On the Optimality of Shortest-path Routing in Information-Centric Networking

RYO NAKAMURA1,a) HIROYUKI OHSAKI2,b)

Received: November 8, 2019, Accepted: June 1, 2020

Abstract: In recent years, Information-Centric Networking (ICN) that mainly focuses on contents that are transferred and received instead of on end hosts that transmit and receive contents has been under the spotlight. A straightforward approach for content routing in ICNs is to utilize a class of shortest-path routing mechanisms. In past research papers, several cache-aware routing mechanisms for ICNs to take advantage of content caches at intermediate routers have been proposed. In past studies, however, only link-level performance metrics (e.g., cache hit ratio and server load) of content routings have been investigated. Hence, it has been still unclear how the shortest-path routing is effective (or ineffective) in terms of application-level performance metrics. In this paper, we try to answer research questions regarding the optimality of the shortest-path routing. We compare the application-level performances with the shortest-path routing and with the optimal routing obtained by searching all detour paths existing in the vicinity of the shortest-path routing (optimal k-hop detour routing). Our findings include that the shortest-path routing is suitable when the network is balanced and cache sizes at routers are homogeneous, and that the optimal k-hop detour routing is suitable when variation in cache sizes is large.

Keywords: ICN (Information-Centric Networking), shortest-path routing, optimal k-hop detour routing, content delivery delay, routing optimality

1. Introduction

In recent years, Information-Centric Networking (ICN) that mainly focuses on contents that are transferred and received instead of on end hosts that transmit and receive contents has been under the spotlight [1], [2], [3]. Two major network architectures for realizing ICNs are CCN (Content-Centric Networking) [2] and NDN (Named Data Networking) [3]. In ICNs, the content request packet sent from an entity (i.e., content consumer) is forwarded to a repository (i.e., content provider) that stores the content. The requested content is returned to the entity from the repository as a response packet by retracing the trajectory of the request packet.

Two major challenges in the ICN architecture design are content caching and content routing [4]. Content caching improves the way a router in a network caches contents for performance improvement in terms of the reduction in the traffic volume transferred through the network. On the other hand, content routing is aimed at effectively discovering the content by appropriately selecting the path from a router to the nearest repository which stores the content.

A large number of studies have been devoted for designing content caching mechanisms in ICNs [5], [6]. In those studies, several content caching mechanisms to effectively utilize caches of intermediate routers on the path are proposed. For instance, authors of Ref. [5] proposed an in-network caching scheme, in which a router caches contents based on the probability calculated from a distance to its destination and the caching capability of other routers along a path. Also, authors of Ref. [6] focused on round-trip times for contents measured at a router, and proposed a caching algorithm based on the probability calculated from observed round-trip times.

On the other hand, not many but several studies have been devoted for content routing in ICNs [7], [8], [9]. A few studies focus on routing mechanisms for request and content packets to achieve better performance than that with the shortest-path routing.

A straightforward approach for content routing in ICNs is to utilize a class of shortest-path routing mechanisms. In similar fashion to the conventional IP network, usage of the shortest-path routing simply based on the number of hops or the link-level metrics between a router and a repository, is considered in CCN [2]. However, since routers on the path cache contents in ICNs, the shortest-path routing might not be always optimal.

In past research papers, several cache-aware routing mechanisms to take advantage of content caches at intermediate routers have been proposed (see, e.g., Refs. [7], [8]). Generally, cache-aware routing mechanisms determine the path through which a request packet is forwarded by taking account of both the proximity of content replicas and the likelihood of cache hits at intermediate routers. Cache-aware routings are expected to reduce the server load as well as the content delivery delay.

In those studies, however, only link-level performance metrics...
Routing is effective (or ineffective) in terms of application-level performance metrics (e.g., content delivery delay and throughput). Also, cache-aware routing mechanisms generally rely on the availability of cache hit ratios and/or cache sizes of routers, which might not be easy to obtain accurately in a timely fashion.

For a given scenario, the shortest-path routing might not provide the best application-level performance. A sophisticated content routing should provide a better application-level performance than that with the shortest-path routing. However, under realistic scenarios where many factors (e.g., network topology, bandwidths and propagation delays of links, cache and buffer sizes of routers, and workloads generated from entities) are varying and/or uncertain, an overly-optimized content routing could result in a poor performance.

In this paper, we investigate the optimality of the shortest-path routing by comparing the performances with shortest-path routing and with an optimal routing obtained by searching all detour paths existing in the vicinity of the shortest-path routing (optimal k-hop detour routing). We focus on the average content delivery delay, which is one of the key application-level performance metrics. Through a number of experiments, we compare average content delivery delays with the shortest-path routing and with the optimal k-hop detour routing.

This paper addresses the following research questions regarding the optimality of shortest-path routing in ICNs.

Q1. Under a given condition, of the shortest-path routing or the optimal k-hop detour routing which is suitable in terms of the average content delivery delay?

Q2. How robust are the shortest-path routing and the optimal k-hop detour routing against measurement errors in cache hit ratios at routers?

In past research papers, there have been several studies that proposed effective routing mechanisms for ICNs and evaluated their effectiveness (see e.g., Refs. [7], [10]). Different from those studies, this paper is aimed at investigating the optimality of the shortest-path routing compared to a near-optimal content routing (i.e., the optimal k-hop detour routing). The optimal k-hop detour routing is prohibitive in practice, but near-optimal in theory.

This paper is an extension of our previous work [11], as here we investigate the effectiveness of the shortest-path routing in ICNs under two simple network topologies.

In this paper, we extend our previous work to further investigate the optimality of the shortest-path routing in diverse scenarios. We quantitatively investigate the optimality of the shortest-path routing in ICNs by comparing the average content delivery delay under the shortest-path routing with that under the optimal two-hop detour routing in several networks (triangular network, grid network, and cluster network). Furthermore, we also investigate the robustness of the shortest-path routing against measurement errors in cache hit ratios at routers.

This paper is organized as follows. First, Section 2 introduces previous works related to content routings in ICNs. Section 3 explains the methodology to investigate the optimality of the shortest-path routing in ICNs. Section 4 presents experiment results and discusses the optimality of the shortest-path routing. Finally, Section 5 provides the summary of this paper and discusses future works.

2. Related Works

In past research papers, it is known that sophisticated content routings including the cache-aware routing achieve better performance than the shortest-path routing in ICN [7], [8], [9], [12], [13]. Authors of Ref. [7] proposed a cache-aware routing that dynamically selects the path so that the number of hops to retrieve the content can be minimized [7]. They reported that, with their cache-aware routing, the server load can be reduced by approximately 18% from the shortest-path routing. Another cache-aware routing is proposed in Ref. [8]. The authors proposed a weight-based cache-aware routing that minimizes the content access delay based on the existence of content cache at routers. Authors of Ref. [9] proposed an efficient content routing by adopting a different approach than cache-aware routing. Specifically, in the proposed content routing, a router measures round-trip times for contents returned from repositories, and it probabilistically determines a next node to forward a request packet based on measured round-trip times. Through simulations, it was shown that, compared to the shortest-path routing, the proposed content routing can improve the content delivery delay because of reducing loads occurred at the repository.

However, a few studies reported that there is no significant difference in the performance between the shortest-path routing and the cache-aware routing [14]. Authors of Ref. [14] compared the shortest-path routing and the nearest-replica routing, which is one of cache-aware routings, while changing several factors such as the network topology and content request pattern. In nearest-replica routing, a request packet for a content from an entity is delivered to the nearest router/repository which is storing the requested content. Although the nearest-replica routing is not a practical routing mechanism, it can be regarded as a baseline for other content routings in ICNs. In Ref. [14], improvements in all performance metrics (e.g., the number of hops required for contents delivery) with the nearest-replica routing is at most 2%, compared to the shortest-path routing.

3. Method

3.1 Optimal k-Hop Detour Routing

In this paper, we perform three types of experiments to investigate the optimality of shortest-path routing. In all experiments, we compare the average content delivery delay under the shortest-path routing with that under the optimal two-hop detour routing.

The optimal k-hop detour routing is defined as the content routing with the least average content delivery delay among all possible k-hop detour paths obtained from the shortest path (Fig. 1). In this paper, k-hop detour path is defined as follows. k-hop detour path $P^k$ for path $P = (s, \ldots, t)$ of length $l$ is a path between node $s$ and node $t$ containing all nodes in path $P (v \in P)$, whose length is no more than $l + k$. In our experiments, we obtained the optimal two-hop detour routing using an exhaustive search: (1) obtain all two-hop detour paths regarding arbitrary single hops in the shortest-path, (2) calculate average content delivery delays of
all two-hop detour paths using our performance analysis, and (3) select the path with the least average content delivery delay.

Note that the optimal $k$-hop detour routing is not for practical purposes, but for theoretical analyses. The optimal $k$-hop detour routing is based on the idea that, even with caches at routers, the optimal path should be more or less similar to the shortest-path. Namely, the optimal $k$-hop detour routing exhaustively searches solution space around the shortest-path to hopefully find a reasonably better path than the shortest-path.

In what follows, the methodologies of three experiments are explained.

### 3.2 E1: Effect of Giant Cache

In this experiment, the cache size of a specific router is varied to examine how the existence of a giant cache affects the effectiveness of the shortest-path routing as well as the optimal two-hop detour routing.

As network topology, we use a triangular network shown in Fig. 2(a). The communication delays of links between an entity and a router are negligibly small and the communication delays of all other links are equally set to 1 [ms].

Cache sizes of all routers are equally set to $C = 5$ or 10 [content]. The cache size of a specific router (router 1, shaded router in Fig. 2(a)) is varied between 0–40 [content] for investigating how the cache size affects the optimality of the shortest-path routing. The cache replacement algorithms at all routers are LRU (Least-Recently Used), which is widely used for the performance evaluation and analyses of cache networks [15].

50 contents are placed at a single repository, and every entity (i.e., content consumer that injects content requests to the network) randomly and continuously requests contents of either the rate $k$ [request/ms] for content $k$ ($1 \leq k \leq 50$), or the popularity following Zipf distribution with the mean of 200 [request/s] and the exponent parameter of 1.0.

Utilizing our ICN performance analysis [16], the average content delivery delay (i.e., the average time required for an entity to retrieve a content) for a given content routing is calculated. In the performance analysis of ICN [16], under several assumptions, for instance, all content sizes are equal and congestion does not occur, effects of content size and bandwidth and processing delays at nodes (i.e., router/repository) are expressed as communication delay between nodes. For evaluating a communication network which is comprised of entities (i.e., end users) and intermediate nodes (i.e., routers), its performance is measured by metrics (e.g., response time and throughput) directly related with the service [17]. In this paper, since we focused on the performance of network rather than that of the router (e.g., cache hit ratio), we used the average content delivery delay, which is one of the major performance metrics for evaluating the network.

### 3.3 E2: Effect of Cache Sparseness

In this experiment, different from experiment E1, the density of cache-equipped routers (cache sparseness) is varied. Namely, instead of changing the cache size of a specific router, cache sizes of all routers in the network are uniformly changed.

As network topologies, we use three network topologies: triangular network, grid network, and cluster network shown in Figs. 3(a), 4(a), and 5(a), respectively. Every network topology has a single repository. Note that there are multiple shortest-paths from an entity to the repository in the grid network (Fig. 4(a)). In our experiments, we selected a shortest-path with an intermediate node whose identifier is smaller. In the grid network, for
instance, the shortest-path from the router 1 to the repository is
the path through the router 2 rather than the router 4. Similar
to experiment E1, the communication delays of links between an
entity and a router are negligibly small and the communication
delays of all other links are equally set to 1 [ms]. The request rate
is given by a Zipf distribution with the exponent parameter of 1.0.

To adjust the density of caches in the network, we define the pa-
rameter $M$ called cache sparseness. Cache sparseness $M$ is a pos-
itive integer that controls the density of cache-equipped routers.
Namely, for a given parameter $k$ ($0 \leq k < M$), every router whose
identifier $i$ satisfies $i \equiv k \pmod{M}$ has the cache size $C_1$, and all
other routers have the cache size $C_2$. For instance, all routers
have the same cache size for $M = 1$, and the one-fourth of routers
have the cache size $C_1$ and others have $C_2$ for $M = 4$. We use
$C_2 = 0$ [content]. For given parameters $M$, $C_1$, and $C_2$, we
obtained the mean of average content delivery delays when $k$ was
varied between $0$–$M − 1$.

All other conditions are the same as those in experiment E1.

3.4 E3: Robustness against Measurement Errors in Cache
Hit Ratios

The third experiment investigates how the effectiveness of the
shortest-path routing and the optimal two-hop detour routing is
affected by measurement errors in cache hit ratios. Effectiveness
of cache-aware routings, including the optimal $k$-hop routing, de-
pends on the accuracy and timeliness of cache hit ratios at routers.
In this experiment, we simulate how the performance of the two-
hop detour routing is degraded when it chooses paths using dirty
cache hit ratios.

We use the same conditions with those in experiment E2 except
the computation of the average content delivery delay, which will
be explained below, and $C_1 = 20$.

For a given condition (e.g., network topology, cache size $C_1$,
cache sparseness $M$, and routing algorithm), we first determine
routes from all entities to the repository, and compute the cache
hit ratios at all routers using our ICN performance analysis [16].
We regard these cache hit ratios as dirty cache hit ratios; i.e., con-
taminated cache hit ratios used by the optimal two-hop detour
routing. As genuine cache hit ratios, we randomly generate 100
sets of cache hit ratios using a parameter $\epsilon$. Specifically, every
cache hit ratio is multiplied by a random number following the
uniform distribution $[1 − \epsilon : 1 + \epsilon]$. By definition, the cache hit
ratio which is larger than 1 is truncated to 1.
4. Results and Discussion

4.1 E1: Effect of Giant Cache

Average content delivery delays under shortest-path routing and optimal two-hop detour routing when changing the cache size of router 1 are shown in Figs. 2 (b) and 2 (c). In these figures, “proportional” indicates the proportional request pattern whereas “Zipf” indicates the Zipf-distributed request pattern.

One can find from these figures that the optimality of the shortest-path routing depends on the cache sizes of routers on the path. It can also be found that the shortest-path routing is optimal regardless of the content request patterns when cache sizes of all routers are identical. Since the triangular network is equally-balanced, it is intuitive that the shortest-path routing is optimal when cache sizes are the same.

The larger the cache size of router 1 becomes, the less optimal the shortest-path routing becomes. From Figs. 2 (b) and 2 (c), it is found that the average content delivery delay under the optimal two-hop detour routing becomes smaller when the cache size ratio (i.e., the ratio of the cache size of router 1 to that of other routers) exceeds approximately 2–3.

These results show that the shortest-path routing is optimal under a balanced network (i.e., the network whose structure is uniform such as triangular network) with comparable cache sizes at routers, and that the optimal two-hop detour routing archives a better application-level performance when the cache size ratio is large.

4.2 E2: Effect of Cache Sparseness

Average content delivery delays under shortest-path routing and optimal two-hop detour routing when changing cache sparseness $M$ are shown in Figs. 3 (b), 4 (b) and 5 (b). To clearly reveal differences in average content delivery delays with the shortest-path routing and the optimal two-hop detour routing, detouring inefficiency (the average content delivery delay with the optimal two-hop routing/that with the shortest-path routing) are plotted in Figs. 3 (c), 4 (c) and 5 (c).

One can find from these figures that the shortest-path routing achieves the best performance when the cache sparseness is very low (i.e., $M = 1$). This observation agrees with our finding in the previous section; i.e., all networks used in experiment E2 are equally-balanced so that the shortest-path routing is the optimal.

On the contrary, if the caches are sparse in routers (e.g., $M = 2$ or $M = 3$), the shortest-path routing shows worse performance than that with the optimal two-hop detour routing in triangular and cluster networks. In particular, the optimal two-hop detour routing is quite effective in the triangular network with a large cache size (e.g., $C_1 = 30$) and modest cache sparseness (e.g., $M = 2$).

Surprisingly, regardless of the cache sparseness and the cache size, the shortest-path routing is always optimal in the grid network. This implies that the cache-aware routing, including the two-hop detour routing, should be carefully deployed since the (generally complex) cache-aware routing does not always achieve better performance than the simplest shortest-path routing.

In the following, we discuss the effectiveness of the shortest-path routing in ICN. From Figs. 3 (c) through 5 (c), except for cases with $M = 2, 3, 4$ in the triangular network, the difference in the average content delivery delay with the shortest-path routing and that with the optimal two-hop detour routing is about 10%, which implies that the shortest-path routing can achieve sufficient performance without using complex routing mechanisms such as the optimal two-hop detour routing. Note that, in our experiments, the ratio of the average content delivery delay with the shortest-path routing to that with the optimal two-hop detour routing (Figs. 3 (c) through 5 (c)) is not affected by the value of link delay. Therefore, even though link delays are varied, our observations are still valid.

4.3 E3: Robustness against Measurement Errors in Cache Hit Ratios

Average content delivery delays under the shortest-path routing and the optimal two-hop detour routing are shown in Figs. 6 (a) and 7 (a). Also, degradation factors for a given content routing, which is defined as the ratio of the average content delivery delays calculated from dirty cache hit ratios to those calculated from genuine cache hit ratios, are shown in Figs. 6 (b) and 7 (b). Figures 6 and 7 show the results for $\epsilon = 0.5$ and $\epsilon = 1.0$, respectively. In those figures, average content delivery delays sorted in descending order are plotted.
One can find from these figures that, even though the cache hit ratios include errors, the shortest-path routing is still better than the optimal-two hop detour routing in the case of \( M = 1 \). Hence, both of the shortest-path routing and the optimal two-hop detour routing are affected by measurement errors. In particular, it is found that the variation in average content delivery delays with the optimal two-hop detour routing is large in the case of \( M = 2, 3 \). On the contrary, the shortest-path routing achieves almost the same performance regardless of measurement errors in cache hit ratios in the case of \( M = 2, 3 \).

**4.4 Discussion**

In the following, we answer research questions described in Section 1 from observations in experiments E1–E3, and discuss the optimality of the shortest-path routing in ICN.

Q1. Under a given condition, of the shortest-path routing or the optimal \( k \)-hop detour routing which is suitable in terms of the average content delivery delay?

The shortest-path routing is suitable when the network is balanced and cache sizes of routers are homogeneously allocated (i.e., cache sizes of all routers are equal). In contrast, the optimal \( k \)-hop detour routing is suitable when variation in cache sizes is large; i.e., a specific router has a giant cache or a part of routers have caches.

Q2. How robust are the shortest-path routing and the optimal \( k \)-hop detour routing against measurement errors in cache hit ratios at routers?

When cache sizes of routers are homogeneously allocated, average content delivery delays with both of the shortest-path routing and the optimal two-hop detour routings increase due to measurement errors in cache hit ratios. In contrast, when cache sizes of routers are heterogeneously allocated, the average content delivery delay with only the two-hop optimal routing is degraded.

**5. Conclusion**

In this paper, we have investigated the optimality of the shortest-path routing in terms of application-level performance metrics. Specifically, we have compared the average content delivery with the shortest-path routing and that with the optimal two-hop detour routings through a number of experiments. Our findings include that the shortest-path routing is suitable when the network is balanced and cache sizes of routers are homogeneously allocated, and that the optimal \( k \)-hop detour routing is suitable when the variation in cache sizes is large. Furthermore, we have investigated the robustness of the shortest-path routing and the optimal \( k \)-hop detour routing against measurement errors in cache hit ratios. Consequently, we have shown that the shortest-path routing achieves almost the same performance regardless of measurement errors in cache hit ratios when cache
sizes of routers are heterogeneously allocated.

Our future challenges include experiments with realistic network topologies and the performance comparison in terms of other application-level metrics such as throughput.

Acknowledgments This work was partly supported by JSPS KAKENHI Grant Number 19H04104 and 18J10278.

References

[1] Xylomenos, G. et al.: A Survey of Information-Centric Networking Research, IEEE Communications Surveys & Tutorials, Vol.16, pp.1024–1049 (2014).
[2] Jacobson, V., Smetters, D.K., Thornton, J.D., Plass, M.F., Briggs, N.H. and Braynard, R.L.: Networking named content, Proc. 5th International Conference on Emerging Networking Experiments and Technologies (CoNEXT 2009), pp.1–12 (2009).
[3] Zhang, L. et al.: Named data networking, ACM SIGCOMM Computer Communication Review, Vol.44, pp.66–73 (2014).
[4] Ioannou, A. and Weber, S.: A Survey of Caching Policies and Forwarding Mechanisms in Information-Centric Networking, IEEE Communications Surveys & Tutorials, Vol.18, pp.2847–2886 (2016).
[5] Psaras, I., Chai, W.K. and Pavlou, G.: Probabilistic in-network caching for information-centric networks, Proc. 2nd ACM SIGCOMM Workshop on Information-Centric Networking (ICN 2012), pp.55–60 (2012).
[6] Yokota, K., Sugiyama, K., Kurihara, J. and Tagami, A.: RTT-Based Caching Policies to Improve User-Centric Performance in CCN, Proc. 30th IEEE International Conference on Advanced Information Networking and Applications (AINA 2016), pp.124–131 (2016).
[7] Sourlas, V., Flegkas, P. and Tassiulas, L.: Cache-Aware Routing in Information-Centric Networks, Proc. IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp.582–588 (2013).
[8] Vassilakis, V.G., AL-Naday, M.F., Reed, M.J. and Alzahrani, B.A.: A Cache-Aware Routing Scheme for Information-Centric Networks, Proc. 9th IEEE International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP 2014), pp.721–726 (2014).
[9] Mori, T., Hirata, K. and Yamamoto, M.: Content-oriented probabilistic routing with measured RTT, Proc. IEEE Communications Society Communications Quality and Reliability Workshop (CQR 2016), pp.1–6 (2016).
[10] Mahdian, M. and Yeh, E.: MinDelay: Low-Latency Joint Caching and Forwarding for Multi-hop Networks, Proc. IEEE International Conference on Communications (ICC 2018), pp.1–7 (2018).
[11] Nakamura, R. and Ohsaki, H.: Performance Comparison of Shortest-Path Routing and Optimal Detour Routing in Content-Centric Networking, Proc. 40th IEEE Signature Conference on Computers, Software, and Applications (Doctoral Symposium) (COMPSAC 2016), pp.494–495 (2016).
[12] Rossini, G. and Rossi, D.: Coupling Caching and Forwarding: Benefits, Analysis, and Implementation, Proc. 1st ACM Conference on Information-Centric Networking (ICN 2014), pp.127–136 (2014).
[13] Mordjana, Y., Senouci, M.R. and Mellouk, A.: Performance Analysis of Caching and Forwarding Strategies in Content-Centric Networking, Proc. IEEE Global Telecommunications Conference (GLOBECOM 2017), pp.1–6 (2017).
[14] Fayazbakhsh, S.K. et al.: Less Pain, Most of the Gain: Incrementally Deployable ICN, Proc. ACM SIGCOMM 2013, pp.147–158 (2013).
[15] Garetto, M., Leonardi, E. and Martina, V.: A Unified Approach to the Performance Analysis of Caching Systems, ACM SIGCOMM Computer Communication Review, Vol.1, pp.1–28 (2016).
[16] Nakamura, R. and Ohsaki, H.: Performance Analysis of Content-Centric Networking on an Arbitrary Network Topology, IEICE Trans. Communications, Special Section on Internet Technologies to Accelerate Smart Society, Vol.E101-B, No.1, pp.24–34 (2018).
[17] Jain, R.: The Art of Computer Systems Performance Analysis, Wiley-Interscience (1991).