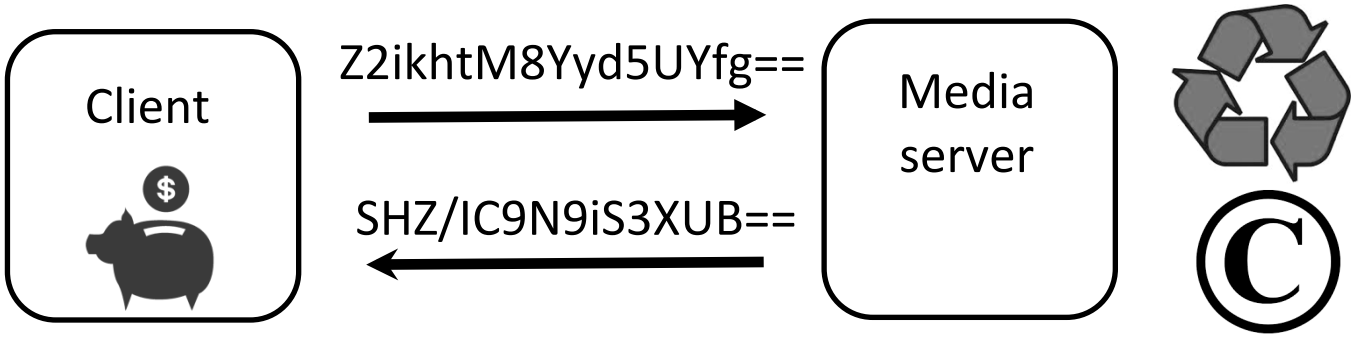
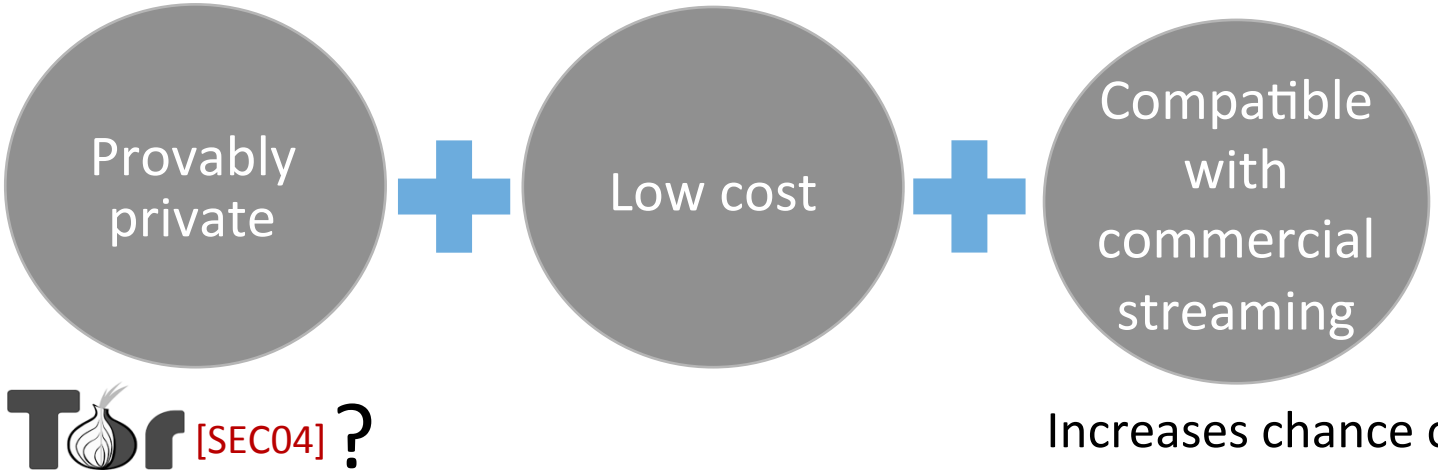
Scalable and private media consumption with Popcorn

Trinabh Gupta*¶, Natacha Crooks*δ, Whitney Mulhern¶,
Srinath Setty₮, Lorenzo Alvisi*, and Michael Walfish¶

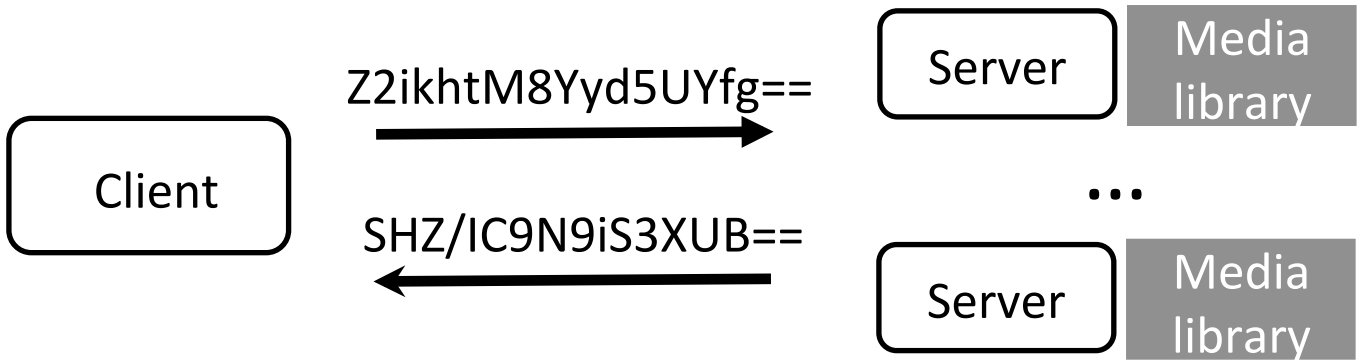
*The University of Texas at Austin
¶New York University
δMPI-SWS
₮Microsoft Research



Why? Because media diet can reveal sensitive information.



Attacks: [NDSS14, ESORICS13, ESORICS12, …]



Private Information Retrieval (PIR) provably hides requests but …

- Each request must touch the entire library.
- There is a tension between overhead and content protection.
- PIR assumes fixed-size objects, but media sizes vary.

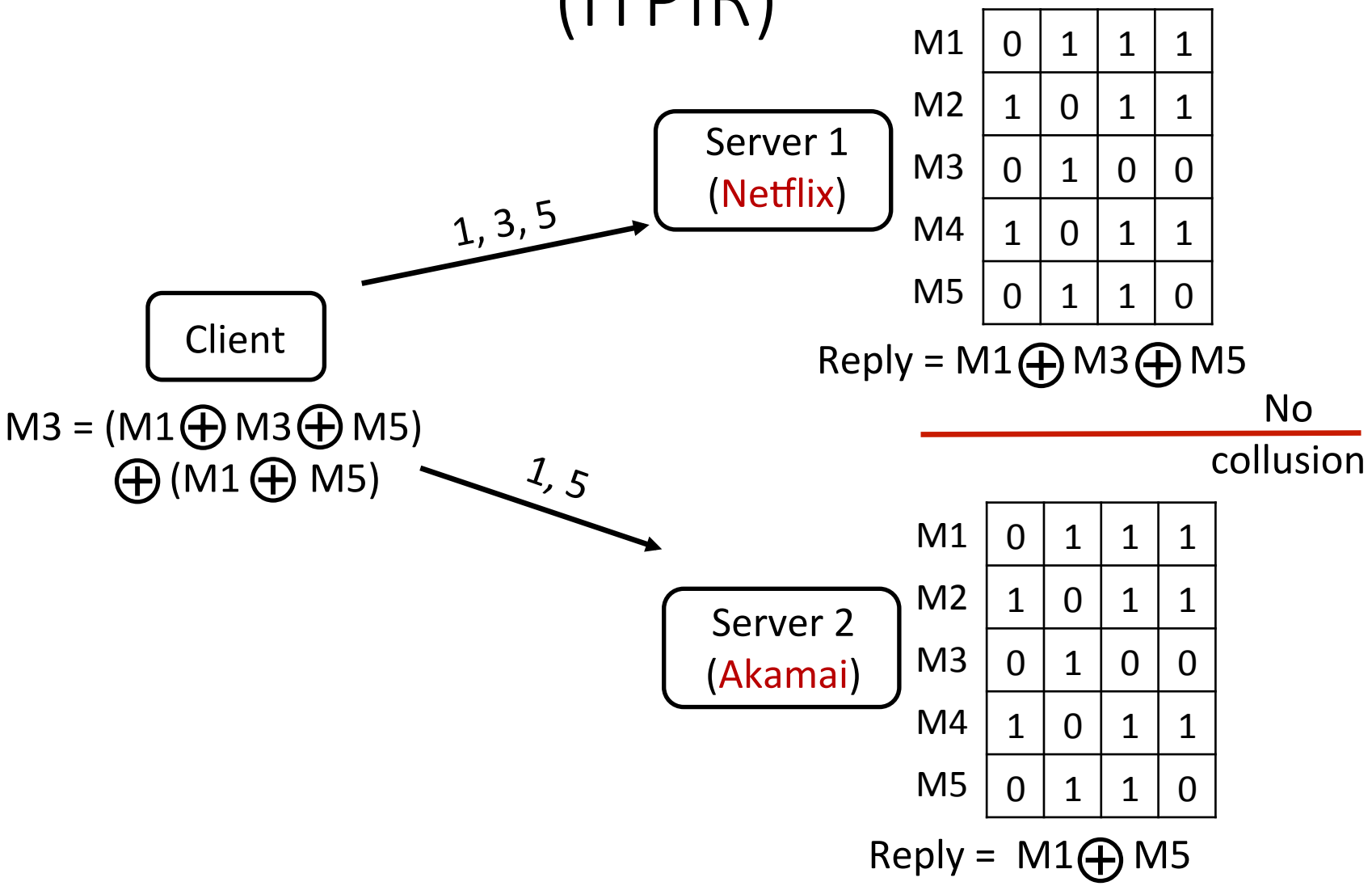
Popcorn tailors PIR for media to meet our three requirements.

Its per-request dollar cost is 3.87x times that of a non-private baseline.

Rest of this talk

- Background on PIR.
- Design (tailoring of PIR) and evaluation of Popcorn.

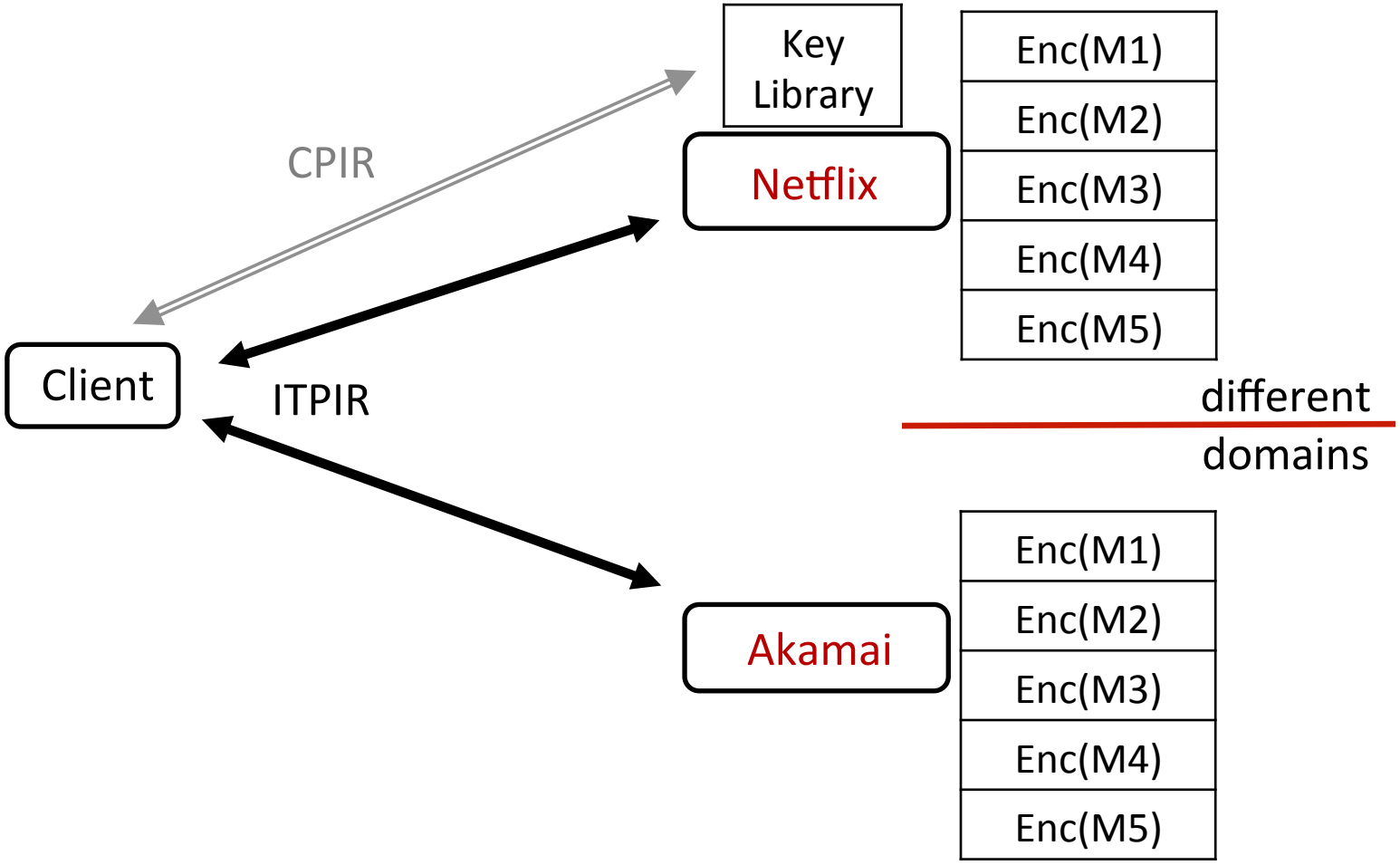
Background on Information-theoretic PIR (ITPIR)



ITPIR	CPIR
cheap operations (XORs)	expensive operations (up to 10x ITPIR)
process entire library per request	process entire library per request
does not respect controls on content dissemination	respects controls on content dissemination

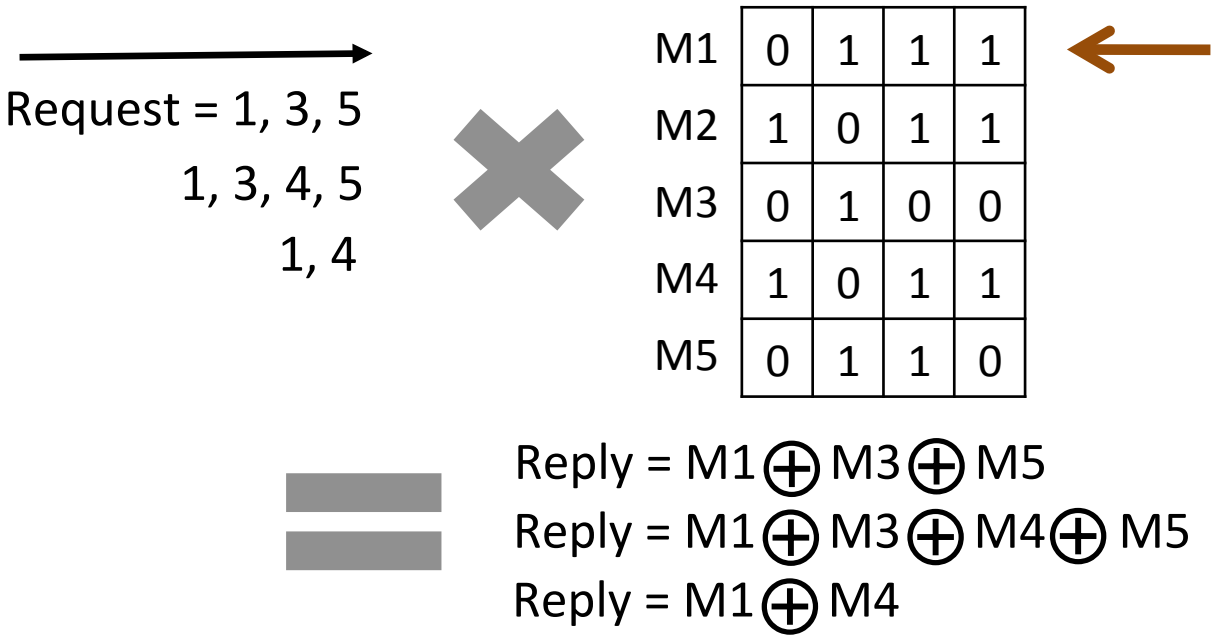
Given these, how can we build a system that is low cost and compatible with commercial streaming?

Popcorn composes ITPIR and CPIR to get desirable properties from both



ITPIR	CPIR	Popcorn
cheap operations (XORs)	expensive operations (5-10x ITPIR)	mostly cheap operations
does not respect controls on content dissemination	respects controls on content dissemination	respects controls on content dissemination
process entire library per request	process entire library per request	?

Popcorn batches requests to amortize the overhead of ITPIR

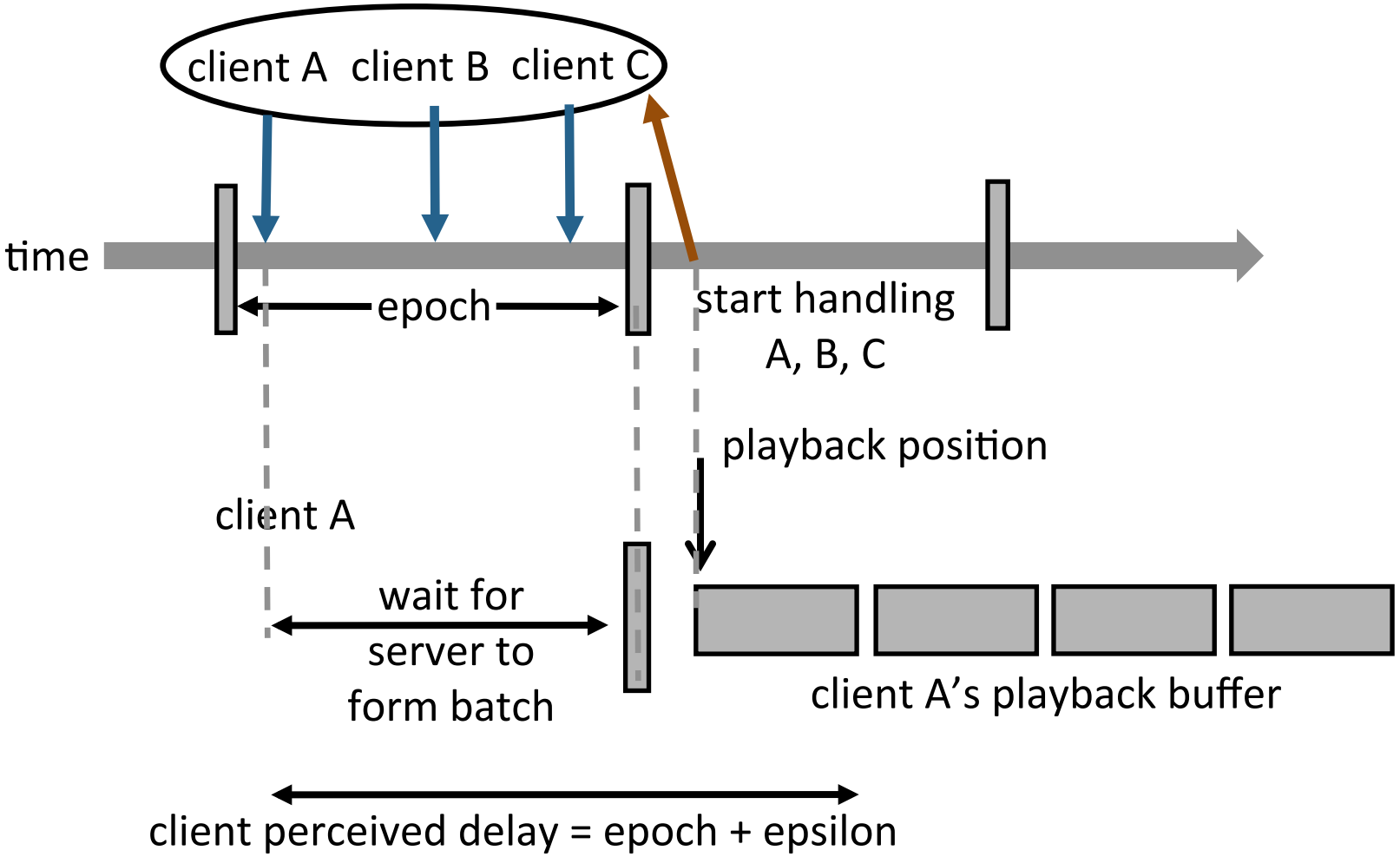


Observation: Same I/O work for each request!

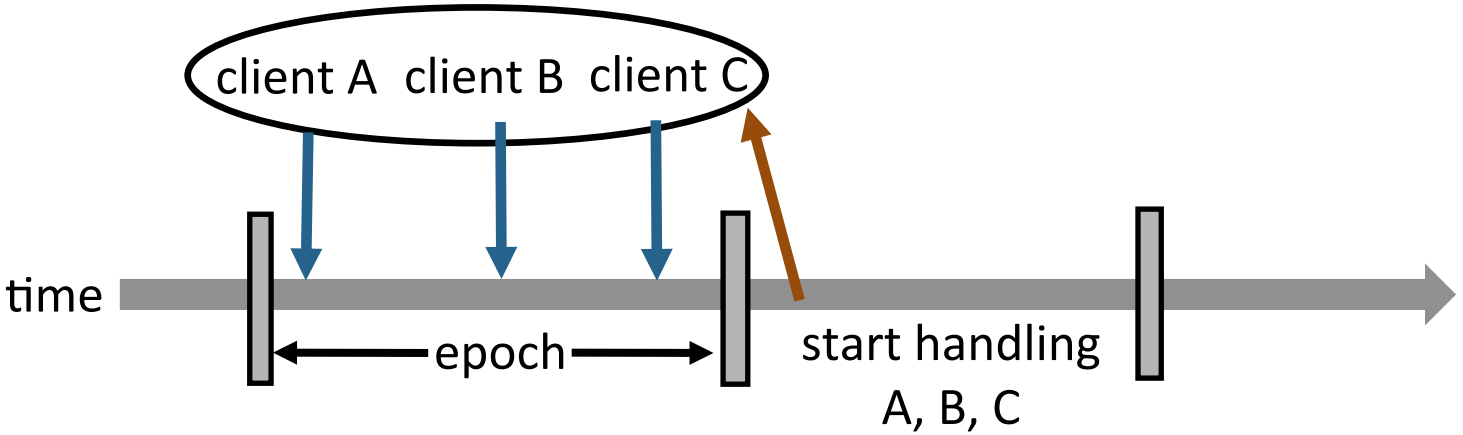
Benefits of batching:

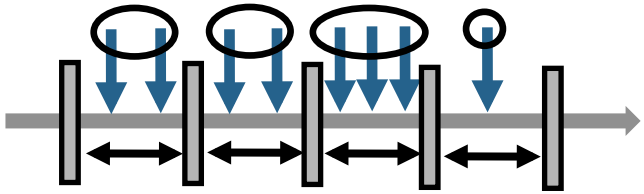
- I/O transfers are amortized.
- CPU cycles are reduced as matrix multiplication algorithms exploit cache locality.

Strawman: Group requests that arrive during an epoch

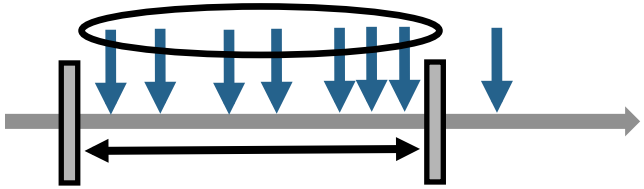


Strawman: Group requests that arrive during an epoch





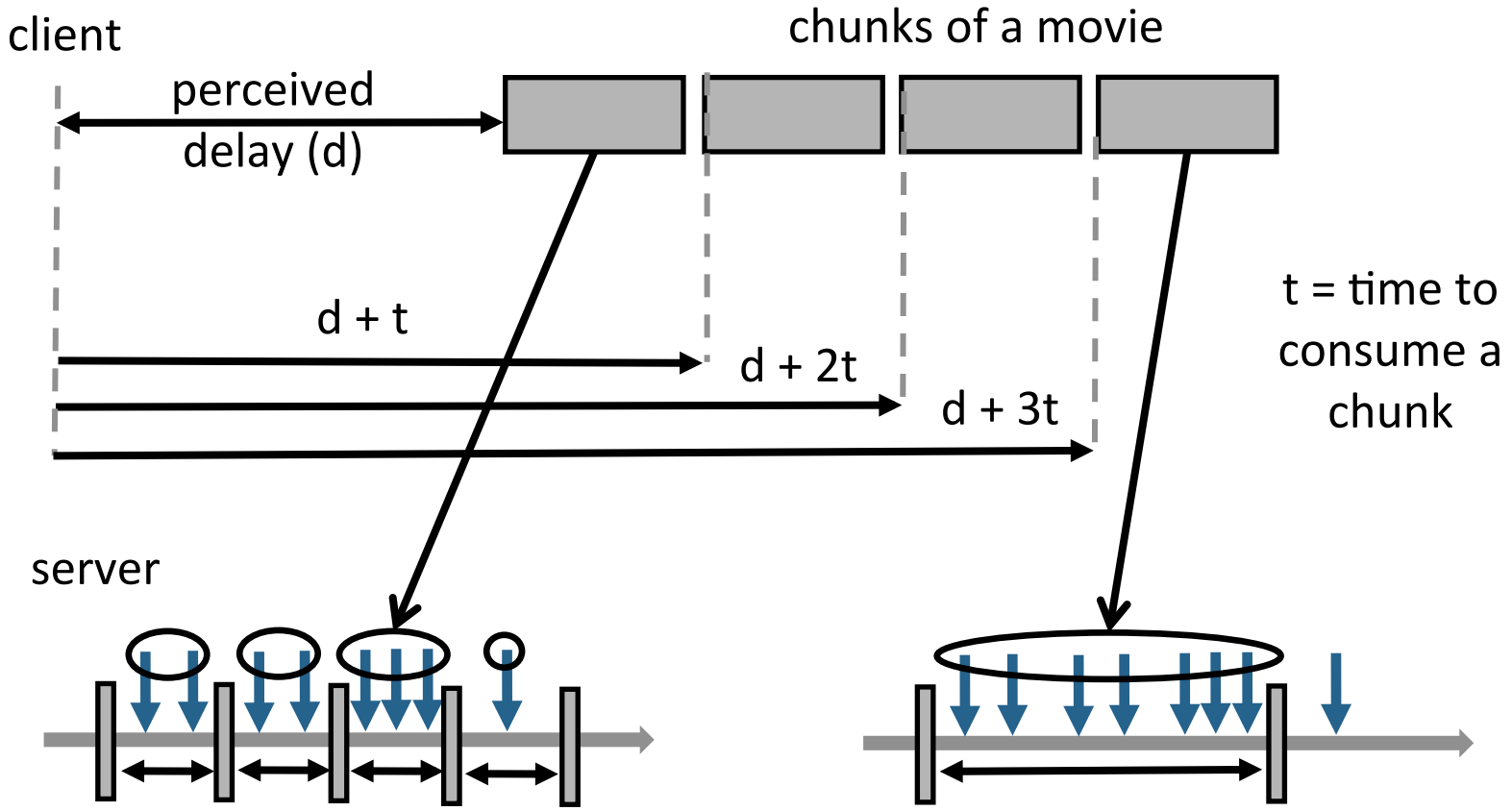
Small epoch, small batch, small delay



Large epoch, large batch, large delay

Issue: Hard to get both a small delay and a large batch

Popcorn exploits streaming to form large batches with small startup delay



- Inspired by pyramid broadcasting [MMCN95]

Other design considerations

- Popcorn must handle variable-sized media objects.

  Response: Change bitrates to make movies of the same size.

Outline

- ✓Background on PIR.
- ✓Design (tailoring of PIR) of Popcorn.
- Evaluation of Popcorn.

Experiment method

- Baselines:
  - Non-private system (Apache server)
  - State-of-the-art CPIR [XPIR PETS16]
  - State-of-the-art ITPIR [Percy++] modified to support streaming
  - CPIR and ITPIR extended with the strawman batching scheme
- Netflix-like library: 8000 movies, 90 minutes, 4Mbps
- Workload: 10K clients arrive within 90 minutes according to a Poisson process
- Estimate per-request dollar cost using Amazon's pricing model
  - CPU: $0.0076/hour
  - I/O bandwidth: $0.042/Gbps-hour
  - Network: $0.006/GB

System	CPUs	I/O (Gbps)	Network (relative to non-private)	$ relative to non-private
Non-private	0	0	1x	1x
CPIR	11.6	64	5x	265x
ITPIR	3.1	64	2x	256x
ITPIR++ (delay 15s)	0.65	3	2x	14x
ITPIR++ (delay 10min)	0.41	0.058	2x	2.5x
Popcorn (delay 15s)	0.74	0.23	2x	3.87x

Related work

- Improving performance of PIR.
  - Distributing work [FC13, TDSC12], cheaper crypto [PETS16, ESORICS14, ISC10, TKDE13, WEWoRC07], bucketing [DBSec10, PETS10], batching [FC15, JoC04], secure co-processors [PET03, FAST13, NDSS08, IBM Systems Journal01]
- Protecting library content in ITPIR [RANDOM98, S&P07, WPES13]
- Handling variable-sized objects [CCSW14, NDSS13]
- Prior PIR implementations [Percy++, PETS16, CCSW14]
- Video-on-demand [MMCN95]

Take-away points

- It is possible to build a private, backwards compatible, and low-cost media delivery system …
- … by tailoring PIR to media delivery.
- The per-request cost in Popcorn is 3.87x that of a non-private baseline.