Computer Networks xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computer Networks

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

Content discovery for information-centric networking

Munyoung Lee [a], Junghwan Song [a], Kideok Cho [a], Sangheon Pack [b], Ted "Taekyoung" Kwon [a,*], Jussi Kangasharju [c], Yanghee Choi [a]

[a] School of Computer Science and Engineering, Seoul National University, Seoul 151-742, Republic of Korea
[b] School of Electrical Engineering, Korea University, Seoul 136-701, Republic of Korea
[c] Department of Computer Science, PL 68 (Gustaf Hallstromin katu 2b) 00014, University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 24 October 2013
Received in revised form 19 August 2014
Accepted 1 October 2014
Available online xxxx

Keywords:
Information-centric networking
Content discovery
Bloom filter

ABSTRACT

The information-centric networking (ICN) adopts a content name as a network identifier and utilizes in-network storages to cache the contents. With the name-based routing and content caching, ICN can provide substantial benefits such as faster content retrieval and network traffic reduction by exploiting a nearby (cached) copy of content and reducing duplicated transmissions for the same content request. Prior researches on ICN usually rely on an opportunistic cache-hit (happen-to-meet) to utilize the in-network storages. In the happen-to-meet fashion, only the content cached on the path towards the content source can be utilized, which limits the network-wide usage of the in-network storages. To exploit cached contents better, we propose a content discovery scheme, dubbed SCAN, which can exploit nearby content copies for the efficient delivery. SCAN exchanges the cached content information among the neighbor routers using Bloom filters for the content discovery. With extensive simulations, SCAN shows better performance than a happen-to-meet cache-hit scheme in terms of average hop counts, traffic volume, and load balancing among links.

1. Introduction

A recent measurement study shows that most Internet traffic is governed by content retrieval applications like P2P file sharing or video on demand (VOD) services [1]. Users only care about the content in these applications instead of the source of the content. Currently, the Internet architecture is based on a host-to-host communication paradigm so that inefficient content deliveries often occur. For instance, a user may retrieve content from a distant server even when there is a copy nearby. A significant amount of traffic is also generated when multiple users repeatedly download the same popular content.

To address such issues, a novel networking paradigm called an information-centric networking (ICN) has been proposed [2–5]. In ICN, the networking architecture is designed to be consistent with the current Internet usage pattern (focusing on the contents). Thus, ICN utilizes a content name as an identifier at the networking level to address and retrieve the content. Also, contents are cached in in-network storages[1] attached to routers in order to deal with later repeated requests. The content copies stored in the in-network storages are retrieved by a name-based routing in order to improve content delivery. With close and

* Corresponding author. Tel.: +82 2 880 9105.
E-mail addresses: mylee@mmlab.snu.ac.kr (M. Lee), jhsong@mmlab.snu.ac.kr (J. Song), kdcho@mmlab.snu.ac.kr (K. Cho), shpack@korea.ac.kr (S. Pack), tkkwon@snu.ac.kr (T."Taekyoung" Kwon), Jussi.Kangasharju@helsinki.fi (J. Kangasharju), yhchoi@snu.ac.kr (Y. Choi).

[1] We use the terms cache and in-network storage interchangeably in this paper.

http://dx.doi.org/10.1016/j.comnet.2014.10.006
1389-1286/

multiple cached copies, user experiences can be improved while network resources are efficiently utilized [4–6].

To exploit the cached contents, prior researches on ICN [3,4,7–10] usually rely on an opportunistic cache-hit (happen-to-meet), which can utilize contents cached only in routers on the path towards the publisher. However, when considering the portion of in-network storages on the path and the spatial diversity of content requests, the opportunistic cache-hit limits the potential advantages of network-wide in-network storages. To fully exploit the in-network storages, a content discovery scheme is needed that can utilize the network-wide cached contents. When devising the content discovery scheme, the following issues should be addressed:

- *How can the information of cached contents be managed?* To discover and retrieve content from a source (i.e., a publisher or in-network storages) at the networking level, the content name information should be maintained in a routing table. However, maintaining the information of all contents in the network is not feasible in terms of scalability [11,12], since the number of contents in the network is enormous. In particular, if non-aggregatable flat names are employed [2], they highly increase the size of the routing table. Even with the hierarchical names [3,4], the routing scalability problem still remains when considering not only original contents by the publishers, but also multiple cached copies of each original content in the network. Therefore, the cached content information should be efficiently managed to not significantly increase the routing table size.
- *How can the information exchange overhead be reduced?* Since cached contents are volatile due to the capacity limitation of the in-network storage, the cached content information is periodically exchanged among the content routers (C-routers) for cache consistency. Considering the large number of cached contents and their dispersed names, the traffic overhead for information exchange cannot be negligible. Thus, a compression scheme to reduce the information is needed in order to lower the exchange overhead among the C-routers.

We propose a content discovery scheme for ICN, dubbed SCAN. To address the first issue, SCAN is designed to manage the permanent original contents and temporary cached contents differently: the information for 'all' the original contents is kept in the routing table,[2] while the information for only the 'subset' of the cached contents is maintained in another table. Using the named data networking (NDN)/content centric networking (CCN) as a reference model for ICN, a cache information base (CIB) for cached contents can be populated in addition to the forwarding information base (FIB) for the original contents. In other words, a routing table (i.e., FIB) indicates the next-hop information for the publishers while a separate CIB is maintained for each interface at a C-router to keep the routing information to the network-wide cached contents. Additionally, the second issue is solved when SCAN stores the content information in the CIB and compresses CIBs using the Bloom filter (BF) to keep the table size small and reduce the overhead of information exchange. Note that we assume C-routers cache the contents in their attached storage module as with [2–5,13].

By referencing CIBs, SCAN can potentially find multiple close copies of the requested content (we call this operation *"scanning"*). However, there might be a reachability problem if the content discovery process relies only on the CIBs to find the content since the cached contents in the in-network storage can be replaced. To guarantee the reachability with a fallback mechanism in SCAN, a content request is forwarded to the publisher by looking up the FIBs, which is in parallel to the *"scanning"* operation that follows the CIBs. In summary, SCAN improves efficiency in content discovery by the scanning operation while guaranteeing the reachability with the name-based routing of NDN.

The main contributions of this paper can be summarized as follows:

- We propose a content discovery scheme 'SCAN' to discover off-path content replicas. To handle a large number of cached contents with small overhead for information exchange, we adopt Bloom filters (BFs) as CIBs and propose a BF-based CIB exchange mechanism using information decay.
- In ICN, a content request (i.e., an interest packet) cannot be sent to a specific location since the location information is not used in the routing. Thus, we devise a content discovery mechanism that can find nearby content sources without requiring the location information.

The remainder of this paper is organized as follows: In Section 2, we survey related work and compare SCAN to other proposals. Sections 3 and 4 describe the operation of SCAN and the CIB maintenance issues, respectively. Section 5 presents the evaluation results and Section 6 concludes this paper.

2. Related work

To solve the inefficiency of content retrieval in the current Internet architecture, there have been several efforts to redesign it in a clean-slate manner. Named data networking (NDN) [3,4] is a representative architecture based on a content-centric paradigm. NDN adopts a hierarchical name structure to aggregate content names for name-based routing. Since SCAN chooses NDN as a reference ICN model and integrates its content discovery scheme with NDN, we will briefly introduce the operation of NDN. However, note that SCAN can coexist with various ICN models since SCAN leverages the underlying ICN routing model (main routing scheme) and adds the scanning operation as a subsidiary mechanism.

2.1. Named data networking (NDN)

NDN uses an interest packet and a data packet for content request and retrieval. The interest packet contains the

[2] By exploiting hierarchical name structures, the routing table entries can be aggregated.

Please cite this article in press as: M. Lee et al., Content discovery for information-centric networking, Comput. Netw. (2014), http://dx.doi.org/10.1016/j.comnet.2014.10.006