Content Placement Learning for Success Probability Maximization in Wireless Edge Caching Networks

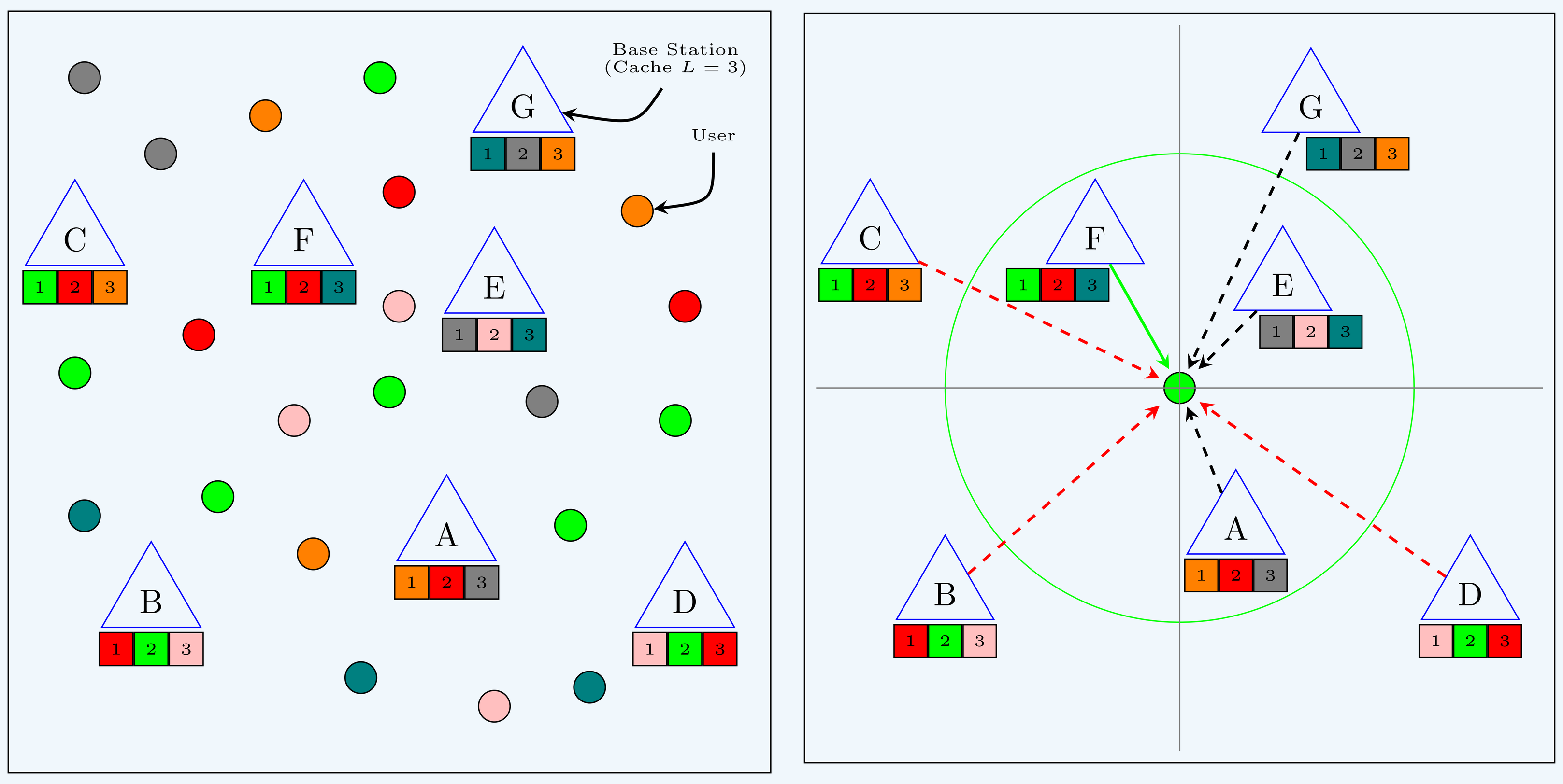
Navneet Garg[†], Mathini Sellathurai[†], Tharmalingam Ratnarajah[‡]

[†]Heriot-Watt University, Edinburgh, UK; [‡]The University of Edinburgh, UK.

Overview

- To handle the repeated requests at the base stations (BS), appropriate contents need to be cached in each time slot based on time-varying content popularity.
- Modeling content popularity as a finite state Markov chain [1] in a network with homogeneous Poisson point process (PPP) distributed BSs and users, reinforcement Q-learning is employed to learn optimal content placement probabilities to maximize the average success probability (ASP).

Physical Layer Model



(a) (b)

Figure 1: Homogeneous Poisson point process (PPP) distributed caching network in (a) and simplified with Silvanyak-Mecke Theorem in (b).

- Let $\mathbf{f} = [f_1, \ldots, f_N]^T$ be the global content popularity profile of the network and $\mathbf{a} = [a_1 \ldots a_N]^T$ be the content placement probabilities such that $\mathbf{a}^T\mathbf{1} \leq L$.
- From Slivnyak-Mecke theorem in Figure 1 (b), ASP at the typical user at the origin can be simplified (for interference limited case) as [2]

$$P_a(\mathbf{f}, \mathbf{a}) = \sum_{l=1}^{N} f_l \mathbb{E}_{\Phi_{BS}} \Pr\left\{W \log_2(1 + \Gamma(a_l)) \geq R_0\right\}$$
$$= \sum_{l=1}^{N} \frac{C f_l a_l}{a_l A + (1 - a_l)B + a_l C},$$

where $\Gamma(a_l)$ is the downlink SINR for the l^{th} file; A, B and C are physical layer constants depending on the density of PPP.

States, Actions and Cost

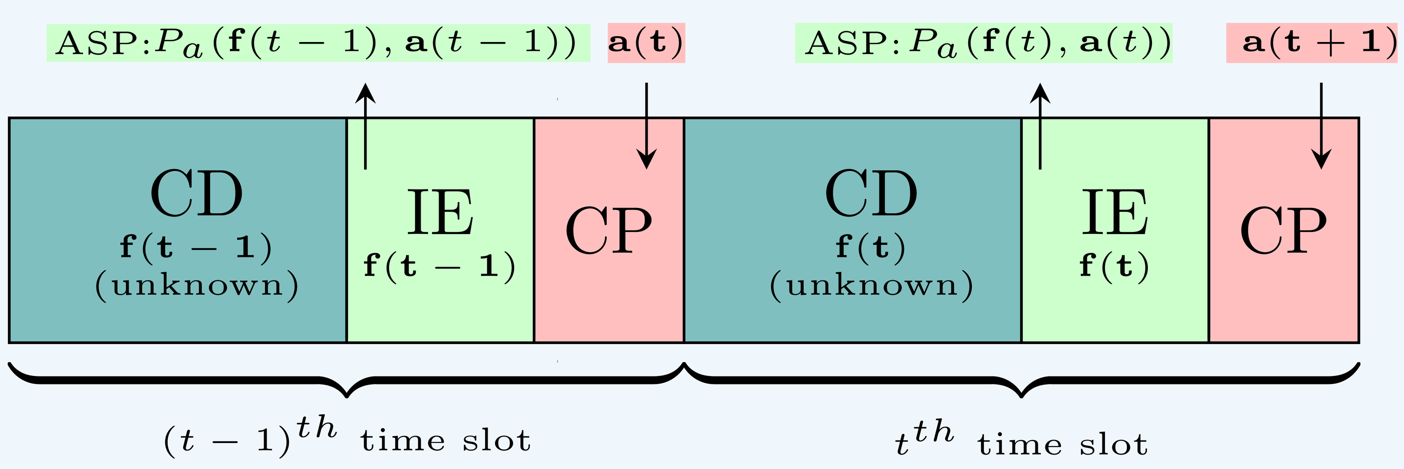


Figure 2: The time slot structure and the evolution of key quantities. The slots can be of unequal length. CD: Content Delivery, CP: Content placement, IE: Information Exchange.

- For t^{th} time slot, *state* of the network is the current content popularity profile and present content in the caches i.e.,
-
$$\mathbf{s}(t) = \begin{bmatrix} \mathbf{f}(t) \\ \mathbf{a}(t) \end{bmatrix} \in \mathcal{S} = \mathcal{F} \times \mathcal{A},$$
where $\mathcal{A} := \left\{\mathbf{a} | \mathbf{a} \in [0,1]^N, \mathbf{a}^T\mathbf{1} = L\right\}$ and $\mathcal{F} := \left\{\mathbf{f}_1, \ldots, \mathbf{f}_{|\mathcal{F}|}\right\}$.
- Based on the state $\mathbf{s}(t-1)$, *action* which is defined as the content placement probabilities of the network, $\mathbf{a}(t)$ is selected for the next time slot t.
- When $\mathbf{f}(t)$ is revealed, the *cost* of the action $\mathbf{a}(t)$ is computed in terms of ASP as
$$c(\mathbf{s}(t), \mathbf{a}(t)) = 1 - P_a(\mathbf{f}(t), \mathbf{a}(t)), \quad (1)$$
where

Acknowledgements

This work was supported in part by the U.K. Engineering and Physical Sciences Research Council under Grant EP/P009549/1 and EP/P009670/1 and in part by the U.K.-India Education and Research Initiative Thematic Partnerships under Grant DSTUKIERI-2016-17-0060 and UGCUKIERI 2016-17-058.

Reinforcement Learning

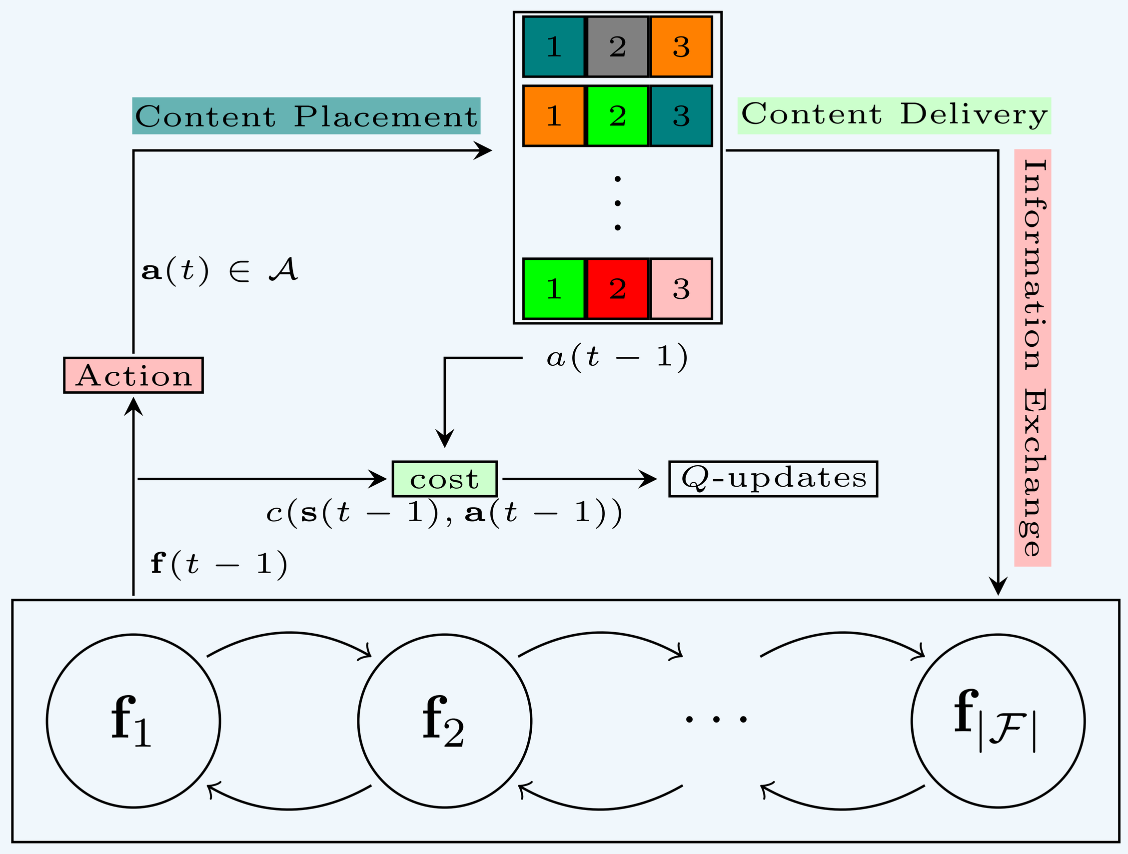


Figure 3: Learning model, where popularity varies according to Markov chain.

Q-learning Algorithm

Initialize state $\mathbf{s}(0)$ randomly and $Q_0(\mathbf{s}, \mathbf{a}) = 0 \forall \mathbf{s}, \mathbf{a}$

For $t = 1, 2, \ldots$

After *content delivery* at t time slot, do the following

Information Exchange: popularity profile $\mathbf{f}(t)$ is revealed based on user requests

Set $\mathbf{s}(t) = [\mathbf{f}^T(t), \mathbf{a}^T(t)]^T$, compute $c(\mathbf{s}(t), \mathbf{a}(t))$ and update

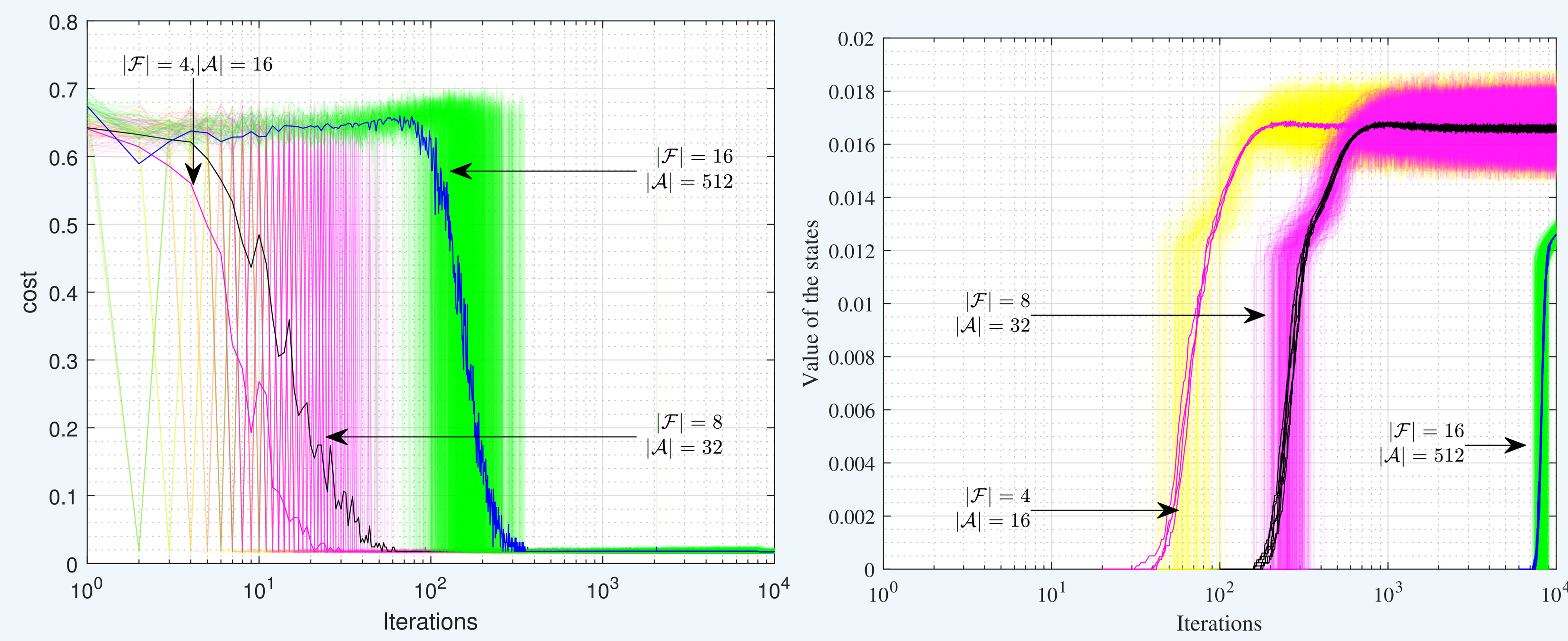
$$Q_t(\mathbf{s}(t), \mathbf{a}(t)) = (1 - \beta_t) Q_{t-1}(\mathbf{s}(t), \mathbf{a}(t)) + \beta_t \left[c(\mathbf{s}(t), \mathbf{a}(t)) + \gamma \min_{\mathbf{a}'} Q_{t-1}(\mathbf{s}(t), \mathbf{a}')\right]$$

Content Placement: take action $\mathbf{a}(t+1)$ for time $t+1$ chosen probabilistically

$$\mathbf{a}(t+1) = \begin{cases} \arg\min_{\mathbf{a}} Q_t(\mathbf{s}(t), \mathbf{a}), & \text{w.p. } 1 - \epsilon \\ \text{random } \mathbf{a} \in \mathcal{A}, & \text{w.p. } \epsilon \end{cases}$$

EndFor

Simulation Results



(a) (b)

Figure 4: (a) Cost and (b) state-value function vs iterations with three set of states and actions for $N = 100$, and $L = 20$.

Conclusion

- For a PPP based cellular network with global content popularities modeled as a finite state Markov chain, Q-learning method has been presented to find the optimal content placement probabilities.
- Simulations show that the Q-learning converges and learns the best content placement.

References

[1] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and scalable caching for 5G using reinforcement learning of space-time popularities," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 180–190, 2018.

[2] N. Garg, V. Bhatia, B. Bettagere, M. Sellathurai, and T. Ratnarajah, "Online learning models for content popularity prediction in wireless edge caching," *arXiv preprints*, 2019.