Reinforcement Learning for 5G Caching with Dynamic Cost

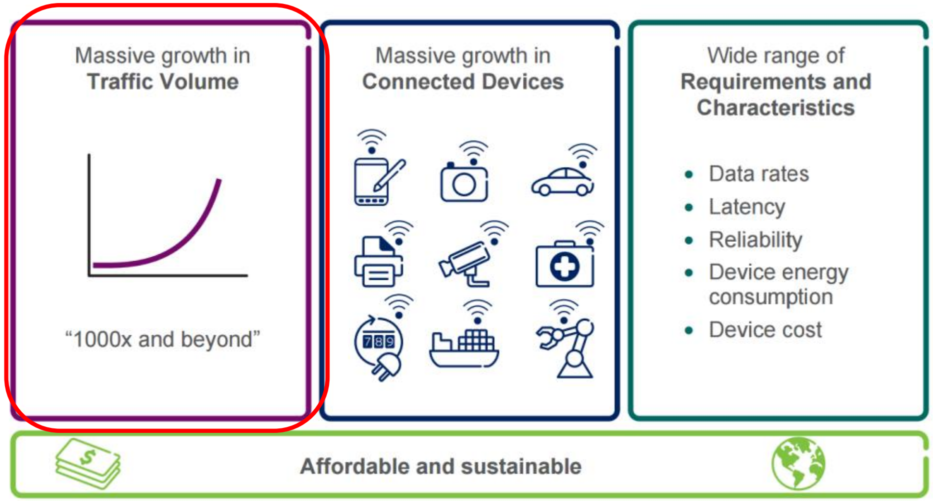
*A. Sadeghi, F. Sheikholeslami, **A. G. Marques** and G. B. Giannakis*

Acknowledgements: NSF grants 1423316, 1508993, 1514056, 1711471
Spanish MINECO grant OMICROM (TEC2013-41604-R)

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)
Calgary, Alberta, Canada, 15-20 April 2018

Challenges for 5G

- ❑ Evolving user profiles
  - ▪ Mobile video streaming
    - ✓ 50 % of traffic volume in 4G
    - ✓ 500-fold increase in 5G
  - ▪ Social networking
    - ✓ 15% of traffic volume in 4G
  - ▪ Music, games …



Massive growth in **Traffic Volume**

"1000x and beyond"

Massive growth in **Connected Devices**

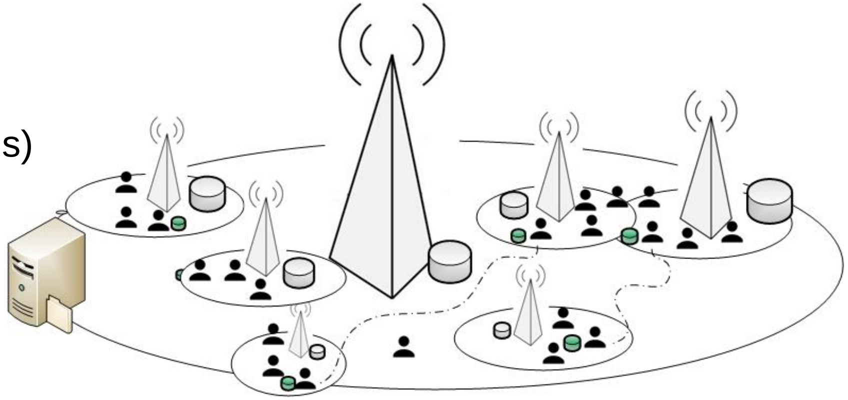
Wide range of **Requirements and Characteristics**

- Data rates
- Latency
- Reliability
- Device energy consumption
- Device cost

Affordable and sustainable

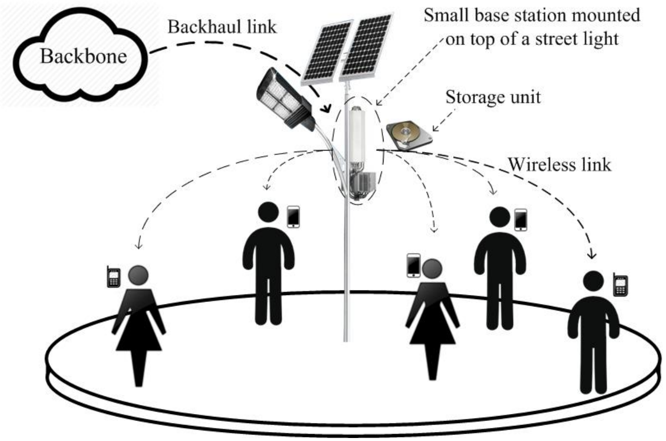
- ❑ 60% of data is reusable, a.k.a. ***contents***

- ❑ Heterogeneous network architecture (HetNet)
  - ▪ Utilization of **storage** units at small base stations (SBs)
    - ➢ Proactively store popular contents (**cache**)
  - ▪ Challenge: ***what*** and ***when*** to store?
    - ➢ Requires **learning** content popularities



G. Paschos, E. Bastug, I. Land, G. Caire and M. Debbah, "Wireless caching: technical misconceptions and business barriers," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 16-22, August 2016.

Caching in wireless networks



- Memory-enabled SBs
  - Cache during off-peak hours
  - Reduce load on backhauls during peak traffic periods
  - Reduce cost for providing service with high QoS

- Generally unknown content popularity profiles

- Prior art
  - Popularity profile learning [Blasco, Gunduz'15], [Baştuğ',Debbah, Saad16], [Bharath'16]
  - Multi-armed bandit (MAB) formulation [Belasco et al'14]
  - Distributed and convexified MAB [Sengupta et al'14]
  - Dynamic popularity profiles [Sadeghi et al'18]
  - Game-theoretic caching [Hamidouche, Debbah, Saad'16]
  - Coded caching [Maddah-Ali, Usir'16], [Alizadeh, Avestimehr'16], [Amiri, Gunduz'17]

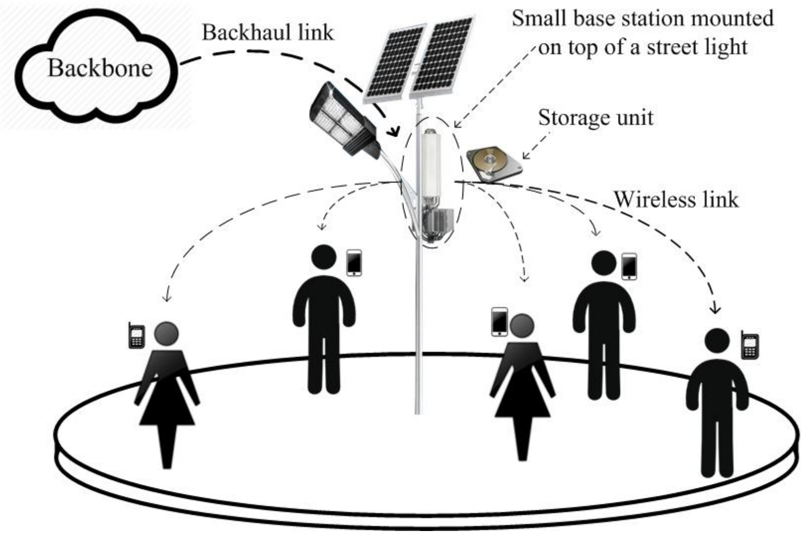
⟹ Unknown static popularity profile

⟹ Unknown dynamic popularity profile

Proposed approach

Caching via reinforcement learning while considering dynamic fetching-caching costs

Problem statement



❑ Discrete-time network

- Access point with storage to cache popular files
- Total number of F contents at the back-bone
- User content requests are served using
  - ✓ Proactively cached contents, or
  - ✓ Reactively fetched (via back-haul link) contents

❑ Pertinent costs for provisioning a request

- Storing locally in access point: processing, storage and energy consumption
- Fetching from cloud: scheduling, routing, transmission through expensive back-haul

Minimize sum-average cost by sequentially making fetching-caching decisions

Content-dependent variable costs

❑ **File- and size-dependent** caching-fetching costs

size dep. file dep.

- Cost of caching file *f* at slot *t* $c_t^f := \sigma_f({c'}_t + {c'}_t^f) + ({c''}_t + {c''}_t^f)$
- Cost of fetching file *f* at slot *t* $\phi_t^f := \sigma_f({\phi'}_t + {\phi'}_t^f) + ({\phi''}_t + {\phi''}_t^f)$

❑ Base-station receives user file request for file *f* at slot *t* $(r_t^f = 1)$

- Fetching-action variable $w_t^f \in \{0, 1\}$
- Caching-action variable $a_t^f \in \{0, 1\}$

Goal: Given pdf of iid $\{r_\tau^f, c_\tau^f, \phi_\tau^f\}_{\tau=t+1}^{\infty}$ and instantaneous values per slot *t*, find $\{w_\tau^{f*}, a_\tau^{f*}\}_{\tau=t+1}^{\infty}$

Fetch and cache via dynamic programing

- ❑ Cache-state variable $s_t^f \in \{0,1\}$ where $s_t^f = a_{t-1}^f$
- ❑ Constraints on fetch-cache decision variables $\{w_\tau^f, a_\tau^f\}_{\tau=t+1}^{\infty}$
  - File requests must be served (no drop-off allowed) $r_t^f \leq w_t^f + s_t^f, \quad \forall f, t$
  - Caching is feasible iff file is available $a_t^f \leq s_t^f + w_t^f, \quad \forall f, t$
- ❑ Fetch-cache cost $C^f(a_t^f, w_t^f; c_t^f, \phi_t^f) = c_t^f a_t^f + \phi_t^f w_t^f$

$$\min_{\{(w_\tau^f, a_\tau^f)\}_{f,\tau\geq t}} \bar{C}_t := \sum_{\tau=t}^{\infty} \sum_{f=1}^{F} \gamma^{\tau-t} \mathbb{E}\left[C^f\left(a_\tau^f, w_\tau^f; c_\tau^f, \phi_\tau^f\right)\right]$$

$$\text{s.t.} \quad (w_\tau^f, a_\tau^f) \in \mathcal{X}(r_\tau^f, s_\tau^f), \quad \forall f, \ \tau \geq t$$

$$\mathcal{X}(r_t^f, s_t^f) := \{(w,a) \mid w, a \in \{0,1\}, s_t^f = a_{t-1}^f, \ r_t^f \leq w + s_t^f, \ \ a \leq s_t^f + w\}$$

- ❑ Optimization separable across files!

Formulating per content optimization

- ❑ Given $(s_t^f, r_t^f, c_t^f, \phi_t^f)$

$$\boldsymbol{\theta}_t^f := \left[r_t^f, c_t^f, \phi_t^f\right]$$

$$\left(w_t^{f*}, a_t^{f*}\right) := \arg\min_{(w,a)\in\mathcal{X}(r_t^f,s_t^f)} \left\{ \mathop{\mathbb{E}}_{r_\tau^f,c_\tau^f,\phi_\tau^f;\tau>t} \left[\min_{(w_\tau,a_\tau)\in\mathcal{X}(r_\tau^f,s_\tau^f)} \left\{ \sum_{\tau=t}^{\infty} \gamma^{\tau-t} \left[C^f(a_\tau^f, w_\tau^f; c_\tau^f, \phi_\tau^f) \middle| a_t^f = a, w_t^f = w, \boldsymbol{\theta}_t^f = \boldsymbol{\theta}_0^f \right] \right\} \right] \right\}$$

$$= \arg\min_{(w,a)\in\mathcal{X}(r_t^f,s_t^f)} \left\{ c^f(a, w; c_t^f, \phi_t^f) + \mathop{\mathbb{E}}_{r_\tau^f,c_\tau^f,\phi_\tau^f;\tau>t} \left[\min_{(w_\tau,a_\tau)\in\mathcal{X}(r_\tau^f,s_\tau^f)} \sum_{\tau=t+1}^{\infty} \gamma^{\tau-t} \left[c^f(a_\tau^f, w_\tau^f; c_\tau^f, \phi_\tau^f) \middle| s_{t+1}^f = a \right] \right] \right\}$$

- ❑ Marginalized value function

$$\bar{V}^f(s^f) := \mathop{\mathbb{E}}_{r^f,c^f,\phi^f} \left[\min_{(w,a)\in\mathcal{X}(r^f,s^f)} \left\{ C^f(a, w; c^f, \phi^f) + \gamma \bar{V}^f(a) \right\} \right]$$

Value iteration algorithm

❑ Value iteration algorithm to find $\bar{V}^f(s^f)$

- **Input:** probability density functions of c^f, ϕ^f and r^f
- **Initialize:** $\bar{V}_0^f(s) = 0, \text{ for } s \in \{0,1\}$
- **While** $|\bar{V}_k^f(s) - \bar{V}_{k+1}^f(s)| < \epsilon; \quad \forall f, s \in \{0,1\}$

$$\bar{V}_{k+1}^f(s) = \mathbb{E}_{r^f, c^f, \phi^f} \min_{(w,a) \in \mathcal{X}(r^f, s)} \left\{ C^f(a, w; c^f, \phi^f) + \gamma \bar{V}_k^f(a) \right\}$$ for $s = 0, 1$

- **Output:** $\bar{V}^f(0), \bar{V}^f(1)$

❑ Optimal cache-fetch decisions given $(s_t^f, r_t^f, c_t^f, \phi_t^f)$

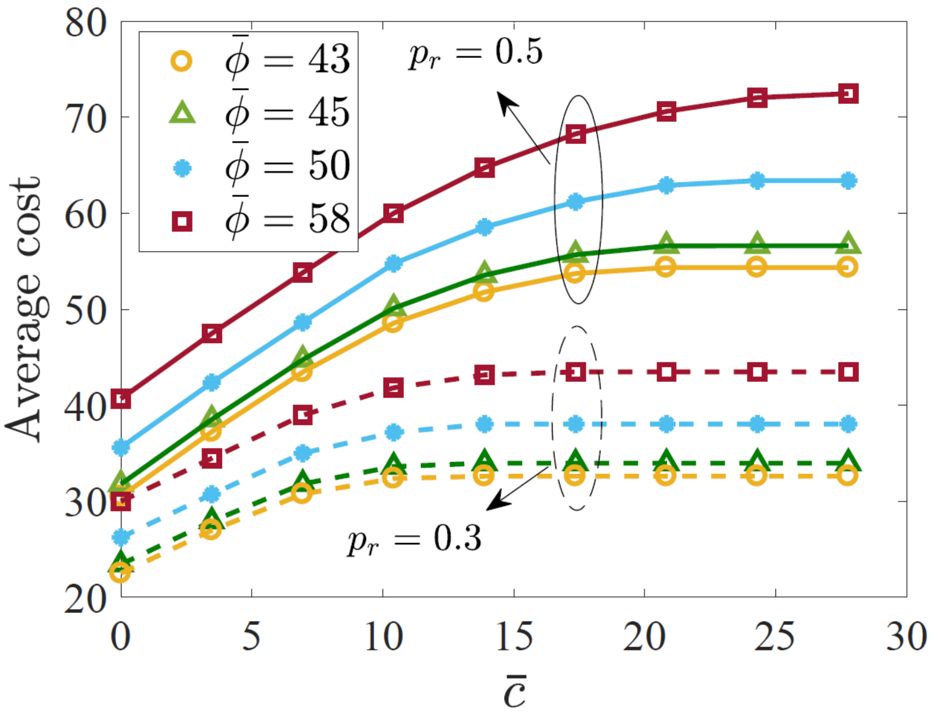
$$\left(w_t^{f*}, a_t^{f*}\right) = \arg\min_{(w,a) \in \mathcal{X}(r_t^f, s_t^f)} \left\{ C^f(a, w; c_t^f, \phi_t^f) + \bar{V}^f(a) \right\}$$

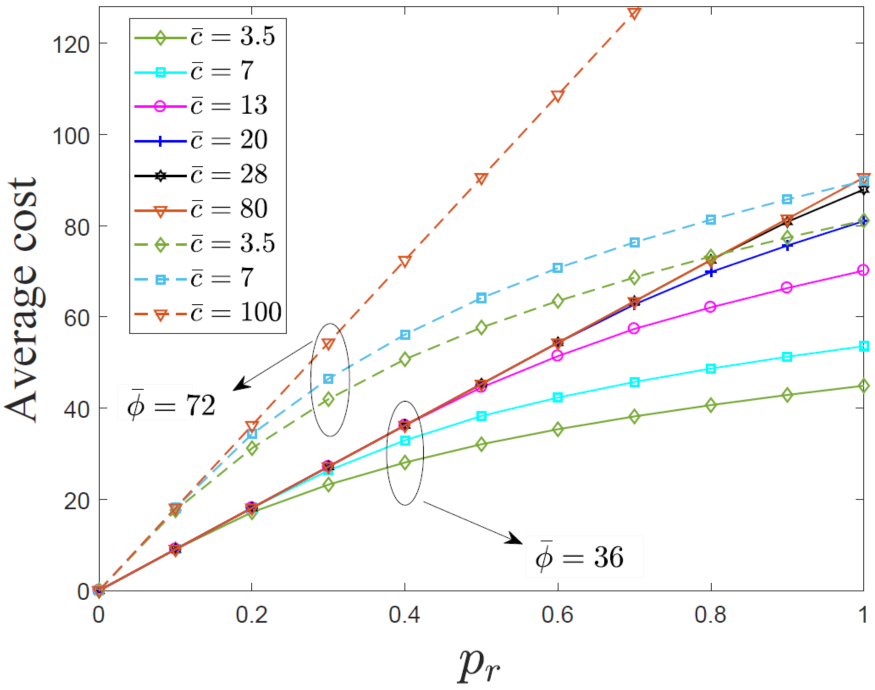
❑ Under iid assumption, marginalized value function takes binary input

$\Longrightarrow$ Fast convergence!

Numerical test with dynamic cost

- Per file *f* pdfs $c_t^f \sim \mathcal{U}(0, 2\bar{c}), \quad \phi_t^f \sim \mathcal{U}(0, 2\bar{\phi}), \quad r_t^f \sim \text{Bernoulli}\ (p_r^f)$



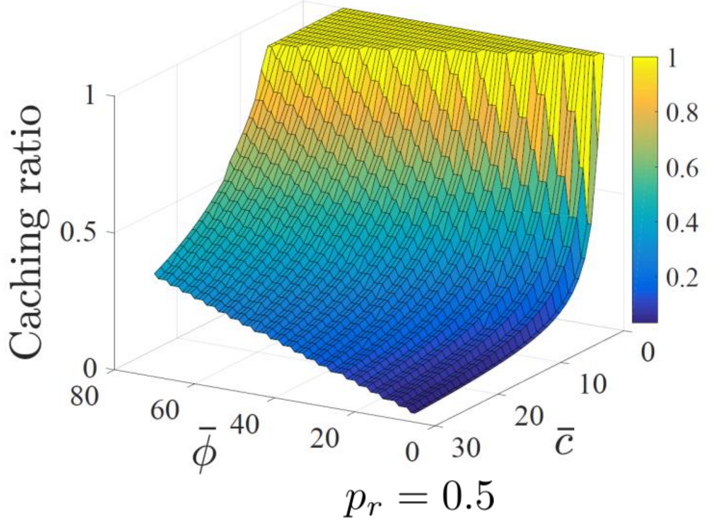


- ✓ For fixed $\bar{\phi}$, up to a certain value of $\bar{c}$, caching is encouraged
- ✓ Large p_r encourages caching for a larger range of $\bar{c}$
- ✓ Increased cost with more requests p_r

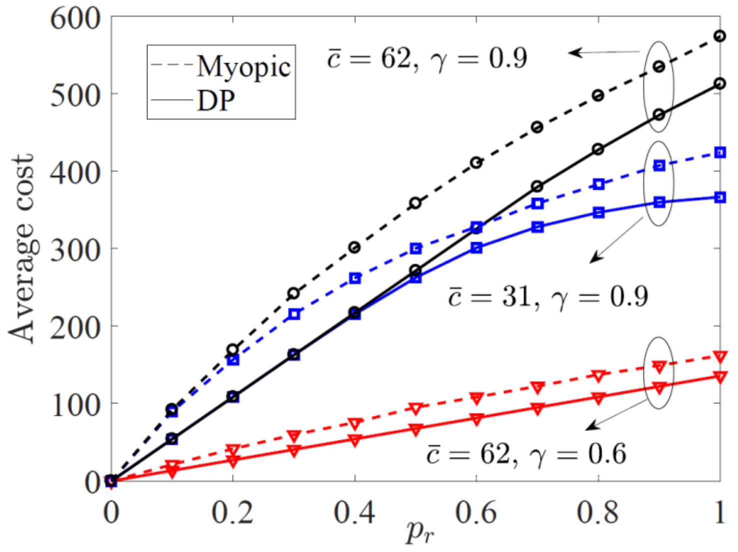
Further comparisons

- $c_t^f \sim \mathcal{U}(0, 2\bar{c}), \quad \phi_t^f \sim \mathcal{U}(0, 2\bar{\phi}), \quad r_t^f \sim \text{Bernoulli}\ (p_r^f)$

$$\text{caching ratio} := \frac{\text{No. of caching decisions}}{\text{Total No. of decisions}}$$



✓ DP vs. myopic caching



✓ Low $\bar{c}$, high $\bar{\phi}$ ⟶ caching ratio = 1 (flat area)

✓ High $\bar{c}$, low $\bar{\phi}$ ⟶ caching ratio = 0

✓ Intermediate values ⟶ 0< caching ratio <1

✓ DP considers future, thus reaches smaller average cost versus myopic caching

Future research and stakeholder analysis

❑ Multi-file caching considering queuing and cache refreshing costs

❑ Cooperative caching across neighboring small cells

❑ Cross-layer design of coded caching

❑ Privacy-preserving, secure, space-time variable caching

Thank you!