Blockchain-Enabled Intelligent Video Content Caching for D2D Mobile Networks

Hardhik Mohanty, Riya Tapwal, and Sudip Misra, *Senior Member, IEEE*

Department of Electrical Engineering, Indian Institute of Technology Kharagpur
Kharagpur 721302, India
Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur
Kharagpur 721302, India
Email: hardhikiitkgp2018@gmail.com, tapwalriya@kgpian.iitkgp.ac.in, sudipm@iitkgp.ac.in

***Abstract*—Modern wireless networks have witnessed an enormous rise in the demand for high-quality video streaming. For reducing the overhead of backhaul networks, device-to-device (D2D) communication enables mobile users to share video content with other devices within a short transmission range to reduce network latency. While mobile edge caching provides computational resources in the vicinity of the end-users, they face the following shortcomings: Mobile users are resource-limited in terms of cache storage. The mobile edge caching framework becomes more complex and dynamic when user mobility is considered. Mobile users lack an incentive to share video content due to the absence of a reward mechanism. To jointly tackle the above-mentioned issues, we utilize a vector auto regression (VAR) based content placement strategy for mobility-aware and blockchain (BC)-enabled D2D content sharing. VAR estimates the popularity of video content by modeling the previous requests as a multivariate time series function. BC network provides security against malicious mobile users and introduces a reward mechanism for building incentives. The mobile user is paid using ethereum coins in exchange for sharing the requested content. Simulation results indicate that the VAR model outperforms the baseline caching strategies on the basis of cache hit rate, latency, and profit. In particular, the VAR model obtains an improvement of 10%-20% in cache hit rate. Also, the VAR model achieves a low latency of around 5-8 milliseconds and obtains maximum profit under different content popularity distributions.**

***Index Terms*—D2D caching, blockchain, content placement, mobility aware, vector auto regression, ethereum.**

I. Introduction

According to the reports by Cisco [1], a dominant 75% part of the future network traffic will consist of high-quality video streaming. The video content requested by mobile users originates from content providers (CPs) situated in remote clouds. However, most of the network traffic is due to video content that is repeatedly requested. Moreover, applications like augmented reality, real-time mobile games, and smart factory manufacturing have low latency requirements. A practical method to achieve low latency is caching the video contents in edge cache nodes located close to the end mobile user. Edge caching of video contents improves the quality of experience (QoE) and reduces network delay in mobile users.

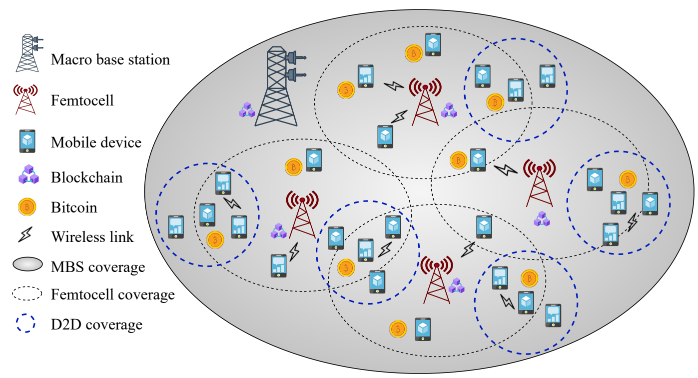


Figure 1: System architecture.

However, edge nodes have limited cache storage while the number of distinct video contents requested by mobile users is vast. As a result, the mobile user must devise an effective content placement strategy to cache video content that mobile users and their neighbors are likely to request. The existing cache replacement techniques include rule-based algorithms like least frequently used (LFU), first-in-first-out (FIFO), and least recently used (LRU). However, these algorithms perform poorly in such a dynamic request scenario due to their oversimplified approach and limited knowledge of content popularity. In this regard, recent works [2] employ deep neural network (DNN) for content popularity prediction due to their ability to understand long-term dependencies better than rule-based algorithms. Nevertheless, training a DNN needs a massive data for the model to converge. In addition, DNN model training is computationally expensive in terms of energy and storage. So the resource constraint characteristic of mobile users restricts the deployment of DNN models to these end devices.

The majority of previous works on video content caching did not consider the mobility patterns of mobile users. Mobility is an inherent characteristic of D2D communication, making the network architecture dynamic. In the mobility-aware network topology, mobile user opportunistically tries to fetch the requested contents through D2D networks. So the mobile user might encounter different mobile users within its D2D transmission range over time. These properties of user mobility aware D2D communication make video content placement a complex and heterogeneous problem. Consequently, it becomes critical to leverage mobility patterns for developing efficient content caching algorithms [3]. Another challenge in deploying edge caching nodes and D2D networks involves securing the framework against cyberattacks. Malicious attackers might compromise some of the edge caching nodes. The attackers might then use these compromised nodes to remove the cached contents or return malware in place of the requested content. Furthermore, edge caching nodes compromised by malicious attackers might intentionally deny paying for the caching service. In general, mobile users are interested in participating in the D2D content sharing process only if there is some advantage or reward. In this regard, blockchain (BC) technology safeguards the edge caching framework from cyberattacks using cryptography techniques. BC can also help build incentives among edge caching nodes to share video content in exchange for ethereum coins.

This work tackles the above issues by investigating the mobility-aware content placement problem in BC-enabled D2D networks. The inter-contact model defines the user mobility patterns in D2D networks. In this mobility model, the user contact process follows a Poisson process. To predict the content popularity in such a dynamical system, we utilize a vector autoregression (VAR) model [4] with a softmax final layer for popularity prediction and selection of the video contents to be placed in the cache. VAR is used for predicting the future popularity of the video content. Initially, if the video content requested by the mobile user is available in one of its neighboring mobile users, it will fetch the video contents via D2D communication. Or else, the request is forwarded to the edge caching node and fetched via wireless transmission. The proposed edge caching framework integrates BC technology employing the Proof of stake (PoS) consensus mechanism to build mobile users' incentives to share video content. The mobile user providing the requested video content is awarded ethereum coins in exchange for the service. The validator node is selected via the consensus mechanism to verify all the caching transactions. In this way, BC can secure the edge caching framework against cyber attacks and efficiently manage the cache transactions. In conclusion, the following are the key contributions of this research work:

1) We devise BC-enabled D2D caching framework considering user mobility to model modern wireless networking conditions. Furthermore, the edge caching nodes provide backup to the D2D content sharing process.
2) We utilize a vector autoregression (VAR) [4] to facilitate the content placement. The VAR model predicts the future change in content popularity using the historical data of content requests.
3) Our solution utilizes BC technology to build incentives among mobile users to participate in the D2D content sharing process. So the mobile users share their cached contents in exchange for ethereum coins.
4) The proposed decentralized framework protects mobile users from malicious attackers with the help of a consensus mechanism. In particular, the Proof of stake (PoS) consensus mechanism is adapted to randomly select one of the validator edge nodes for verifying the caching transactions.

II. Related Work

A. Intelligent D2D Caching

Li *et al.* [5] proposed a robust and efficient caching algorithm to minimize transmission delay in D2D networks. The non-parametric kernel density estimator was utilized to learn the intensity function of the content requests. With the objective to develop an improved D2D communication framework, Liu *et al.* [6] proposed an optimal caching strategy for video contents to reduce latency, outage probability and maximize the cache hit probability. The authors employed a recommendation algorithm to determine the video content that should be cached based on the individuals' preferences.

B. Mobility-Aware D2D Caching

Zhang *et al.* [7] investigated the cache hit maximization problem in D2D content sharing. In addition, the authors modeled the user mobility and developed a methodology for predicting user interest based on social vicinity and dynamic popularity of the contents. To study the mobility-aware cache placement problem in D2D networks for delay minimization, Sun *et al.* [8] proposed dynamic programming (DP) based solution to obtain the optimal caching. To minimize the handovers and energy consumed by the users in a wireless networking setup, Zohreh *et al.* [9] proposed a mobility-aware femtocaching scheme that limited the users with high velocity to only fetch desired contents from the femto access points (FAPs) instead of downloading the content via D2D communication.

C. BC-Enabled D2D Caching

BC is similar to a distributed ledger system for securely storing the transactions of mobile users. Recent research utilizes BC for building incentives among mobile users for D2D content sharing. For instance, Cui *et al.* [10] proposed a BC-based solution to build incentives among users for D2D caching and sharing of contents. Similarly, Zhang *et al.* [11] proposed a BC-based D2D network for mobile edge caching. The node selection and content placement problem were modeled as a Markov decision process (MDP). Deep Q-Network (DQN) was employed to solve complex and dynamic problems. Other research works utilized BC to provide a secure caching framework to mobile users. For instance, Adhikari *et al.* [12] enhanced the D2D

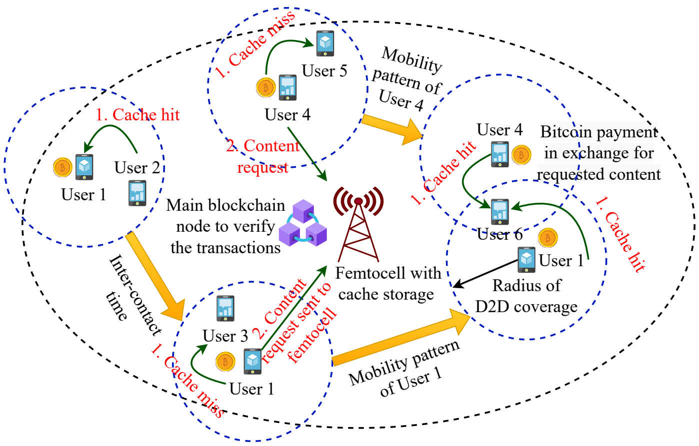


Figure 2: Network architecture.

content sharing in 5G and B5G networks by integrating BC-based smart contracts and priority-based content placement.

III. System Model

As shown in Fig 1, the 5G network architecture consists of a macro base station (MBS) that fetches requested content directly from the content providers (CPs) via the core network.

A. Network Architecture

We consider a densely deployed 5G network architecture with a central MBS. Let $\mathcal{N} = \{1, 2, \cdots, N\}$ denote the set of femtocells that are connected to the central MBS via optical fiber networks. The femtocell $i \in \mathcal{N}$ has limited caching storage of C_i to place popular contents. In addition, the entire network supports the communication of M mobile users represented by the set $\mathcal{M} = \{1, 2, \cdots, M\}$. Each mobile user $m \in \mathcal{M}$ consists of its own local cache storage of capacity C_m. The mobile users change location following a random walk process. At a discrete time step, the mobile user m requests content from the neighboring mobile users present within the D2D coverage radius. If the content is found in the neighboring mobile users, then it is considered a cache hit. Else, if the content is not found, then the mobile user sends the content request to its nearest femtocell.

As demonstrated in Fig 2, mobile user 1 is able to fetch the requested content from mobile user two during their contact time via D2D communication. However, mobile user 1, during its contact period with mobile user 3 is not able to find its requested content resulting in a cache miss. So the content request of mobile user 1 is then forwarded to its nearest femtocell, which inspects its cache storage. If the requested content is found in the femtocell's cache storage, then it is transmitted to the mobile user via wireless communication.

Algorithm 1: Blockchain-**E**nabled intel**LI**g**E**nt **V**id**E**o content caching algorithm (BELIEVE).

Input: Cumulative video content request count n
Output: Video content popularity P
Parameters: User identity i, Time horizon $\mathcal{T}$
Procedure:
for *each time step* $t \in \mathcal{T}$ **do**
  requested video content $c \sim$ MZipf(q, γ)
  ethereum transaction list $\mathcal{E} \leftarrow \{\}$
  if $c \in \mathcal{C}_i$ **then**
    $n_c^i \leftarrow n_c^i + 1$
  end
  else
    $\mathcal{D} \leftarrow$ **FindD2DNeighbours**(i)
    for *neighbouring users* $j \in \mathcal{D}$ **do**
      if *requested video content* $c \in \mathcal{C}_j$ **then**
        $n_c^j \leftarrow n_c^j + 1$
        $\mathcal{E} \leftarrow$ **AddTransaction**(i, j, c)
      end
    end
  end
  $\mathcal{C}_i \leftarrow$ **VectorAutoRegression**(n^i)
  BC$(\mathcal{E})$
end
FindD2DNeighbours(i)**:**
$\mathcal{D} \leftarrow \{\}$
for *mobile user* $u \in \mathcal{U}$ **do**
  if *CalcDist*$(i, u) \leq r$ **then**
    $\mathcal{D} \leftarrow \mathcal{D} \,||\, u$
  end
end
VectorAutoRegression(n^i)**:**
for *video content* $c \in \mathcal{C}$ **do**
  $P_t^c \leftarrow$ softmax(n^i)
  $P_{t+1}^c \leftarrow \alpha_c + \sum_{c=1}^{C} \beta_c P_t^c + \epsilon_c$
end
BC$(\mathcal{E})$**:**
Validator node confirms all transactions $\mathcal{E}$
Creation of new block using PoS consensus
Addition of new block to the BC

B. BELIEVE Algorithm

Algorithm 1 summarises the BC-enabled intelligent video content caching (BELIEVE) algorithm for D2D mobile users. The mobile user initially searches for the requested video content in its local storage. If unavailable, the neighbouring mobile users within the D2D communication range are discovered and searched for the requested video content. Next, the requester pays the mobile user to provide the video content in ethereum coins. Meanwhile, the VAR algorithm records the request count of the video contents and feeds the changing popularity trend to the softmax layer. The final decision of content caching is taken by sampling from the probability values obtained from the final softmax layer. Furthermore, the BC securely records the caching transactions using cryptographic keys. The validator edge node is selected at random using the PoS consensus mechanism to approve the transactions contained in the block. Finally, the validator node pushes the verified block to the BC.

C. Content Popularity Model

Let the set of available video contents be denoted by the set $\mathcal{C} = \{1, 2, \cdots, C\}$. The video contents are provided by the CPs situated in the cloud servers. The mobile user predicts the popularity of the video contents according to their request pattern and caches the most popular contents in its local cache.

Further, let the popularity of video content $c \in \mathcal{C}$ be represented by P_c. Consequently, $\sum_{c \in \mathcal{C}} P_c = 1$ and the set $P = \{P_1, P_2, \cdots, P_C\}$ is assumed to follow the Mandelbrot–Zipf (MZipf) [13] probability distribution. The probability P_c can be mathematically represented as follows:

$$P_c = \frac{(i+q)^{-\gamma}}{\sum_{i=1}^{C} (i+q)^{-\gamma}} \tag{1}$$

where i denotes the rank of the video content c. Whereas, q and γ represent the plateau factor and skewness factor, respectively. Consequently, rank = 1 is assigned to the most popular content which has the maximum probability of being requested by the mobile users.

D. Mobility and Contact Model

In a real-life scenario, mobile users opportunistically come in contact with other mobile users, depending on their mobility pattern. Our work adapts the inter-contact model [8] to characterize the mobility pattern of mobile users. The inter-contact mobility model consists of two phases – 1) contact time and 2) inter-contact time. The contact time is the duration for which the mobile user comes under the D2D communication range of another mobile user. Within this time period, the mobile user can fetch the requested content by establishing a D2D link. Whereas the inter-contact time represents the duration between two successive contacts. The contact process between mobile user $i \in \mathcal{M}$ and $j \in \mathcal{M}$ can be modelled as a Poisson process with rate $\lambda_{i,j}$. Consequently, the inter-contact time follows an exponential distribution with parameter $\lambda_{i,j}$. Further, the contact process for different mobile user pairs can be considered independent of one another due to the randomness of the mobility model.

E. Transmission Delay Model

The time taken for a femtocell to deliver the requested content to the mobile user is called transmission delay. Let the delay experienced by mobile user $m \in \mathcal{M}$ after the content request is sent to the femtocell be denoted as d_m. Further, the rate of data transfer through the wireless link be represented as r_m. Similar to [14], the data transmission rate r_m can be mathematically represented as follows:

$$r_m = b_m \log_2 \left(1 + \frac{p_m h_{m,n}}{N_0 + \sum_{i \in \mathcal{M} \setminus \{m\} : a_i = a_m} p_i h_{i,n}} \right) \tag{2}$$

where, b_m denotes the bandwidth assigned to mobile user m, p_m denotes the average transmit power, $h_{m,n}$ denotes the channel gain between mobile user m and femtocell n, which is assumed to be equal to $r^{-\beta}$ with β signifying the path loss exponent and r denotes the distance between mobile user and femtocell. N_0 indicates the power spectral density corresponding to Gaussian noise. Further, a_m denotes the wireless channel assigned to mobile user m by the femtocell. The transmission delay incurred via D2D communication is considered negligible when compared to wireless transmission.

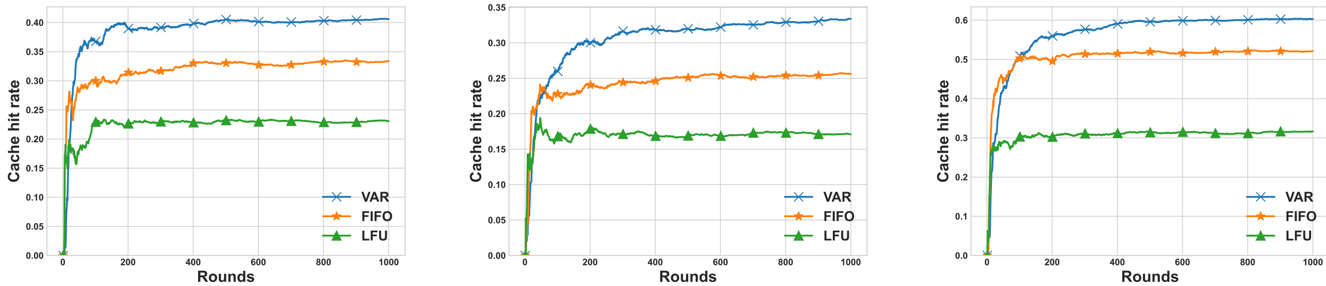
IV. Performance Evaluation

A. Experiment Setup

For simulating the proposed framework, we consider 30 mobile users with local cache storage. These mobile users are supported by 4 femtocells which perform edge caching. Femtocell is a base station that can support the communication of 8-10 mobile users within a radius of 10 meters. We assume that each mobile user can cache a maximum of 3 video contents in its local cache storage. Furthermore, we assume that a total number of 100 video contents are hosted by the CP. The popularity of the video contents is modeled by the MZipf probability distribution. We evaluate the proposed solution and the baselines on three different content popularity scenarios which follow MZipf(q = 1, γ = 1), MZipf(q = 1, γ = 2), and MZipf(q = 2, γ = 4) probability distribution. According to the inter-contact model, the contact points between two mobile users follow a Poisson process with a rate λ. We consider the rate λ to be uniformly distributed according to $\mathcal{U}(5, 15)$ and the total number of rounds to be equal to 1000. In addition, the VAR model has time lag $l = 1$. Regarding the wireless transmission model, the transmit power is considered to be equal to $p = 10$ dB, and the bandwidth is set as $B = 20$ MHz. Whereas the path loss component is set to $\beta = 2$. This implies that the channel gain can be expressed as $h = r^{-2}$, where the distance of the mobile user to the access points follows a uniform distribution of $\mathcal{U}(10, 50)$ meter. The size of each video content is fixed and equal to 1 MB. Furthermore, the power spectral density of the noise is assumed to be equal to $N_0 = -174$ dBm/Hz. The BC framework is utilized to build incentives among mobile users to share video content via D2D communication in exchange for ethereum coins (ETH). In particular, we fixed the amount of money exchanged in sharing 1 video content to be equal to 0.001 ETH.

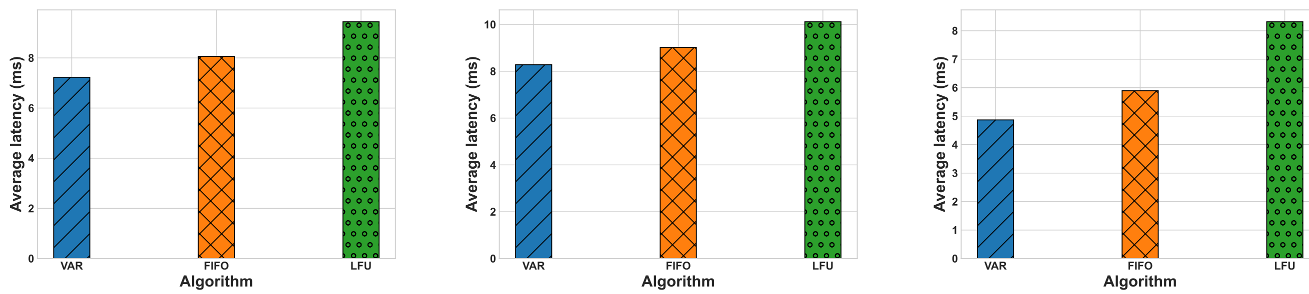
B. Results

1) Cache hit rate: In this section, we observe the number of hits in one second (cache hit rate). Fig. 3 demonstrates the



(a) MZipf(q = 1, $\gamma = 1$) (b) MZipf(q = 1, $\gamma = 2$) (c) MZipf(q = 2, $\gamma = 4$)

Figure 3: Improvement in cache hit rate with number of rounds



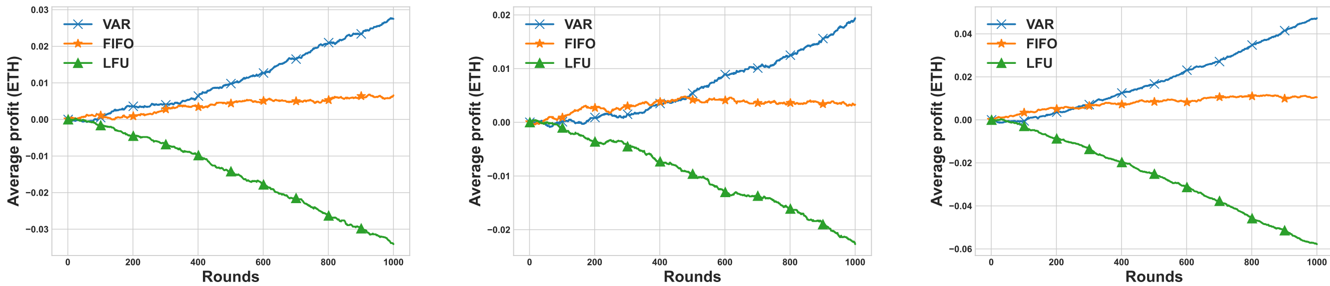
(a) MZipf(q = 1, $\gamma = 1$) (b) MZipf(q = 1, $\gamma = 2$) (c) MZipf(q = 2, $\gamma = 4$)

Figure 4: Cumulative transmission delay in fetching the requested content

improvement of cache hit rate with an increasing number of rounds. It can be observed that the VAR model outperforms the other baselines cache replacement algorithms by a better cache hit rate. Specifically, the VAR model obtains a cache hit rate of 0.4, 0.34, and 0.6 as shown in Fig. 3a, 3b, and 3c, respectively. The VAR model forecasts the future popularity of the video contents by modeling it as a linear function of the previous requests of all the video contents. In this way, the VAR model is able to capture the popularity trend of the video content better than the simplified rule-based algorithms such as LFU and FIFO. Furthermore, it can be inferred from Fig. 3, that the VAR model achieves an improvement of more than 10% and 20% than the FIFO and LFU model, respectively.

2) Latency: Fig. 4 presents the average end-to-end transmission delay incurred while retrieving the requested video content over 1000 rounds. We can observe that the VAR model performs the best among the other baseline algorithms by attaining the lowest latency. In particular, from Fig. 4a, 4b, and 4c we can see that the VAR model has the lowest latency of 7.2 ms, 8.4 ms, and 4.9 ms, respectively. The VAR model intelligently predicts the popularity of the video contents and stores the most popular ones in its local cache storage. Due to this reason, most of the video contents requested by mobile users can be found on neighboring devices and fetched via D2D communication. Consequently, the latency in obtaining the requested video content is less compared to other rule-based algorithms. This is because less number of video content requests are fetched from access points which reduces delay due to wireless transmission.

3) Profit: Fig. 5 illustrates the profit earned by mobile users in terms of ethereum coins. It can be observed that the VAR model maximizes its profit over the number of rounds. Moreover, the VAR model outperforms the baseline algorithms by obtaining a large profit margin. Specifically, the VAR model achieves an average profit of 0.03 ETH, 0.02 ETH, and 0.05 ETH as demonstrated in Fig. 5a, 5b, and 5c, respectively. The mobile users employing the VAR model intelligently cache the popular contents, which is used to serve the video content requester. Consequently, in exchange for the service provided by mobile users employing the VAR model, they are paid in ethereum coins, which increases their profit. If the requester mobile user continually does not locally cache the most popularly requested video content in many rounds, then it might go into loss.



(a) MZipf(q = 1, $\gamma = 1$) (b) MZipf(q = 1, $\gamma = 2$) (c) MZipf(q = 2, $\gamma = 4$)

Figure 5: Average profit earned in exchanging ethereum coins with requested content

V. Conclusion

This paper utilized a VAR model for content placement in a mobility-aware and BC-enabled edge caching framework. Using the past content request data, the VAR model can predict the future popularity of the video content. The BC provided a secure framework for the requester to exchange ethereum coins for the requested video content. Also, the PoS consensus mechanism verified the authenticity of the edge nodes to prevent malicious attacks. Through extensive simulations, we compared the performance of the VAR model against two baseline algorithms, namely, LFU and FIFO. The simulation results illustrated the efficacy of the VAR model in respect of cache hit rate, latency, and profit earned. The VAR model enhanced the cache hit rate by 10-20% and minimized the latency to 5-8 ms. Moreover, the VAR model profited in the D2D content exchange process by earning an average profit of around 0.02-0.05 ETH.

In the future, we plan to investigate the resource constraint characteristic of edge nodes in mobility-aware and BC-enabled edge caching frameworks. Another extension to our work will include cooperation among mobile users to maximize their combined profit.

References

[1] C. V. Networking, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper," *Cisco White Paper*, 2016.

[2] R.-X. Zhang, T. Huang, and L. Sun, "A Long-Short-Term Fusion Approach For Video Cache," in *Proceedings of the IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[3] B. Gabr, S. Hosny, and M. Nafie, "Content Delivery In Mobility-Aware D2D Caching Networks," in *Proceedings of the IEEE Global Communications Workshop*. IEEE, 2018, pp. 1–7.

[4] H. Lütkepohl, *New Introduction To Multiple Time Series Analysis*. Springer Science & Business Media, 2005.

[5] Y. Li, C. Zhong, M. C. Gursoy, and S. Velipasalar, "Learning-Based Delay-Aware Caching In Wireless D2D Caching Networks," *IEEE Access*, vol. 6, pp. 77 250–77 264, 2018.

[6] Z. Liu, H. Song, and D. Pan, "Distributed Video Content Caching Policy With Deep Learning Approaches For D2D Communication," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 644–15 655, 2020.

[7] W. Zhang, D. Wu, W. Yang, and Y. Cai, "Caching On The Move: A User Interest-Driven Caching Strategy For D2D Content Sharing," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2958–2971, 2019.

[8] R. Sun, Y. Wang, L. Lyu, N. Cheng, S. Zhang, T. Yang, and X. Shen, "Delay-Oriented Caching Strategies In D2D Mobile Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8529–8541, 2020.

[9] Z. HajiAkhondi-Meybodi, J. Abouei, M. Jassemuddin, and A. Mohammadi, "Mobility-Aware Femtocaching Algorithm In D2D Networks Based On Handover," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10 188–10 201, 2020.

[10] H. Cui, Z. Chen, N. Liu, and B. Xia, "Blockchain-Driven Contents Sharing Strategy For Wireless Cache-Enabled D2D Networks," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2019, pp. 1–5.

[11] R. Zhang, F. R. Yu, J. Liu, R. Xie, and T. Huang, "Blockchain-Incentivized D2D And Mobile Edge Caching: A Deep Reinforcement Learning Approach," *IEEE Network*, vol. 34, no. 4, pp. 150–157, 2020.

[12] A. Adhikari and D. B. Rawat, "Leveraging Smart Contracts For Priority Based Caching In D2D Networks For 5G And Beyond," in *Proceedings of the IEEE Globecom Workshops*. IEEE, 2021, pp. 1–6.

[13] M. Hefeeda and O. Saleh, "Traffic Modeling And Proportional Partial Caching For Peer-to-Peer Systems," *IEEE/ACM Transactions on Networking*, vol. 16, no. 6, pp. 1447–1460, 2008.

[14] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient Multi-User Computation Offloading For Mobile-Edge Cloud Computing," *IEEE/ACM transactions on networking*, vol. 24, no. 5, pp. 2795–2808, 2015.