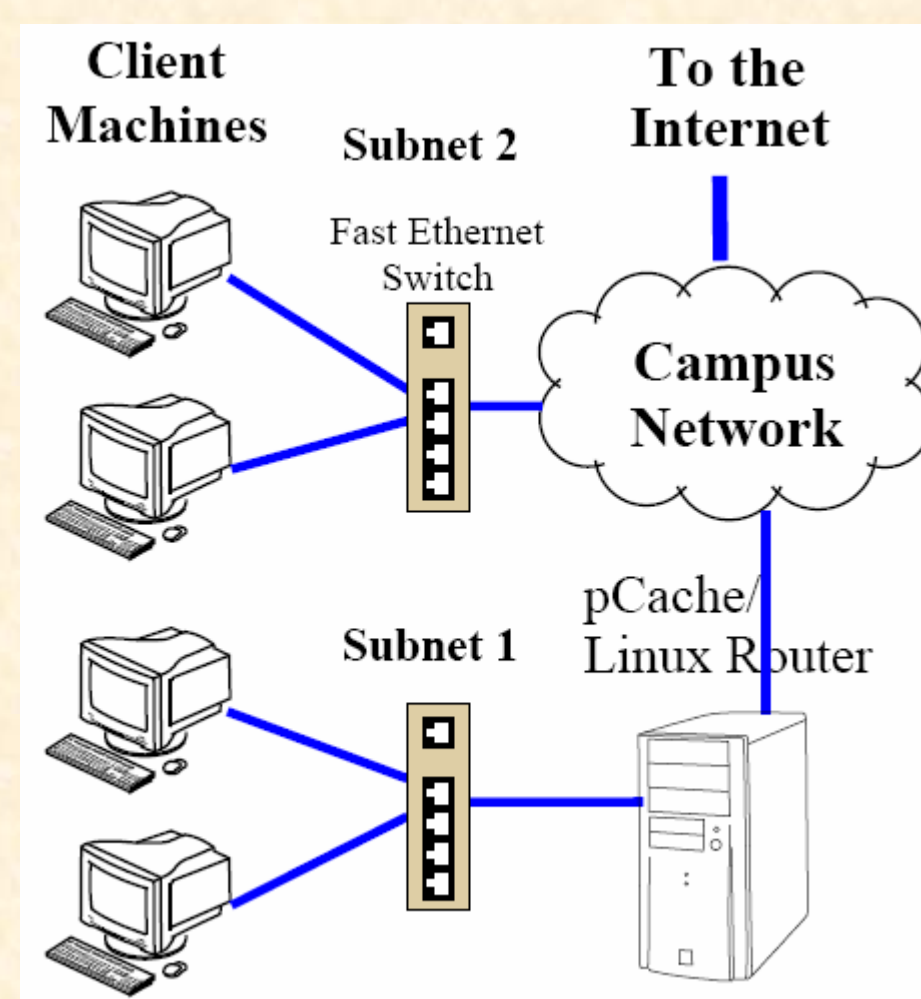
- **Parser: extracts messages (control/data)**
- **Composer: prepares protocol-specific messages**
- **Analyzer: placeholder for auxiliary functions**
 - Piece length inference for BitTorrent objects using:
 - o First few requests issued by clients
 - o A priori knowledge about piece lengths
 - Quantifiable confidence as input
 - Incorrect inference: minor performance penalty

7. Future Works

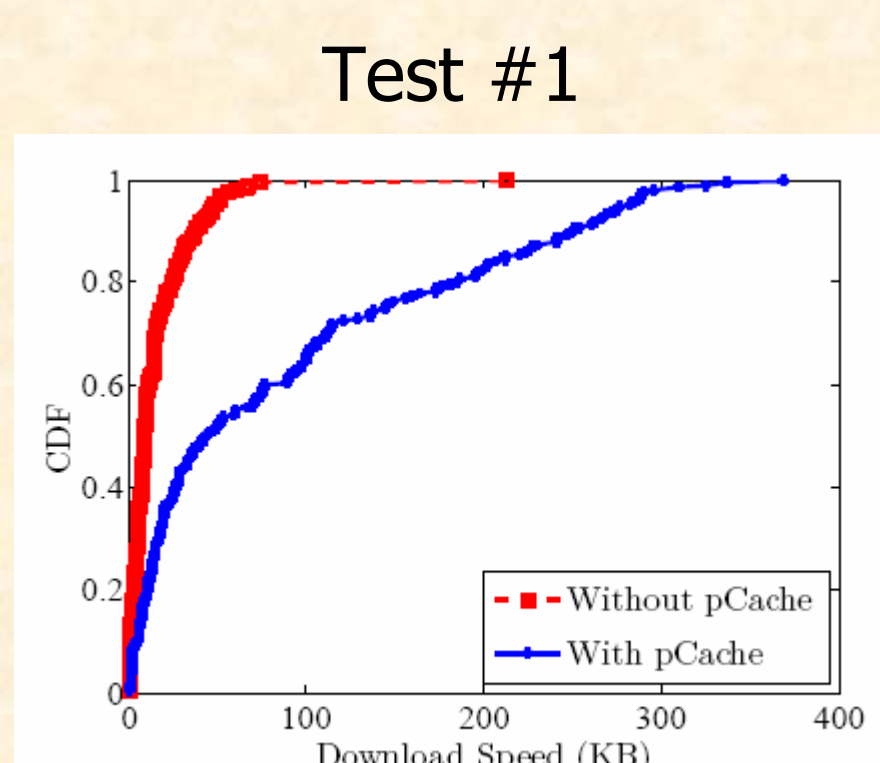
- **Encrypted P2P traffic**
 - Cache as a man-in-the-middle
- **Cross-system caching**
- **New replacement policies**
- **More info at**
<http://nsl.cs.sfu.ca/wiki/>

6. Evaluation of pCache

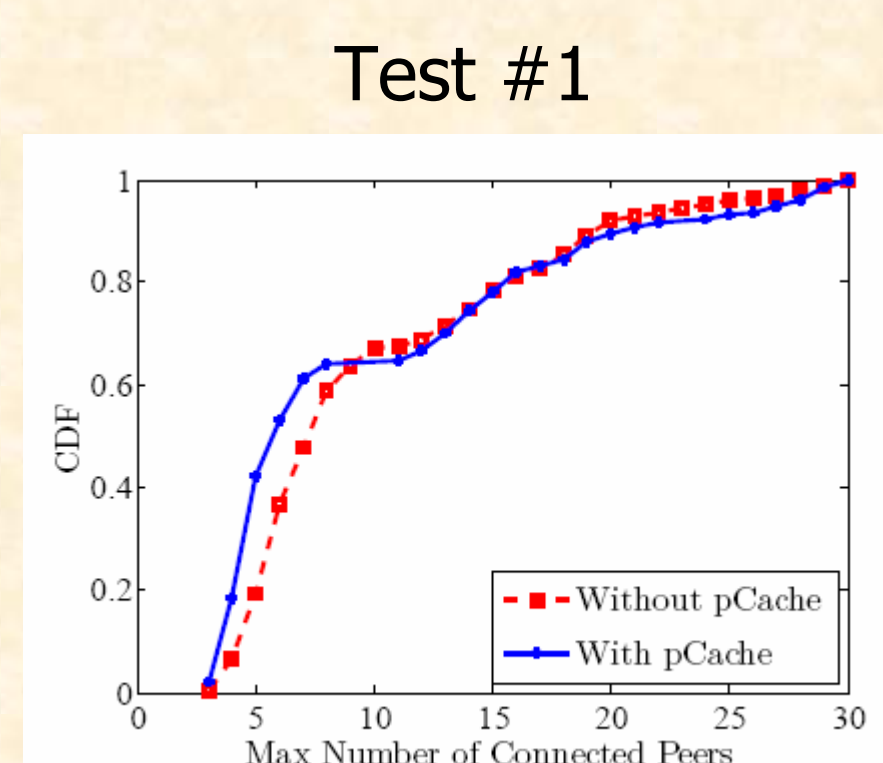
- **With real P2P traffic**
- **Test #1: the whole system**
 - 2700 downloads, 500 GB
 - pCache software passed all validation tests
 - pCache Benefits the ISPs
 - o 90% of traffic was served from cache
 - pCache Benefits the P2P clients
 - o No impact on connectivity
 - o Higher download speed (With no higher upload speed)



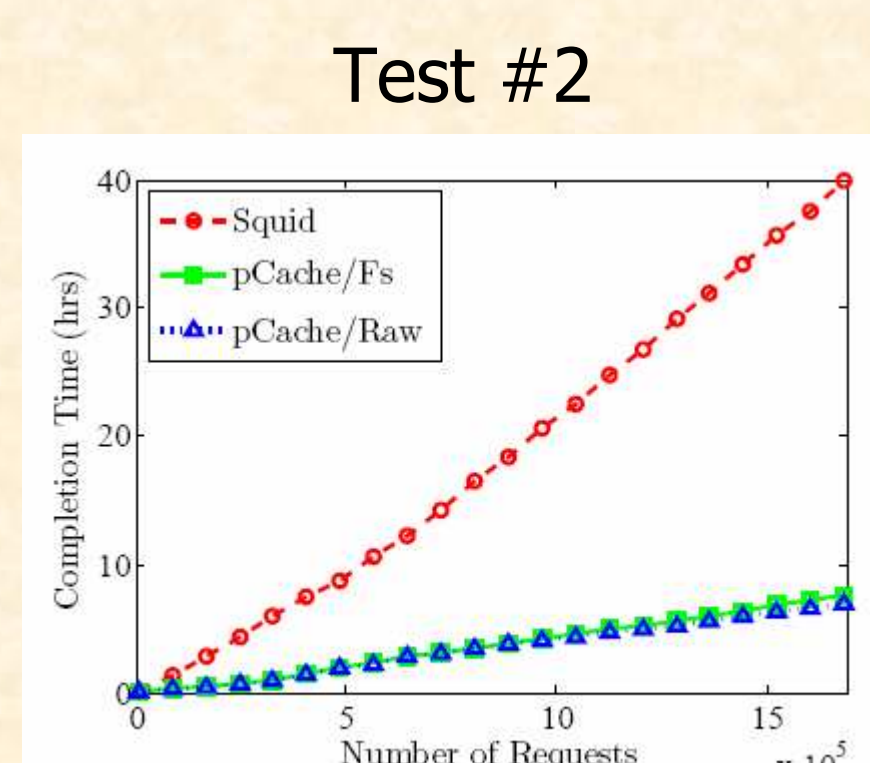
- **Test #2: the storage system**
 - Storage implemented in two versions
 - o On a large file on ext2 Linux file system
 - o On a raw partition (direct I/O)
- **Test #3: scalability**
 - Scalability of the bottleneck: storage system
 - 200 Mbps throughput for P2P traffic
- **Test #4: Piece length inference**
 - Tested for 2100 downloads: 99.7% accuracy by observing only 3% of download traffic of a file



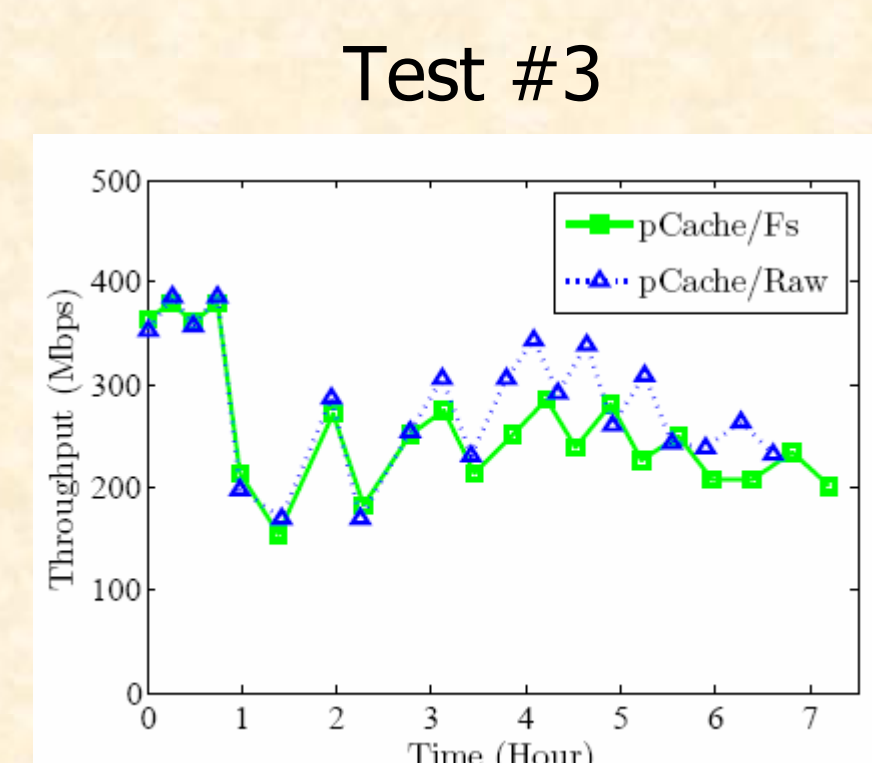
Higher download speed



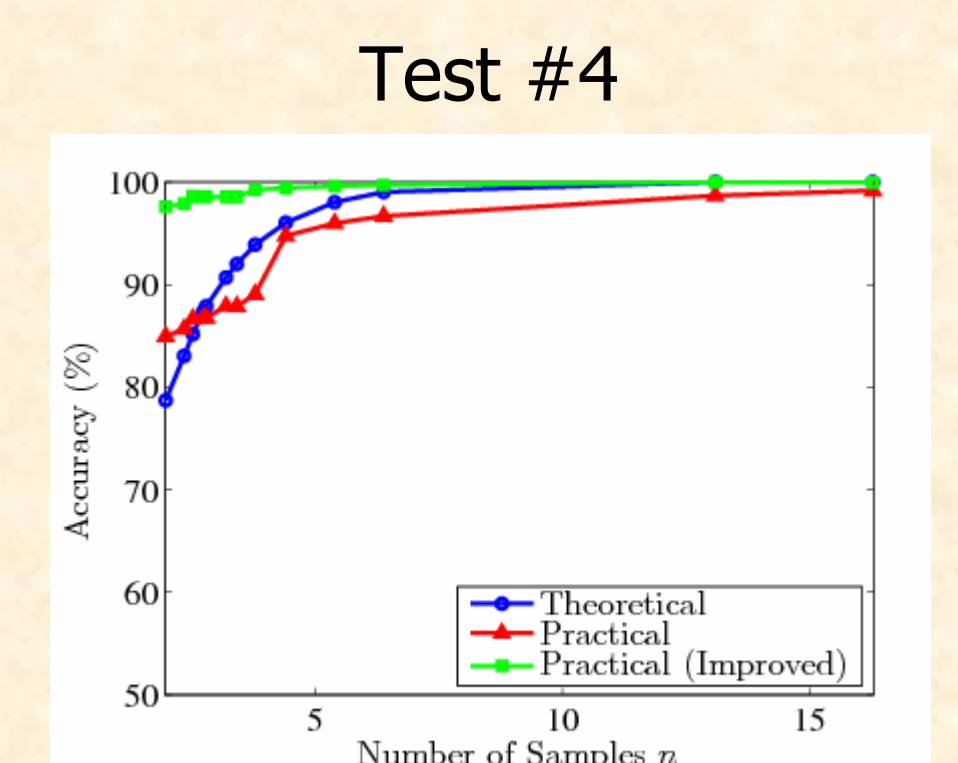
No impact on connectivity



Performance vs. squid



Throughput achieved



Inference accuracy