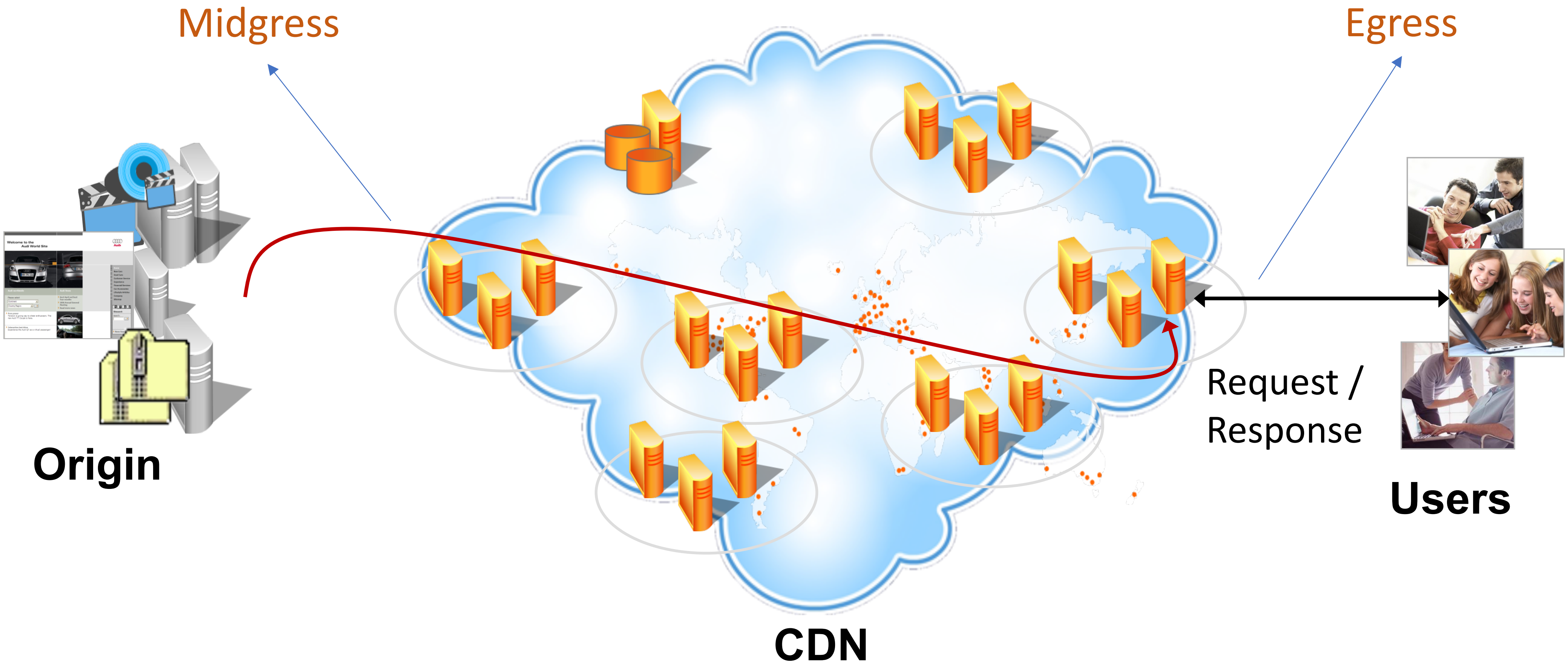
Midgress-aware traffic provisioning for content delivery

Aditya Sundarrajan, Mangesh Kasbekar,
Ramesh K. Sitaraman, Samta Shukla

CDNs serve more than 50% of content



Performance and cost metrics

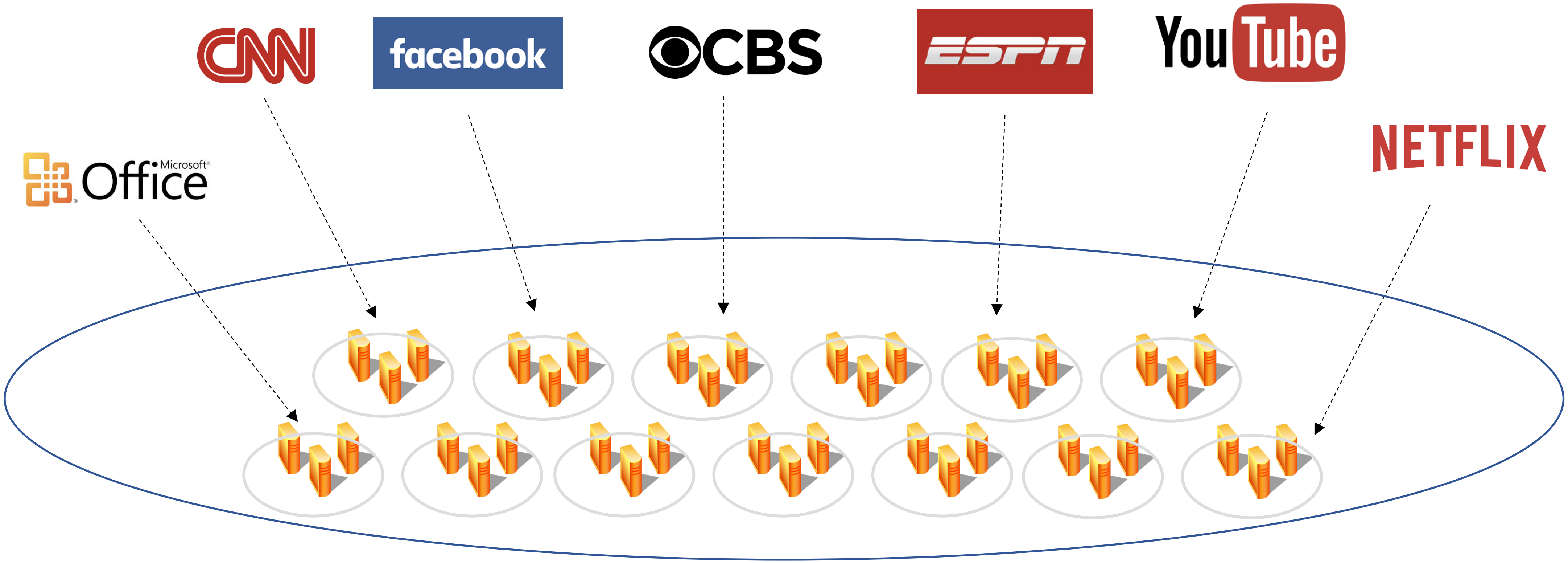
End-user latency

Origin offload ratio

Bandwidth cost

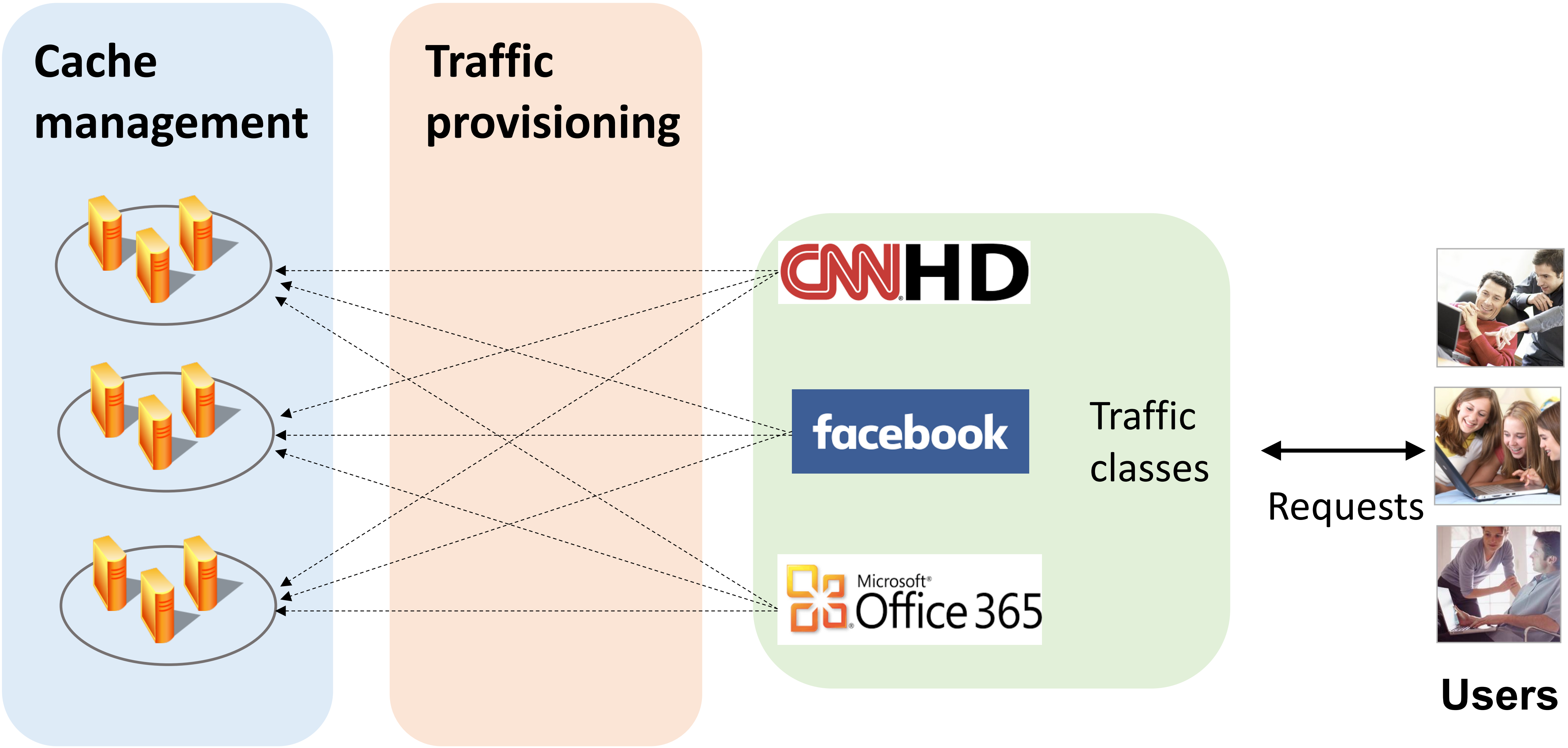
Cache hit rate

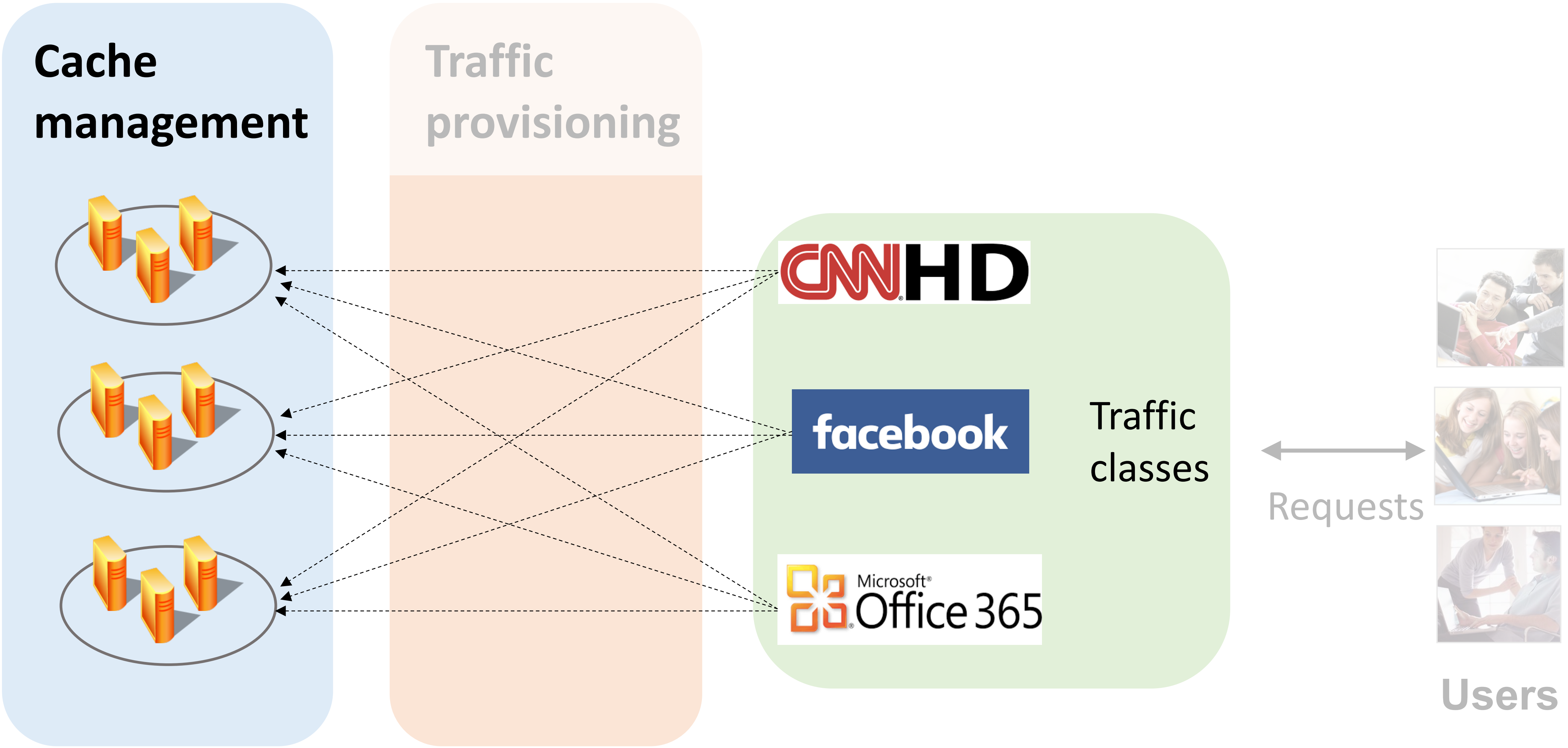
100s of content providers



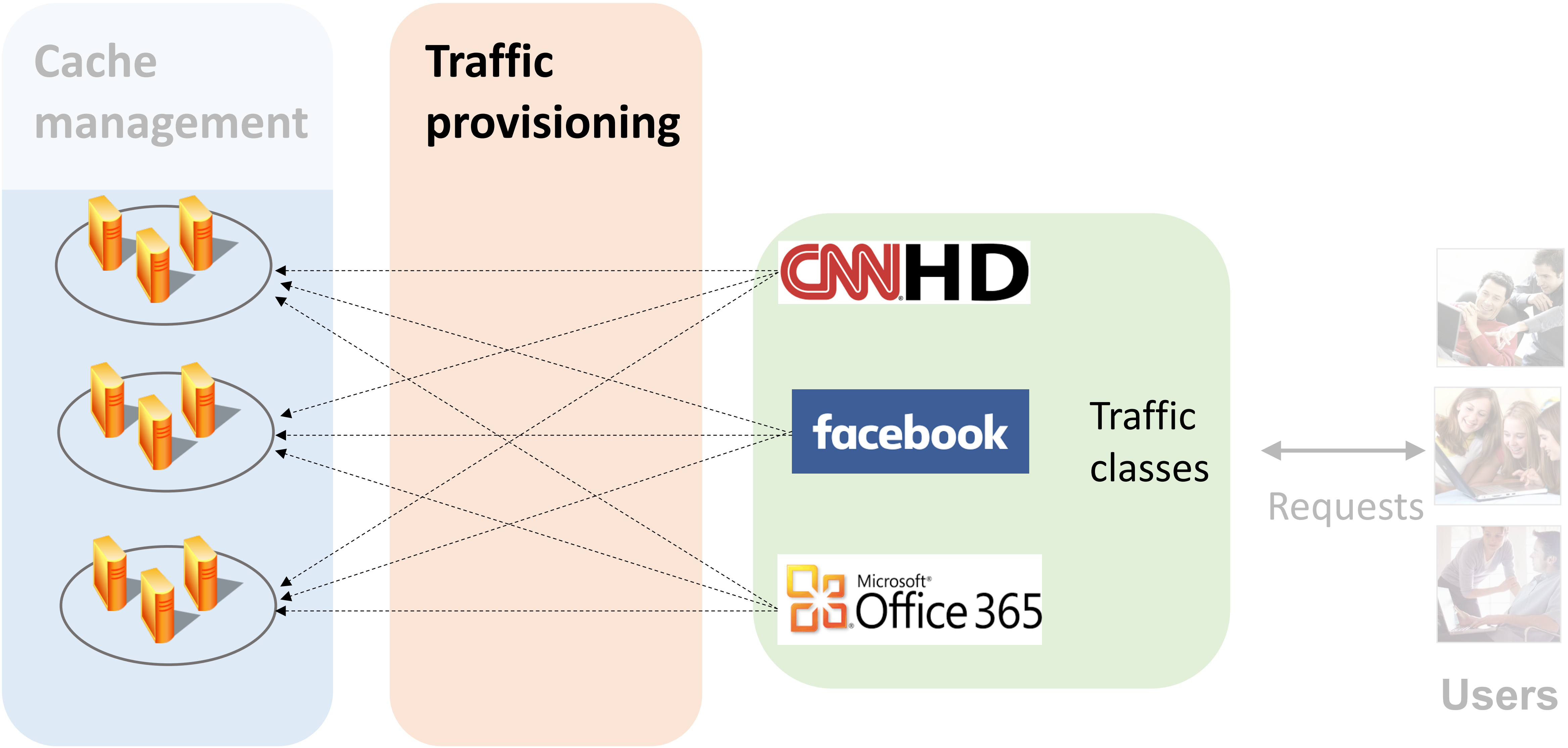
100s of 1000s of servers

Millions of users


Cache management
Traffic provisioning
Traffic classes
Requests
Users

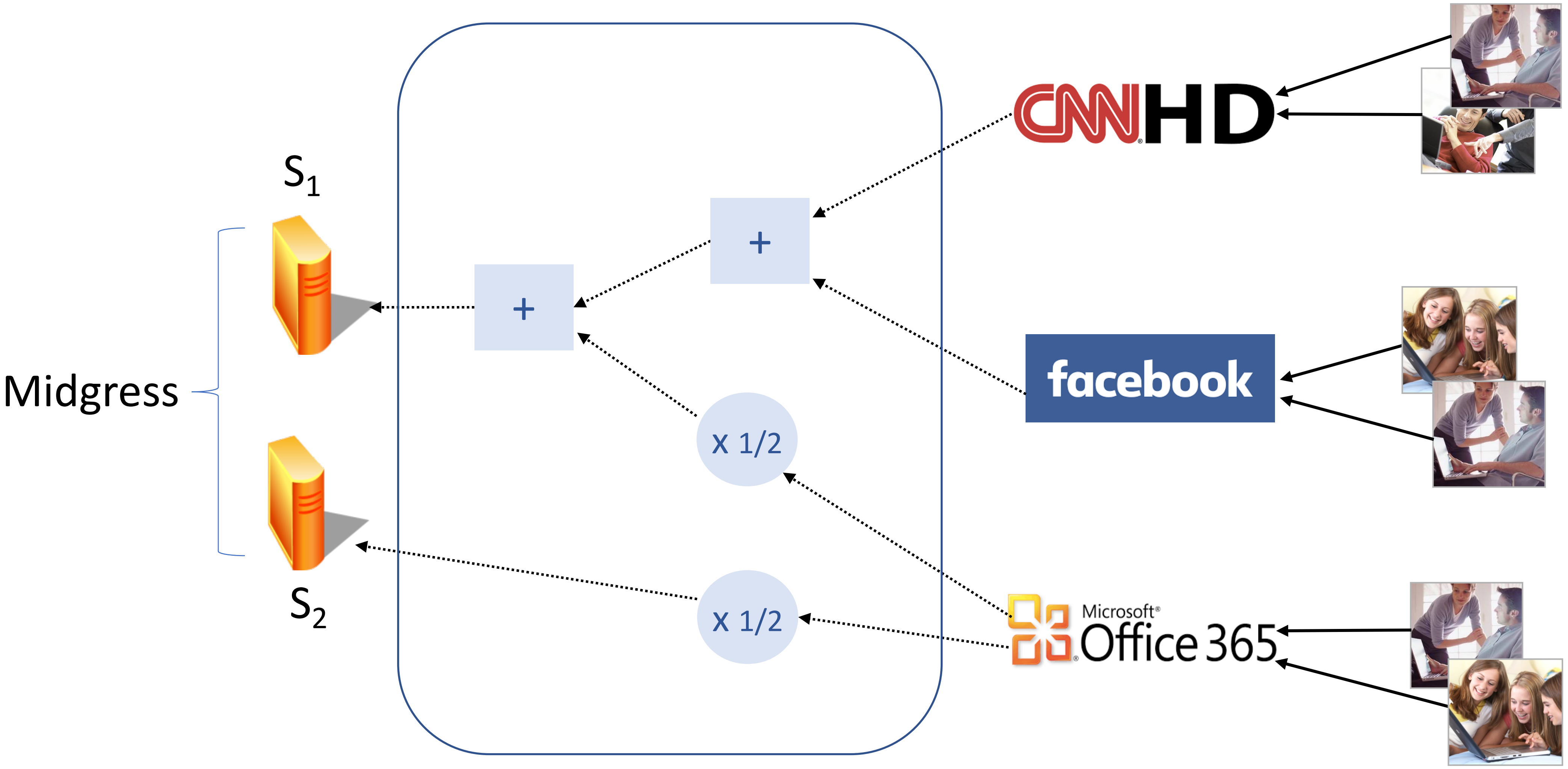


Past work has focused on cache management



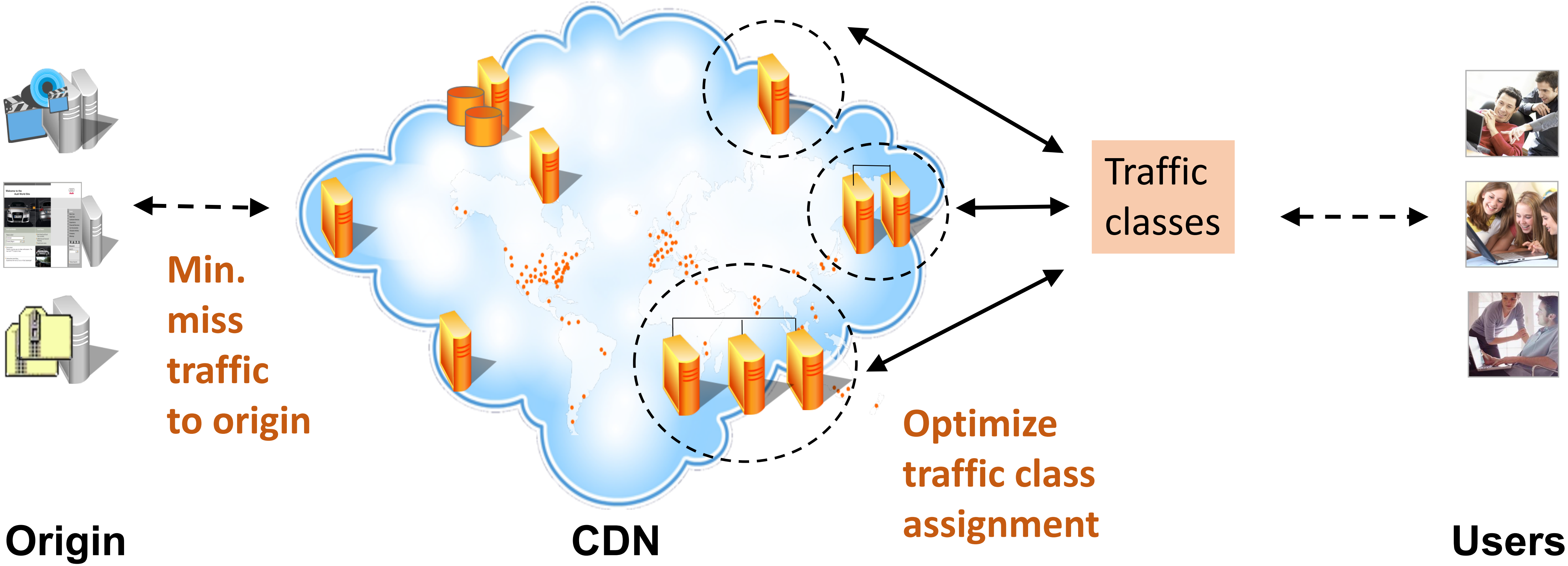
How can we assign traffic classes to reduce midgress?

Traffic provisioning to reduce midgress

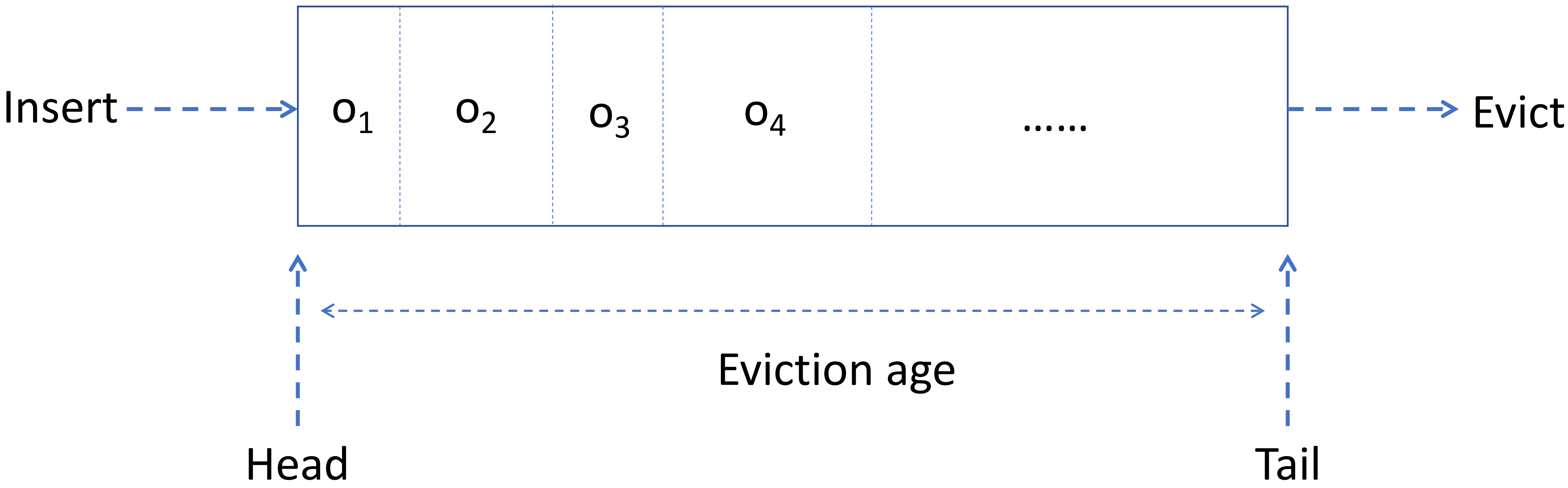


100s of traffic assignment scenarios!

Traffic provisioning to minimize midgress

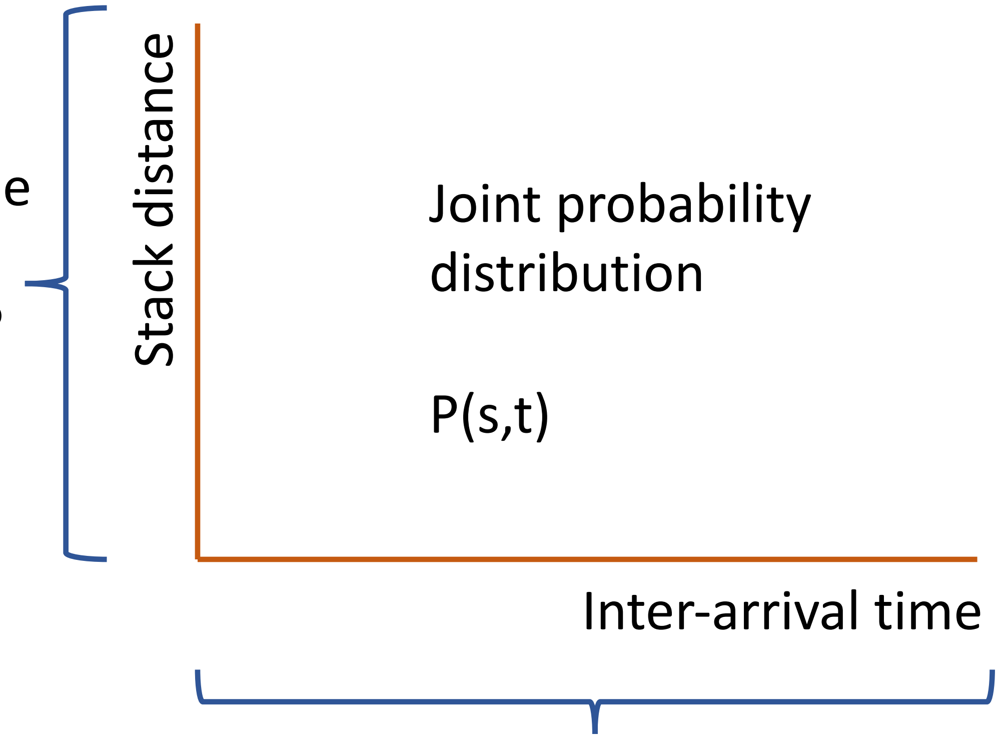


Eviction age equality



Footprint descriptors*

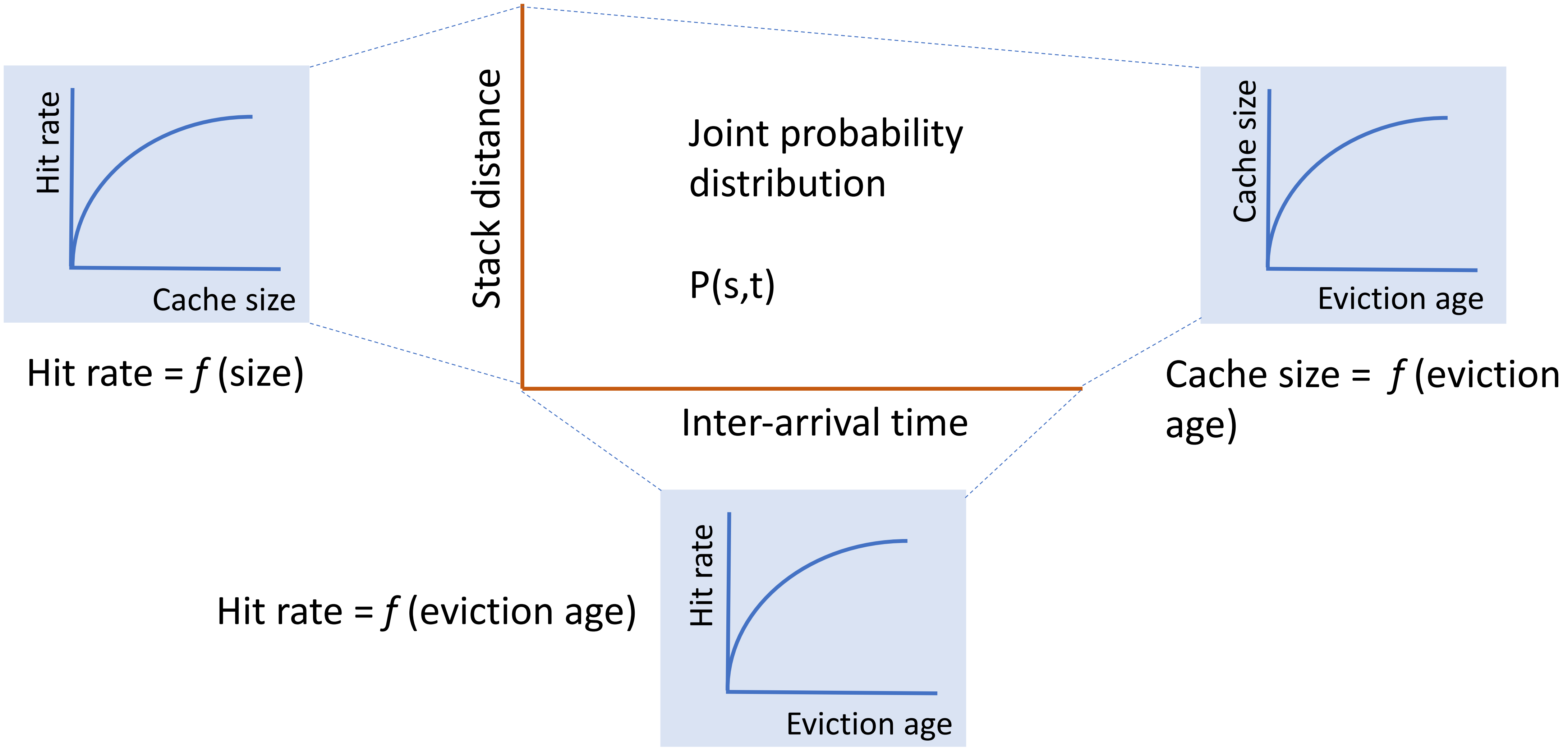
Spatial locality: How many unique bytes are requested between successive requests of an object?



Temporal locality: How often is an object requested?

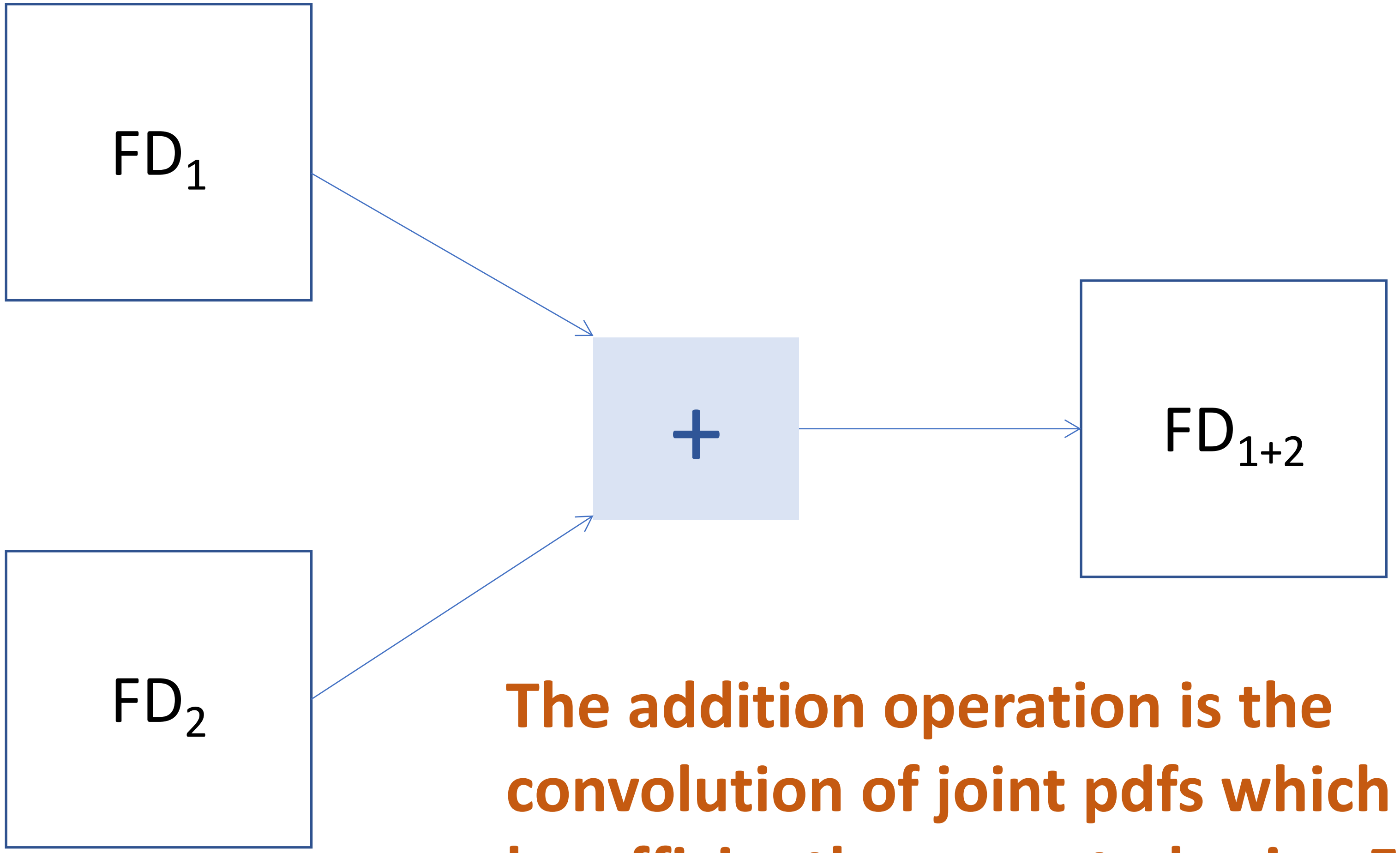
* Footprint descriptors: Theory and practice of cache provisioning in a global CDN, A. Sundarrajan et al. in ACM CoNEXT 2017

Caching properties from FDs



* Footprint descriptors: Theory and practice of cache provisioning in a global CDN, A. Sundarrajan et al. in ACM CoNEXT 2017

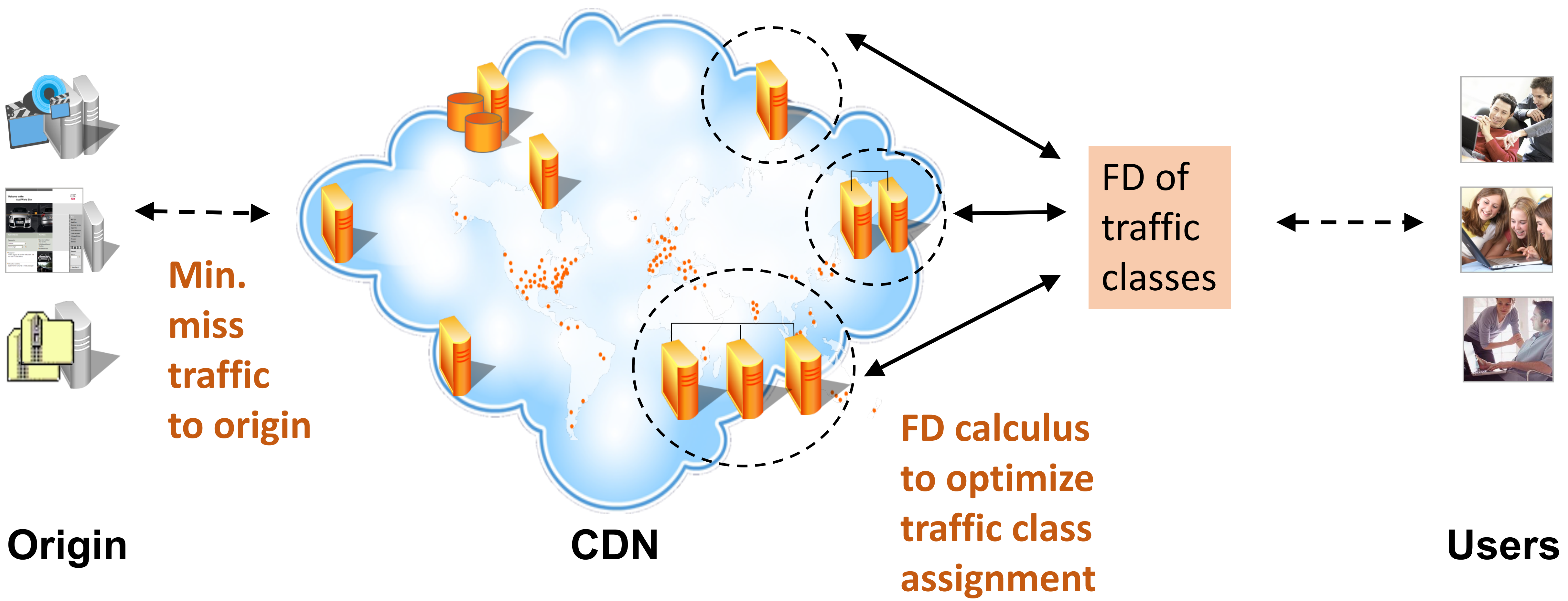
Traffic mixing using FD calculus



The addition operation is the convolution of joint pdfs which can be efficiently computed using FFT

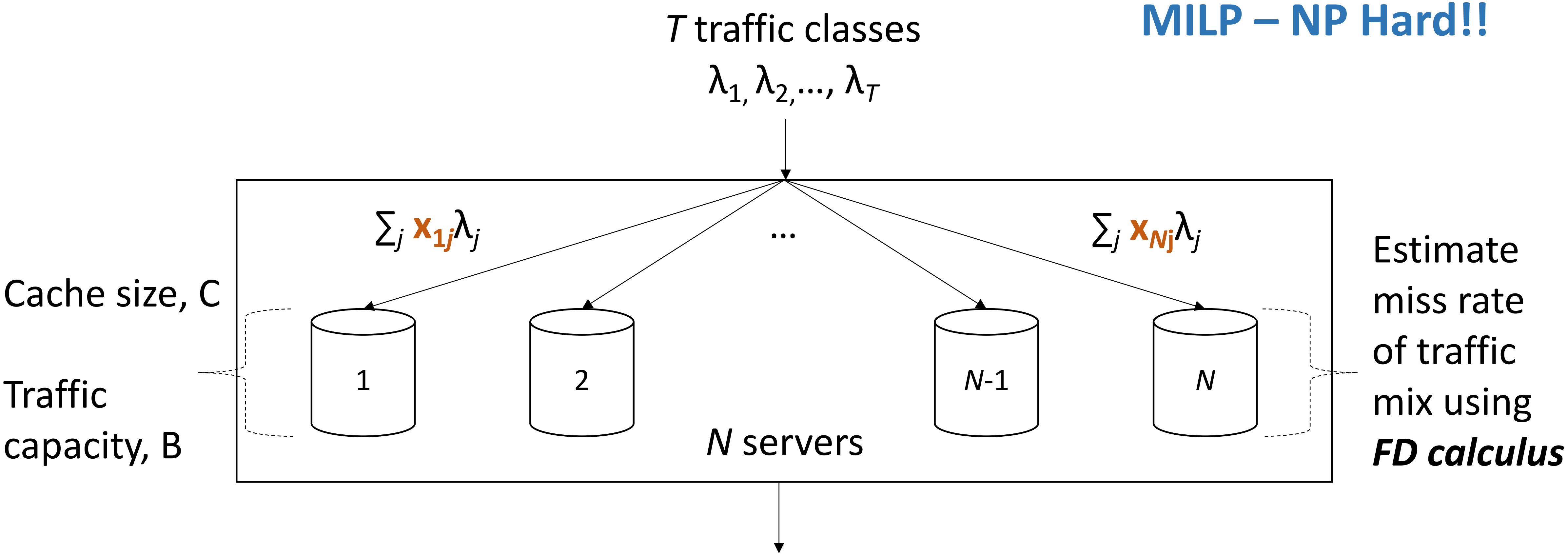
* Footprint descriptors: Theory and practice of cache provisioning in a global CDN, A. Sundarrajan et al. in ACM CoNEXT 2017

Traffic provisioning to minimize midgress



Traffic provisioning as an optimization problem

MILP – NP Hard!!

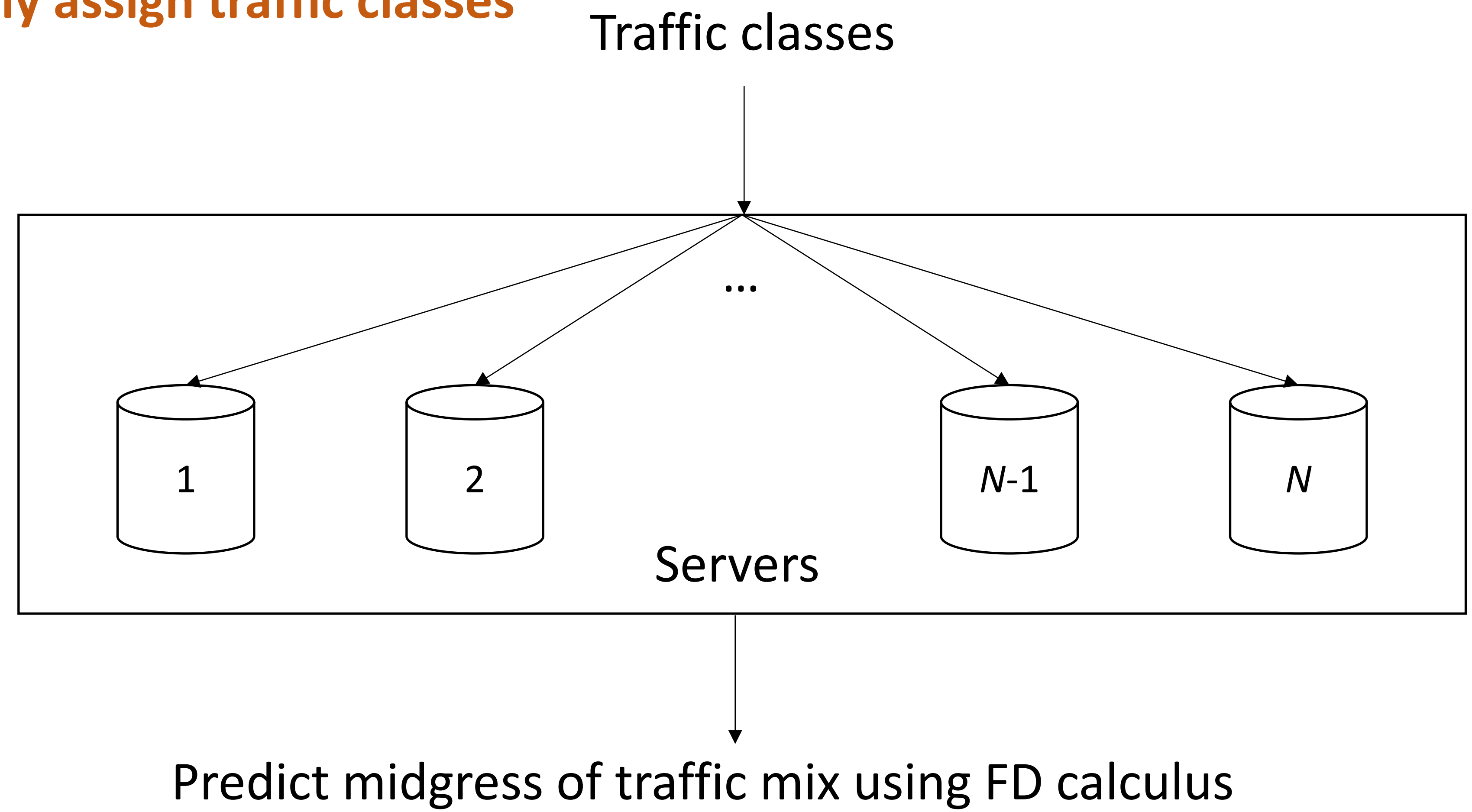


Min. $\sum_{ij} \mathbf{x_{ij}} \lambda_j \mathbf{m_j(c_{ij})}$

Total miss traffic from cluster

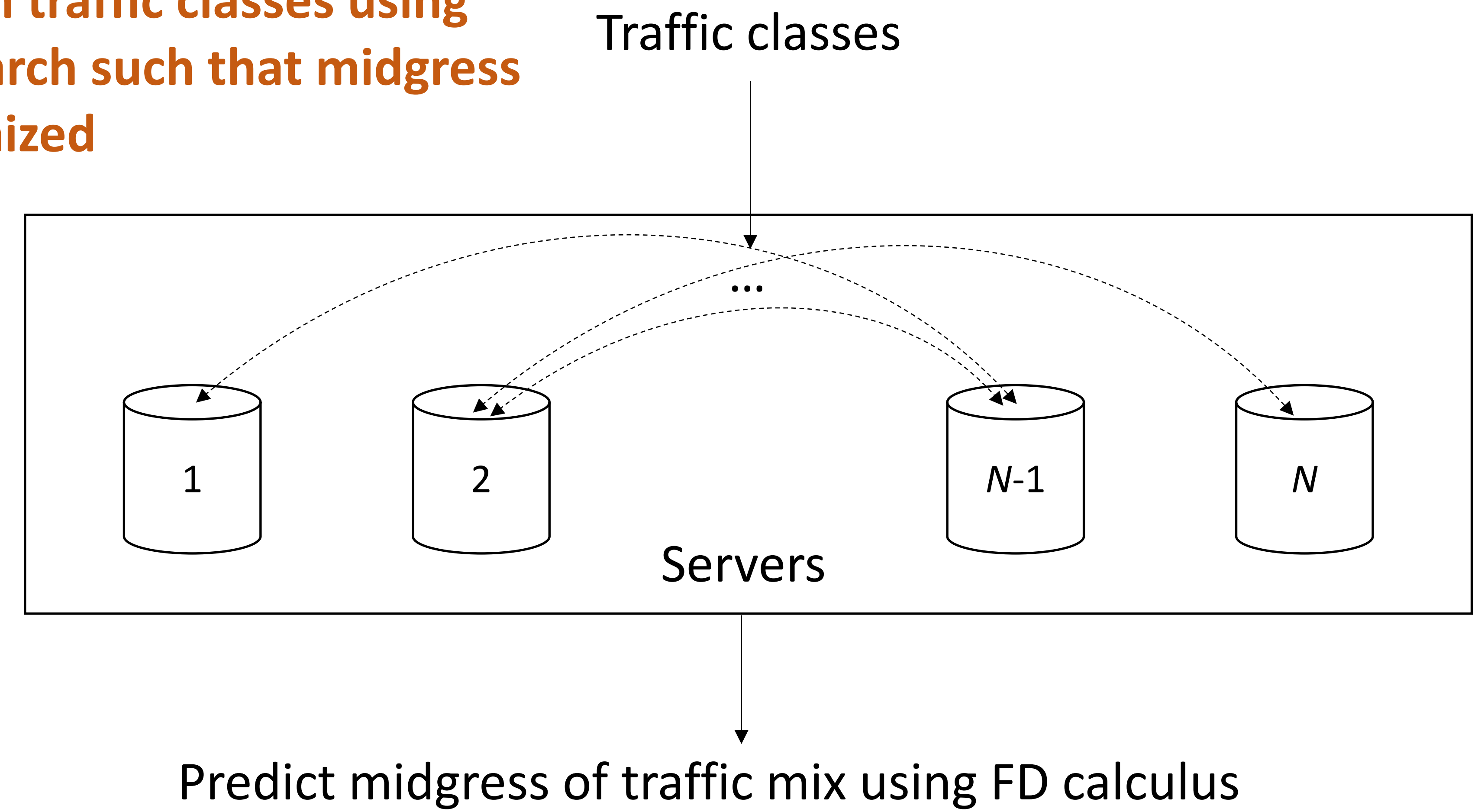
FD-based local search is faster than MILP

1. Randomly assign traffic classes

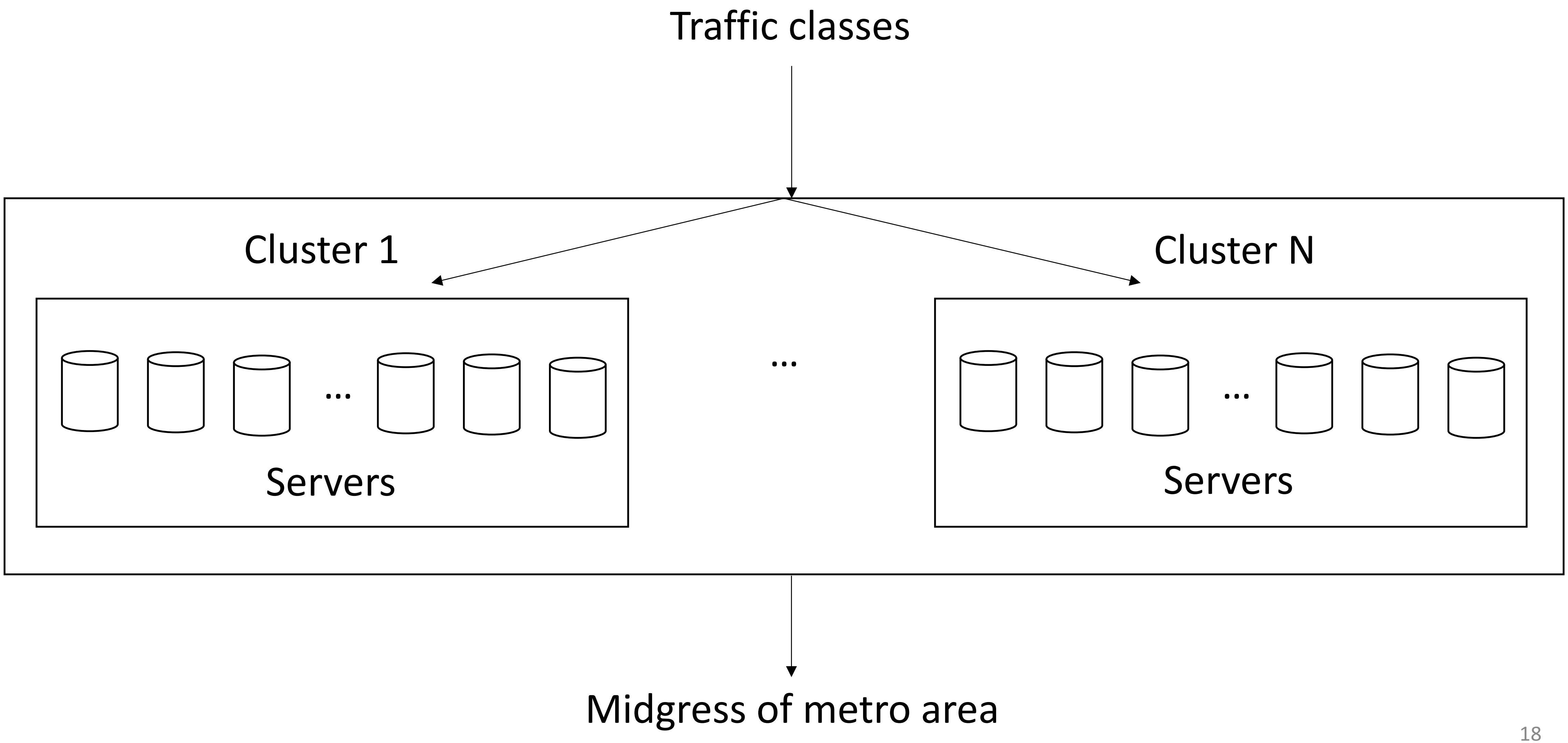


FD-based local search is faster than MILP

2. Reassign traffic classes using local search such that midgress is minimized



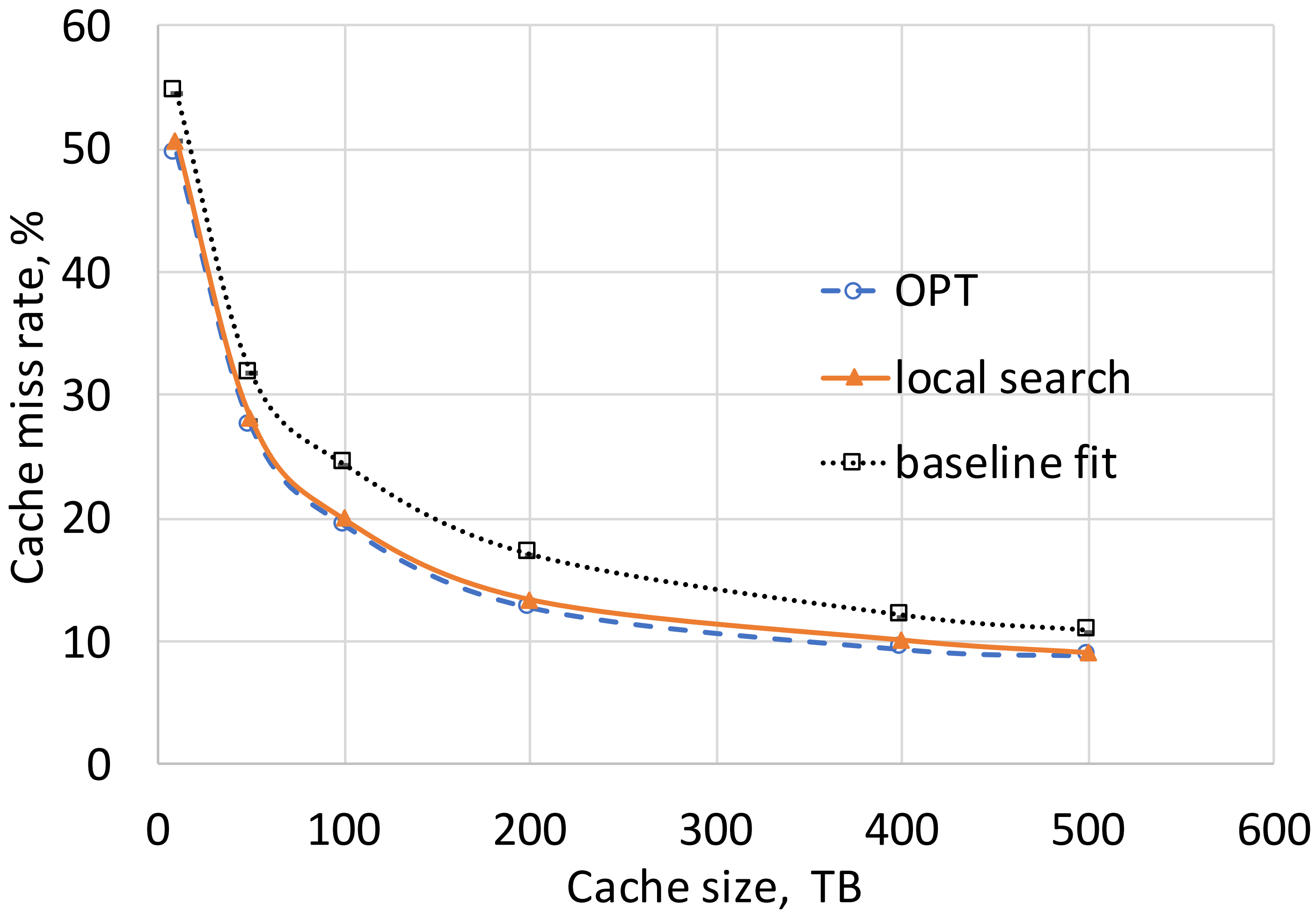
Metro-level traffic provisioning



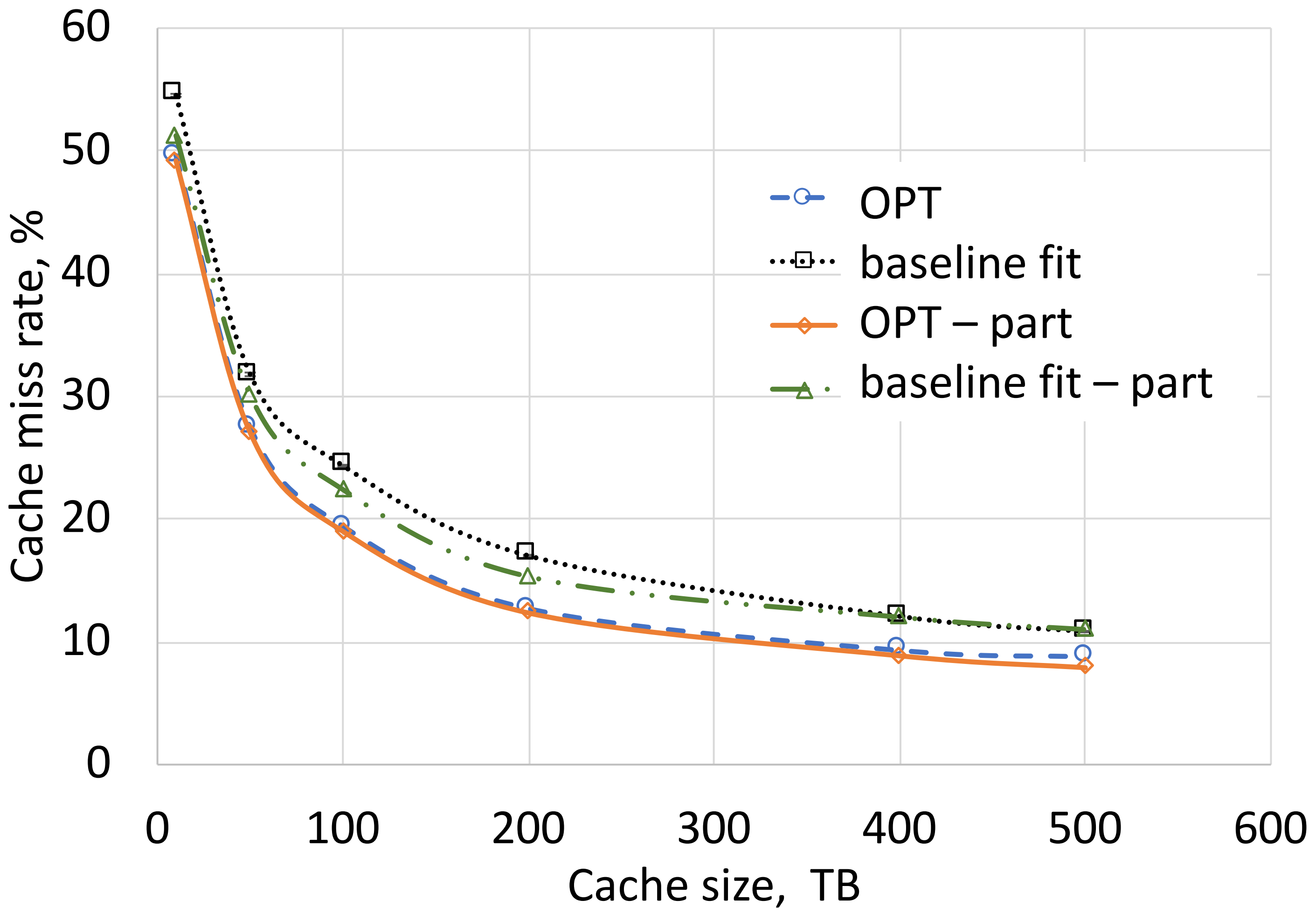
Trace characteristics

Number of traffic classes	25
Length of trace	16 days
Traffic types	Web, media, download

Metro-level midgress reduced by 20%



Traffic provisioning in partitioned caches



Conclusions

Midgress-aware traffic provisioning reduced midgress by almost 20% in metro area

Midgress-aware heuristic performs within 1.1% of OPT but is much faster

Midgress-aware traffic provisioning can be extended to work with additional constraints such as minimum redundancy and maximum midgress, any cache management algorithm, and with partitioned caches

Thank you!

Email: asundar@cs.umass.edu