EdgeBuffer: Caching and Prefetching Content at the Edge in the MobilityFirst Future Internet Architecture

RUTGERS
WINLAB | Wireless Information Network Laboratory

Feixiong Zhang, Yanyong Zhang, K.K. Ramakrishnan, Dipankar Raychaudhuri, ect

Motivation

- Dense deployment of wireless access networks: 3G, 4G, Wi-Fi, etc
- Transition across multiple access networks within one session, especially in urban vehicular environment
- How can we facilitate mobile content delivery across multiple networks? Can we do better than traditionally caching?

Our Approach

- MobilityFirst future Internet architecture
- EdgeBuffer: content caching and prefetching at the edge
  - Utilize edge storage as distributed caching system
  - Partition cache storage space for caching buffer and prefetch buffer
  - Network-level mobility prediction: guide prefetching

MobilityFirst

- Scalable, Flat Content Name Space
  - Human readable name → flat globally unique identifiers (GUID)
- Efficient Content Discovery
  - Content GUID → Content Location
  - GNRS: dynamic binding between GUID and locations
- Reliable Content Transport
  - Hop-by-hop storage-aware transport scheme

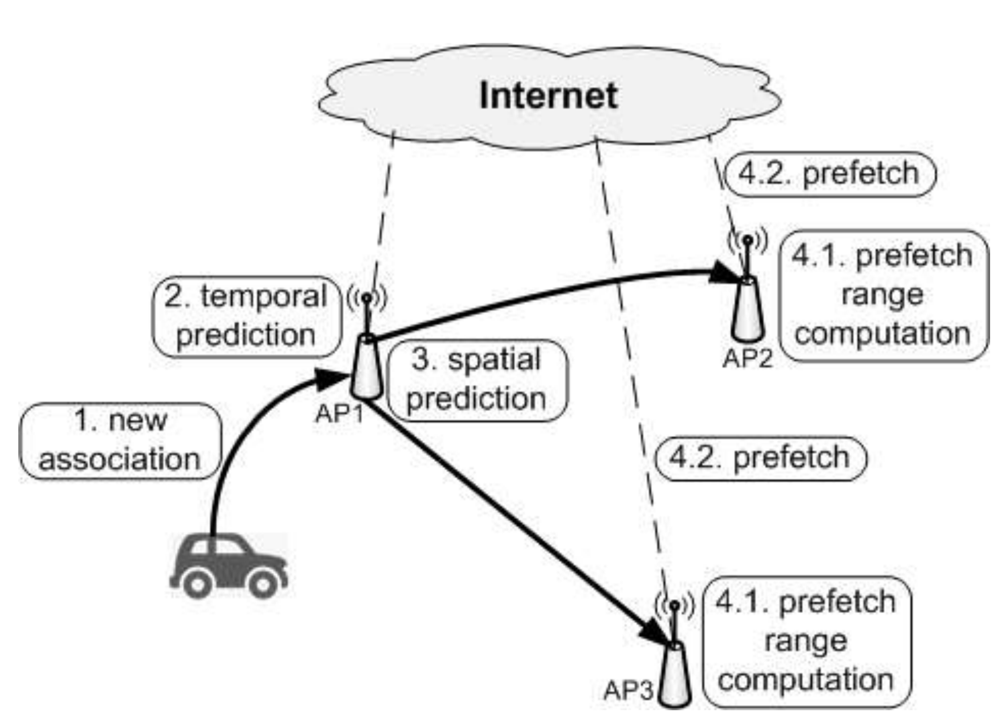

Fig. 2: Content prefetching protocol

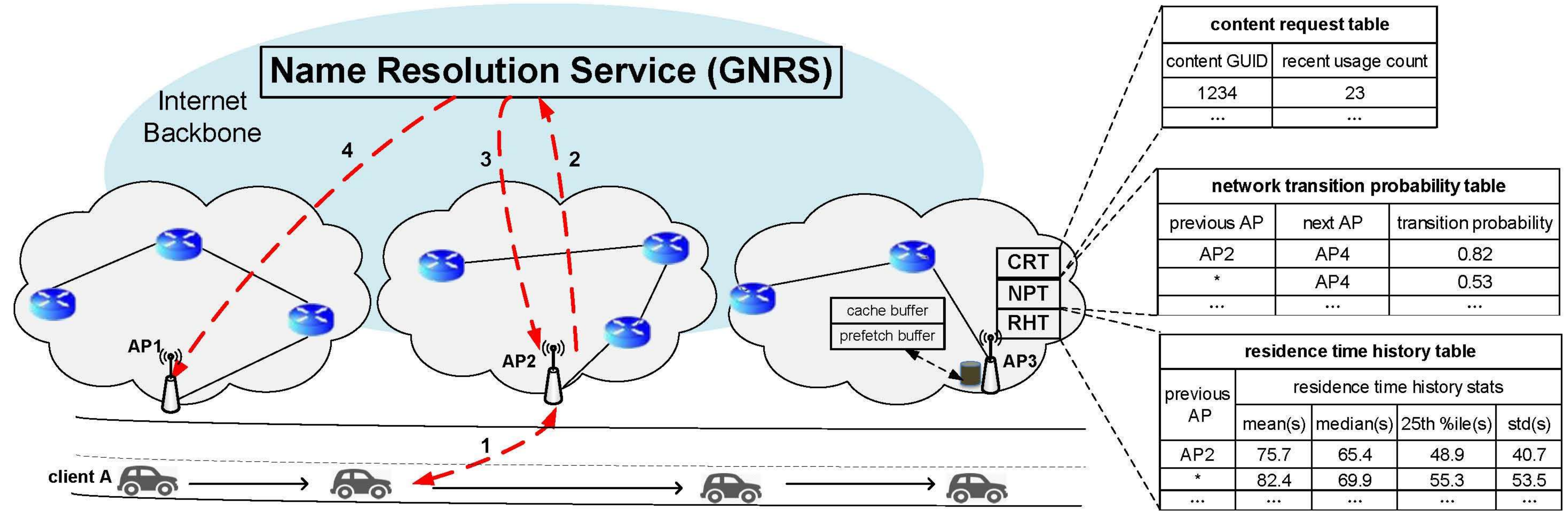

Fig. 1: EdgeBuffer: mobility prediction and content caching/prefetching framework in MobilityFirst

EdgeBuffer

- **Content caching and prefetching**
  - Utilize mobility information to enhance traditional caching
  - Separate cache buffer and prefetch buffer at AP
    - popularity-based content caching: capture long-term aggregated content access pattern
    - mobility prediction-based prefetching: capture short-term user access pattern utilizing its mobility information
- **Network-level mobility prediction**
  - Aggregate mobility information from all connected mobile devices, adapt to traffic variation.
  - Utilize GNRS
- **Mobility predictor at AP**
  - Each AP has its own predictor
  - Spatial prediction
    - Network transition probability table (NPT): second order Markov chain
    - Previous AP, current AP → next AP
  - Temporal prediction
    - Residence time history table (RHT)
    - Previous AP, current AP → residence time at current AP
- **Content prefetching protocol**
  - Association
  - Temporal prediction: AP1 predicts the client's residence time.
  - Spatial prediction: AP1 predicts the next AP(s) → sends out prefetching message with content ID and initial chunk ID.
  - Prefetch: next AP computes the prefetching range → conducts the prefetching.

Evaluation

- **Evaluation setting**
  - Discrete-event simulation
  - Mobile access workload:
    - San Francisco taxi trace + WiGLE AP database
    - 536 taxis, 3 weeks
    - Merge AP within certain distance: SF-250, SF-100
- **Evaluation metric**
  - Prediction accuracy
    - Spatial prediction accuracy
    - Temporal prediction accuracy
  - Hit ratio at the edge buffer
    - characterizes the performance gain of caching and prefetching regardless of different network settings
- **Different strategies**
  - LRU: pure LRU cache
  - popCache: pure popularity cache
  - popCache+userPredict: popularity cache + user-level prediction based prefetching
  - popCache+netPredict: our strategy

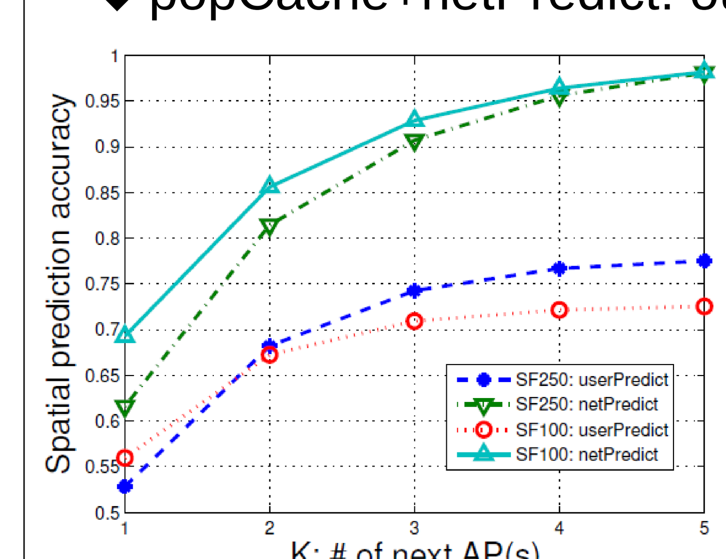

Fig. 3: Spatial prediction accuracy.

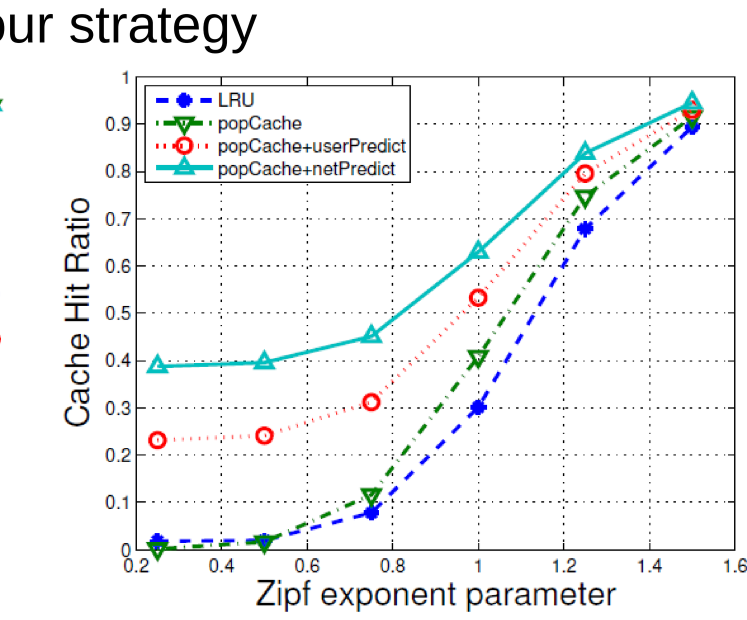

Fig. 4: Cache hit ratio for four edge caching strategies

WINLAB

Reference: EdgeBuffer: Caching and Prefetching Content at the Edge in the MobilityFirst Future Internet Architecture. IEEE WoWMoM 2015