Performance Assessment of Routing Strategies in Named Data Networking

Michele Tortelli, Luigi Alfredo Grieco and Gennaro Boggia
DEI, Politecnico di Bari (Italy), {m.tortelli,a.grieco,g.boggia}@poliba.it.

***Abstract*—Information Centric Networking (ICN) architectures are currently being investigated to orient the Future Internet towards a content centric paradigm, thus allowing the provisioning of more secure, efficient, and scalable services. In this work, we focus on the Named Data Networking (NDN) proposal to analyze the impact of several routing and forwarding strategies, which play a fundamental role in ICN. In particular, thanks to the recently devised ns-3 based NDN simulator, namely *ndnSIM*, we conduce an extensive simulation campaign using the GEANT topology as a core network. We monitor different distinctive metrics, such as file download time, server load reduction, hit ratio, hit distance, and communication overhead, in different topologies and traffic conditions. Simulation results show that the election of a single best forwarding strategy is a difficult task. Indeed, the pros and cons of each strategy are heavily influenced by the popularity distribution of contents, which, in turn, affects the effectiveness of the distributed caching mechanisms typically used in the NDN architecture.**

***Index Terms*—Named Data Networking, Routing Strategies, Performance Evaluation.**

I. Introduction

The Information Centric Networking (ICN) paradigm [1], which is expected to play a key role in the Future Internet, encompasses different proposals and projects, such as [2]–[9], with a shared common ambition: introducing content-driven networking operations to enhance and/or replace the current Internet model, which is entirely based on IP addresses. Albeit with different characteristics, all ICN architectures aim at pursuing the following main achievements [1]: the resolution of some inefficiencies of the current Internet, such as security and mobility; an improved capability to handle disruptions and flash crowds; the maximization of the Quality of Experience (QoE) perceived by end-users, and the reduction of the servers' load. One of the novelties introduced by ICN architectures is the capability of nodes to cache copies of retrieved contents. The way this distributed caching capacity is implemented can be different from one ICN proposal to the other [1], but the common goal is to migrate content copies nearer to the end users, thus satisfying subsequent requests without reaching the original content providers.

The entire process of generating content requests and forwarding them towards an available copy is heavily affected by the adopted naming convention and by the popularity of the contents. Indeed, the name space design directly influences scalability, security, routing and forwarding strategies [10], which in turn shape the way content copies are cached in the network. The ICN proposals conceived so far adopt two different schemes for naming contents: hierarchical names are used in NDN [11] and CONVERGENCE [12], while flat self-certifying ones are used in Scalable and Adaptive Internet Solutions (SAIL) [5], Publish Subscribe Internet Technology (PURSUIT) [4] and Data-Oriented Network Architecture (DONA) [2], to name a few. Hierarchical names easily lend themselves to aggregation and longest Prefix Match (LPM) operations, and they can be human-readable; nevertheless, they could lack in global uniqueness. On the contrary, flat self-certifying names guarantee global uniqueness, but they require an external resolution entity to bind them to human-readable names [10]. For what concerns contents popularity, instead, the ICN literature ([13]–[16]) largely adopts the Zipf's probability distribution as a reference model. This choice is in line with the consolidated results of the past literature which elected the Zipf's probability model as the best candidate to represent the pattern of clients requests in several scenarios, such as Web, File Sharing, Video on Demand (VoD) and User Generated Content (UGC), as observed in [17]. As already pointed out in [14], the choice of the Zipf's parameter (see Section II) heavily influences the effectiveness of caching contents in intermediate routers, in terms of cache dimensioning and hit ratio. As a consequence, the evaluation of possible forwarding strategies to be used in ICN is also influenced by the supposed popularity distribution of contents.

In this paper, focusing on the NDN architecture, we present an evaluation study on the performances of different forwarding strategies under different hypotheses. By means of a recently devised ns-3 based simulator, called "ndnSIM" [18], we compare, in realistic scenarios, three forwarding strategies: *flooding*, *best-route with caching* (i.e., every node knows the best path to reach the original content providers) and *best-route without caching*. We evaluate pros and cons of each one by supposing a Zipf-like content probability distribution and by varying its parameter in the range [0.8,1.4]. The comparison is based on different metrics, such as file download time, hit distance, hit ratio, communication overhead and servers' load reduction. The obtained results (Section III) show that the Zipf's parameter, because of its strong influence on the effectiveness of caching contents in the networks, also polarizes the evaluation of the forwarding strategies under comparison. This means that it is not possible to elect a best strategy, and future efforts should try to catch the benefits of both Flooding and Best Route.

The remainder of this paper is organized as follows. In Section II we revisit the basic aspects of the NDN architecture, and we report a small description of the Zipf's law and of its implications. In Section III, we describe the settings used in our simulations, the evaluated metrics, as well as the obtained results. In the end, in Section IV, we remark our conclusions and our objectives for future works.

II. NDN and Zipf's Law Background

In this section we report a brief overview of the NDN architecture and we introduce the basic principles of the Zipf's law.

One of the main differences between the Content Centric Networking (CCN) architecture [3], now in the NDN project [11], and the other ICN proposals lies in the adoption of hierarchical and human readable names to identify contents. Every content name, in NDN, is formed by several components arranged hierarchically, so that every name prefix identifies a sub-tree in the name space. Each component, moreover, can be any string of arbitrary length. In NDN, communications are receiver driven and they are based on only two types of messages: *Interest* and *Data*. The former is generated by a client that wants to fetch a particular content, while the latter encapsulates the requested content item, and it can be generated by every node having the requested content in its cache. An *Interest* is said to be satisfied by a *Data* when the content name in the *Interest* is either the same or a prefix of the name contained in the generated *Data* packet. Three main data structures are used in a NDN node: the Content Store (CS), which is the cache memory; the Forwarding Information Base (FIB), which stores the next-hop information for every known content prefix (more than one interface can be associated to a single prefix entry); the Pending Interest Table (PIT), which keeps track of the arrival faces associated to the *Interest* packets which have been previously forwarded but that are still unsatisfied. This architecture allows routing and forwarding operations to be made only on *Interest* packets; indeed, *Data* packets can simply follow back the road to the requesters by exploiting the states created along the path (i.e. PIT entries) by the forwarded *Interests*. Furthermore, this stateful forwarding plane avoid *Interests* loop and it permits their aggregation (*Interests* for the same content can be aggregated in the same PIT entry, recording the respective incoming faces and forwarding only the first one).

The NDN architecture, along with the other ICN proposals, is supposed to bring benefits to several services based on the Internet, such as Web, File Sharing (like BitTorrent), UGC (with the YouTube leadership) and VoD [17]. There have been several studies in past years which stated that the Zipf's distribution is the discrete distribution that best represents the request frequency of contents in all the aforementioned scenarios, as stated in [17]. That is why the Zipf's distribution is also adopted when evaluating forwarding or caching performances in NDN. The logic of the Zipf's distribution is that the frequency of a content request is inversely proportional to the rank of the content itself (i.e., the smaller the rank the higher the request frequency). If we denote with M the content catalog cardinality and with $1 \leq i \leq M$ the rank of the i-th most popular content, the probability of requesting the content with rank i can be expressed as:

$$P(X = i) = \frac{1/i^{\alpha}}{C}. \tag{1}$$

with $C = \sum_{j=1}^{M} 1/j^{\alpha}$, where $\alpha > 0$. The percentage of requests directed to the k most popular contents, i.e., $P(X = i) = k$, is representative of the impact that the α parameter has in shaping content requests. If we consider a content catalog with cardinality 10^5, which is quite bigger than the catalog estimated for a VoD service [17], we obtain the results shown in Tab. I, which reports the percentage of the content catalog representative of the 95% of the content requests for different values of α. It can be noticed that with a small variation in the α parameter, passing from 1.2 to 1.4, there is a considerable reduction in the percentage of the content catalog to which clients requests are directed. In particular, with $\alpha = 1.4$, the 95% of requests are directed

TABLE I
PERCENTAGE OF THE CONTENT CATALOG REPRESENTING THE 95% OF REQUESTS

α	[%]
0.8	79.1
1	54.5
1.2	12.6
1.4	0.7

towards only the 700 most popular contents.

III. PERFORMANCE EVALUATION

A. Simulator and Routing Strategies Under Comparison

In this work the *ndnSIM* [18] simulator is used to evaluate the performance of considered algorithms. It is a recently devised NDN simulator based an ns-3 [19], released by the Internet Research Lab of UCLA. It reproduces the basic structures of a NDN node (i.e., CS, PIT, FIB, strategy layer, and so on), and it lets the users simulate different scenarios according to their needs.

We compare three forwarding strategies in our tests:

- *Flooding*; every node forwards the received *Interests* towards all of its interfaces except for the incoming one (i.e., the interface it received the *Interest* from.).
- *Best-Route with Caching*; in this scenario, we apply the Dijkstra's algorithm [20] to calculate the best paths (in terms of the minor number of hops) to reach every permanent content copy; then, we set up the FIB of each node accordingly.
- *Best-Route without Caching*; in this scenario we eliminate the in-network caching capacity, so the *Interests* expressed by clients are satisfied only by repositories which store the seed copies. This will let us evaluate the effective benefits introduced by a distributed caching capacity, as fostered by NDN, under different values of the α parameter, and correlate these benefits with the pros and cons of the forwarding strategies under comparison.

B. Topology, Content Catalog and Caches

For what concerns the topology used in tests, we simulate the GEANT network [21], reported in Fig. 1, as done in [22] or [13]. It is a core network, with a diameter of 6 hops, composed by 22 routers, which interconnects several European research institutes and universities. At each core router, we consider a variable number of attached clients, ranging form 3 to 6, thus obtaining a total number of nodes equal to 95. The number of repositories which store permanent content copies, as well as the points of attachment to the core network, is randomly varied in each simulated run. In particular, these repositories share the seed copies in a way that only a single permanent copy for each content is present in the network.

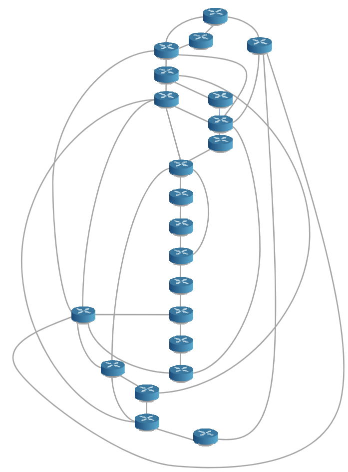

Fig. 1. GEANT Topology.

We use a catalog of 10^5 unique contents, which are supposed to have a geometrically distributed size, with an average size of 100 chunks (as in [13]). Given that *ndnSIM* does not support fragmentation at the time of our experiments, we fixed the payload of each chunk at 1500 bytes (i.e., the Ethernet MTU). For what concerns content catalog, as largely done in the ICN literature ([13]–[16]), contents are characterized by a Zipf-like probability distribution, as expressed by Eq. (1). To span over multiple cases, we evaluate the three aforementioned forwarding strategies using different values of the paramater α in the interval $[0.8, 1.4]$.

We dimension the cache size of each node in terms of the number of chunks it can store; specifically, we fixed the cache size to 10^4 chunks, corresponding to the 0.1% of the catalog size. Although this ratio is quite larger with respect to the values considered in the ICN literature [14], chosing a smaller ratio would have been worthless with the simulated catalog, because each node would have cached less than 10 contents.

C. Evaluated Metrics

Our analysis is based on the following metrics.

- *Hit Distance*: it represents to number of hops that an *Interest* has to travel to find the desired data item.
- *Hit Ratio*: it is the ratio between the total number of *Interests* received by a node and the number of

Data packets generated by it (i.e., the number of satisfied *Interests* at that node).
- *File Download Time*: it is the sum of the download times of all the chunks that form a content object.
- *Server Load Reduction*: it is calculated as 1 minus the ratio between the number of content objects generated by repositories in the flooding or in the best-route scenario, and the number of content objects generated by repositories in the best-route with no caching scenario, which is the worst case.
- *Overhead*: it represents the amount of bytes injected in the network by core routers for each Interest issued by clients. More specifically, it is a ratio where the numerator is obtained by summing the size of both the *Data* packets, generated and forwarded, and of the *Interests* packets forwarded by core routers, whereas the denominator is equal to the sum of the *Interests* expressed by all the clients.

A summary of all the parameters adopted in the simulated scenarios is reported in Tab. II.

D. Results

File download time, hit distance, hit ratio and communication overhead are shown in Fig. 2, while the reduction of the servers' load with respect to the Best Route without caching scenario is reported in Fig. 3.

In particular, we average the results of eight runs for each evaluated metric and for each forwarding strategy; furthermore, we report the average values with the 95% confidence intervals.

For what concerns the *mean file download time*, which is reported in Fig. 2.(a), the advantages of using a Flooding strategy with respect to a Best Route one (with or without caching) are clear. Indeed, comparing Flooding with the Best Route with caching case, it is possible to note a reduction of the file download time that goes from the 27% (with $\alpha = 0.8$) to the 50% (with $\alpha = 1.4$). This result remarks the importance of finding a routing and forwarding stategy capable of discovering extra path content copies. When $\alpha = 1.4$, the union of the caching capacity with a Flooding strategy leads to a reduction of the 64% of the file download time with respect to the Best Route case without caching. Furthermore, we can note, frome Fig. 2.(a), that the advantage of caching contents in intermediate nodes between clients and servers becomes more and more evident as α increases. This reflects in the progressive decreasing of the file download time for the Best Route with caching scenario.

The Flooding strategy is the best choice by observing also the *mean hit distance*, reported in Fig. 2.(b). If we average among the results associated with the different values of α, we obtain a mean hit distance of 3.7, 4.2, and 4.8 hops for the Flooding, Best Route with caching, and Best Route without caching strategy, respectively. With $\alpha = 1.4$, the path stretch (i.e., the ratio between the length of the used path and the length of the known best path) between Flooding and Best Route without caching is equal to 0.6. The same considerations about the connection between α and the benefits of caching reported above for the file download time are also valid also for the hit distance.

However, the advantages that characterize the Flooding strategy, like a smaller file download time and hit distance for all the values of α, are counterbalanced by the negative effects of a high communication overhead, as shown in Fig. 2.(c), and a small hit ratio, as can be seen from Fig. 2.(d). First of all, the higher amount of bytes injected in the network to retrieve a single chunk, when using the Flooding strategy, is due to both an uncontrolled and excessive propagation of the *Interest* massages, and to the *Data* packets which are generated in response to them. This makes the overhead introduced by the Flooding strategy always higher with respect to Best Route with/without caching. An exception is present only for $\alpha = 1.4$, where the chunk retrieval overhead of the Flooding is even smaller than the overhead introduced by a Best Route strategy without caching.

The hit ratio, instead, is degraded, in case of using a Flooding strategy, because most of the *Interests* are forwarded toward paths which do not contain any copy of the requested contents. Indeed, the hit ratio remains always below the 15%, as it can be noticed from Fig. 2.(d). Furthermore, the fact that clients find the requested contents more and more closely as α increases is the cause of the increment of the hit ratio for the Best Route with caching scenario.

The impact of the parameter α on the efficiency of caching appears clearly from Fig. 3, which reports the servers' load reduction. Here, we measure the percentage of *Interests* which are not satisfied by repositories, and we compare the Flooding and the Best Route with the Best Route without caching scenario. Introducing a distributed caching capacity in the network becomes worthless if the Zipf's parameter is less than 1. In this case, clients express the 95% of their requests for a number of contents which is almost the 80% of the entire catalog, as reported in Tab. I. Indeed, we measured a reduction which is at most equal to the 10%, both for

TABLE II
SUMMARY OF SIMULATION PARAMETERS

Parameter	Value
# Nodes	95
Content Catalog	10^5
Contents Size	Geometrically Distributed with Mean value equal to 100 Chunks
Chunk Payload	1500 bytes
Popularity Distribution	Zipf with $\alpha = [0.8, 1, 1.2, 1.4]$
Cache Size	10^4 chunks
Simulation Time	28 hours

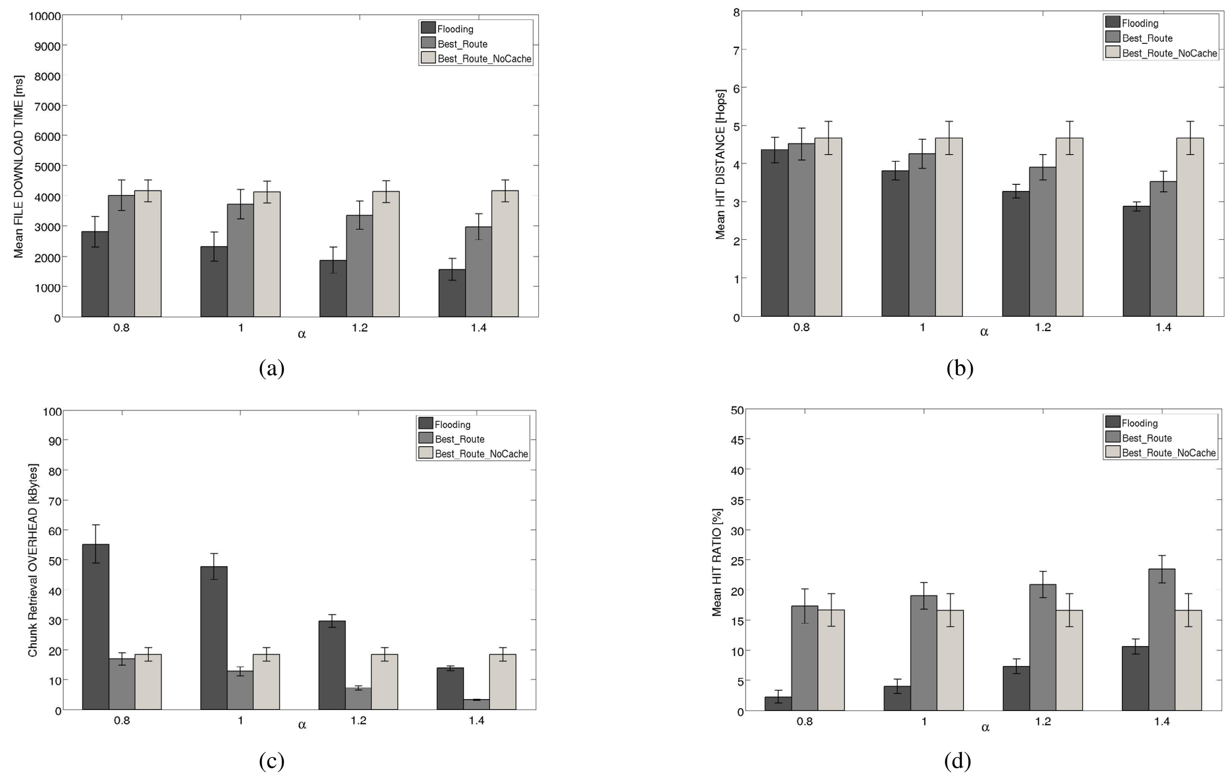

Fig. 2. Evaluated Metrics under different values of α: (a) Mean File Download Time; (b) Mean Hit Distance; (c) Chunk Retrieval Overhead; (d) Mean Hit Ratio.

Flooding and for Best Route, as it can be noticed from Fig. 3. On the contrary, when $\alpha = 1.4$, the reduction of the servers' load reaches the 86%.

In conclusion, it follows that the Flooding strategy is always the best choice in terms of file download time and hit distance, but:

- if $\alpha \leq 1$, it presents a high communication overhead and a low hit ratio due to the minimal benefits introduced by caching contents along the paths toward the original content providers;
- if $\alpha > 1$, caching contents brings some benefits that could justify the adoption of a Flooding strategy, as for the case $\alpha = 1.4$ in Fig. 2.(c), where

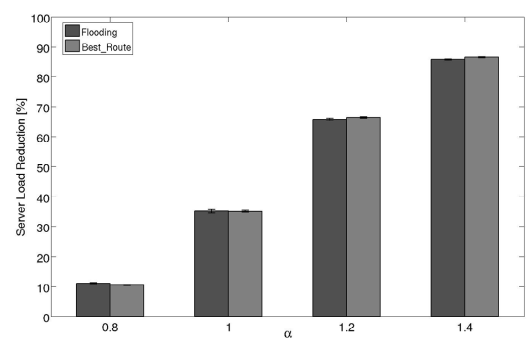

Fig. 3. Server Load Reduction.

the overhead associated to the Flooding strategy is smaller than the overhead generated by the Best Route without caching scenario.

However, it is worthwhile to note that a forwarding strategy based on the availability of the best paths towards every content involves the use of some protocols to calculate them. One possible way to discover the best paths is using IP routing strategies, which must be adopted as fall-back or primary mechanisms to be run in addition to the name based forwarding. Adapting classical Internet Protocol (IP) routing techniques to a content centric network requires big efforts [23] and it could lead to potential problems, such as enormous routing tables and longer lookup times [10].

IV. Conclusion and Future Work

The performance evaluation presented in this work has regarded two opposite forwarding strategies for content centric networks: Flooding and Best Route. We, also, considered a Best Route strategy with the absence of caching. We made a comparison by evaluating different metrics, such as hit ratio, hit distance, chunk retrieval overhead, servers load reduction and file download time, under different values of the Zipf's parameter α. It emerged that the election of a single best forwarding strategy to be used in a NDN network is impossible. Indeed, both Flooding and Best Route present pros and cons that vary with the values of α, which, in turn, affects the effectiveness of caching contents in the network along the paths towards the servers that store permanent copies. It is evident that the evaluation of future forwarding strategies must be strongly contextualized to the probability distribution used to model the pattern of requests expressed by clients.

This study can support the research for a forwarding strategy which could encompass the positive aspects of both Flooding (i.e., small hit distance and file download time) and Best Route, such as a low communication overhead and a high hit ratio. In the near future we aim at proposing our own routing and forwarding strategy for content centric network, and extending our performance evaluation to multiple and new alternatives.

References

[1] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *Communications Magazine, IEEE*, vol. 50, no. 7, pp. 26–36, Jul. 2012.

[2] T. Koponen, M. Chawla, B. Chun, A. Ermolinskiy, K. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 181–192, Aug. 2007.

[3] V. Jacobson, D. Smetters, J. Thornton, M. Plass, N. Briggss, and R. Braynard, "Networking named content," in *In Proc. of ACM CoNEXT*, Rome, Italy, Dec. 2009.

[4] N. Fotiou, P. Nikander, D. Trossen, and G. Polyzos, "Developing Information Networking Further: From PSIRP to PURSUIT," in *In Proc. of ICST BROADNETS*, Oct. 2011.

[5] B. Ahlgren and al., "Final NetInf Architecture," SAIL FP7 Project, Deliverable D.B.3, Nov. 2012.

[6] "COMET project website." [Online]. Available: www.comet-project.org/

[7] "COAST project website." [Online]. Available: http://www.coast-fp7.eu/

[8] "CONVERGENCE project website." [Online]. Available: www.ict-convergence.eu/

[9] I. Seskar, K. Nagaraja, S. Nelson, and D. Raychaudhuri, "MobilityFirst future internet architecture project," in *In Proc. of the 7th ACM AINTEC*, Bangkok, Thailand, Jul. 2011.

[10] M. Bari, S. Chowdhury, R. Ahmed, R. Boutaba, and B. Mathieu, "A survey of naming and routing in information-centric networks," *Communications Magazine, IEEE*, vol. 50, no. 12, pp. 44–53, Dec. 2012.

[11] L. Zhang and al., "Named data networking (NDN) project," PARC Technical Report NDN-0001, Oct. 2010.

[12] A. C. Anadiotis and et. al., "Final protocol architecture," CONVERGENCE Deliverable. D5.3, 2012.

[13] R. Chiocchetti, D. Rossi, G. Rossini, G. Carofiglio, and D. Perino, "Exploit the known or explore the unknown?: Hamlet-like doubts in ICN," in *In Proc. of ACM ICN Workshop*, Helsinki, Finland, Aug. 2012.

[14] D. Rossi and G. Rossini, "On sizing CCN content stores by exploiting topological information," in *In Proc. of IEEE NOMEN Workshop*, Orlando, FL, USA, Mar. 2012.

[15] I. Psaras, W. K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," in *In Proc. of ACM ICN workshop*, Helsinki, Finland, Aug. 2012.

[16] V. Ciancaglini, G. Piro, R. Loti, L. A. Grieco, and L. Liguori, "CCN-TV: a data-centric approach to real-time video services," in *In Proc. of IEEE AINA*, Barcelona, Spain, Mar. 2013.

[17] C. Fricker, P. Robert, J. Roberts, and N. Sbihi, "Impact of traffic mix on caching performance in a content-centric network," in *In Proc. of IEEE NOMEN 2012*, Orlando, USA, Mar. 2012.

[18] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN simulator for NS-3," PARC Technical Report NDN-0005, Oct. 2012.

[19] "ns-3 project website." [Online]. Available: http://www.nsnam.org/

[20] E. W. Dijkstra, "A note on two problems in connexion with graphs," *NUMERISCHE MATHEMATIK*, vol. 1, no. 1, pp. 269–271, 1959.

[21] "Geant project website." [Online]. Available: http://www.geant.net/

[22] D. Rossi and G. Rossini, "Caching performance of content centric networks under multi-path routing (and more)," Telecom ParisTech Technical Report, 2011.

[23] L. Wang, A. Hoque, C. Yi, A. Alyyan, and B. Zhang, "OSPFN: An OSPF Based Routing Protocol for Named Data Networking," Tech. Report NDN-0003, Jul. 2012.