Journal of Network and Computer Applications 56 (2015) 48–59

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Survey on caching approaches in Information Centric Networking

Ibrahim Abdullahi [a,b,*], Suki Arif [a], Suhaidi Hassan [a]

[a] InterNetWorks Research Laboratory, School of Computing, Universiti Utara Malaysia, 06010 UUM Sintok, Malaysia
[b] Department of Mathematics/Computer Science, Ibrahim Badamasi Babangida University, Lapai, Niger State, Nigeria

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form
18 April 2015
Accepted 2 June 2015
Available online 10 July 2015

Keywords:
Information Centric-Networks
Cache strategies
In-network caching
Off-path caching
Network communication

ABSTRACT

Information Centric Network (ICN) is increasingly becoming the alternative paradigm to the traditional Internet through improving information (content) dissemination on the Internet with names. The need to reduce redundancy and frequent access to a host (provider of information) has raised an alternative of a man-in-middle concept of ICN. This has necessitated the introduction of some ICN popular architectures (such as Named Data Network (NDN), Content Centric Network (CCN), to name a few) to manage the salient advantages incorporated in ICN. Despite all efforts and issues in naming, security, routing and mobility, power consumption; caching has become the leading variable to fully actualize the future Internet dream by carefully solving the problems in frequency and recency (in objects). Determining what part of the content is to be cached? When is the most appropriate time for caching? How would the object be cached (placed and replaced) and also what path would the object be cached? Thus, this paper span through some selected ICN architectures and projects to investigate and suggest forms of caching in minimizing the total bandwidth consumption, enhanced Delivery of Service (DoS), reduced upwards and downward streaming. In conclusion, pointing out some of the future probable ways to improve caching in ICN. This survey also highlighted the top sensitive issues that influence the active deployment of caches in ICN through recency, frequency, content size, cost of retrieval and coordination, update in caches and replacements. Several cache characteristics were further presented in ways that would improve cache techniques, deployments as research suggestions for content placement, replacement and quick scan on nodes on and off-path of the network.

1. Introduction

Information Centric Network (ICN) aims at achieving the possibilities of bringing new dimension and improved information dissemination platform on the Internet. Its major targeted distinction between the current Internet will be its ability to use names and not the host content addressing like the traditional Internet Protocol (IP) Internet. Several researches and studies conducted previously, have proven that the aspiration will soon be a success (Jacobson et al., 2009; Xylomenos et al., 2013). The advantages of this new paradigm cannot be over emphasized as bandwidth consumption will be reduced, through neighbor node information/interest distribution. In the paper Jacobson et al. (2009, 2012), it was argued that the advantages of dissociating host and Internet Protocol (IP) will also improve fast delivery of services, better scalability among others. However, achieving these advantages like any other concepts has its pros and cons. Studies conducted in the area have proven that users have grown not to be so concerned about where to get information, but rather what to get at the most demanding time (Xylomenos et al., 2013; Ahlgren et al., 2012; Bari et al., 2012). Answering this vital issue of interest (needed data) spans through challenges such as the availability of what is demanded for, the security of what has been found and the authenticity of trusting the location the resulted interest is coming from. ICN has become an attractive subject as researchers identified many issues in designing architectures, frameworks and algorithms to drive the quest to the dream land of content centric networking. A strong motivation behind the paradigm can be based on the report by Gantz et al. (2007) and Cisco (2014) on the rapid growth of contents on the Internet of about 500 exabytes in 2008. IP traffic will be increasing exponentially, thus causing aggregation from 2009 to 2016 which could be a big problem as the years pass on. It further opens the need to find an alternative to the current information dissemination platform that the Internet community benefits.

However, the studies described in Xylomenos et al. (2013) and Ahlgren et al. (2012) identified other issues that would need quick attention to have an acceptable framework which include confidentiality, data integrity, accountability (owner/publisher authentication

* Corresponding author at: InterNetWorks Research Laboratory, School of Computing, Universiti Utara Malaysia, 06010 UUM Sintok, Malaysia. Tel.: +60 169290804.
E-mail address: ibrahim@internetworks.my (I. Abdullahi).

http://dx.doi.org/10.1016/j.jnca.2015.06.011
1084-8045/

and identification), availability and control access (read, write, execute) and storage (Jacobson et al., 2009). This paper is posed at clearly reviewing some selected studies in ICN particularly focusing our attention towards caching as an open area of great importance.

The paper further selected some popular ICN architectures, projects and approaches namely Content Centric Network (CCN) (Jacobson et al., 2009), Named Data Networking (NDN) (Jacobson et al., 2009), Publish-Subscribe Internet Routing Paradigm (PSIRP)/PURSUIT) (Dimitrov and Koptchev, 2010; Dominguez et al., 2011), Network of Information (NetInf) (Ahlgren et al., 2008; Dannewitz et al., 2012), Data-Oriented Network Architecture (DONA) (Koponen et al., 2007), as the test bed comparison in relation to how it intends and implements the caching on nodes, in-network and off-network (in-path and off-path). Good surveys and highly analytical papers have explained ICN in wider perspectives. Examples were also described in Jacobson et al. (2009), Xylomenos et al. (2013), Ahlgren et al. (2012) and Tyson et al. (2013). Cache is mostly influence to achieving some of its advantages which include good use of the bandwidth by reducing wastage. Thus enhancing the prompt delivery of information (reduced delay) and reducing the overall loads on the main source (host). The caching influencing factors can further be elaborated as:

(a) Frequency – in number terms, how many requests are posted or how frequent is an object requested for?
(b) Recency – the time an object or content was referred to or demanded for
(c) Size – the size of a content
(d) Cost of retrieval – the cost incurred to retrieve the content or object
(e) Time of update – a modification in the cache
(f) Replacement – the best time a content becomes less relevant

The structure of this paper is as follows: Section 2 carefully looked at the issues in ICN related to caching, mentioning the challenges and efforts by researchers in the areas of naming, name resolution, mobility and security. Section 3 presented some selected ICN architectures describing their approaches and relating the framework to each cache practices. Section 4 on the other hand presented the generality of caching mentioning cache frameworks and architectural builds in ICN. Section 5 presents the open research issues in ICN caching looking at the in-path and off-path caching. Section 6 concludes the paper.

2. Issues of caching in ICN

This highlights the major issues in ICN design and architectures. The issues explain the principles in naming, name resolution, mobility and security. The four listed issues would be better achieved when an optimal cache description and form is in place. However, the caching in ICN now becomes so relevant and important as to aid the future implementation of the future Internet. ICN concepts generally has predicted the step-by-step manifestation of the current Internet by its wide advantage of enabling the closest nodes specifying and granting request that are similar when interests are sent. The earlier mentioned issues as explained in (Jacobson et al., 2009) can be depicted and illustrated in Fig. 1.

Note that the advantages through caching are that any time a request is posed by client 2 (see Fig. 1), for a similar interest, the closer nodes that has previously initiated such request offer the data requested. This practice is only possible when a chunk of the data is cached in the content store (CS) of the former thus enhanced delivery of service alongside lower upstream and downstream is also incurred (Jacobson et al., 2009). From Fig. 1, client 1 initiates the request channeled through its closest router node B. The routers in some ICN approaches have incorporated the content store (CS), forwarding information base (FIB) and the pending interest table (PIT) as seen on router A. Router B caches the data obtained from Router A and thus passes it onto Router C when it request similar data. To actualize the positive contribution of ICN deployment, one would notice that all information (content) needed (interest) are referred by name, which only can be served after caching. Name resolution therefore becomes paramount because when an interest is sent, there has to be a resolve operation from the content store as in the case of the figure above when client 2 sent out its interest. Mobility and security that will be discussed in the next subsection requires an effective caching mechanism to actualize the goals of ICN (Jacobson et al., 2009, 2012).

(a) Naming
(b) Name resolution
(c) Mobility
(d) Security.

2.1. Naming

Naming in ICN has given the ideas of trying to answer the questions related to the huge amount of information on the traditional Internet. This is due to the fact that in ICN, the clients and nodes are less concerned about the host location address but rather the request of their interest. ICN as a thread process of the current Internet and will have to borrow some ideas of information retrieval and naming like IP. Naming now becomes a vital point of the underlying design of ICN. Hierarchical naming and Distributed Hash Tables (DHTs) are suggested forms of naming in Jacobson et al. (2009); Tyson et al. (2013); Zhang et al. (2011); Kubiatowicz et al. (2000). Naming is thus seen as the way the requester (client 1 and 2) sends an interest. The information is then received by the closest node/router A or station (as depicted in Fig. 2) which, if it has in its Content Store (CS), it forwards through the functions defined in the Forwarding Information Base (FIB) otherwise the interest is placed in a special table known as the Pending Interest Table (PIT) as presented in Jacobson et al. (2009). The motivation here is that the content is cached alongside the interest for further request of similar data. For the purpose of this survey, the paper shall discuss the caching in the remaining sections. An advantage also incurred in ICN is that information are not automatically deleted like in the IP forwarding, thus decreasing the upward streaming of bandwidth and making information fast available for users that subsequently send interests of the same information.

2.2. Name resolution

Name Resolution action normally occurs at a router-like station for SAIL (Xylomenos et al., 2013), 4WARD (Ahlgren et al., 2012), NetInf (Ahlgren et al., 2008,Dannewitz2012), DONA (Koponen et al., 2007), and the rendezvous as in the case in PSIRP/PURSUIT (Xylomenos et al., 2013). The rendezvous in PSIRP architecture work in a fashion of a negotiator and a settler by receiving the interest from the subscriber. The rendezvous finds a match through an opposite provider (Publisher) of the said interest and data. A good number of algorithms have been proposed and discussed in Jacobson et al. (2009) and Xylomenos et al. (2013), which are yielding results in different ICN project and research implemented tests (Ahlgren et al., 2012; Karila and Team, 2008). It will be good to mention that unlike IP, the request and resolution could be settled by a node that has the information that belongs to a close name resolution station. This flexible advantage of retrieving information from multiple span Name Resolution System