

ОБРАБОТКА И ПЕРЕДАЧА ИНФОРМАЦИИ

УДК 316.776

G. Jaber, N. V. Patsei, F. Rahal
Belarusian State Technological University

SEMANTIC INFORMATION-CENTRIC NETWORKING NAMING SCHEMA

The article describes a new semantic-base naming schemer. This proposal takes into consideration the problem of data communication types that traverse the ICN. The legacy proposals in ICN have weaknesses in dealing with some type of communication. In order to deal with this problem, a three-dimension addressing scheme was presented. It includes Geographical, Semantic, and Publisher ID addresses. The article discusses the process of forming a Semantic address on the basis of Network Universal Language with the construction of a semantic graph. We used the IPv6 extension header to define a new routing scheme that can work with a three-dimension address. In conclusion, the routing scheme and tables are briefly described. As a result, the proposed scheme will evolve the interests of Subscribers to a higher abstract level and will reduce the name resolution brokers and delays in some cases.

Key words: routing, information-centric networks, semantic, geographical, address, publisher, subscriber request, IPv6.

Introduction. Information-Centric Networking (ICN) or its other names including Data-Oriented Networking, Content-Based Networking or Content-Centric Networking/Named Data Networking, is a substitute paradigm for the present architecture of the Internet that focuses on naming data for its model of communication [1]. There are some problems in the present architecture of internet for which the ICN is able to find resolutions. The problems include ineffective use of resources, Distributed Denial of Service (DDoS) attacks, lack of security, and problems in the fields of mobility, scalability, routing protocol as well as economic problems [2].

The routing protocol defines the manner of communication between network routers. This protocol sends required information to routers and enables them to select possible routes between two existing nodes in the network. On the other hand, routing algorithms are responsible to make decision about the appropriate selection of the route [3]. Each router is equipped with the knowledge of specific networks with direct connections to it. The routing protocol distributes this information to adjacent neighbors in the first place, and to the whole network in the second place. That is how the routers gain knowledge about the topology of the network [4]. The routing approach can be considered the heart of any ICN architecture, in this regard, each ICN routing protocol tries to find one or more copies of the distributed information within the network [5]. There are different routing protocols offered in different ICN architectures, from which name resolution and data routing are the most common protocols.

There are two roles defined for routers in the ICN architectures at the time of a request for a particular Named Data Object (NDO). The first task of the routers is finding a node that has a copy of the required piece of information, and forwarding a request the node. The second task is finding a route from the node to the user who had asked for the information piece. A method of doing these two tasks is called *name resolution*. This method includes finding one or more lower-layer locators for the name of NDO. These locators are able to call back the requested NDO. The other way to do the routing tasks is called *name-based routing*. In this method, the request for the NDO is directly routed to the node that has a copy of the content (based on the NDO's name). The name resolution phase in the name-based routing is removed [6, 7]. Fig. 1 displays the types of routing in the popular ICN architectures.

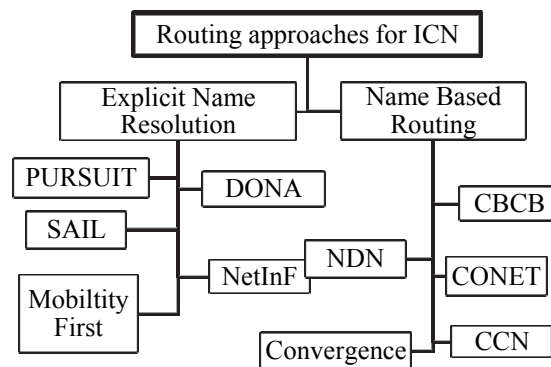


Fig. 1. Types of routing in different ICN architectures

Challenges in the name resolution and routing process are the following: ensured delivery and detection of the nearest copy of required content, scalability (it includes the development of an information model and a naming framework which support efficient information dissemination with improved security properties; it also includes the development of a world-wide scalable name resolution mechanism for a new namespace), excessive current on routing tables (if an overflow takes place, the router rejects request packets, the user experiences a low transmission rate, and the whole network will crash as a result), single point for failure (this problem happens when a great number of published and registered NDOs in the Name Resolution System (NRS) go unavailable), security and filtering.

Most of the proposed techniques in ICN are not suitable to deal with all data transmission types between Publishers/Subscribers. Another problem that resides in the proposed ICN schemas (fig. 1) is the limitation to deal with knowledge searching.

To solve these problems the types of data communication were examined and there was suggested the classification of data transmission into four types based on the number of subscriptions and frequency of data object use [8]. There are four scenarios for working with data, conditionally named A, B, C and D:

- type A: one subscriber – one use (voice call);
- type B: one subscriber – reusable (cloud storage);
- type C: several subscribers – one use (video streaming);
- type D: multiple subscribers – reusable (YouTube).

Besides, we also classified subscribers' requests into types. Subscribers' Requests may be of four types: *R1* – requesting any data content from a specific Publisher; *R2* – requesting specific data content from a specific Publisher; *R3* – requesting specific data content from a specific Publisher; *R4* – requesting information with any data content from any Publisher.

Theoretical base. Due to the high mobility of terminals in nowadays networks, the Publisher and Subscriber should hold a *Dynamic Address* that may be changed according to their geography in the network. In addition, the name should represent the content (an intuitive address) to serve Subscribers' requesting information (*R4* type) and should be unique to serve *R3* Subscribers. Thus, three dimensions for a naming scheme are proposed in a model of ICN network called Semantic Information-Centric Networking (SINC).

SINC naming scheme is based on the principle that the user (Publisher/Subscriber) should label

the data with at least one dimension. The three dimensions (3D-address) are: Publisher ID, Semantic name, and Geographical ID.

Geographical address. A geographical address is a 128-bit unique address assigned by local host itself to route data towards a particular known location in the network based on the hierarchal structure. The IP address is an application on the geographical address that routes data from a source to a destination in a very flexible and fast way. This address is used here since it will facilitate routing towards the Publisher and the Subscriber taking into consideration the mobility of the Subscriber or the Publisher.

When the Subscriber moves from one sub network to another, his geographical address (IPv6) should change based on his new sub network (location/geography), so in proposed scheme we suggest to use EUI64 addressing technique to all mobile users (Publishers/Subscribers). This address allocation technique will allow each user in the network to have a unique address suffix due to the fact that the last part of EUI64 address is based on the MAC address of the user interface.

Considering a user interface with the following MAC address: 20-68-9D-94-77-1E moving to a subnet 2000::/64 will automatically assign the following IPv6 address: 2000::22:68:9D:FF:FE:94:77:1E/64.

Suppose that the user changes his sub network, it could be easily reached by his EUI64 suffix. A suggestion to reach this user is through the packet broadcast to all the nearest sub networks by changing the subnet address prefix part and fixing its EUI suffix address thus fixing the suffix and changing the prefix (table 1). This process will ensure the roaming of the users (Publishers or Subscribers) between subnets even in case of high mobility.

Table 1

Geographical Address Structure

Prefix: Subnet (mobile)	Suffix: Mac Address (fixed)
2000::/64	20-68-9D-94-77-1E
2000::22:68:9D:FF:FE:94:77:1E/64	

Publisher ID address. It is a set of addresses, built on the root or main unique address which is assigned by a central authority (Assigned Names and Numbers) ICANN. ICANN authorizes domain name registrars, through which domain names may be registered and reassigned. Publisher ID address is a 128-bit hierarchal address. This address is flat human friendly address that is readable by human (Domain Name Space). Each content within the Publisher can be addressed with other sub address that is assigned locally by the Publisher itself.

Let’s take “BELSTU” as an example. It is a Publisher, that has a global unique 128-bit address assigned from ICANN. “BELSTU” will give each content (faculties and departments) it publishes a 128-bit sub address. This address is important to be used as in *R1* type request. Another example where this address shows high significance is the necessity to verify the publisher’s ID. In case of *R1* and *R3* Subscriber’s request (e.g. voice call, video call), a central agent (e.g. WhatsApp sever) should have a public address and manage the data transmission between two Subscribers.

Semantic address. A semantic address is formed on the basis of Universal Networking Language (UNL). UNL is a declarative formal language specifically designed to represent semantic data extracted from natural language texts [9]. The pivot paradigm is used: the representation of an utterance in the UNL interlingua is a hypergraph where normal nodes bear UWs (Universal Words) with semantic attributes (@a), and arcs bear semantic relations (R) as it shown on fig. 2 [10, 11].

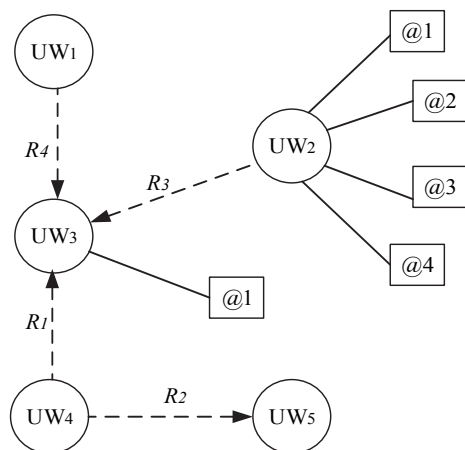


Fig. 2. UNL structure

The term “Universal Word” represents simple or compound concepts. There are three types of UWs: basic, restricted and extra UWs. A UW can have an UW ID. It is used to refer to some information and there are thirty-six UW-IDs (numbers from 0 to 9 and letters from A-Z).

Compound UWs represents a set of binary relations that are grouped together to express a concept. A sentence itself is considered a compound UW. For the graph fig. 2 the UNL representation is the following:

- R1 (ID4: UW4: —, ID3: UW3: @1 —)
- R2 (ID4: UW4: —, ID5: UW5: —)
- R3 (ID2: UW2: @1@2@3@4, ID3: UW3: @1 —)
- R4 (ID1: UW1: —, ID3: UW3: @1)

The sentence [BSTU held a conference] could be presented with UW as follows:

```
agt(held(ic1>do), BSTU(ic1>organization))
obj(held(ic1>do), conference(ic1>event))
```

Relations are binary connecting two UWs. There are forty labels that represent the relations between UWs in binary relation. They can be *ontological* (such as “ic1” and “iof”, referred to above), *logical* (such as “and” and “or”), and *thematic* (such as “agt” = agent, “ins” = instrument, “tim” = time, “plc” = place, etc.).

Attributes of UWs are used to describe subjectivity of sentences. Attributes represent information that cannot be conveyed by UWs and relations. UNL attributes shows view, aspect, time of event, etc. Normally, they represent information concerning *time* (“@past”, “@future”, etc.), *reference* (“@def”, “@indef”, etc.), *modality* (“@can”, “@must”, etc.), *focus* (“@topic”, “@focus”, etc.), and so on. There are 58 attributes in UNL [12].

For example:

```
agt(held(ic1>do).@entry.@past, BSTU
(ic1>organization))
obj(held(ic1>do).@entry.@past,
conference(ic1>event).@indef)
```

The attribute @entry denotes the main predicate of the sentence, @past – the present tense, and @indef – a non-specific class.

In proposed name scheme for SINC, UNL is adapted to create Semantic addresses.

In SINC name we assign for R (relation) 6 bits. 12 bits for the weight of the relation between two Universal Words (fig. 3). In SINC scheme we assign for each UW 31 bits, 6 bits for UW-ID and in each UW up to three attributes for each – 6 bits.

As you can see on fig. 3 for every relation we have 128 bits. So, a semantic address is a set of relations and descriptions of a semantic graph.

SINC Header Format. In accordance with added addresses, the IPv6 header structure has been redesigned (fig. 4). We left the fixed part of the header unchanged: *Version* (4 bits) indicates version of Internet Protocol; *Traffic Class* (8 bits) indicates class or priority of IPv6 packet, it helps routers to handle the traffic based on priority of the packet; *Flow Label* (20 bits) is used by source to label the packets belonging to the same flow; *Payload Length* (16 bits) – indicates total size of the payload which tells routers about the amount of information a particular packet contains in its payload; *Next Header* (8 bits) – indicates type of extension header; *Hop Limit* (8 bits) is same as TTL in IPv4 packets and indicates the maximum number of intermediate nodes IPv6 packet is allowed to travel; *Source Address* (128 bits) – an address of the original source of the packet; *Destination Address* (128 bits) – field indicates the IPv6 address of the final destination [13].

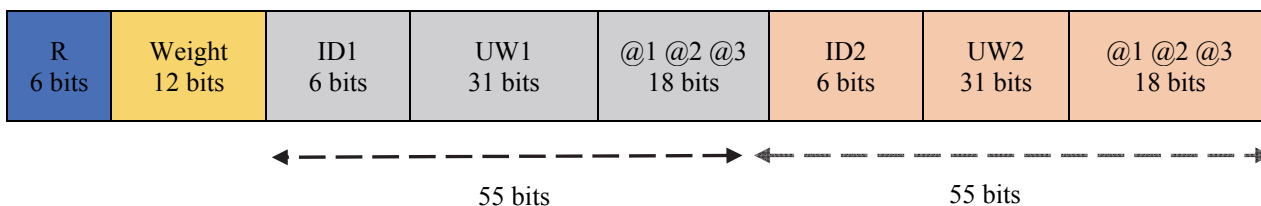


Fig. 3. Structure of UNL based semantic relation

Then we will use the *Extension Headers* for storage of Metadata and a three-dimension-naming scheme (3D-address): Geographical, Publisher ID and Semantic address.

Metadata Addressing fields are used for the address classification. 128 bit field divided into 12 parts of 10 bits each and with a remaining part of 8 bits as shown in fig. 4. Thus, each 10 bits part of the 12 parts in this field is classified into two sub parts where the mask part has 7 bits and the logic relation part has 3 bits.

<i>Version</i> (4 bits)	<i>Priority/Traffic class</i> (8 bits)	<i>Flow label</i> (20 bits)	
<i>Payload length</i> (16 bits)		<i>Header SICN</i> (8 bits)	<i>Hope limit</i> (8 bits)
<i>Source address</i> (128 bits)			
<i>Geographic Destination address</i> (128 bits)			
<i>Metadata Addressing fields</i> (128 bits)			
<i>Geographical addresses</i> (128 bits)			
<i>Publisher ID addresses</i> (128 bits)			
<i>Semantic addresses</i> (128 bits)			
...			

Fig. 4. SICN Header Format

Note that the number of Semantic addresses will be variable. The total number of addresses (Geographical, Publisher ID and Semantic) should be no more than 12. If you need to set the number of addresses to less than 12, it should complete the address set with zeroes, which indicates the end of the address list.

Routing schema. Conventional IP networks routing schemes works on network layer and do not take into consideration any aspect related to the content of the routed data. To reach a data destination, packets are labeled with Geographical Address (IPv6) that it is easy to reach by the help of routing tables and where these labels are learned dynamically by routing protocols or predefined statically by network administrators.

Literature proposed routing schemes [7] are based on Publisher/Subscriber scheme, which uses filters to match rendezvous points between Subscriber Interests, and Publisher advertisement compared to the conventional IP. These routing schemes work well with data type C and D where many Subscribers are interested in the published

data. However, there is a shortage in dealing with the case of one Subscriber interested in the data from a specific Publisher (type A), i.e. when the Subscriber needs the data from a specified Publisher, whatever the data content. The proposed schemes in the literature would cost many network resources compared to the conventional IP network. Even in case of type B data if the Subscriber has a low reuse factor (frequency of usage), the conventional IP network will solve the problem in a better way. This fact is due to large amounts of filters that will be registered at the routers using the Publisher/Subscriber technique data. In addition, the latency time for the path, that is due to the path initiation required to match between the information and the Subscriber is inconvenient to provide quality of service (QoS) with type A data.

Routing scheme, should deliver data to Subscribers with an effective path cost (served from the nearest node that caches the data).

Fast routing is a necessity, which insists on designing a simple scheme that do not exhaust the router with much process complexity (e.g. solving first order logic filters or searching huge list of flat addresses).

Our proposed SICN takes into consideration the types of data usage. In addition, it can match between the Publisher knowledge and the Subscriber interested knowledge. Currently, the search engine holds this role. In other words, the whole network will work as a big routing search engine that matches Subscriber interest to data, and the Subscriber's interest to Publishers. This is done with the help of a three-dimension-naming scheme (Geographical address, User (Publisher/Subscriber) unique address, Semantic address) that is done in the extremes not in the core.

Routing tables. Routers will hold three tables with three address dimensions combined in them. The first one is the *Semantic-ID* that connects the Semantic address to the Publisher ID address. The second one is the *Geo-ID* that connects the Publisher ID address and the Geographical address and the third table is the *Geo-Semantic* matches the Semantic address to the Geographical address.

These three address dimensions will allow the matching between the Publisher and Subscriber based on naming scheme that includes any Publisher ID, Semantic or Geographical address in

the network and will be designed to include the four types of data and the four types of Subscriber's requests. A Subscriber interested in one of the three address dimensions can find a match to the other two address dimensions using the proposed routing tables. For example, an interest message containing only a Semantic address can easily be matched to Publisher IDs and their Geographical location using these tables. Considering another example where a subscriber having a phone call with a specific Publisher ID can follow the Geographical location of the Publisher using the second table.

Each table includes two parts. The first part, which is the address part (Publisher ID, Geographical and Semantic addresses) that names the data and are learnt or defined from the Publisher's advertisement. The second part of each table, which is the orientation part (cache (TTL) and

Interface) that directs the data toward the Subscriber and are learnt from the Subscriber's interest message. The interface is an input-output port, which connects network nodes.

Conclusion. This article presents a new scheme in ICN. Through this project, we addressed the problem of Naming and Routing in the field of Information-Centric Networking where a new semantic-based scheme is proposed to solve the obstacles facing IP networks. We presented a new architecture scheme SICN and detailed its naming and a part of routing designs. An important contribution is classifying data into four types and classifying the Subscriber's request into four classes where the new system can cope with these different types and classes. In addition, three naming schemes were detailed. Furthermore, we designed the SICN Header format.

References

1. Jaber G., Patsei N. V. Information Centric Networking for web-based content distribution and manipulation. *Trudy BGTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2017, no. 2, pp. 88–91.
2. Alzahrani B. A., Vassilakis V. G., Reed M. J. Key management in information centric networking. *Int. J. Comput. Networks Commun.*, 2013, vol. 5, no. 6, pp. 153–156.
3. Pepper R. Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update. *Tech. Rep.* Berlin, 2013, p. 245.
4. Olsen L. J. Services for substance abuse-affected families: The Project Connect experience. *Child Adolesc. Soc. Work J.*, 1995, vol. 12, no. 3, pp. 183–196.
5. De Brito M. A. G., Galotto L., Sampaio L. P. Evaluation of the main MPPT techniques for photovoltaic applications. *IEEE Trans. Ind. Electron.*, 2013, vol. 60, no. 3, pp. 1156–1167.
6. Lee J.-C., Lim W.-S., Jung H.-Y. Scalable domain-based routing scheme for ICN. *Information and Communication Technology Convergence (ICTC): International Conference*. Jeju, 2014, pp. 770–774.
7. Navrotsky Y., Patsei N. Caching Control and Optimization in Information-Content Networks. *Open Conference of Electrical, Electronic and Information Sciences (eStream): Proceedings of the Conference*. Vilnius, 2019, pp. 70–74.
8. Jaber G., Patsei N., Rahal F. Different Naming in Information-Centric Networks (ICN). *Scholars Journal of Engineering and Technology*, 2019, no. 7 (8), pp. 235–237.
9. UNL web community portal. Available at: <http://www.unlweb.net/unlweb/> (accessed 18.11.2019).
10. UNL portal. Available at: <http://www.undl.org/http://www.unlweb.net/unlweb/> (accessed 18.11.2019).
11. Uchida H., Zhu M. The universal networking language beyond machine translation. *International Symposium on Language in Cyberspace*. Seoul, 2001, pp. 26–27.
12. Alansary S., Nagi M., Adly N. The universal networking language in action in English-Arabic machine translation. *Proceedings of 9th Egyptian Society of Language Engineering Conference on Language Engineering (ESOLEC 2009)*. Cairo, 2009, pp. 23–24.
13. Hinden R., Deering S. IP Version 6 Addressing Architecture. Network Working Group. DOI: 10.17487/RFC4291. Available at: <https://tools.ietf.org/html/rfc4291> (accessed 18.11.2019).

Information about the authors

Jaber Ghassan – PhD student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: ghassanjaber@hotmail.com

Patsei Nataliya Vladimirovna – PhD (Engineering), Associate Professor, Head of the Department of Software Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: n.patsei@belstu.by

Rahal Fatima – PhD student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: Fatimarahaljaber@gmail.com

Received after revision 19.11.2019