On the Delay of Content-Centric Mobile Multihop Networks Using File Segmentation

Trung-Anh Do†, Won-Yong Shin‡, and Hoai Bac Dang‡
†Electronics Engineering, Posts and Telecommunications Institute of Technology, Hanoi 100000, Vietnam
‡Computer Science and Engineering, Dankook University, Yongin 16890, Republic of Korea
Email:anhdt@ptit.edu.vn; wyshin@dankook.ac.kr; bacd@ptit.edu.vn

Abstract—In this paper, we study the delay performance in a content-centric mobile ad hoc network, where each node moves using multihop according to the random walk mobility model and requests a content object from the library independently at random, according to a Zipf popularity distribution. In our network model, we assume that each content object consists of $K$ distinct segments of equal size such that each of $n$ mobile nodes is able to transmit completely one segment to one of its neighbor cells in one time slot. Using multihop transmission, the delay scaling law is analyzed based on the following two reception strategies to determine how $K$ distinct segments are fully delivered to the requesting node in turn to rebuild the desired content object: A sequential reception and a random reception. Then, we do intensive computer simulations to characterize the influence of $K$ on the delay and find the optimal cache allocation strategies, which minimize the delay. Our main result indicates that the delay obtained from the random reception strategy outperforms the sequential reception case.

Index Terms—Caching, content-centric network, mobile ad hoc network, file segmentation, scaling law.

I. INTRODUCTION

Recently, data caching [1], which brings contents closer to users, has emerged as a promising technique that deals with the exponential growth of internet traffic [2]. Thanks to the benefits of replicating popular contents across networks, caching techniques play an important role in maintaining the sustainability of future wireless networks.

As the number of users continues to grow dramatically, the capacity scaling law behavior has been widely studied in large wireless networks. In [3], Gupta and Kumar showed that, for a static ad hoc network with $n$ source–destination pairs randomly distributed in a unit area, the per-node throughput of the order of $\frac{1}{\sqrt{n \log n}}$ is achievable using the nearest neighbor multihop transmission. Besides multihop transmission, there have been various research directions to improve the per-node throughput up to a constant scaling by using hierarchical cooperation [4] and node mobility [5], [6].

Contrary to the studies on the conventional ad hoc network model in which source–destination pairs are given and fixed, investigating content-centric ad hoc networks would be quite challenging. As content objects are cached by numerous nodes over a network, finding the closest content holder of each request and scheduling between requests are of crucially importance for the overall network throughput. The scaling behavior of content-centric ad hoc networks has received a lot of attention in the literature [7]–[10]. In static ad hoc networks, throughput scaling laws were analyzed using multihop communications [7], [8], which yields a significant performance gain over the single-hop caching scenario. In addition, in [9], performance on the throughput and delay was investigated for mobile ad hoc networks under various mobility models by showing that increasing the mobility degrees of nodes leads to worse performance. Such an analysis was extended to static and mobile hybrid networks with multiple base stations, [10]–[12], where the fluid model [6] allowing the size of each content object to be arbitrarily small is adopted. Thus, the time required for transmitting a content object is assumed to be much smaller than a time slot.

Meanwhile, a different caching framework, termed coded caching [13]–[15], has received a great deal of attention in content-centric wireless networks, where a single transmitter simultaneously deals with several different demands using common coded multicast transmission. Based on this approach, a global caching gain can be achieved by finding the optimal content placement such that multicasting opportunities are exploited simultaneously for all possible requests in the delivery phase. In addition, caching using file segmentation, which is a popular configuration of coded caching, was introduced in [16], [17], where each file is formed by $K$ encoded segments and each user downloads a part of $K$ segments to rebuild the original file.

In this paper, we consider a content-centric mobile ad hoc network using file segmentation, where each node moves according to the random walk mobility model (RWMM) and requests a content object from the library independently at random, according to a Zipf popularity distribution. Specifically, instead of using the fluid model, we take into account the case where the size of each content object is so large that it can not be successfully transmitted in one time slot. Then, each content object is divided into $K$ uncoded distinct segments of equal (unit) size such that each of $n$ mobile nodes is able to transmit completely one segment to one of its neighbor cells in one time slot. We introduce the following two reception strategies to decide how $K$ distinct segments of a content object are fully delivered to the requesting node: A sequential reception and a random reception. Based on the two reception strategies, the delay scaling law is shown. We perform numerical evaluation via computer simulations to find the optimal cache allocation strategies in terms of minimizing the delay, which shows how $K$ affects the delay performance. Our main result indicates
that the delay obtained from the random reception strategy is superior to that of the sequential reception strategy.

We use the following asymptotic notation: i) \( f(x) = O(g(x)) \) means that there exist constants \( C \) and \( c \) such that \( f(x) \leq C g(x) \) for all \( x > c \), ii) \( f(x) = o(g(x)) \) means that \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = 0 \), iii) \( f(x) = \Omega(g(x)) \) if \( g(x) = O(f(x)) \), iv) \( f(x) = \Theta(g(x)) \) if \( f(x) = O(g(x)) \) and \( f(x) = \Omega(g(x)) \) [18].

II. SYSTEM MODEL

A. System Model

We consider a content-centric mobile ad hoc network, where \( n \) mobile users move according to the RWMM in a unit area. At each time slot, every node requests content objects in the library of size \( M \), where \( M = o(n) \), independently at random. In our work, instead of assuming that the sizes of content objects are arbitrarily small so that the time required for transmitting a content object is much smaller than one time slot as in the fluid model, we assume that the size of each content object is so large that it cannot be successfully transmitted in one time slot. In our model, each content object is assumed to be divided into \( K = o(n) \) distinct uncoded segments of equal (unit) size such that each of \( n \) mobile nodes is able to completely transmit one segment to one of its neighbor cells in each time slot. Each node is assumed to be equipped with a local cache, which is installed to store at most \( S = o(n) \) distinct segments.

We assume that the request probability of content object \( m \in M \), where \( M = \{1, \ldots, M\} \), following a Zipf popularity distribution is given by \( p_m = \frac{m^{-\alpha}}{K M^\alpha} \), where \( \alpha > 0 \) is the Zipf exponent and \( H_M^{(\alpha)} = \sum_{i=1}^M i^{-\alpha} \) is a normalization constant formed in the Riemann zeta function.

In cache-enabled wireless networks, a caching problem is generally partitioned into the caching phase and the delivery phase. Equivalently, the problem consists of storing content objects in the caches and establishing efficient delivery routing paths for the requested content objects. We first describe the caching phase, which determines the cache segments of content objects in the storage of \( n \) nodes. Let \( X_{m,i} \) denote the number of replicas of segment \( i \) of content object \( m \in M \) for \( i \in \{1, \ldots, K\} \). We assume that \( X_{m,i} \) replicates are distributed uniformly and independently over the network. Suppose that \( X_{m,i} = X_m \) for all \( i \in \{1, \ldots, K\} \), where \( X_m \) is the number of replicas of content object \( m \) stored at nodes. In order for a cache allocation to be feasible, \( \{X_m\}_{m=1}^M \) should satisfy the following total caching constraint:

\[
\sum_{m=1}^M KX_m \leq nS. \tag{1}
\]

Similarly as in [7], [9], [11], [12], we employ the random caching strategy such that replicas of each segments are stored uniformly at random over the caches in \( n \) nodes. Note that in the delivery phase, the positions of \( X_m \) nodes storing the segment will be changed independently and randomly over time due to the node mobility. Now, let us describe the delivery phase of the requested content objects, which allows the requested content objects to be delivered to the corresponding nodes over wireless channels. During the delivery phase, each node downloads its requested content object (possibly via multihop) from one of the nodes storing the requested content object in their caches. We adopt the protocol model [3] for successful content delivery. In particular, let \( d(u, v) \) denote the Euclidean distance between nodes \( u \) and \( v \). Then, content delivery from node \( u \) to node \( v \) is assumed to be successful if and only if \( d(u, v) \leq r \) and there is no other active transmitter in a circle of radius \( (1 + \Delta) r \) from node \( v \), where \( r \) and \( \Delta > 0 \) are given protocol parameters.

B. Performance Metric

To fully receive the desired content object, the requesting nodes need to collect all of \( K \) separate uncoded segments of the content object distributed over the network. In our work, we make a slight modification to the definition of delay in [11], [12] to fit into our content-centric mobile network setup, which is provided as follows.

**Definition 1 (Delay):** Let \( D(i, k) \) denote the delay of the \( k \)th requested content object of node \( i \), which is measured from the moment that the requesting message for the first segment leaves the requesting node until all of \( K \) distinct segments of the corresponding content object fully arrives at the node. For a particular realization of the network, the delay for node \( i \) is \( \limsup_{r \to \infty} \frac{1}{r} \sum_{k=1}^r D(i, k) \). Then, the delay is defined as the expectation of the average delay of all nodes over all network realizations, i.e.,

\[
D(n) \triangleq \mathbb{E}\left[ \frac{1}{n} \sum_{i=1}^n \limsup_{r \to \infty} \frac{1}{r} \sum_{k=1}^r D(i, k) \right].
\]

III. RECEPTION STRATEGIES AND SEGMENT DELIVERY ROUTING PROTOCOL

A. Reception Strategies

In our work, we describe on the two reception strategies illustrated in Fig. 1, which represents a way of how \( K \) segments of the desired content object are delivered to the requesting node.

- **Sequential reception:** All \( K \) segments of a content object are delivered in sequence. As illustrated in Fig. 1(a), the requesting node receives the first segment, denoted by \( X_{m,1} \), of content object \( m \), the second segment \( X_{m,2} \), and the third segment \( X_{m,3} \) in sequence until it receives all \( K \) distinct segments of the desired content object.

- **Random reception:** The requesting node receives the segments in an arbitrary way. As illustrated in Fig. 1(b), it first receives the second segment \( X_{m,2} \) of content object \( m \), which is the nearest one. Then, the second nearest segment (i.e., \( X_{m,1} \)) is delivered to the requesting node. This continues until the requesting node receives all \( K \) distinct segments of the desired content object.
multihop in backward direction.

IV. DELAY ANALYSIS AND NUMERICAL EVALUATION

A. Delay Analysis

It is obvious that the delay of a transmission depends entirely on the distance between two partners. Similarly to [10]–[12], we establish the following lemma, which characterizes the delay performance.

Lemma 1: For any mobile node requesting segment \( i \) of content object \( m \in \mathcal{M} \), the average initial distance between any requesting node and its closest holder of segment \( i \) of content object \( m \) is \( \Theta \left( \frac{1}{\sqrt{X_{m,i}}} \right) \), where \( X_{m,i} \) is the number of replicas of segment \( i \) of content object \( m \) stored at nodes.

Now, let us turn to analyzing the delay of cache-enabled mobile networks according to the two reception strategies.

1) Sequential reception: In this case, all \( K \) segments of a content object are delivered in sequence. From Lemma 1, using the segment delivery routing protocol shown in the previous section, one can see that there is no need to cache more than \( a(n)^{-1} \) replicas of the segment \( i \) of content object \( m \) over the network. Thus, using the fact that \( X_{m,i} = X_m \) for all \( i \in \{1, \cdots, K\} \), we impose the following individual caching constraints:

\[
\begin{aligned}
X_m &\leq a(n)^{-1}, \\
X_m &\geq 1.
\end{aligned}
\] 

for all \( m \in \mathcal{M} \).

We are ready to show the first theorem, which shows the delay performance in our cache-enabled model using the sequential reception strategy.

Theorem 1: In a content-centric mobile network where \( K \) segments of a content object are delivered using sequential reception, the delay is given by

\[
D(n) = \Theta \left( \sum_{m=1}^{M} \frac{p_m K}{\sqrt{a(n)X_m}} \right).
\]

Proof: Using Definition 1 and Lemma 1, the time that the requesting node receives a certain segment of content object \( m \) via multihop can be computed as \( \Theta \left( \frac{1}{\sqrt{a(n)X_m}} \right) \), which means that the duration that the requesting node receives content object \( m \) consisting of \( K \) segments is \( \Theta \left( \frac{K}{\sqrt{a(n)X_m}} \right) \). Hence, for \( \forall m \in \mathcal{M} \), we finally obtain (3), which completes the proof of the theorem.

Note that assuming \( X_m = a(n)^{-1} \) for all \( m \in \mathcal{M} \) leads to \( D(n) = \Theta(K) \), which corresponds to the minimum delay we can hope for. Using (1), (2), and Theorem 1, we formulate the
following optimization problem.

**Problem 1 (Sequential reception case):**

\[
\begin{align*}
\min_{\{X_m\}_{m \in M}} & \sum_{m=1}^{M} \frac{p_m K}{\sqrt{a(n)X_m}} \\
\text{subject to} & \sum_{m=1}^{M} X_m \leq nS/K, \\
& 1 \leq X_m \leq a(n)^{-1} \quad \text{for } m \in M.
\end{align*}
\]

2) **Random reception:** In this case, a requesting node receives \(K\) segments of its desired content object in an arbitrary way. For content object \(m\), there are \(KX_m\) segments over the network. From Lemma 1, for the first segment of content object \(m\), the average distance between the requesting node and the closest holder is \(\Theta\left(\frac{1}{\sqrt{X_m}}\right)\). Since the number of the remaining segments is \((K-1)X_m\), the average distance between the requesting node and the closest holder is \(\Theta\left(\frac{1}{\sqrt{(K-1)X_m}}\right)\).

Let us first consider the case where \(KX_m \geq a(n)^{-1}\). The number of the remaining segments over the network is \((K - l_m + 1)X_m\) for the \(l_m\)-th segment, where \(l_m\) denote the smallest index of the segment such that \((K - l_m + 1)X_m \leq a(n)^{-1}\). Using Lemma 1, the time that it takes to collect all of \(K\) segments of content object \(m\) is given by 

\[
(l_m - 1) + \sum_{i=1}^{K-1} \Theta\left(\frac{1}{\sqrt{a(n)(K-i)X_m}}\right).
\]

Here, it follows that

\[
\sum_{i=1}^{K-1} \Theta\left(\frac{1}{\sqrt{a(n)(K-i)X_m}}\right) = \Theta\left(\frac{1}{\sqrt{a(n)X_m}}\sqrt{K/l_m}\right).
\]

Thus, we have \((K - l_m + 1)X_m = \Theta(a(n)^{-1})\). Then, the delay for content object \(m\) is \(\Theta(l_m - 1) + \Theta(K - l_m + 1) = \Theta(K)\), which corresponds to the minimum bound on the delay. Since there is no need to cache more than \(X_m = a(n)^{-1}K\) replicas of content object \(m\) at nodes, we impose the following individual caching constraint:

\[
\begin{align*}
X_m & \leq a(n)^{-1}K, \\
X_m & \geq 1.
\end{align*}
\]

We now establish our second theorem, which shows the delay performance for the random reception case.

**Theorem 2:** In a content-centric mobile network where \(K\) segments of a content object are delivered using random reception, the delay is given by

\[
D(n) = \Theta\left(\sum_{m=1}^{M} p_m \sqrt{\frac{K}{a(n)X_m}}\right).
\]

**Proof:** Using Definition 1 and Lemma 1, the duration that the requesting node receives all \(K\) distinct segments of content object \(m\) is

\[
\sum_{i=1}^{K-1} \frac{1}{\sqrt{a(n)(K-i)X_m}} = \Theta\left(\frac{1}{\sqrt{a(n)X_m}}\right)\sqrt{K/l_m}.
\]

Hence, for \(\forall m \in M\), we finally obtain (6), which completes the proof of the theorem.

Form the fact that the total cache storage of the network is \(nS\), if \(M = O\left(S/n a(n)^{\alpha}\right)\), then the minimum delay \(D(n) = \Theta(K)\) is always achieved by the replication strategy \(X_m = \Theta\left(a(n)^{-1}K\right)\) for all \(m \in M\). Otherwise, using (1), (5), and Theorem 2, we formulate the second optimization problem.

**Problem 2 (Random reception case):**

\[
\begin{align*}
\min_{\{X_m\}_{m \in M}} & \sum_{m=1}^{M} \frac{p_m K}{\sqrt{a(n)X_m}} \\
\text{subject to} & \sum_{m=1}^{M} X_m \leq nS/K, \\
& 1 \leq X_m \leq a(n)^{-1}K, \quad \text{for } m \in M.
\end{align*}
\]

**B. Numerical Evaluation**

In this subsection, to find the optimal delay (i.e., the optimal cache allocation strategies), we perform numerical evaluation to solve Problems 1 and 2 via intensive computer simulations according to finite values of the system parameters in Table I. As depicted in Fig. 2, the results show that \(X_m^*\) monotonically decreases as \(m\) increases for Problem 1; for Problem 2, \(X_m^*\) is fixed to a certain value for small \(m\) and then decreases as \(m\)
increases. This difference comes from the individual caching constraints imposed in two reception strategies in (2) and (5).

The delay obtained from the two reception strategies is illustrated in Fig. 3. It is seen that the delay increases when $K$ increases due to the fact that the requesting nodes need to find and collect a larger number of segments to rebuild the desired content object when $K$ increases. In addition, it is also shown that using random reception yields better performance on the delay than the sequential reception case. This is because random reception tends to operate more flexibly in the sense of collecting $K$ segments as there is no a priori reception order of the segments, thus leading to the reduction of delay.

V. Conclusion

In this paper, we studied the delay of a content-centric mobile ad hoc network, where each content object is divided into $K$ distinct segments of equal size such that each of $n$ mobile nodes is able to completely transmit one segment to one of its neighbor cells in one time slot. By introducing the sequential and random reception strategies, we analyzed the delay scaling law and then found the optimal cache allocation strategies in terms of minimizing the delay. As our main results, it was shown that the random reception strategy outperforms the sequential reception one.

Acknowledgment

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1A09000835) and by the Ministry of Science, ICT & Future Planning (MSIP) (2015R1A2A1A15054248).

References