Poster: Enabling Mobile Content-Oriented Networking in the MobilityFirst Future Internet Architecture

Feixiong Zhang, Yi Hu, Yanyong Zhang, Dipankar Raychaudhuri
WINLAB, Rutgers University, North Brunswick, NJ, USA
{feixiong, yihu, yyzhang, ray}@winlab.rutgers.edu

ABSTRACT

The prevalence of mobile devices has attracted research on mobile content delivery techniques. The MobilityFirst(MF) project, discussed in this paper, proposes a clean-slate Internet architecture that enables mobile content-oriented operations at the network level. We describe the design details of the architecture in realizing this and provides a preliminary evaluation on scalability and performance.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design

Keywords

Mobile content; Content-oriented networking; MobilityFirst

1. INTRODUCTION

Given that Internet usage is dominated by content distribution and retrieval, content-oriented networking (CON) [1] is proposed to facilitate such operations from a network architecture perspective. Moreover, mobile data demands is growing rapidly and mobile Internet is gaining increasing prosperity [2]. Such trends motivate consideration of a content-oriented networking with inherent mobility support.

The MobilityFirst project [3] presents a clean slate Internet architecture that supports large-scale, efficient and robust network services with mobility as the norm. It accommodates content-oriented applications and services in mobile scenarios based on the following features:

- *Flat Content Name Space.* A flat, self-certifying global unique ID (GUID) space is used to name content at the network level and is decoupled from network address.

- *GUID to Address Resolution.* A distributed global name resolution service (GNRS) is utilized to maintain and resolve the binding between content GUID and address(es). The GNRS is designed to ensure scalable and fast content discovery.

MobiHoc'14, August 11–14, 2014, Philadelphia, PA, USA.
ACM 978-1-4503-2620-9/14/08.
http://dx.doi.org/10.1145/2632951.2635943.

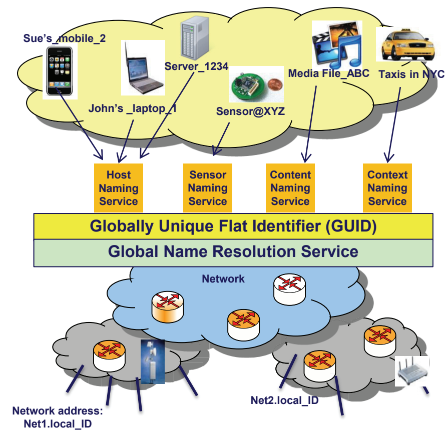

Figure 1: Name-locator separation and mapping services in MF

In MF, due to name and locator separation, users can request content directly by content name, totally oblivious to its current location. Routers query the GNRS dynamically to get updated content location(s) in a timely fashion, and further conduct routing. On top of the base solution, a popularity-based content caching scheme is proposed to further facilitate content distribution.

2. DESIGN

2.1 Content Naming and Addressing

In MobilityFirst, content is considered a first-class network-attached object, a status equal to that of a host. As shown in Figure 1, MF uses flat globally unique identifiers (GUIDs) to name all network-attached objects at the network level . Particularly, content GUID is derived through a cryptographic hash of the associated content, thus enabling self-certifying communication when passed in packet headers.

A GUID is assigned to a network object and maintained by domain specific name certification service (NCS). For example, content publishers such as Netflix and YouTube can maintain their own NCS that contains a directory of their content and corresponding GUID, and the GUID is returned when providing the content's human readable name or keywords to NCS.

GUID is further separated from the object's locator or network address(NA). The dynamic mapping of GUIDs to NAs is maintained by a logically centralized, but physically distributed infrastructure, namely GNRS.

2.2 Content Publishing and Locating

In MF, GNRS provides the basis for content operations. GNRS is structured in two levels: a Local Name Resolution Service(LNRS) at each domain/AS, and the global-level GNRS (gGNRS). LNRS contains the mappings of local objects within the AS which is composed by one or multiple central servers, while gGNRS is designed as an in-network 1-hop DHT with servers distributed all across the Internet. Each server in this distributed service advertises responsibility to a portion of an orthogonal name-space, which is used to designate (using consistent hashing) a host for mappings of a particular GUID. The structuring design of GNRS is aimed to support response latencies lower than 100ms for all these operations in order to enable real-time applications for mobile hosts. Such objective is validated in one manifestation of GNRS scheme [4].

To make a content available over the network, the publisher inserts its name-to-locator mapping to GNRS after assigning a GUID to the content. Specifically, a GNRS *insert message*: ⟨GUID, Addr⟩ is sent to an LNRS server. The LNRS server on receiving the insert message first initiates an update at the local service plane. It then applies a consistent hash function on the GUID to determine the gGNRS server that will host the mapping ⟨GUID, NA⟩. An insert message is then forwarded towards that server.

Resolution of content location is initiated by issuing a GNRS *lookup message*: ⟨GUID, options⟩ to LNRS. The LNRS server first checks with the local service plane whether the content requested is located within the local network. If so, the request need not be forwarded to the global plane and a response is returned from the LNRS server. Otherwise, a consistent hash determines the gGNRS server that hosts the mapping, and the lookup message is forwarded. Address received in the lookup response can then be used to address a content retrieval request.

When a content or its host moves, the host initiates an *update message*: ⟨GUID, Addr, options⟩ to LNRS. If the update is only to the local address, the update is handled at LNRS plane alone. If not, the LNRS forwards an update message to the gGNRS server determined by the hash to update the global mapping.

2.3 GNRS-assisted Popularity Caching

When a client requests a content, a GNRS lookup message is issued for location resolution. Since the message is received by LNRS first before sent to global-level GNRS, the LNRS can build a content request table capturing local content request pattern. The table maintains a recent usage count (RUC) for each content requested in the network. RUC value is managed using an explicit aging method with a periodic aging function:

$$C(x) \leftarrow \alpha * C(x), 0 < \alpha < 1 \qquad (1)$$

At the beginning of each time slot (e.g. 1 hr), each RUC value $C(x)$ is decreased to α times its original value. Within a time slot, each new content request causes its RUC value to be increased by one.

The content request table is used for popularity-based caching, i.e., most popular contents are cached in the AS. Specifically, when a content is retrieved to the requesting AS, the content is cached if space is available in the AS proxy cache; otherwise, cache replacement is performed by comparing RUC values. If the content is cached, the cached copy is updated at LNRS, thus following request of the content will be resolved and retrieved locally from the AS. Such GNRS-assisted caching scheme can dynamically capture popular content and facilitate content retrieval.

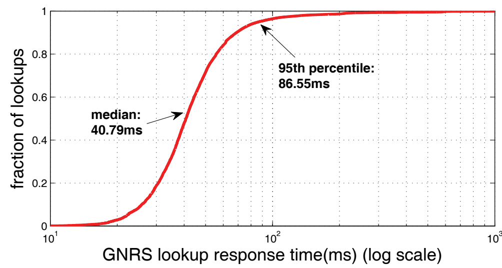

Figure 2: CDF of GNRS lookup response time

3. EVALUATION

We first analyze the feasibility of requiring GNRS to store the name-locator mapping of content in terms of its storage overhead. We assume a total of 1 trillion content [5] published to GNRS, with each GUID of length 20 bytes. We further assume a GUID is associated with 10 NAs (accounting for content replicas) of length 8 bytes each (including extra information along with each NA). With all GUID mappings distributed among the whole Internet (∼25000 ASes), the average storage requirements per AS is roughly $10^{12} * (20 + 10 * 8)/25000 = 4GBytes$, a small value considering current technology.

Next, We build a discrete-event simulator to evaluate the lookup response time of our two-level GNRS. The simulator uses an AS-level topology of current Internet measured in the DIMES project [6]. We evaluate the GNRS lookup response time by having 1 million content distributed across all ASs and generating 10 million lookups according to Zipf popularity distribution. Figure 2 plots the cumulative distribution function (CDF) of GNRS lookup response time. The results show that for an Internet-scale GNRS deployment the location resolution cost is bound to ∼100ms.

4. CONCLUSION AND ON-GOING WORK

In this paper, we have presented our mobility-centric content-oriented networking scheme to meet the challenge of mobile content delivery. Our ongoing work includes designing a mobility prediction and content prefetching framework to enhance seamless delivery, and conducting a thorough evaluation of our mobile CON scheme.

5. REFERENCES

[1] Van Jacobson et al. Networking named content. In *Proc. of ACM CoNEXT*, 2009.

[2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018.

[3] MobilityFirst project. `http://mobilityfirst.winlab.rutgers.edu/`.

[4] Tam Vu et al. Dmap: A shared hosting scheme for dynamic identifier to locator mappings in the global internet. In *IEEE ICDCS'12*, pages 698–707, June 2012.

[5] We knew the web was big. `http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html`.

[6] Y. Shavitt and E. Shir. Dimes - letting the internet measure itself. `http://www.netdimes.ogr/`.