A Cache-Aware Routing Scheme
for Information-Centric Networks

Vassilios G. Vassilakis*, Mays F. AL-Naday**, Martin J. Reed**, Bander A. Alzahrani**,
Kun Yang**, Ioannis D. Moscholios***, and Michael D. Logothetis****

*Dept. of Electronic Engineering, University of Surrey, Guildford, U.K.
**School of Computer Science & Electronic Engineering, University of Essex, Colchester, U.K.
***Dept. of Informatics & Telecommunications, University of Peloponnese, Tripolis, Greece
****WCL, Dept. of Electrical & Computer Engineering, University of Patras, Patras, Greece

Abstract—**In recent years, the information-centric networking (ICN) concept has been attracting increasing attention of the research community. The aim is to overcome intrinsic inefficiencies of the existing host-to-host communication paradigm, as well as to provide new and enhanced services to mobile and fixed users. A key feature of ICN is the support for in-network content caching. In this paper, we present a new cache-aware routing scheme for ICN. Our scheme takes into account the information about the locations of caches in the network and constructs delivery paths for efficient content dissemination. The proposed approach does not impose additional signaling overhead in the network; while at the same time it is agnostic of the cached contents. The performance of the proposed scheme is verified by simulation studies, which show an up to 50% delay reduction compared to traditional routing approaches.**

Keywords—information-centric networks; in-network caching; cache-aware routing.

I. INTRODUCTION

It has been widely acknowledged that today the Internet is mainly used for delivering multimedia content to end users, rather than connecting hosts. The user is interested in the content itself, while the *location* of the content is usually of minor importance. However, the current Internet has been built and has evolved using the host-centric communication paradigm. According to recent studies, many of the current Internet's limitations in the areas of security, mobility, efficient content dissemination, efficient use of communication, computational, and storage resources, etc. lie in its host-centric nature (e.g. [1], [2]).

Information-centric networking (ICN) is an emerging technology that provides access to named information/content/data objects as a native network service, aiming at higher efficiency and better security than today's host-based communication mechanisms. To put it simply, in a traditional host-centric network, the user sends requests containing the IP address of the host/server that stores the desired information object. On the other hand, in an ICN, the user issues a request containing the unique *name* (or identifier) of the object. It is then the duty of the network to identify the most appropriate location to retrieve the object from.

The ICN concept has recently been explored by a number of research initiatives, including DONA [3], NDN/CCN [4], SAIL [5], COMET [6], and PURSUIT/PSIRP [7]. For an in depth study of existing ICN approaches see [8]-[10].

Despite the variety of proposed ICN approaches, they share one common characteristic. Each information object (e.g. a movie file) has a unique name bound to it (and there are typically mechanisms to validate and secure this binding). This feature enables the use of in-network caching to speed up the content distribution [11]. Though caching has already been used in the current Internet (e.g. Web caching, peer-to-peer (P2P) caching, content delivery networks (CDN) caching), the absence of a unified naming system for identical information objects makes it inefficient. For example, two copies of the same object held at different servers, will be treated as two different objects by the Web caching system.

Our proposed cache-aware scheme exploits the in-network caching capabilities of ICN to enable a more efficient content distribution. While the locations of the caches are exploited, no signaling or computational overhead is introduced for disseminating the information about the cached objects, as in other existing schemes.

The rest of the paper is organized as follows. In Section II, we present the related work. In Section III, we describe our considered ICN architecture. In Section IV, we present our proposed weight-based cache-aware routing algorithm. We evaluate the performance of our proposed scheme in Section V. Finally, we conclude in Section VI.

II. RELATED WORK

Research on ICN has touched and already provided notable solutions in the areas of network efficiency and traffic engineering [12], [13], network coding [14], [15], satellite communications [16], [17], network security [18], [19] and mobility support [20].

Especially regarding the support for caching, there are a number of important research studies [21]-[26]. Some of the proposed solutions are specific to a particular ICN architecture (e.g. CCN), whereas others are more generic ICN solutions. In [21], the authors propose the MultiCache architecture for ICN. This approach exploits a multicast overlay as a means for

content delivery and anycast for locating the caches in the network. In [22], an analytical study of caching trees in CCN is presented. In particular, the authors model the proportion of the time a given content is cached. In [23], the authors propose the ProbCache, a probabilistic caching scheme for ICN. The aim is to balance the available caching space in the network. In [24], a thorough evaluation of the caching space in CCN is performed. In [25], an extensive study of caching performance in ICN is presented. The tradeoff between the caching space and content access latency is evaluated. A study of optimal cache selection and content placement in caches is shown in [26].

While the aforementioned works have made significant progress in terms of efficient caching policies, cache placement and management strategies, little attention has been given to exploit caching for developing efficient routing mechanisms [27], [28]. In [27], a cache-aware routing scheme is proposed. The authors use a dynamic programming approach for the computation of the delivery paths with minimum transportation cost. The path determination is based on the information object demands and the caching capabilities in the network. In [28], a cooperative cache-aware routing scheme is proposed. The aim is to increase the availability of the cached objects through implicit cooperation of in-network caches.

It is evident that, the deployment of caching mechanisms in ICN opens up the possibility for joint optimization of caching and routing. For instance, routing decisions could be based on the cached contents, the location of caches, etc. This optimization is greatly facilitated by the development of ICN-specific simulation tools in recent years [29], [30].

Compared to the above mentioned cache-aware routing schemes, our scheme, while leveraging in-network caches, introduces no signaling overhead and requires no information on the locations of cached objects. Therefore, our scheme is particularly suitable (but not only) for environments with restricted computational and communication resources, such as disruption-tolerant networks (DTNs).

III. NETWORK ARCHITECTURE

A. Basic Network Entities

Consider the simple network of Fig. 1. We distinguish three types of network nodes, namely, *publishers* (Pub), *subscribers* (Sub), and *content routers* (CR). Publishers possess some *content objects* (or simply *objects*) that they wish to share with the subscribers.

Each content object in the network has a unique identifier, the *object ID*. If the object is large (e.g. a video file), it can be divided into a number of chunks. For example a 100 MB video file may consist of 200 chunks of 500 KB each. Each chunk is distinguished by a unique (within the object) *chunk ID*. Therefore, the pair (object ID, chunk ID) uniquely identifies a chunk in the network.

Publishers advertise the availability of the content objects by issuing the *publication messages* in the network (green arrows in Fig. 1). A typical way of disseminating the publication messages is *flooding*. A content router upon receiving a publication message, records state (content ID and incoming interface), in order to serve any future requests for this object.
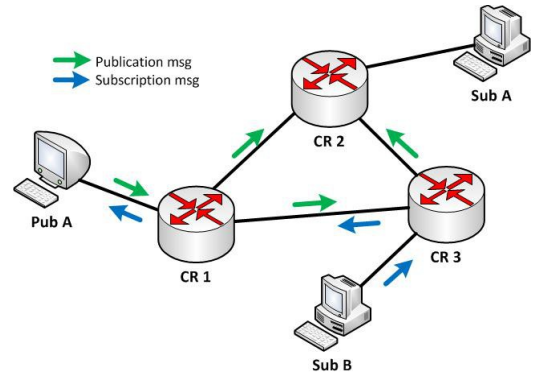

Fig. 1. Issuing publication and subscription messages.

Subscribers issue the *subscription messages* in the network to indicate their interest in receiving a particular object. In the example of Fig. 1, the Sub B sends the subscription message (blue arrow) to its attached router CR 3. The CR 3 having previously received a publication message from Pub A through CR 1, will forward the subscription message to CR 1. Next the CR1 will forward the subscription message to Pub A. Finally, Pub A will send the requested object to Sub B following the reverse path. The object delivery is not shown in Fig. 1.

We distinguish two types of subscriptions, *file-level* and *chunk-level.* In the former, a single subscription message is used to request the whole file, while in the latter, each chunk requires a dedicated subscription massage. There are pros and cons to both approaches. For example, the file-level subscription approach introduces less traffic in the network, but suffers from larger delays in case of lost subscription messages.

We consider two strategies of disseminating the subscription messages, *flooding* and *mediation-based*. According to the mediation-based strategy, the subscriber consults the *mediation system* about the "best" interface to send the subscription message to. For instance, the best interface could be the one that minimizes the number of hops (shortest path) from the subscriber to the publisher or meeting any other optimization objectives.

For example, in the PURSUIT architecture [7] the mediation system consists of the *topology management function* [31], responsible for constructing the delivery tree between publisher(s) and subscriber(s), and the *rendezvous function* [32], responsible for matching publisher(s), subscriber(s), and content objects. The mediation system can either determine the shortest path or a path based on predefined traffic engineering objectives [13]. However, contrary to our approach, in the PURSUIT architecture the mediation system will send the encoded path/tree to the publisher, so that the publisher can send the requested object to the subscriber(s). One of the limitations of this approach is the possibility of distributed denial-of-service (DDoS) attacks, as shown in [33].

In this study, we do not explicitly consider the impact of different traffic models on the network performance. However, a number of existing traffic models (e.g. [34]-[38]), developed for traditional host-centric networks, can be incorporated into our routing scheme. A detailed study of different traffic models in ICN is left for future work.

B. Support for In-Network Caching

Recall that each chunk can be uniquely identified in the network. Therefore, it is possible for the content routers to cache some of the passing through chunks, in order to serve any future requests for these chunks. This concept is known as *in-network caching* and it is an essential feature in most of the ICN architectures [11]. For example the NDN architecture [4] assumes that *every* router in the network has caching capability. In our considered ICN architecture, we relax this condition and permit also non-cache-equipped low-cost routers. Therefore, content routers can be further categorized into *cache-equipped* and *non-cache-equipped*.

Continuing the example of Fig. 1, assume that the CR 1 is equipped with a cache, and it cached the object that was previously delivered from Pub A to Sub B. Next, the Sub A issues a subscription message for the same object (Fig. 2). The CR 1 after receiving this subscription message will discard it (without forwarding towards the Pub A) and will send the cached object back to the Sub A. As a result, the round-trip time has been reduced by 33%.
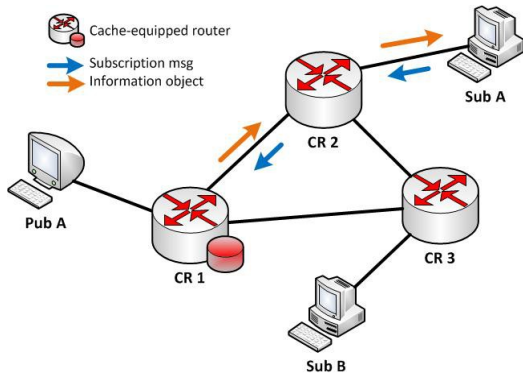

Fig. 2. Getting the content object from a cache.

IV. A CACHE-AWARE ROUTING ALGORITHM

A. Performance Metrics

Based on the network architecture of Section III, we propose a weight-based cache-aware routing algorithm. Our objective is to reduce the *content access delay*, which consists of the

- *content resolution delay*: time elapsed from the moment of issuing the subscription message at the subscriber side, until the moment of receiving the message at the publisher or cache holding the requested content object.

- *content delivery delay*, time elapsed from the moment of sending the requested object from the publisher or cache, until the moment of receiving the object at the subscriber side.

Recall that the requested content objects/chunks follow the reverse path of the subscription messages. Therefore, intuitively, when reducing the resolution delay, we would expect that the delivery delay will be reduced as well. While this is true in most of the cases, there could be situations where the delivery delay is not reduced or even increased, for example, due to highly overloaded links in the reverse path. Therefore we believe that the access delay is a more appropriate performance metric rather than the resolution delay.

B. Content Router Functionalities

Each content router performs the following tasks

- Receive/forward subscription messages

- Receive/forward content objects

- Record state for subscriptions and publications

In addition to the above, a cache-equipped content router performs the tasks of caching/retrieving content chunks.
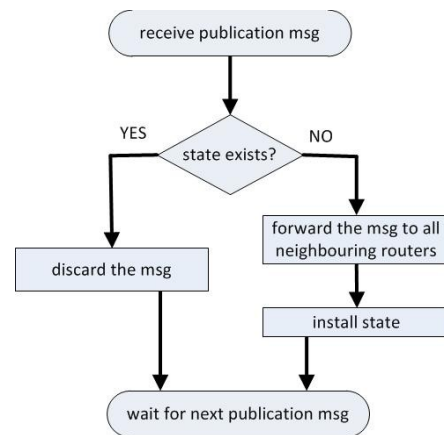

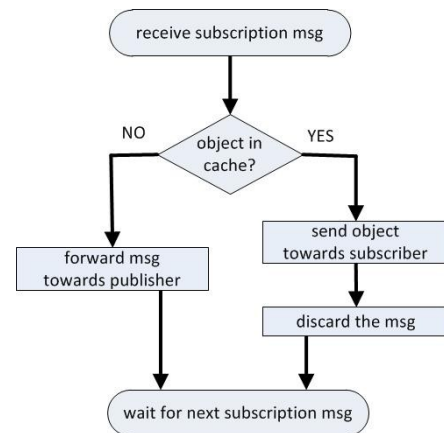Fig. 3. Receiving publication message at the content router.

Fig. 4. Receiving subscription message at the content router.

In Figs. 3 and 4, we show the flowchart for a router when receiving publication and subscription messages. Recall that a content router upon receiving a publication message for an object, records the state for this object. If the state already exists, the publication message is being discarded. For example, in Fig. 1 the CR 2 receives the same publication twice.

C. The Mediation System

As discussed earlier, the mediation system is responsible for identifying the resolution path for every subscription request. In this paper we consider two algorithms for the path determination, the *shortest path algorithm* and the proposed *weight-based cache-aware algorithm*. In the former, all link weights, w_{SP}, are assumed to be the same. In the latter, the weight, w_{WB}, of a link depends on the number of attached cache-equipped routers. In particular

$$w_{WB} = \begin{cases} w_{SP} & \text{if both routers are non-cache-equipped} \\ a \times w_{SP} & \text{if exactly one router is cache-equipped} \\ b \times w_{SP} & \text{if both routers are cache-equipped} \end{cases} \quad (1)$$

$$\alpha + \beta = \chi. \quad (1)$$
$$(1)$$

where a, b are constants such that $1 \le a \le b$.

Note that the weights as defined above do not depend on the cached content objects or any other temporary information at the routers. Therefore, no additional signaling overhead is required to communicate/update these weights to the mediation system. In Section V, we experimentally determine the parameters a and b to achieve the lowest content access delay.

In the following we present a simple example to illustrate the proposed cache-aware routing scheme. Let us consider the network of Fig. 5. In this example, we have two publishers, three subscribers and seven content routers. Two of the routers (CR 5 and CR 6) are equipped with caches. Consider two information objects, *video conference* and *YouTube video*, available at the Pub A and Pub B, respectively. Assume that Sub A has subscribed to the video conference, whereas Sub B to the YouTube video.

Our goal is to select delivery paths that minimize the average content access delay in the network, by exploiting the caches. To this end, in some cases the chosen delivery paths may be longer than shortest, with the aim of passing via caches and reducing the access delay of any future subscriptions. On the other hand, we would not wish to increase the access delay of some delay-sensitive services, like a video conference.

In Fig. 5 we observe that the shortest path to deliver the YouTube video object from Pub B to Sub B is 4 hops (via the CR4-CR7 link). However, if this object is very popular, there is a high probability that it will be requested in the near future by another subscriber (e.g. Sub C). Therefore, we may select a slightly longer, 5-hop path, to deliver this object to Sub B (as shown in Fig. 5). This new path passes through the cache-

equipped CR 6 that may cache the video object and serve later the Sub C. In that case the delivery delay for Sub C will be 1 hop as opposed to a 4-hop delay of getting this object from Pub B.

On the other hand, for the delivery of the video conference object to Sub A, we would prefer the shortest path. This object is delay-sensitive and also it can be considered as non-cacheable, since there is a very low probability of it being requested by other subscribers. For this reason, as shown in Fig. 5, the selected delivery path is the shortest one, rather than the one passing through the CR 5.
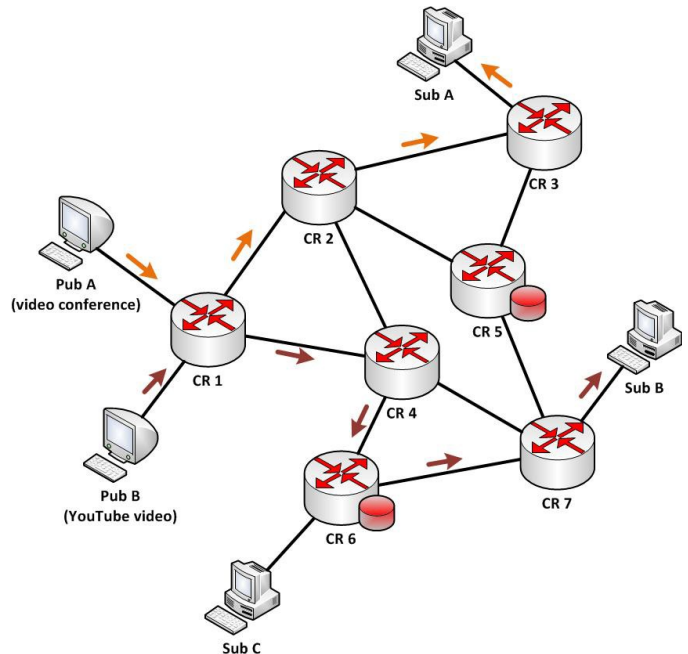

Fig. 5. Illustration of the proposed cache-aware routing scheme.

V. PERFORMANCE EVALUATION

In this section we compare our proposed routing algorithm with the shortest path algorithm. The comparison is performed by means of computer simulations with NS-3 [39].

A. Simulation Setup

We consider the Rocketfuel network topology [40] consisting of 300 content routers and 2,000 links. Each link has a bandwidth of 10 Gbps and a delay of 1 ms. The number of cache-equipped content routers is initially 100. Then we increase this number to 200. The rest of the routers are non-cache-equipped.

In the first scenario, we consider 6 different video files. These files are placed randomly in the network. We consider 8 file requests originated at random places in the network. Four files are requested only once, whereas the two other files are requested twice. In the second scenario we increase the number of files to 12 and the number of requests to 16. Hence, 8 files are requested once and 4 files are requested twice.

The weights for the shortest path routing algorithm are $w_{SP} = 1$. The weights for the proposed *weight-based cache-aware algorithm* are determined according to (1). The parameters a and b are variable with the aim to select them in a way that decreases the content access delay. The presented simulation results are mean values from 6 runs.

B. Simulation Results

In Fig. 6 we present the simulation results for the first scenario (8 content requests). We show the content access delay for each request. We observe that, for the first two requests, the proposed weight-based cache aware (WBCA) routing algorithm has similar performance with the existing shortest path (SP) algorithm. The reason is that at the beginning the caches are empty and the cache-aware routing has almost no effect. However, in the subsequent requests, we observe that the WBCA routing has up to 30% performance gain compared to the SP routing.
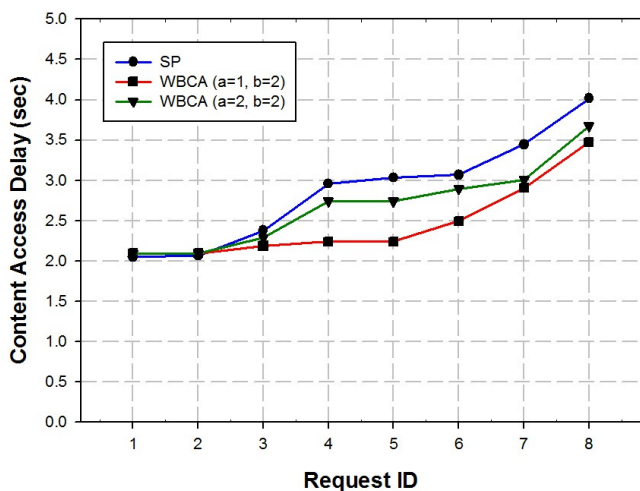

Fig. 6.   Content access delay in the first scenario.

In Figs. 7-8 we present the simulation results for the second scenario (16 content requests). In Fig. 7 the number of cache-equipped routers is 100 (out of 300), whereas in Fig. 8 the number of cache-equipped routers is 200. Similarly to the first scenario, here also we observe no effect of the WBCA routing to the first requests and then the performance improves up to 50% in some cases. We also notice that the performance gain depends on the selection of the parameters a and b. In both scenarios we observe that the choice $a = 1, b = 2$ gives the best performance.

VI.   CONCLUSIONS

In this paper we presented a weight-based cache-aware routing algorithm for information-centric networking. The proposed approach exploits the knowledge of the placement of cache-equipped routers in the network. However, it does not require any additional information regarding the cached content. This feature makes the proposed routing scheme easy to implement and does not impose any signaling overhead for communicating the dynamically changing caching information. Simulation results show an up to 50% reduction of the content access delay compared with the shortest path routing.
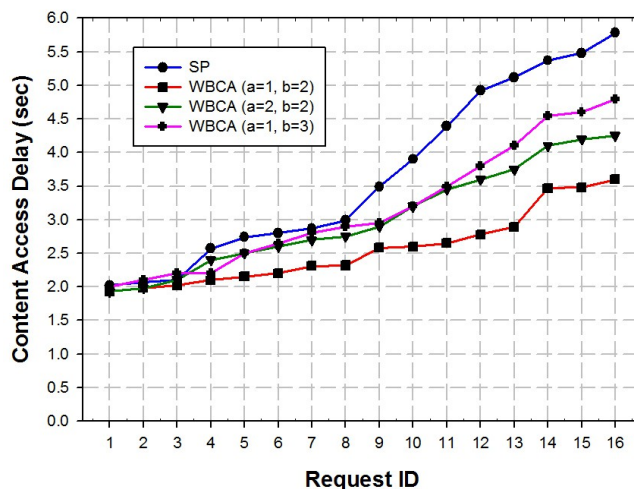

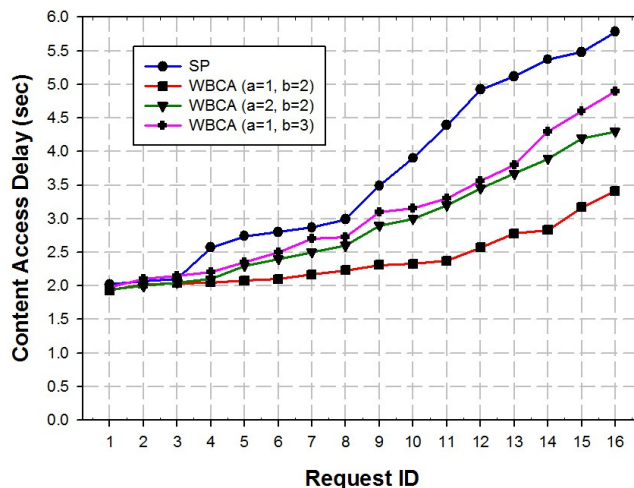Fig. 7.   Content access delay in the second scenario (100 cache-equipped routers).

Fig. 8.   Content access delay in the second scenario (200 cache-equipped routers).

REFERENCES

[1]   A. Feldmann, "Internet clean-slate design: what and why?," ACM SIGCOMM Computer Communication Review, vol. 37, pp. 59-64, 2007.

[2]   M. Handley, "Why the Internet only just works," BT Technology Journal, vol. 24, no. 3, pp. 119-129, July 2006.

[3]   T. Koponen et al., "A data-oriented (and beyond) network architecture," Proc. SIGCOMM '07, Kyoto, Japan, Aug. 2007.

[4] V. Jacobson et al., "Networking named content," Proc. CoNEXT, Rome, Italy, Dec. 2009.

[5] M. Brunner, "Scalable & adaptive Internet solutions (SAIL)," Future Internet Assembly (FIA), Ghent, Belgium, Dec 2010.

[6] G. Pavlou, N. Wang, W. K. Chai, and I. Psaras, "Internet-scale content mediation in information-centric networks," Annals of Telecommunications, Special Issue on Networked Digital Media, vol. 68, no. 3-4, pp. 167-177, April 2013.

[7] N. Fotiou, G. C. Polyzos, P. Nikander, and D. Trossen, "Developing information networking further: from PSIRP to PURSUIT," Proc. The 7th International ICST Conference on Broadband Communications, Networks and Systems (BROADNETS), Athens, Greece, Oct. 2010.

[8] J. Choi, J. Han, E. Cho, T. Kwon, and Y. Choi, "Survey on content oriented networking for efficient content delivery," IEEE Communications, vol. 49, no. 3, pp. 121–127, March 2011.

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

[10] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos, "A survey of information-centric networking research," IEEE Communications Surveys & Tutorials, no. 99, pp. 1–26, 2013.

[11] G. Zhang, Y. Li, and T. Lin, "Caching in information centric networking:a survey," Computer Networks, vol. 57, pp. 3128–3141, 2013.

[12] L. Carrea, A. Vernitski, and M. Reed, "Optimized hash for network path encoding with minimized false positives," Computer Networks, vol. 58, pp. 180-191, 2014.

[13] M. J. Reed, "Traffic engineering for information-centric networks," Proc. IEEE ICC 2012, Ottawa, Canada, June 2012.

[14] J. Llorca, A. M. Turino, K. Guan, and D. Kilper, "Network-coded caching-aided multicast for efficient content delivery," Proc. IEEE ICC 2013, Budapest, Hungary, June 2013.

[15] M. J. Montpetit, C. Westphal, and D. Trossen. "Network coding meets information-centric networking," Proc. 1st ACM Workshop on Emerging Name-Oriented Mobile Networking Design, Hilton Head, USA, June 2012.

[16] V. A. Siris, C. N. Ververidis, G. C. Polyzos, and K. P. Liolis, "Information centric networking (ICN) architectures for integration of satellites into the future internet," Proc. IEEE ESTEL 2012, Rome, Italy, Oct. 2012.

[17] S. D'Oro, L. Galluccio, G. Morabito, and S. Palazzo, "SatCache: a profile-aware caching strategy for information-centric satellite networks," Trans. on Emerging Telecom. Tech., Wiley Online Library, vol. 25, no. 4, pp. 436-444, April 2014.

[18] B. A. Alzahrani, V. G. Vassilakis, and M. J. Reed, "Securing the forwarding plane in information centric networks," Proc. 5th Computer Science and Electronic Engineering Conference (CEEC), Colchester, U.K., Sep. 2013.

[19] B. A. Alzahrani, V. G. Vassilakis, and M. J. Reed, "Key management in information centric networking," International Journal of Computer Networks & Communications (IJCNC), vol. 5, no. 6, pp. 153-166, Nov. 2013.

[20] X. Vasilakos, V. A. Siris, G. C. Polyzos, and M. Pomonis, "Proactive selective neighbor caching for enhancing mobility support in information-centric networks," Proc. ICN'12, Helsinki, Finland, August 2012.

[21] K. Katsaros, G. Xylomenos, and G. C. Polyzos, "MultiCache: an incrementally deployable overlay architecture for information centric networking," Proc. IEEE INFOCOM 2010, San Diego, USA, March 2010.

[22] I. Psaras, R. G. Clegg, R. Landa, W. K. Chai, and G. Pavlou, "Modelling and evaluation of CCN-caching trees," Proc. IFIP Networking 2011, Valencia, Spain, May 2011.

[23] I. Psaras, W. K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," Proc. ICN'12, Helsinki, Finland, August 2012.

[24] D. Rossi and G. Rossini, "On sizing CCN content stores by exploiting topological information," Proc. IEEE INFOCOM Workshop on NOMEN, Orlando, USA, March 2012.

[25] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache less for more in information-centric networks," Proc. IFIP Networking 2012, Prague, Czech Republic, May 2012.

[26] V. Sourlas, P. Flegkas, G. S. Paschos, D. Katsaros, and L. Tassiulas, "Storage planning and replica assignment in content centric publish/subscribe networks," Computer Networks, vol. 55, no. 18, pp. 4021-4032, 2011.

[27] V. Sourlas, P. Flegkas, and L. Tassiulas, "A novel cache aware routing scheme for information-centric networks," Computer Networks, vol. 59, pp. 44-61, 2014.

[28] S. Saha, A. Lukyanenko, and A. Ylä-Jääski, "Cooperative caching through routing control in information-centric networks," Proc. IEEE INFOCOM 2013, Turin, Italy, April 2013.

[29] L. Saino, I. Psaras, and G. Pavlou, "Icarus: a caching simulator for information centric networking," Proc. of SIMUTools 2014, Lisbon, Portugal, March 2014.

[30] N. Vastardis, A. Bontozoglou, K. Yang, and M. Reed. "Simulation tools enabling research on information-centric networks," Proc. 5th International Workshop on the Network of the Future (FutureNet V), in IEEE ICC 2012, Ottawa, Canada, June 2012.

[31] B. Gajic, J. Riihijärvi, and P. Mähönen, "Intra-domain topology manager for publish-subscribe networks," Proc. 18th International Conference on Telecommunications (ICT 2011), Traverse City, USA, July 2011.

[32] N. Fotiou, G. Marias, and G. Polyzos, "Towards a secure rendezvous network for future publish/subscribe architectures," Proc. Future Internet Symposium (FIS 2010), Berlin, Germany, Sep. 2010.

[33] B. A. Alzahrani, V. G. Vassilakis, and M. J. Reed, "Mitigating brute-force attacks on Bloom-filter based forwarding," Proc. Conference on Future Internet Communications (CFIC 2013), Coimbra, Protugal, May 2013.

[34] I. D. Moscholios, J. S. Vardakas, M. D. Logothetis, and A. C. Boucouvalas, "QoS guarantee in a batched poisson multirate loss model supporting elastic and adaptive traffic," Proc. IEEE ICC 2012, Ottawa, Canada, June 2012.

[35] M. Glabowski, M. Sobieraj, and M. Stasiak, "A full-availability group model with multi-service sources and threshold mechanisms," Proc. 8th International Symposium on Communication Systems, Networks & Digital Signal Processing (IEEE CSNDSP 2012), Poznan, Poland, July 2012.

[36] M. Glabowski and M. Stasiak, "Internal blocking probability calculation in switching networks with additional inter-stage links and engset traffic," Proc. 8th International Symposium on Communication Systems, Networks & Digital Signal Processing (IEEE CSNDSP 2012), Poznan, Poland, July 2012.

[37] I. D. Moscholios, V. G. Vassilakis, M. D Logothetis, and M. N. Koukias, "QoS equalization in a multirate loss model of elastic and adaptive traffic with retrials," Proc. 5th International Conference on Emerging Network Intelligence (EMERGING 2013), Porto, Portugal, Oct. 2013.

[38] M. Sobieraj, P. Zwierzykowski, M. Stasiak, and J. Weissenberg, "Single hysteresis model for limited-availability group with BPP traffic," Journal of Telecommunications and Information Technology (JTIT), vol.3 pp. 89-96, Sept. 2013.

[39] The NS-3 network simulator. http://www.nsnam.org/ [May 2014].

[40] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with Rocketfuel," IEEE/ACM Trans. Netw., vol. 12, no. 1, pp. 2–16, Feb. 2004.