3-2017

ISBORD: Internet Searching Based on Resource Description

Zhiwei Yan
China Internet Network Information Center, China

Sherali Zeadally
University of Kentucky, szeadally@uky.edu

Guanggang Geng
China Internet Network Information Center, China

The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 3 (2017) 48–51

www.elsevier.com/locate/icte

ISBORD: Internet Searching based on Resource Description☆

Zhiwei Yan[a], Sherali Zeadally[b], Guanggang Geng[a,*]

[a] China Internet Network Information Center, NANEL, Beijing, 100190, PR China
[b] College of Communication and Information, University of Kentucky, Lexington, KY, 40506, USA

Received 19 September 2016; received in revised form 8 December 2016; accepted 12 December 2016
Available online 23 December 2016

Abstract

Based on the Information-Centric Networking (ICN) concept and the mature model of the current TCP/IP-based Internet, we propose content searching based on universal and scalable resource description, namely ISBORD (Internet Searching based on Resource Description). This novel concept aims to improve the efficiency of content searching and simplifies the end-user functionality to support the evolution of the content-centric Internet.

Keywords: Internet searching; Information-Centric Networking; Resource description

1. Introduction

The current Internet architecture cannot effectively and efficiently meet various requirements, including security, mobility, scalability, and quality of service because of several shortcomings in its original design [1]. A key goal of the original Internet design was to efficiently interconnect mainframes and minicomputers, and to provide efficient remote access to them. This end-to-end approach and especially its specific practical implementation, however, have been identified as the root causes of several key limitations of the current Internet architecture. Then various add-ons, such as Network Address Translation (NAT), Mobile IP (MIP), Content Distribution Network (CDN), Peer-to-Peer (P2P) overlay, etc., all violate, in various ways, several aspects of the original Internet architecture in order to provide features that were not part of its original requirements.

Moreover, it has been observed recently that information is at the heart of almost all communications and particularly Internet uses [2]. This information-centric usage of the Internet raises various architectural challenges, many of which are not effectively handled by the current Internet architecture, even with various add-ons mentioned above. One direct and unified approach to address this problem is to replace the "where is the information?" with "what is the information?". Then the Information-Centric Networking (ICN) concept was proposed to address the question of "what is the information?" [3].

In the past few years, various ICN projects [4–7] have proposed ICN architectures that leverage in-network storage for caching, multiparty communication through replication, and interaction models that decouple senders and receivers. The common goal of these proposals is to achieve efficient and reliable distribution of content by providing a general platform for communication services that are today only available in dedicated systems such as P2P overlay and proprietary CDN. The ICN approach is also expected to be able to better cope with disconnections, disruptions, and flash crowd effects while supporting various types of communication services [8,9].

A naming system is also used in ICN to efficiently manage all the contents as in the TCP/IP-based Internet. However, in this naming system, it is difficult for the user to identify the name of the content directly and to invoke content addressing, although name-based routing is efficient and secure.

In contrast, the Domain Name System (DNS) which was created for mapping names to addresses of hosts and networks, and keeps evolving with the explosive extension of the global Internet. Today, there are more than 300 million second level

* Corresponding author.
E-mail address: gengguanggang@cnnic.cn (G. Geng).

Peer review under responsibility of The Korean Institute of Communications Information Sciences.

☆ This paper has been handled by Prof. Dong-Soo Han.

http://dx.doi.org/10.1016/j.icte.2016.12.001

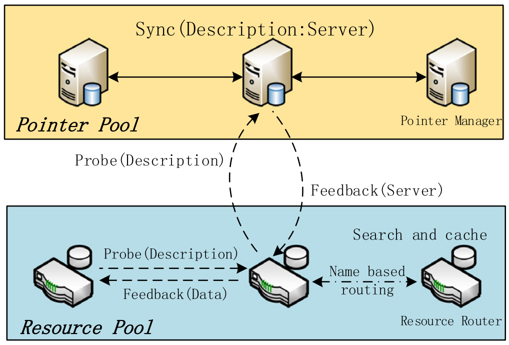

Fig. 1. ISBORD communication model.

domain names covering more than 1000 top level domains. DNS offers more functionalities than the original "host table" because there could be multiple records associated with a given fully-qualified name. This flexibility and scalability demonstrate how the DNS can act as the basic and core naming system in the current Internet.

But it is also impossible for current Internet users to remember so many DNS names to retrieve the related information. Fortunately, search engine is more user-friendly to map the DNS name with the key words which can be input by Internet users as the information profile, without being aware of the DNS names that are actually present. It is conceivable that search engine use becomes so ubiquitous that ready access to DNS-based navigation mechanisms will disappear entirely from user interfaces. But as an application-layer solution, search engine is another add-on, which poses additional security, scalability and efficiency challenges to the current Internet.

Based on the above considerations, in this work we propose a novel Internet architecture, Internet Searching based on Resource Description (ISBORD), to support the direct fuzzy search of content based on the description, which is quickly becoming an emerging Internet trend.

2. ISBORD

For scalability and efficiency, ISBORD consists of two layers: the management layer and the routing layer. To differentiate ISBORD from the current Internet architecture design, the communication model consists of two layers namely, the *Pointer Pool* and the *Resource Pool* as shown in Fig. 1.

Pointer pool: This layer should be involved with the management of the registration information of Internet resources. When a publisher wants to push some content to the Internet, the related information such as the <Description: Server> pair will be registered with a Pointer Manager. "Description" is the keyword to identify this resource, and "Server" is the server name of the content provider. The Pointer Managers will synchronize this information as the CDN nodes do for example.

Resource pool: The Resource Routers in this layer are in charge of content addressing and caching based on the request from the users. Normally, two types of routing should be supported: description-based routing and name-based routing. In the former case, the router checks its local cache first and responds back with the matched data as today's search engine does. If the request cannot be satisfied within the requested hops, the Pointer Manager requests the content directly from the server. In the latter case, the content name is used to fetch the content directly from the specified server as the router in Named Data Networking (NDN) [5] does. The difference between ISBORD and NDN is that description-based routing is used in ISBORD to improve the user experience and avoid the name resolution overlay in NDN.

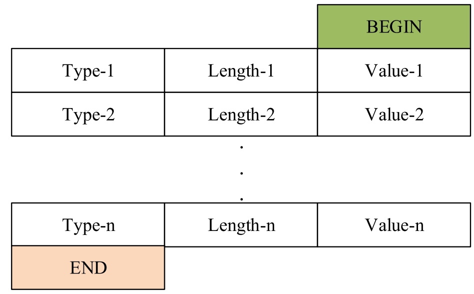

Fig. 2. Structure of the resource description.

2.1. Signaling messages

To support this communication model, the following signaling messages should be used:

Probe: This message is used to request the content based on the description contained inside it. Besides, a Time-To-Live (TTL) should be set in the Probe message to specify the hop limit of this message. When the Resource Router receives the Probe message, it checks its local cache with the description information. If there is no positive match, the Resource Router propagates the Probe message with the TTL decreased by one. When the value of TTL is zero, the Resource Router should send the Probe message to its default Pointer Manager.

Feedback: This message is the response to the Probe message. For the Resource Router, the related Probe message carries the content that matches with the description. For the Pointer Manager, the name of the content server is contained in order to enable the Resource Router to fetch the requested content directly from the server.

Sync: This message is multicast or broadcast based and it is used to synchronize the content description and the related server's name among the Pointer Managers.

2.2. Resource description

The way we describe the resource in a scalable and universal manner is crucial for the large-scale deployment of the ISBORD. The structure of the resource description is shown in Fig. 2.

Starting with BEGIN and ending with END, part of the description consists of one or more keywords. The keyword, being the basic element, is organized by the Type-Length-Value (TLV) field. For example, if a user wants to fetch a picture of

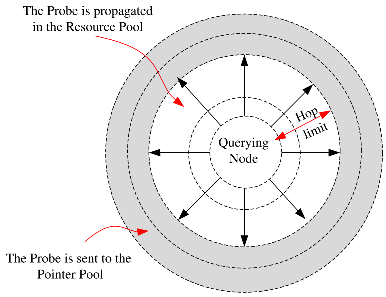

Fig. 3. Network model.

Bruce Lee, who is the movie star, the request should contain at least three keywords:

$T1 =$ name,	$L1 = 9$,	$V1 =$ Bruce Lee
$T2 =$ profession,	$L2 = 10$,	$V2 =$ Movie Star
$T3 =$ format,	$L3 = 7$,	$V3 =$ Picture.

3. Performance evaluation

Fig. 3 shows the network model we used to evaluate the performance of ISBORD and compare it with content searching in the TCP/IP-based Internet by using content searching latency as the performance metric. In this model, the querying node is at the center of the network and the hop limit to propagate the Probe message is H. The total number of routers in the domain we consider is (within the hop limit) N. The topology of the routers is a binary tree, which is a simplified network model and is widely used in NDN cache management analysis [10]. The number of routers in the first hop ring is N_1 and the number of routers in the second hop ring is $2^1 N_1$. With the same logic, the number of routers within the Hth hop ring is:

$$N = N_1 + 2^1 N_1 + 2^2 N_1 + \cdots + 2^{H-1} N_1 = (2^H - 1)N_1.$$

We also assume that the popularity of the content is q which represents the cache probability of the content within N routers. Then,

$$q = \frac{X}{N}$$

where X is the number of routers with the cached content and these routers are distributed evenly within the N routers. This means that when all the routers cache the content, $q = 1$; in contrast, when none of the routers cache the content, $q = 0$.

If the one-hop Probe/Feedback round-trip latency is α and the latency from any router to the resource source or Pointer Manager is β, the average latency for the content searching under ISBORD is:

$$\begin{aligned} L1 &= q \times \alpha + (1-q)q \times 2\alpha + \cdots + (1-q)^{H-1} q \times H\alpha \\ &\quad + (1-q)^H \times 2\beta \\ &= \alpha \times \frac{1-(1-q)^{H-1} - H \times (1-q)^H}{1-q} \\ &\quad + (1-q)^H \times 2\beta \\ &= \alpha \times \frac{1-\left(1-\frac{X}{N}\right)^{H-1} - H \times \left(1-\frac{X}{N}\right)^H}{1-\frac{X}{N}} \\ &\quad + \left(1-\frac{X}{N}\right)^H \times 2\beta \\ &= \alpha \times \frac{1-\left(1-\frac{X}{(2^H-1)N_1}\right)^{H-1} - H \times \left(1-\frac{X}{(2^H-1)N_1}\right)^H}{1-\frac{X}{(2^H-1)N_1}} \\ &\quad + \left(1-\frac{X}{(2^H-1)N_1}\right)^H \times 2\beta. \end{aligned}$$

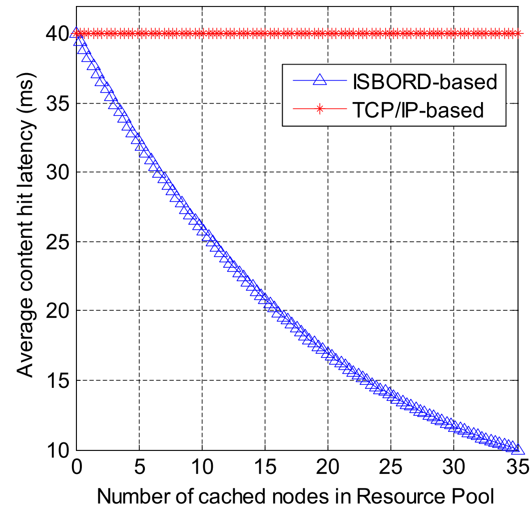

Fig. 4. Average content searching latency.

In the basic TCP/IP Internet, the average latency for the content searching is $L2 = 2\beta$. Then the comparison results between TCP/IP based process and ISBORD based process are shown in Fig. 4, in which we set N_1, H, α and β to be 5, 3, 10 ms and 20 ms respectively [8].

As shown in Fig. 4, as the number of routers which cache the content increases, ISBORD can significantly decrease the content searching latency because the server–client model is changed to a flat model which focuses only on the content name and its description. In contrast, the TCP/IP model cannot differentiate the popularity of the content and always fetches the data from the source which is likely to be remote. However, in order to support the efficient content searching in ISBORD, the content caching approach used in NDN should be supported by the routers in the Resource Pool. As a result, ISBORD will also introduce some additional costs compared with the traditional TCP/IP network.

4. Conclusion

We propose and briefly evaluate the performance of a novel Internet model, namely ISBORD. ISBORD combines the flexible content description used by the search engine in the current Internet together with the efficient NDN routing scheme.

Acknowledgments

We thank the anonymous reviewers for their valuable comments which helped us improve the quality of this paper.

This work was supported by the National Natural Science Foundation of China under Grant No. 61303242.

References

[1] A. Feldmann, Internet clean-slate design: what and why? SIGCOMM Comput. Commun. Rev. 37 (3) (2007) 59–64.

[2] D. Trossen, G. Parisis, Designing and realizing an information-centric internet, IEEE Commun. Mag. 50 (7) (2012) 60–67.

[3] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, B. Ohlman, A survey of information-centric networking (draft), in: Proc. Dagstuhl Seminar on Information-Centric Networking, Dagstuhl, Germany, February 2011.

[4] T. Koponen, et al. A data-oriented (and beyond) network architecture, in: Proc. of ACM SIGCOMM, Kyoto, Japan, Aug. 27–31, 2007.

[5] V. Jacobson, et al. Networking named content, in: Proc. of ACM CoNEXT, Rome, December 1–4, 2009.

[6] M. Ain, et al. D2.3 — Architecture definition, component descriptions, and requirements, deliverable, PSIRP 7th FP EU-funded project, February 2009.

[7] B. Ahlgren, et al. Second NetInf architecture description, 4WARD EU FP7 Project, 2010.

[8] Z.W. Yan, J.-H. Lee, X.D. Lee, Y.-J. Park, Utilizing the domain name system for scalable and secure content-centric networking, Springer WPC J. 84 (3) (2015) 2135–2149.

[9] Z. Yan, S. Zeadally, Y. Park, A Novel Vehicular Information Network Architecture based on Named Data Networking (NDN), IEEE Internet Things J. 1 (6) (2015) 525–532.

[10] M. Li, Recent advances in named data caching and routing, Project Reports, December 2013.