Cache Node Placement for Maximum Traffic Minimization in Content-Centric Networking

Shohei Nakajima and Nattapong Kitsuwan

Department of Computer and Network Engineering,
The University of Electro-Communications, Tokyo, Japan.

***Abstract*—Content-centric networking (CCN) is a network architecture based on content. The conventional network architecture uses IP address to route packets based on location. In CCN, which routes packets based on content name, a cache function can be attached to nodes. Getting contents from cache nodes alleviates traffic congestion in the network. In the case that the cache function is not attached to all nodes in the network, it is necessary to decide where cache nodes are placed. Congestion tends to occur around cache nodes when using the conventional method. We formulate a problem of cache node placement as an integer linear programming (ILP) considering the population of each node. The ILP formulation determines the cache node placement to minimize the value of the maximum traffic. This result shows the proposed ILP approach reduces the maximum traffic by 59% on average.**

***Index Terms*—Content-centric networking, maximum traffic, optimization**

I. Introduction

The rapid progress of digitization in recent years makes huge traffic on the Internet to get variable content such as video streaming. By using cache in Content-centric networking (CCN), network traffic can be reduced. CCN is a network architecture that names contents transferred over the Internet [1]. It enables to transfer contents without consideration of physical location like an IP address. The cache function on nodes allows the network to store contents. A node with this function is called a cache node. If several users request the same content, such as streaming of a sport, the current network architecture needs to transmit the same content several times from one server to different users. The congestion occurs in the network, especially around the server. In CCN, cache nodes placed in the network enable clients to get content data from them. Server processing load and network congestion are then reduced. However, replacing all nodes on the core network with cache nodes is costly. Thus, it is necessary to decide which nodes to replace. The conventional method to solve this placement problem uses betweenness centrality (BC) [2]. However, this paper does not consider the population size. To address this issue, we propose an integer linear programming (ILP) formulation to determine the appropriate cache node placement and compare it with the BC approach.

II. Content-Centric Networking

CCN is a network architecture based on content names to transfer data. In other words, it is not based on where the contents are but based on what contents clients request. Communication in CCN uses two types of packets, interest packets and data packets. Clients who request content send an interest packet with the content name. In CCN, packets are routed based on content name in the same way that IP network routes packet based on IP address. If the interest packet arrives at a node that has the content specified in the interest packet, the node replies data packets with the content. The data packets are routed on the same path as the interest packets in reverse order.

CCN nodes have three types of data structures, forwarding information base (FIB), pending interest table (PIT), and content store (CS). The FIB is a routing table to transfer an interest packet to nodes that have the corresponding content. The PIT is used to temporarily record where the interest packet comes from. CCN nodes create a PIT entry when the interest packet is routed to nodes that have the content. If a data packet arrives, CCN nodes look up the PIT entry and transfer the packet to the nodes to clients, and finally the data packet arrives at the clients. The CS is a memory to store contents. Nodes with the CS are called cache nodes. The client can obtain contents from cache nodes using these data structures, and network traffic is distributed.

An appropriate node placement scheme is necessary to distribute traffic congestion efficiently. This paper models the traffic considering the population in a node and proposes ILP formulation to place cache nodes.

III. Betweenness Centrality Approach

An approach using betweenness centrality (BC) was proposed to decide cache node placement to decrease the hop count for the client to obtain contents [3]. This is called the BC approach.

BC is one of the metrics to measure network centrality. BC of node $v \in V$, where V is a set of nodes, is defined by Eq. (1). Higher values indicate greater centrality.

$$C_B(v) = \sum_{s,t \in V: s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

σ_{st} is the number of the shortest path between node s and node t. $\sigma_{st}(v)$ is the number of the shortest path through node v between node s and node t. Given the number of cache nodes as K, the BC approach chooses K nodes with the highest centrality as cache nodes. However, cache nodes placed by the

BC approach are concentrated in the center of the network. It concentrates traffic on links around cache nodes. Fig. 1 shows congested links in pink and empty links in cyan when all nodes in the network request a content provided by a source node and cache nodes. It is assumed that interest packets are routed in such a way as to minimize the maximum traffic.

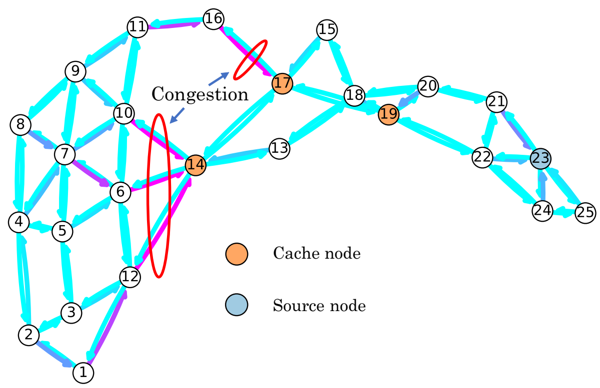

Fig. 1: Traffic concentration in BC approach

IV. Proposed Integer Linear Programming Approach

As mentioned in the previous section, the BC approach tends to have higher maximum traffic. A proposed ILP approach finds the minimum value of maximum traffic and the placement at that time. To formulate the problem of cache node placement as an ILP, the following are assumed. The network is modeled as connected undirected graphs. A few popular contents which dominate the majority of network traffic is considered because considering several contents requires a long computation time to solve the ILP. Cache nodes have all of the few contents considered to simplify the problem. Traffic for cache nodes to get contents from source nodes is not considered because the traffic only happens once when a content is cached and does not have significant effect on traffic in the network. Traffic loads of interest packets are ignored because they are much smaller than traffic loads of data packets. The given constant $D(p)$ is defined the number of requests by node p based on the population. Traffic in the network is distributed to minimize the maximum traffic.

The decision variables and parameters used in the ILP formulation are summarized in Tables I and II.

TABLE I: Definitions of Variables

x^c_{ij}	Traffic for content c through link (i,j)
y_p	Whether node p is a cache node
v	Maximum traffic in the network

$x^c_{ij} \geq 0$ and $v \geq 0$ are integer decision variables that means $x^c_{ij} \in \mathbb{Z}$, $v \in \mathbb{Z}$. Binary decision variable y_p is 1 if node p is a cache node, 0 otherwise. $s^c_p \in \{0,1\}$ is 1 if node p is a source node of content c, 0 otherwise.

Given the source nodes placement of each content in the network, a proposed ILP approach output the maximum traffic

TABLE II: Definitions of Parameters

V	Set of nodes
E	Set of directed links
C	Set of contents
$D(p)$	The number of requests at node p for a type of content
s^c_p	Whether node p has an original of content c

and the cache nodes placement. The objective function is the following.

$$\min \quad Mv + \sum_{c\in C}\sum_{(i,j)\in E} x^c_{ij} \tag{2}$$

M is a constant that satisfies $M \geq |E||C|\sum_{p\in V} D(p)$. The objective function minimizes the maximum traffic v primarily and secondly minimizes the total number of hops to route Interest packets. To show constraints in this narrow field, $R(c,p)$ which indicates the number of requests node p sends for content c is introduced. Constraints follow the definition of $R(c,p)$.

$$R(c,p) := D(p)(1-s^c_p)(1-y_p) \tag{3}$$

$$R(c,p) - L\cdot(s^c_p + y_p) \leq \sum_{j:(p,j)\in E} x^c_{pj} - \sum_{j:(j,p)\in E} x^c_{jp} \quad \forall p\in V,\, c\in C \tag{4}$$

$$\sum_{j:(p,j)\in E} x^c_{pj} - \sum_{j:(j,p)\in E} x^c_{jp} \leq R(c,p) \quad \forall p\in V,\, c\in C \tag{5}$$

$$v \geq \sum_{c\in C} x^c_{ij} \qquad \forall (i,j)\in E \tag{6}$$

$$\sum_{p\in V} y_p \leq K \tag{7}$$

L is a constant that satisfies $L \geq \sum_{p\in V} D(p)$. Constant K is the number of cache nodes. Constraints (4) and (5) specify the amount of traffic generated in node p for content c. Constraint (6) makes variable v the maximum traffic. Constraint (7) limits the number of cache nodes. As an output of the ILP approach, the maximum traffic v and cache nodes placement y_p are obtained.

Cache nodes in Fig. 2 are placed by the ILP approach in the same case as Fig. 1. It shows the ILP approach distributes traffic and alleviates congestion compared to the BC approach.

V. Performance Evaluation

To compare the maximum traffic of the ILP approach with the BC approach, it is necessary to calculate the maximum traffic when cache node placement is decided by the BC approach. Following ILP formulation is used to calculate it.

$$\min \quad Mv + \sum_{c\in C}\sum_{(i,j)\in E} x^c_{ij} \tag{8}$$

$$R(c,p) := D(p)(1-s^c_p)(1-f^c_p) \tag{9}$$

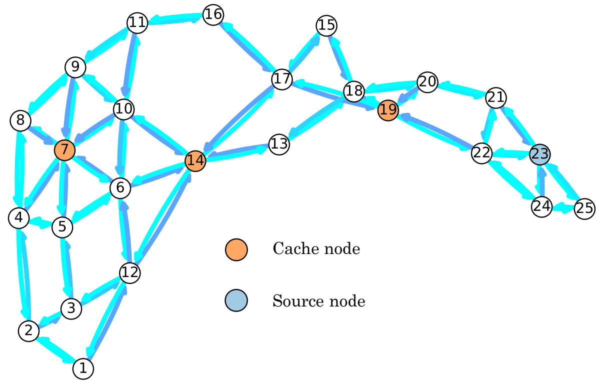

Fig. 2: Traffic in ILP approach

$$R(c,p) - L\cdot(s_p^c + f_p^c) \le \sum_{j:(p,j)\in E} x_{pj}^c - \sum_{j:(j,p)\in E} x_{jp}^c \quad \forall p\in V,\, c\in C \tag{10}$$

$$\sum_{j:(p,j)\in E} x_{pj}^c - \sum_{j:(j,p)\in E} x_{jp}^c \le R(c,p) \quad \forall p\in V,\, c\in C \tag{11}$$

$$v \ge \sum_{c\in C} x_{ij}^c \qquad \forall (i,j)\in E \tag{12}$$

Parameter $f_p^c \in \{0,1\}$ is 1 if a cache node p has a content c, 0 otherwise. Note that the definition of $R(c,p)$ is different from that of the ILP approach.

The maximum traffic for each of the ILP approach and the BC approach was calculated and compared in Japan photonic network [4] which consists of 25 core nodes. The topology is the same as Figs. 1 and 2. Parameter $D(p)$ was set to 1/10000 of the population of node p. Figs. 3 and 4 show the maximum traffic for each number of cache nodes for the BC approach and the ILP approach. Fig. 3 considers one content and one source node that means $|C| = 1$ and $\sum_{p\in V} s_p^c = 1$ ($\forall c \in C$). In Fig. 4 $|C|$ is set to 3 and $\sum_{p\in V} s_p^c = 1$ ($\forall c \in C$). Maximum traffic for each number of cache nodes in Fig. 3 is the average of the maximum traffic calculated for all source node placements. Maximum traffic for each number of cache nodes in Fig. 4 is the average of the maximum traffic calculated for random 150 source node placements because it takes a long time to calculate ILP problems for all source node placements.

In the BC approach, increasing the number of cache nodes may not reduce the maximum traffic because congestion on links around cache nodes becomes the bottleneck. For example, the bottleneck occurs when the number of cache nodes is between 8 and 19, as shown in both figures. The figures show maximum traffic of the ILP approach is less than that of the BC approach for all the number of cache nodes except in the case that all nodes are cache nodes. This is because the ILP approach gives optimal solutions in given situations. The ILP approach reduces maximum traffic by 61% on average compared to the BC approach for the $|C| = 1$ case, and reduces it by 59% on average and by 80% at maximum for the $|C| = 3$ case.

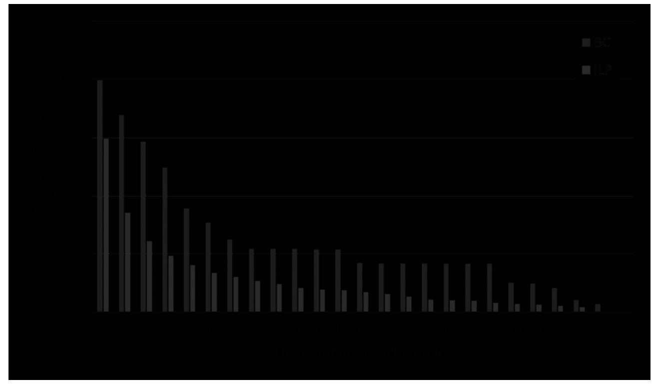

Fig. 3: Maximum traffic ($|C| = 1$)

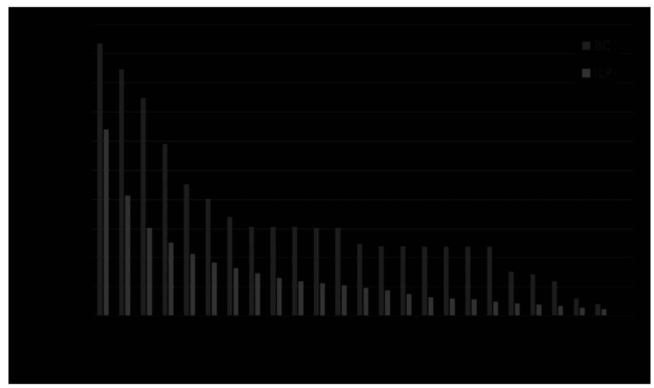

Fig. 4: Maximum traffic ($|C| = 3$)

VI. Conclusion

This paper proposed an ILP approach to reduce the maximum traffic in the network. The conventional approach using BC to solve a problem of cache node placements in CCN has the problem of high maximum traffic. In contrast, the proposed ILP approach considers population to distribute the traffic. We compared the maximum traffic in the network that cache nodes placed by the ILP approach with the maximum traffic calculated in the network that cache nodes are placed by the BC approach. We show that the ILP approach reduces maximum traffic compared to the BC approach. In the case that the number of cache nodes is 3, the maximum traffic is reduced by 59% on average.

References

[1] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in Proc. ACM 5th Int. Conf. Emerg. Netw. Experim. Technol., pp. 1-12, 2009, Rome, Italy.

[2] U. Brandes, "A faster algorithm for betweenness centrality," J. Math. Sociol., vol. 25, no. 2, pp. 163-177, 2001.

[3] N. Kitsuwan, H. Tahara, and E. Oki, "Heuristic Approach to Determining Cache Node Locations in Content-Centric Networks," IEEE Access, vol. 6, pp. 1542-1549, 2018.

[4] (April 2022). Japan Photonic Network Model. [Online]. Available: http://www.ieice.org/cs/pn/jpn/jpnm.html