Mode Selection of Content Delivery At Edge Nodes Based On Learning

Yanli Xu (✉ ylxu@shmtu.edu.cn)

Shanghai Maritime University

Hao Tang

Shanghai Maritime University <https://orcid.org/0000-0001-5441-3183>

Research Article

Keywords: Content delivery, delivery mode selection, learning, multi-armed bandit (MAB)

Posted Date: December 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-848230/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Mode Selection of Content Delivery at Edge Nodes Based on Learning

Yanli Xu^{1*} and Hao Tang¹

^{1*}College of Information Engineering, Shanghai Maritime University, No.1550 Haigang Avenue, Pudong District, Shanghai, 201306, China.

*Corresponding author(s). E-mail(s): ylxu@shmtu.edu.cn;
Contributing authors: jack295072236@163.com;

Abstract

Pre-caching at edge nodes may improve resource efficiency for future content-centric cellular networks by bringing contents closer to users. However, which mode (multicast or unicast) should be selected for the efficient delivery of those cached contents is still not well addressed, especially at the situation that some key parameters such as content popularity and transmission environments are unknown. To solve this problem, the criterion of delivery mode selection is studied based on the learning policy of multi-armed bandit (MAB). According to the criterion, a mode selection algorithm is proposed. In this algorithm, edge nodes can choose the better delivery modes in the current slot only dependent on existing observations in the previous slot. Performance evaluation results validate our analyses and proposals on the mode selection.

Keywords: Content delivery, delivery mode selection, learning, multi-armed bandit (MAB)

1 Introduction

Recently, a dramatic increase of traffic load over cellular networks stimulates the investigation on offloading techniques. One of valid approach is content caching at network edge nodes, such as base stations (BSs) of macro and small cells (i.e., micro-cell, pico-cell, and femtocells) and user equipments (UEs).

Different from traditional cloud-based content delivery networks (CDN) [1], where the CDN components are located in the cloud and it may degrade the quality-of-experience (QoE) through initial delay and stall, edge (fog) stratum between the cloud and the end-user provides potential solution to efficient content retrieval by caching content near to users.

After caching, the most important step is content delivery which usually has two transmission modes, unicast and multicast. For unicast mode, content delivery follows a request-to-response mode where an edge node responses one user at a time. For example, in [2], contents are pre-cached at macro BSs (MBSs). The probability that video requests experience low initial delays is improved by video-aware backhaul and wireless channel scheduling. In [3], through content delivery by small cell BSs (SBSs), the authors design a content delivery strategy to find the best SBS to serve the requested UE. In [4], users requesting content are served by connecting to local wireless access points (APs) with cached content. Caching-and-delivery scheme is designed to get approximate optimality of tradeoff between transmission rate, storage, and access cost. In [5], contents can be pre-cached and delivered by UEs. The content delivery problem is formulated based on a flow model and delivery mechanisms are studied to provide quality-of-service (QoS) guaranteed service. For multicast mode, content delivery achieves high spectral efficiency by serving multiple users simultaneously when those users request the same content in the serving scope. In [6], content delivery is realized with selected peers receiving multicast content from some selected agents. The selection problem is formulated as an integer linear programming problem to be optimized. In [7], an optimal dynamic multicast method is proposed to balance the delay and spectral efficiency with the consideration of delivery cost such as the average network delay, backhaul cost, and transmission power. In [8], the authors propose a multicast delivery strategy to maximize the successful transmission probability. Content sharing schemes via device-to-device (D2D) multicast are studied to provide contents for users with worse cellular links to reduce energy consumption in [9, 10]. MBSs-based multicast delivery is studied in [11], where a BS pre-caches all contents that might be requested by the UEs and each UE caches a part of each file. Then, different requests can be satisfied by decoding the content from multicast streams.

Both of delivery modes have their advantages and disadvantages. Hence, a better understand of delivery mode selection is a preliminary step for efficient content delivery, which has few solutions. In our previous work [12], we compare two modes and indicate which one is better given concrete parameters such as network parameters, physical parameters and requested parameters. In practical, each edge node cannot or is reluctant to collect so many parameters before making a decision. In addition, closed-form expressions of performance metrics cannot be given when the distributions of some parameters are unknown. Therefore, we design online learning strategies for the mode selection of content delivery in this letter, which can select a set of delivery modes of requests without knowing content popularity and network parameters.

2 System Model

We consider a cellular network where edge nodes provide high data rate service to its users with cached files. When a user requests files not cached at its served edge node, the edge node will retrieve this file from data center and send it to its user. Since this retrieval will not affect the comparison between the delivery modes discussed here, we assume requested files are cached at edge nodes.

Denote the set of requested files to be $\mathcal{F}$. The size of file $f \in \mathcal{F}$ is S_f . Time is slotted and each time slot spans t . The instantaneous demand of f is d_f , which is an independent identically distributed (iid) random variable with mean θ_f . For simplicity, we assume identical file popularity over users. Considering that the file popularity distribution may change with spatiotemporal variation and some widely known models such as zipf are only suitable for describing the static characteristic of mass of nodes, we assume that the distribution of popularity is unknown at edge nodes. For each request, the served node responds it via two alternative delivery modes, i.e., multicast delivery and unicast delivery. There is a controller in each node, which decides the delivery mode for a file request. The objective is to maximize the expected total scheduling reward during an observed time period T by assigning the appropriate delivery mode to each request.

At per slot t , a controller selects an action x_t (a set composed by the delivery modes of all requests) from a convex set $X \in \mathbf{R}_d$ based on some given mode selection method (to be discussed in Section 3). At each time slot, the instantaneous reward is expressed by $h_{\mathcal{F}}(x_t)$. Denoting the instantaneous reward associated with file f as r_f , $h_{\mathcal{F}}(x_t)$ can be calculated by $h_{\mathcal{F}}(x_t) = \mathbf{E}[\sum_{\mathcal{F}} r_f]$. The available delivery resource such as power and bandwidth at per slot are bounded, which is denoted by a constant $\mathcal{R}_0$ here. Then the optimal problem of maximal expected total reward can be modeled by

$$\begin{aligned} & \max_{\{x_t \in \mathbf{X}, \forall t\}} \sum_{t=1}^T h_{\mathcal{F}}(x_t) \\ & \text{s.t.} \quad \sum_{f=1}^{|\mathcal{F}|} g_{\mathcal{F}}(x_t) \leq \mathcal{R}_0 \end{aligned} \tag{1}$$

where $g_{\mathcal{F}}(x_t)$ is the cost function which describes the total cost of delivering files for all requests with action x_t .

3 Learning on Content Delivery Mode

In this section, the online learning selection strategy is designed to obtain the maximal reward. Since the distribution of expected reward (shown in Eq. (1)) is unknown, a popular learning method, multi-armed bandit (MAB) [13] is used here to help make transmission decisions. MAB assumes that a machine has some arms. At each slot, the machine selects to pull one or several arms

to maximize the expected accumulated rewards overtime based on its current knowledge, while simultaneously acquires new knowledge. In our problem, an edge node can be seen as a machine and a delivery mode of content request can be seen as an arm. At each time slot, the machine chooses some arms to play, i.e., chooses the delivery mode of each request file, and acquires random rewards of the played arms. Which arms should be played in the given time slot are based on the outcomes of the arms played in the previous time slots, so that the expected reward during a time period is as close to that of the optimal arms as possible. Since the reward model is unknown, the essence of the MAB learning problem lies in the well-known tradeoff between exploitation (aiming at gaining the immediate reward by choosing arms that are suggested to be the best by past observations) and exploration (aiming at learning the unknown reward distribution by choosing new arms). Here the tradeoff policy uses combinatorial upper confidence bounds (CUCB), which chooses a set of arms to minimize the difference between its expected accumulated reward and that of the policy always pulling the best arms) [14]. Based on CUCB, an index is given to select a set of arms from all, which will be discussed latter in detail.

For each request, multicast mode m and unicast mode u have different cost (e.g., energy consumption and bandwidth occupation) and different rewards (e.g., the number of simultaneously satisfied requests). Under the delivery mode $\varphi = \{m, u\}$, we denote the reciprocal of cost and the reward of content delivery for f via action x_t as $\mathcal{C}_f^\varphi(x_t)$, and $r_f^\varphi(x_t)$, respectively. Then, per time slot t , the instantaneous reward can be written as $r_f(x_t) = \mathcal{C}_f^\varphi(x_t) d_f S_f$. Then the objective in Eq. (1) can be written as follows.

$$h_{\mathcal{F}}(x_t) = \mathbf{E} \left[\sum_{f \in \mathcal{F}} \mathcal{C}_f^\varphi(x_t) d_f S_f \right] = \sum_{f \in \mathcal{F}} \mathcal{C}_f^\varphi(x_t) \mathbf{E}[d_f] S_f = \sum_{f \in \mathcal{F}} \mathcal{C}_f^\varphi(x_t) \theta_f S_f \quad (2)$$

At each slot, $\mathcal{C}_f^\varphi(x_t)$ is determined by the selected transmission mode of f , which can be expressed by

$$\mathcal{C}_f^\varphi(x_t) = \beta_{f,t}^m \mathcal{C}_f^m + \beta_{f,t}^u \mathcal{C}_f^u \quad (3)$$

where $\beta_{f,t}^\varphi$ is the delivery factor, which is a binary number. For example, $\beta_{f,t}^m = 1$ if multicast delivery is used for the content request of f , otherwise $\beta_{f,t}^m = 0$. At each slot, we have $\beta_{f,t}^m + \beta_{f,t}^u = 1$.

At per time slot, edge node collects data requests from all its serving devices. Then the action set x_t will be determined, which can be expressed by $x_t := [x_{1,t}^\varphi, \dots, x_{f,t}^\varphi, \dots, x_{|\mathcal{F}|,t}^\varphi]$. The cost of f can be denoted by $1/\mathcal{C}_f^\varphi(x_t)$. Then the cost function $g_{\mathcal{F}}(x_t)$ can be written as

$$g_{\mathcal{F}}(x_t) = \frac{\beta_{f,t}^m}{\mathcal{C}_f^m} + \frac{\beta_{f,t}^u}{\mathcal{C}_f^u} \quad (4)$$

Then the constrain of Eq. (1) is written as follows.

$$\sum_{f=1}^{|\mathcal{F}|} \left(\frac{\beta_{f,t}^m}{\mathcal{C}_f^m} + \frac{\beta_{f,t}^u}{\mathcal{C}_f^u} \right) - \mathcal{R}_0 \leq 0 \quad (5)$$

Based on Eqs. (2) and (5), the optimization problem is re-written as follows.

$$\max_{\{x_t \in X, \forall t\}} \sum_{t=1}^T \sum_{f \in \mathcal{F}} \mathcal{C}_f^\varphi(x_t) \theta_f S_f \quad (6a)$$

$$\text{s.t.} \quad \sum_{f=1}^{|\mathcal{F}|} \left(\frac{\beta_{f,t}^m}{\mathcal{C}_f^m} + \frac{\beta_{f,t}^u}{\mathcal{C}_f^u} \right) - \mathcal{R}_0 \leq 0 \quad (6b)$$

$$\beta_{f,t}^m + \beta_{f,t}^u = 1, \forall f \in \mathcal{F} \quad (6c)$$

$$\beta_{f,t}^m = (1 - \beta_{f,t}^m) = 0, \forall f \in \mathcal{F} \quad (6d)$$

$$\beta_{f,t}^u = (1 - \beta_{f,t}^u) = 0, \forall f \in \mathcal{F} \quad (6e)$$

From the perspective of MAB, a set of arms x_t are chosen, and their instantaneous rewards are observed at each slot. To maximize obtained rewards during T , we can obtain an index set based on CUCB as $R(X) = \{\bar{r}_f^\varphi(x_t) : \forall f \in \mathcal{F}, x_t \in X\}$, where $\bar{r}_f^\varphi(x_t)$ is the index of f and can be expressed by

$$\bar{r}_f^\varphi(x_t) = \widehat{r}_f^\varphi(x_t) + \sqrt{\frac{3 \ln t}{2N_f^\varphi}} \quad (7)$$

where $\widehat{r}_f^\varphi(x_t)$ is the observed rewards of f delivered via φ . N_f^φ is the times of using φ for f .

Next, based on the estimated $\bar{r}_f^\varphi(x_t)$, the set of arms which will maximize the accumulated reward during T is selected at each slot. According to the CUCB policy, the set is composed by arms which maximize the estimated reward $\beta_{f,t}^\varphi \bar{r}_f^\varphi(x_t)$ at each slot. Thereby, the original optimization problem during T can be transformed into the optimization problem at slot t as follows.

$$\begin{aligned} & \max \sum_{f \in \mathcal{F}} \beta_{f,t}^\varphi \bar{r}_f^\varphi(x_t) \\ & \text{s.t.} \quad (6b) - (6e) \end{aligned} \quad (8)$$

It is a knapsack problem and NP-hard, which can be solved optimally by using branching algorithms, such as branch and bound (B&B). However, the complexity of B&B in the worst case is exponential, same as exhaustive search. Thus, we use relaxation here for this integer linear programming (ILP) problem by relaxing the binary constraints on $\beta_{f,t}^\varphi$ to $0 \leq \beta_{f,t}^\varphi \leq 1$. Then Eq. (8) becomes a linear program (LP), and can be solved in polynomial time. We can sort $R(X)$ such that $\bar{r}_i^\varphi > \bar{r}_j^\varphi$, if $i < j$, $\forall i, j \in \mathcal{F}$ and the corresponding

coefficient set $\mathbf{B}(X) = \{\beta_{1,t}^u, \beta_{1,t}^m, \dots, \beta_{i,t}^u, \beta_{i,t}^m, \dots, \beta_{2|\mathcal{F}-1|,t}^u, \beta_{2|\mathcal{F}|,t}^m\}$. Since only one transmission mode is used for a file request, we firstly refine $\mathbf{B}(X)$ by removing redundant transmission mode for each file. Then the mode with larger reward will be left for each file and new coefficient set can be expressed by $\tilde{\mathbf{B}}(X) = \{\tilde{\beta}_{1,t}, \dots, \tilde{\beta}_{i,t}, \dots, \tilde{\beta}_{|\mathcal{F}|,t}\}$. The corresponding reward of i can be obtained by $\bar{r}_i = \max\{r_i^u, \bar{r}_i^m\}$.

Solving the LP-relaxation problem, the solution has the following structure $\tilde{\mathbf{B}}(X) = \{1, 1, \dots, 1, \alpha, 0, 0, \dots, 0\}$, where $\alpha = (\mathcal{R}_0 - \sum_{i=1}^{i=M-1} \bar{r}_i) / \bar{r}_M$ and M is the coordinate of α in $\tilde{\mathbf{B}}(X)$. Here α is the single non-integer element in $\tilde{\mathbf{B}}(X)$, it will be approximated integer to approximate the solution to Eq. (8). Let $\mathbf{B}'(X)$ approximate to the optimal solution, which greedily delivery files sequentially, starting from the file with the higher reward, where $\bar{r}_f(x_t) = \max\{r_f^u(x_t), r_f^m(x_t)\}$, until the available transmission resources at the current slot has been assigned completely. As defined in [15], the ratio between the optimal performance and greedy performance is δ , which satisfies $\delta \leq 1 + \max\{r_f^u(x_t)\} / \sum_{f \in \mathcal{F}} \beta_{f,t}^\varphi r_f^\varphi(x_t)$. We can see that $\delta \rightarrow 1$ for the larger number of files, i.e., the performance of the given approximate scheme will be very similar to that of optimal scheme. With the calculated $\mathbf{B}'(X)$, the selected arms, i.e., delivery modes of request files at a slot can be determined by edge nodes. The selection algorithm is given in Algorithm 1.

Algorithm 1 The Mode Selection Scheme of Content Delivery

Initialize: Delivery each content at least once, observe the rewards

Output: The delivery mode set x_t of requests in the current slot

- 1: **for** a slot $t \leq T$
 - 2: Observe $r_f^\varphi(x_{t-1})$
 - 3: Update $\widehat{r}_f^\varphi(x_t) \leftarrow \frac{\widehat{r}_f^\varphi(x_{t-1}) * (t-1) + r_f^\varphi(x_t)}{t}$ and $t \leftarrow t + 1$
 - 4: Calculate $\overline{r}_f^\varphi(x_t) = \widehat{r}_f^\varphi(x_t) + \sqrt{\frac{3 \ln t}{2N_f^\varphi}}$
 - 5: Substitute the updated $\overline{r}_f^\varphi(x_t)$ in Eq. (8) and calculate $\tilde{B}(X)$, i.e., obtain the delivery mode set
 - 6: Deliver request contents via the selected modes according to $\tilde{B}(X)$
 - 7: **end for**
-

4 Performance Evaluation

In this section, we present simulation results to evaluate our proposed theory and algorithm. We assume that the size of each content S_f is randomly chosen from $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ units. In each time slot, requests are produced according to a Zipf distribution. Generally, multicast may use more resource

(e.g., resource blocks (RBs)) than unicast to cover the request with the worst communication channel. Here the cost of unicast and multicast are randomly selected from sets $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $\{0.3, 0.4, 0.5, 0.6\}$, respectively. To evaluate the proposed algorithm, two upper bound algorithms are compared. In the first upper bound algorithm (UB1-exhaustive), we assume that the popularity profile is known and exhaustively search is used for optimization in each slot. In the second upper bound algorithm (UB2-greedy), we assume that the popularity profile is known and greedily search is used for optimization. For comparison, random algorithm is also considered, which randomly delivers requested files until the delivery resource is full.

Figure 1 compares instantaneous rewards among different schemes. It is shown that the proposed algorithm achieves 90% performance of exhaustive algorithm and 95% performance of greedy algorithm. The excellent performance of proposed algorithm do not need complicated computation and file popularity profile like exhaustive and greedy algorithms. Compared to the random scheme which also does not need pre-knowledge of popularity and complicated computation, the performance of proposed algorithm is about 30 times better than that of random scheme.

Fig. 1 Instantaneous rewards of different delivery algorithms. The user number is 100, the popularity exponent is 1 and the file number is 20

Figures 2 and 3 compare rewards based on unicast, multicast and the proposed scheme. Compared Figures 2 and 3, it is shown that both unicast and multicast have their appropriate applicable situations, e.g., multicast is more efficient when there are some most popular files. However, we can see that the proposed scheme performs much better than only using unicast or multicast for delivery (e.g., the performance of the proposed scheme can be more than 10 times better than multicast and unicast). It is because that the proposed scheme can decide a better delivery mode for each requested content by learning without knowing concrete channel quality and network deployment.

Fig. 2 Instantaneous rewards of different delivery algorithms. The user number is 100, the popularity exponent is 1 and the file number is 20

Fig. 3 Instantaneous rewards of different delivery algorithms. The user number is 1000, the popularity exponent is 2 and the file number is 20

5 Conclusion

In this letter, we have studied learning-based mode selection to make the content delivery in cellular networks more resource efficient. The solution to selecting a set of delivery modes for requests during a time period is presented, which only depends on the observed instantaneous demands and rewards of content requests without knowing many affecting parameters such as network parameters and file popularity distribution. In the solution, a selection criterion based on multi-armed bandit (MAB) is given. With this criterion, the content delivery of proposed scheme can achieve a sub-optimal performance without exhaustive or greedy search. The performance evaluation shows that the performance of proposed scheme can achieve 90% and 95% performance

of optimal solution and greedy-search solution. In addition, it performs better than multicast or unicast content delivery, e.g., the performance of the proposed scheme can be more than 10 times better than multicast and unicast.

Acknowledgement. This work was partially supported by the National Natural Science Foundation of China (No.61601283, U1701265, No.61472237 and No.61271283).

Declarations.

Conflict of interest. The authors declare that they have no conflict of interest.

References

- [1] Wang, M., Jayaraman, P.P., Ranjan, R., Mitra, K., Zhang, M., Li, E., Khan, S., Pathan, M., Georgeakopoulos, D.: An overview of cloud based content delivery networks: Research dimensions and state-of-the-art. *Lecture Notes in Computer Science* **9070**, 131–158 (2016)
- [2] Ahlehagh, H., Dey, S.: Video-aware scheduling and caching in the radio access network. *IEEE/ACM Transactions on Networking* **22**(5), 1444–1462 (2014)
- [3] Pantisano, F., Bennis, M., Saad, W., Debbah, M.: Match to cache: Joint user association and backhaul allocation in cache-aware small cell networks. In: *IEEE International Conference on Communications*, pp. 3082–3087 (2015)
- [4] Hachem, J., Karamchandani, N., Diggavi, S.: Content caching and delivery over heterogeneous wireless networks. In: *2015 IEEE Conference on Computer Communications (INFOCOM)*, pp. 756–764 (2015)
- [5] Xu, Y., Liu, F.: Qos provisionings for device-to-device content delivery in cellular networks. *IEEE Transactions on Multimedia* **19**(11), 2597–2608 (2017)
- [6] Zulhasnine, M., Huang, C., Srinivasan, A.: Exploiting cluster multicast for P2P streaming application in cellular system. In: *Proc. 2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 4493–4498 (2013)
- [7] Zhou, B., Cui, Y., Tao, M.: Optimal dynamic multicast scheduling for cache-enabled content-centric wireless networks. *IEEE Transactions on Communications* **65**(7), 2956–2970 (2017)

- [8] Cui, Y., Jiang, D., Wu, Y.: Analysis and optimization of caching and multicasting in large-scale cache-enabled wireless networks. *IEEE Transactions on Wireless Communications* **15**(7), 5101–5112 (2016)
- [9] Militano, L., Condoluci, M., Araniti, G.: When d2d communication improves group oriented services in beyond 4g networks. *Wireless Networks* **21**(4), 1363–1377 (2015)
- [10] Xia, Z., Yan, J., Liu, Y.: Energy efficiency in multicast multihop d2d networks. In: 2016 IEEE/CIC International Conference on Communications in China (ICCC), pp. 1–6 (2016)
- [11] Maddah-Ali, M.A., Niesen, U.: Fundamental limits of caching. *Information Theory IEEE Transactions on* **60**(5), 1077–1081 (2012)
- [12] Xu, Y., Wang, S.: Mode selection for energy efficient content delivery in cellular networks. *IEEE Communications Letters* **20**(4), 728–731 (2016)
- [13] Bubeck, S., Cesabianchi, N.: Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations & Trends in Machine Learning* **5**(1), 101–112 (2012)
- [14] Chen, W., Wang, Y., Yuan, Y.: Combinatorial multi-armed bandit: general framework, results and applications. In: International Conference on International Conference on Machine Learning, p. 151 (2013)
- [15] Korbut, A.A., Sigal, I.K.: Exact and greedy solutions of the knapsack problem: the ratio of values of objective functions. *Journal of Computer & Systems Sciences International* **49**(5), 757–764 (2010)

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Yanli Xu. received the Ph.D. degree from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2012. From 2012 to 2015, she worked as a research scientist in Bell labs of Alcatel-Lucent, China, with a focus on the standardization of 5G cellular networks. Now she is an associate professor of Shanghai Maritime University. She is an associate editor of *IEEE Access*, and serves as a TPC member or reviewer for many conferences and journals, such as *IEEE JSAC*, *TCOM*, *TWC*, *TSP*, *IET Com*, *Globecom*, etc. Her research interests are in 6G, maritime communications, IoT and intelligent driving, with a current focus on edge computation.

Hao Tang. received the B.S. degree in computer science and technology from Hainan Normal University, Haikou, China, in 2017. He is currently pursuing the M.S. degree in communication and information engineering with Shanghai Maritime University. His research interests include time-sensitive networking, and intelligent driving.