ICNRG
Internet Draft
Intended status: Informational
Expires: October 5, 2014

Hongke Zhang
BJTU
Wei Quan
Jianfeng Guan
Changqiao Xu
BUPT
Fei Song
BJTU

April 5, 2014

Uniform information with a hybrid naming (hn) scheme
draft-zhang-icnrg-hn-00.txt

Abstract

This document defines a hybrid naming scheme for unifying all kinds of information. As many proposals of novel network architectures emerge, such as DONA, ICN, NDN, the location-based routing starts to transfer to the content-based ones. Currently, it is incompatible that many different information naming schemes are adopted in different network proposals, respectively, i.e. flat names in DONA, hierarchical names in NDN. The naming format defined is to identify different routing information uniformly. The format adopts a hybrid structure including hierarchical component, flat component and attribute component, providing great compatibility and advantages.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Copyright Notice

Table of Contents

1. Introduction

1.1. Hierarchical naming

Some emerging network architectures (i.e. Content-Centric Network (CCN)[1]/Named Data Networking (NDN)[2]) have proposed a readable naming mechanism based on the hierarchical structure. This hierarchical name is very similar as identifying a web with a URL, for example "/www.bupt.edu.cn/content/a.avi". In this example, "/" is the separator between adjacent components of the name.

We acknowledge that there are some advantages in this naming scheme. First, it has a good compatibility with current applications or systems based on URL, which can reduce the difficulty of deploying the novel network. Second, it has a good aggregation to reduce the number of routing information, and improve lookup efficiency of routing information. Besides, its lookup mechanism has a good compatibility with the existing classless inter-domain routing (CIDR)[3].

However, the hierarchical name also has some fatal disadvantages. It consists of a series of unlimited components. The number of components is variable, and the length of each component is not restricted. All these features cause the length of names is variable and relatively long [4]. In this way, the routing table and forwarding table may be very huge, which results in the lookup efficiency becomes very low.

In addition, when users search for a resource, they might not remember the long name of the resource. For example, users need the resource a.avi, but they might not know the official name "/www.bupt.edu.cn/content/a.avi" or "/www.bupt.edu.cn/movie/a.avi". Hierarchical naming structure is difficult to support a fuzzy matching based on the attributes of names.

1.2. Flat naming

The flat naming mechanism has been used in other novel network architectures, such as DONA [5] and NetInf [6]. This flat name can be produced by cryptographic hashing of the content itself or its attributes.

Due to the flat name has not any structure restriction, it can be obtained and used more flexibly. Any string with a fix length, no matter whether it is unreadable or readable, can be used as a flat name.

However, the flat name has a low degree of aggregation, which will increase the number of the routing entries and reduce the expandability of routing table. Besides, most of flat names are not readable, which increase the probability of users' forgetting the official names of the desired information. When users want to obtain contents, it needs a mapping between readable names and unreadable names for users by means of an additional mapping system.

1.3. Attribute naming

The naming mechanism based on attributes of content is used in the CBCB [7]. It enumerates the attribute information of a resource, such as the category, format, date, feature, level and so on. This name is non-uniqueness which is different from the former two mechanisms. The related content can be searched and located by means of the key properties of resource.

The advantage of this naming is that it supports searching key words and provides benefits for the fuzzy matching for searching resources. However, there may be many similar properties for a set of certain resources. The uniqueness is hardly guaranteed by a limited number of attributes. Thus, to guarantee the uniqueness, the attributes stored in routing system will be very huge.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

In this document, the characters ">>" preceding an indented line(s) indicates a compliance requirement statement using the key words listed above. This convention aids reviewers in quickly identifying or finding the explicit compliance requirements of this RFC.

3. Novel hybrid naming (hn) format

According to the analysis of above three kinds of naming mechanisms in terms of advantages and disadvantages, a hybrid naming is

suggested to highlight the advantages of them and weaken their disadvantages.

Most importantly, three mainstream different naming schemes are adopted in different novel network architectures, which make the networks be hardly compatible and implemented complexly.

One easy and all-benefit solution is the integrated method for them, taking each of them as a part of the hybrid naming solution. In other words, each of them takes some weight of the novel naming scheme.

We proposed a hybrid naming mechanism (named by "hn"), which organizes three kinds of naming mechanisms in a sequence, and builds a more powerful and universal naming format.

The hybrid naming format should include three components:

- Hierarchical component
- Flat component
- Attribute component

Each part carries different information of name in different formats, which produce an entire name. The hybrid name is started by a symbol "hn://". The order of three parts should be as follows:

1. The first part of a name is very important for the aggregation of routing entries. A hierarchical structure is adopted in the first part. The symbol "/
" is used to split the hierarchical levels in
this part.

2. The second part of a name is very important to identify the content uniquely. A flat structure is used in the second part. A string with a fix length can be used by a hash computing.

3. The third part of a name is used to represent the extensive information of resources. The attribute-based structure is selected in the third part, which is composed of a set of attribute words.

An example of the hybrid name for a movie is shown in Figure 1.

Figure 1 An example of hn for a movie

An example of the hybrid name for a picture is shown in Figure 2.

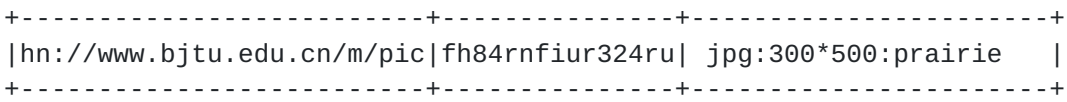

Figure 2 An example of hn for a picture

3.1. Hierarchical component generating

Hierarchical component is the first part of the hn naming format. This part is suggested to be generated following a reference standard to be followed. This standard should define the string set in top level, string set in second level and so on. This reference standard is very useful to promote its aggregation greatly. One available but not complete reference standard for naming hierarchical component is the naming scheme of DNS.

3.2. Flat component generating

Flat component is the second part of hn naming scheme. This part is suggested to identify the information using a string with a limited length. This part must identify the information uniquely by combing with the first part.

Flat component can be generated by cryptographic hash algorithm by the information itself or some characters of the information. This part has a low probability of aggregation, but it highlights and ensures the uniqueness of name.

3.3. Attribute component generating

Attribute component is placed as the third part of hn naming scheme. This part will take charge of the fuzzy matching and some advanced

search, i.e., QoS guarantee. This part will also contribute to conduct some potential advanced application based on the useful attributes. It can be generated by extracting the features of the information, such as the format, issue time, file size, catalog, location, popularity, privacy level and so on.

4. Advantages

4.1. High aggregation

The aggregation of names is very important for the name lookup and storage. According to Google's report, the number of URLs it indexed was 26 million in 1998, which reached to one billion in 2000, and is currently 1 trillion [8]. In July 2011, these URLs could be aggregated to about 280 million domain names, among which 86 million are active.

It is a fact that there is a great aggregation for the first few levels of the hierarchical tree. Therefore, the hierarchical structure is used in the first part of the hn. By this way, the routing entries can be reduced obviously and the aggregation of route can be improved. For example, there are two routing entries "/www.bjtu.edu.cn/m/movie/fhk562nfgjru056:kongfu:avi:1024p:part1 3" and "/www.bjtu.edu.cn/m/picture/fh84rnf213gjrru:jpg:300*500:prairie 3" which have the same forwarding port "3" and prefix "/www.bjtu.edu.cn/m". Therefore, the forwarding port and "/www.bjtu.edu.cn/m" can only be stored in routing table. It not only reduces the entries of routing table, but also reduces the length of each routing entries. An example of aggregation process is shown in Figure 3.

Figure 3 An example of aggregation

4.2. Limited length

The length of name based on hierarchical structure is variable and relatively long because it must be formed by several parts and the number of component is variable. Kelvin [9] has selected 6627999 URL in 78764 different domain names, and the statistics shows that the average length of URL is 76.97 bytes. In the architecture of ICN, the name must be extracted to query in forwarding table or routing table and a long name entry will lead to the query speed becoming low, hance, affects the performance of routing.

The hn naming scheme use a part of flat component in the name to ease this problem. A fix length flat part is embedded behind the hierarchical part. This design not only can restrict the length of names not too long, but also will affect the aggregation not much. For example, if the average length of hierarchical part is controlled within 30 bytes, adopting a flat part with a fix length of 20 bytes, the whole average length will be restricted within 50 bytes. Comparing to 76.97 bytes, the length is shortened by nearly 35%, which will improve the query speed of name greatly using the length-dependent algorithms.

4.3. Suffix holes remission

The suffix hole is a well-known problem for the route of prefix matching. For example, a routing entry "/www.bjtu.edu.cn/movie/3" is stored in the route table for prefix matching. In fact, it is aggregated by "/www.bjtu.edu.cn/movie/a.avi/part1 3"and "/www.bjtu.edu.cn/movie/b.avi/part1 3". In this way, the forwarding packets will be forward from port 3, only if the prefix of name is "/www.bjtu.edu.cn/movie/". However, if packets with a name of "/www.bjtu.edu.cn/movie/c.avi" arrive in the router, it will be forwarded from port 3. Actually, the network that port 3 connects only has a.avi and b.avi. This causes the so-called suffix holes [10].

In the proposed hn scheme, the flat part can solve the problem of suffix holes efficiently. For example, there are two resource names "/www.bjtu.edu.cn/movie/s83hho90oxn2783nde4r:kongfu:avi:1024p:part1 3" and "/www.bjtu.edu.cn/movie/8uh723k9ng556sgaesgs:love:rmvb:720p:part2:2012-3-4 3". After route aggregation, the routing entry will become

"/www.bjtu.edu.cn/movie/ 3". The routing entry will be matched when an packet whose name is "/www.bjtu.edu.cn/movie/a932jfdjf2032942-jdd:control:avi:1024p:part1:part2" arrives at this router.

However, it can not be forwarded from the port 3 based on hn scheme due to the incomplete prefix matching. There is a suffix list in each aggregating prefix, and the packet will be forwarded only when the requesting suffix exists in the suffix list. In hn scheme, it must assort a suffix list for each routing entries like "/www.bjtu.edu.cn/movie/ 3" to store the flat parts of names. Although the name of new packet has been matched to the routing entries, its flat part "a932jfdjf2032942-jdd" does not exist in the suffix list "/www.bjtu.edu.cn/movie/ 3". The plat part will be used to confirm whether it forward the request packet when the prefix is matched. By this way, the problem of suffix holes can be resolved effectively. The lookup process of hn names is shown in Figure 4.

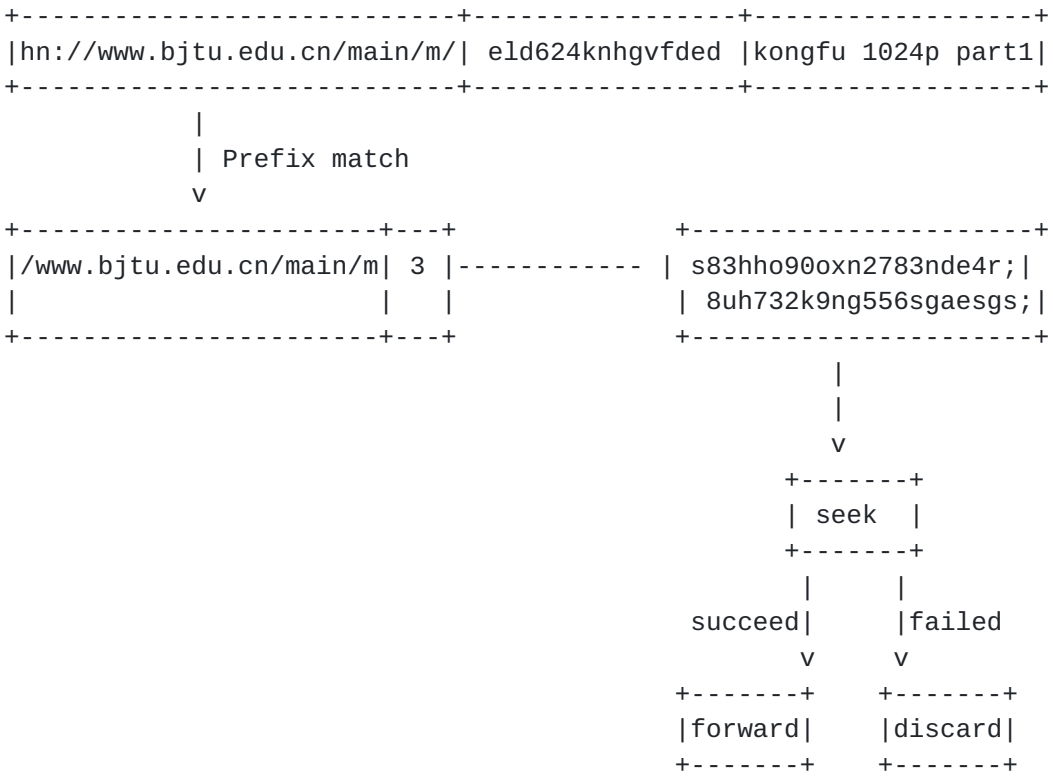

Figure 4 The hn lookup process

4.4. Fuzzy matching support

In the practical, one important situation is that the users may not know the full official resource name when they search a resource. The hn naming scheme support the fuzzy matching thanks to the function of the attribute component. For example, the users need the resource a.avi, they need not know the official name "hn://www.bjtu.edu.cn/m/|u584uuj89324ru|kongfu:movie:avi:1024p:part1". In this case, users only publish the information of video "kongfu" and the resolution ratio is "1024p", the related resources can be found intelligently by fuzzy matching based on the attribute component matching. This is the benefit about embedding attribute of resource in the end of name.

4.5. Good compatibility

This naming scheme provides a good compatibility for all three mainstream naming schemes, which are the subset of the hn naming scheme.

5. Transition from IPv4 and IPv6

In TCP/IP networks, IPv4 and IPv6 addresses are used to represent the resource locations. Combing with the port information and content directory, IPv4 and IPv6 addresses can also be used to fetch the desired information uniquely. We consider the hybrid naming scheme transiting from IPv4 and IPv6 networks.

The IPv4 or IPv6 address is the hierarchical as the first part of the hybrid name. The port number is flat as the second part of the hybrid name. The content directory is a set as the third part of the hybrid name. An illustration of transition from IPv4 and IPv6 is shown in Figure 5.

```
+--------------------+----+-------------------------------------+---+
|hn://192.168.100.100|8080|m:picture:library:west:computer:book | 3 |
+--------------------+----+-------------------------------------+---+

+-----------------------------------------+----+---------------+---+
|hn://2001.da8.215.a815.c492.d445.3489.ec8c|8080|m:picture:book | 3 |
+-----------------------------------------+----+---------------+---+
```

Figure 5 Example of transition from IPv4 and IPv6

6. Formal Syntax

The following syntax specification uses the augmented Backus-Naur Form (BNF) as described in RFC-2234 [RFC2234].

7. Security Considerations

The proposed hn naming scheme has potential benefits for the security. The hierarchical prefix has a high aggregation, which can avoid the security issues of rapid expansion in routing or forwarding table, such as DoS attack. The flat component can protect the users' privacy and the content secrets from readable names. The attributes component can improve the management for the secure contents by using of some encryption key.

8. Conclusions

This document defines a novel hybrid naming scheme for unifying all kinds of information. This hybrid naming scheme owns many advantages, which can provide a good compatibility for existing naming schemes.

9. References

[1] Jacobson, V., Smetters, D., Thornton, J., et al. "Networking named content", Proceedings of the 5th international conference on Emerging networking experiments and technologies. ACM 2009 pp. 1-12.

[2] Zhang, L., Estrin, D., Jacobson V., et al., "Named Data Networking (NDN) project," Technical Report, NDN-0001, 2010.

[3] Yu, J., Varadhan, K., Li, T., et al, "Classless inter-domain routing (CIDR): an address assignment and aggregation strategy", RFC 1519, September 1993.

[4] Ding, S., Chen, Z. and Liu, Z., "Parallelizing FIB Lookup in Content Centric Networking", Networking and Distributed Computing (ICNDC), 2012 Third International Conference on. IEEE, 2012 pp. 6-10.

[5] Koponen, T., Chawla, M., Chun, B., et al, "A data-oriented (and beyond) network architecture", ACM SIGCOMM Computer Communication Review. ACM, 2007 pp. 181-192.

[6] Dannewitz, C., "NetInf: An Information-Centric Design for the Future Internet," Proc. 3rd GI/ITGKuVS Workshop on The Future Internet, Munich, Germany, May 2009.

[7] Carzaniga, A., Rutherford, M. and Wolf, A., "A routing scheme for content-based networking", INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE, 2004 pp. 918-928.

[8] https://observatorio.iti.upv.es/resources/new/542

[9] http://www.supermind.org/blog/740/average-length-of-a-url-part-2

[10] Perino D. and Varvello M., "A reality check for content centric networking", in Proc. ACM SIGCOMM workshop on Information centric networking, 2011 pp. 44-49.

[11] Liu, H. and Zhang, D., "A TLV-structured data naming scheme for content-oriented networking", Communications (ICC), 2012 IEEE International Conference on. IEEE, 2012 pp. 5822-5827.

10. Acknowledgments

Meng Zhang and Liang Zhu contributed to discussion and revision of this document whilst working at Beijing University of Posts and Telecommunications, Beijing, China.

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Hongke Zhang
Beijing Jiaotong University (BJTU)
Beijing, 100044, P.R.China

Email: hkzhang@bjtu.edu.cn

Wei Quan
Beijing University of Posts and Telecommunications (BUPT)
Beijing, 100876, P.R.China

Email: quanwei@bupt.edu.cn

Jianfeng Guan
Beijing University of Posts and Telecommunications (BUPT)
Beijing, 100876, P.R.China

Email: jfguan@bupt.edu.cn

Changqiao Xu
Beijing University of Posts and Telecommunications (BUPT)
Beijing, 100876, P.R.China

Email: cqxu@bupt.edu.cn

Fei Song
Beijing Jiaotong University (BJTU)
Beijing, 100044, P.R.China

Email: fsong@bjtu.edu.cn