УДК 316.776

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

SEMANTIC INFORMATION CENTRIC NETWORK MODEL EVALUATION

The article describes the evaluation result of SICN (Semantic Information Centric Network) model scheme. First, it describes the abstract levels of contents. It divides data and request classification into four types based on the number of subscriptions and frequency of data object use. These scenarios of data flow are described for SICN model illustrating. Time Delay, Flooding or traffic and Efficiency reuse factor of data are used as parameters for evaluation. The paper represents a simple network example with quantitative results. In addition, a simulation model for IP, DONA, PURSUIT, CBCB, KBN, and SICN has been created to understand their performance with some assumptions. Four different scenarios have been applied. Each scenario represents a certain abstract content level knowing that the scenarios represent the data types and request types. The relevant results are shown and they proof the SICN good performance for searching information from different sources and downloading files.

Key words: scenario, information centric networks, semantic, model, performance, routers, subscriber, publisher, node, data type, time delay.

Introduction. The series of articles [1–5] introduced the architecture of Information Centric Network (ICN) named Semantic Information Centric Network (SICN). We detailed its naming, caching and routing designs [4, 5]. SICN naming scheme depends on three addresses: semantic address, geographic and publisher ID address. Furthermore, we designed the names in SICN header format.

SICN could solve the major problems in the current ICN field [5]:

– reduce brokers between information and end-users by compensating search engine;

– reduce redundancy, traffic and overflow of network;

– access semantically, which is better than using brokers as search engines;

– propose a one new scheme that could deal with all data connection types.

An important contribution of the SICN work is the classification of data into the four types based on the number of subscriptions and frequency of data object use [4, 5].

Type A: one subscription and one usage. A subscriber interested in the type A data will select specific unique publisher address from any semantic and any geographic location. Since the subscriber is interested in a specific publisher data source regardless of the content semantic, publisher monitors the mobility of the user in the network. This select query will use the Geo ID routing table [2].

Type B: one subscription and many usages. A subscriber interested in the type B data will select specific semantic from a specific publisher, thus the Semantic ID routing table is suitable to match the subscriber interest to the published data and to locate the publisher.

Type C: many subscriptions and one usage. A subscriber interested in the type C data will select a specific semantic from any publisher where the Semantic ID routing table is suitable for such selection.

Type D: many subscriptions and many usages. A subscriber selects any semantic from any geographic location. The type D in the network is most important that ICN can deal with.

Besides, SICN also classifies subscribers' requests into types. Subscribers' requests may be of four types as shown in table 1.

Table 1

Subscribers' request types

Class	Description	Example
Rq1	Requesting *any data content* from *specific publisher*	Voice call
Rq2	Requesting *specific data content* from a *specific publisher*	Cloud storage
Rq3	Requesting *specific data content* from *any publisher*	Downloading a file
Rq4	Requesting information with *any data content* from *any publisher*	Searching information with Google search

To compare and analyze the work of the proposed SINC scheme, it is necessary to build a model of the semantic information network and compare its work with analogs naturally taking into account the type of information and requests.

Theoretical base. According to the classes and data types we can define four communication scenarios.

Scenario 1 (type A). The communication components are publisher and two subscribers who contact each other. Publisher has two addresses, for example **Pub ID**: *Viber*; **Geo ID**: *2B::01:2C*. The first subscriber is the calling node and the second one is the called node. Subscriber sends *IRM* (Interest Request Message) having the following 3D-address: **Pub ID**: *Viber*; **Geo ID**: *none*; **Semantic ID**: *none*.

IRM propagates (broadcasts) to the routers and searches in each Geo ID table till it reaches the publisher *"Viber"*. Matching will occur between *IRM* and the **Pub ID**. *IRM* will be updated by the new IP at each router. The publisher router will send *ARM* (Address Reply Message) with the following 3D-address: **Pub ID**: *Viber*; **Geo ID**: *2B::01:2C*; **Semantic ID**: *none*. *ARM* will propagate on the same path as IRM but in an inverse direction until reaching the source of *IRM*. The Geo ID table of each router passed by *ARM* will be added by a new record. The calling subscriber router will send *CRM* (Content Reply Message) to the called subscriber router passing through the publisher router. *CRM* will have the following 3D-address: **Pub ID**: *Viber: Private ID: +375 171234567*; **Geo ID**: *2B::01:2C*; **Semantic ID**: *none*.

Scenario 2 (type B). Suppose there is a subscriber phone/computer that needs certain information access in the cloud. The publisher router will send *IRM* to find the information with the following 3D-address: **Pub ID**: *drive.google*; **Geo ID**: *none*; **Semantic ID**: *none*. *IRM* will propagate until it reaches the publisher. The publisher router will send *ARM* with the addresses: **Pub ID**: *drive.google*; **Geo ID**: *2B::02:2C*; **Semantic ID**: *none*. The publisher router will send *CRM* for the subscriber with the requested file.

Scenario 3 (type C). Let's take a case of a video stream where Facebook is a publisher and a certain phone is a subscriber. The subscriber will send *IRM* with these addresses: **Pub ID**: *Facebook/channel/video1*; **Geo ID**: *32::2C:1A*; **Semantic ID**: *none*. *IRM* will reach publisher router where the latest will send *ARM*. When the publisher router sends *ARM* in all scenarios, Time To Live (*TTL*) in router tables will be increased and tables will be updated. If *TTL* reaches threshold, then the addresses will be sent to cache.

Scenario 4 (type D). Let's take a case of a certain subscriber who needs to search the following information in the network: *"diameter of the moon"*, as an example on the type D scenario and suppose that only two publishers have this information. The subscriber will send *IRM* having the following 3D-address: **Pub ID**: *none*; **Geo ID**: *none*; **Semantic ID**: *atb (moon, diameter)*. *IRM* propagates to reach publishers where matching will occur. Each publisher router will send back *ARM* in an inverse direction of *IRM*. Each router will be learnt by the three addresses. *ARM* will reach the subscriber. Subscriber will send the second *IRM2* having addresses: **Pub ID**: *Libgen*; **Geo ID**: *2B::01:2C*; **Semantic ID**: *atb (moon, diameter)*. *IRM2* will reach the needed publisher with a previously known path since the first *IRM* saved all routers IPs needed along the path. The publisher will send *CRM* having the data and the 3D-address.

Comparative analysis. The results of SINC compare with DONA (Data-Oriented Network Architecture), PURSUIT (Publish-Subscribe Internet Technology), CBCB (Combined Broadcast and Content-Based), KBN (Knowledge-Based Networking) schemas according to many criteria including the routing approach, naming structure, caching, and backward comparability shown in table 2.

Table 2

ICN project analysis

ICN	Routing approach	Naming structure	Routing	Caching	Abstract level	Backward compatibility
DONA	Name resolution	Flat naming self-certifying	Pull	On-path; off path caching	Data	Yes, work over IP
CBCB	Name based routing	Set of paired attribute value; don't ensure uniqueness	Pull-push	On-path; off path caching	Information	No
PURSUIT	Name resolution	Flat naming	Pull	On-path; off path caching	Data	No
KBN	Name based routing	Set of keywords; ontological categorization	Pull-push	On-path; off path caching	Knowledge	No
SICN	Name based routing	Human friendly; hierarchal (Geo address associated with IP)	Pull	On path caching	Knowledge	Yes, work over IP

Three different metrics over this ICN schemas have been computed: *Time Delay* (*TD*) – minimum numbers of hops; *Flooding* or *traffic* (*F*) – total number of links occupied starting from the data requested to the data received; *Efficiency reuse factor of data* (*ER*) – ratio of the interest numbers of a certain data to the data travel distance.

In order to compare different schemas (table 2), we have built six simulation models in Python programming language composed of: not fully connected routers (R_i); three network subscribers (NS_i); network publisher (NP_i).

The main content source is connected with private cache (*PC*). Resolution server (*DNS*) is needed for some of the schemas to find data source IP. A search engine is needed for some of the ICN schemes to translate the data from an informal form to a formal one.

Let's consider as a work model example the content transmission in *Scenario 1* with no caching data. It is the type A data request where subscribers ask for any data but from a specific publisher (voice call). In SICN scheme, the message will be passed as follows from the subscriber NS_1 to the publisher as follows:

$$NS_1 \to R_6 \to R_4 \to R_1 \to NP_1.$$

Then back from the publisher to the subscriber as follows:

$$NP_1 \to R_1 \to R_4 \to R_6 \to NS_1.$$

The message transmitted from the subscriber NS_2 as follows:

$$NS_2 \to R_7 \to R_4 \to R_1 \to NP_1.$$

Then back:

$$NP_1 \to R_1 \to R_4 \to R_7 \to NS_2.$$

The message transmitted from the subscriber NS_3 as follows:

$$NS_3 \to R_8 \to R_4 \to R_1 \to NP_1,$$

then

$$NP_1 \to R_1 \to R_4 \to R_8 \to NS_3.$$

TD that represents the number of links from the subscriber to the publisher then from the publisher to the subscriber have 8 links (fig. 1).

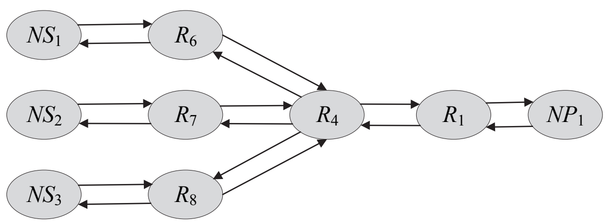

Fig. 1. SICN message flow in *Scenario 1*

F is the total links occupied from requesting data by the subscriber to receiving it. In SICN model it is the number of subscribers multiplied by *TD*. In our scenario, $F = 3 \cdot TD = 24$. The *ER* factor represents the number of reusing data compared to data travel distance. *ER* in SICN is ¼ for each subscriber, thus totaling ER = ¾.

The content transmitted in *Scenario 2* is the type B cashed data. Subscribers ask for specific data from specific publisher. We suppose that in this model users send enough requests for the file to reach the threshold and cache in R_4, and R_5. Each subscriber sends his request by an *IRM* carrying Pub ID the semantic addresses and also the Geo ID address as it is supposed that all tables are converged. It means that the request has been made before.

The data sources are cached and they send *CRM* to the subscribers, thus the message flow in SICN will be as follows (fig. 2). From the subscriber 1 to data source in R_5:

$$NS_1 \to R_6 \to R_5,$$

then, from data source R_5 to the subscriber 1:

$$R_5 \to R_6 \to NS_1.$$

From the subscriber 2 to data source R_4:

$$NS_2 \to R_7 \to R_4,$$

then

$$R_4 \to R_7 \to NS_2.$$

From the subscriber 3 to data source R_4:

$$NS_3 \to R_8 \to R_4,$$

then

$$R_4 \to R_8 \to NS_3.$$

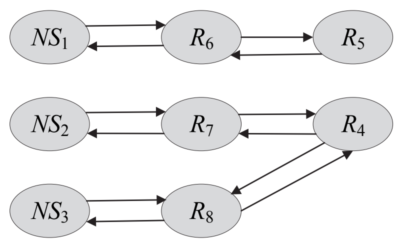

Fig. 2. SICN message flow in *Scenario 2*

The content transmitted in *Scenario 3* is information. It is the type C transmission where subscribers ask for specific information but from any publisher. The content transmitted in *Scenario 4* is knowledge.

Table 3 shows the values of three metrics in each ICN schemes and each scenarios.

For further modeling let's do the following assumptions:

Table 3

ICN projects evaluations

ICN projects	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
	TD	*F*	*ER*	*TD*	*F*	*ER*	*TD*	*F*	*ER*	*TD*	*F*	*ER*
IP	16	48	75	16	48	0.3	32	96	25	32	96	25
SICN	8	24	75	4	12	100	4	12	100	4	12	100
DONA	16	48	75	12	36	100	24	72	100	24	72	100
PURSUIT	16	48	75	12	36	100	24	72	100	24	72	100
CBCB	8	47	75	4	16	90	4	16	90	12	36	90
KBN	8	47	75	4	16	90	4	16	90	4	16	90

u: number of users ($u = 10$);

n: publisher depth (defined as number of extended branches from a subscriber to data source; n is variable);

e: search engine depth (defined as number of branches from a subscriber to search engine; lets $e = n$);

d: DNS depth (defined as number of branches from a subscriber to DNS; lets $d = n / 2$);

c: cache depth (number of branches from a subscriber to cache, we supposed that $c = n / 2$);

s: sharing coefficient (defined as the ratio of shared links by subscribers to total links, lets $s = 0.25$);

r: sharing factor (defined as the utilization factor from sharing paths between subscribers, $r = 1 + s(u - 1)$;

L: total number of extended branches for each subscriber to data source, $L = 2^{(n+1)} - 2$, it is supposed that each node has two branches.

We have modeled the work of six schemes in *Scenario 1*. Fig. 3 illustrates time delay (*TD*) versus the number of links to the data source in six schemes. As shown in the fig. 3, in case when the content schemas use name resolution routing (SICN, CBCB and KBN) outperforms schemas using name-based routing (IP, DONA, PURSUIT). This is obviously shown as *TD* for the first group is less than *TD* for the second group. The results are justified as schema needs DNS in the second group whilst the first group has connection to the publisher.

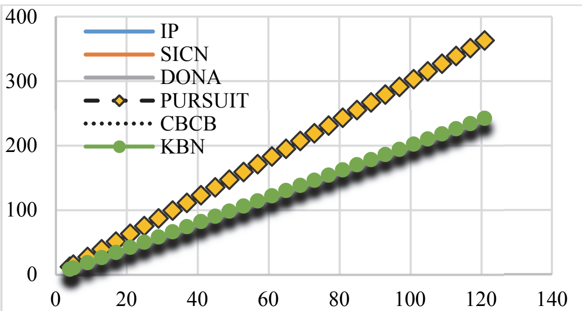

Fig. 3. Time delay versus number of links to the data source

Fig. 4 illustrates that SICN obtains much better performance of SICN compared to other schemas in terms of flooding parameter. This is clarified by a lower *F* for SICN.

The name resolution routing schema (CBCB and KBN) shows high values of *F* (i. e. a lot of flooding) leading to high traffic as they need NRS and make extraction for a tree to find the publisher. This extraction occupies a big part of a tree in case of voice call. The named based routing schema (IP, DONA, and PURSUIT) does a lot of flooding as well. Therefore, SICN might be suited to reduce network traffic in case of *Scenario 1*.

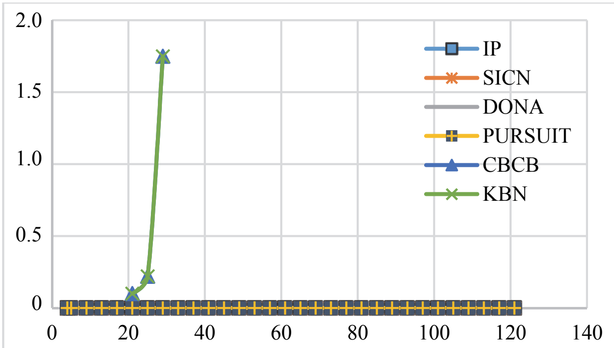

Fig. 4. Flooding versus number of links in the data source

Reuse efficiency fig. 5 shows the data, where all the schemas have the same performance in *Scenario 1*.

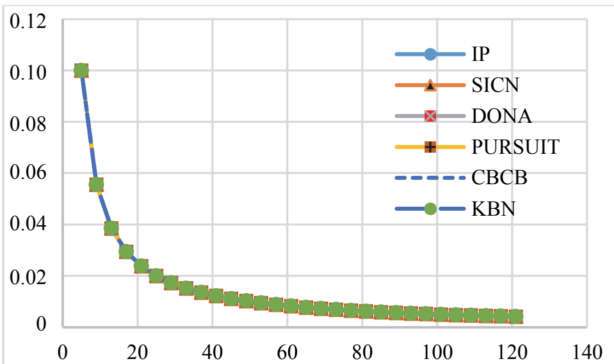

Fig. 5. Efficiency of data versus number of links to the data source

Conclusion. However, none of ICN proposal solutions fit perfectly to all types of requests [6]. For example, Combined Broadcast and Content-Based (CBCB) project can route requests to the content to solve *Rq4* requests, but it consumes a lot of processing at the node to serve *Rq1* and *Rq2* requests.

A main difference between considered in current modeling six different schemes is the method or level each of them deals with the different levels of content concerning their abstract form. Thus, we start up from differentiating between three levels of contents: data, information and knowledge. Thus, empirical results were tested over four scenarios: *Scenario 1* (S1), where the content transmitted is from non-cached data level; *Scenario 2* (S2), where the content transmitted is from cached data level; *Scenario 3* (S3), where the content transmitted is from information level; *Scenario 4* (S4), where the content transmitted is from knowledge level.

According to modeling results SICN, CBCB and KBN outperform IP, DONA and PURSUIT in *Scenario 1* and *2* in terms of *TD*. In terms of *F*, SICN shows the lowest flooding in *Scenario 1* and *2* equals 4 and 12 respectively.

IP shows the highest *F* as it cannot benefit from caching. SICN, DONA and PURSUIT have the highest efficiency in data shown by their efficiency reuse factor (*ER* = 100) for *Rq2*, *Rq3* and *Rq4*. The low *ER* factor in IP is justified as there is no cached data used.

SICN benefits from the absence of DNS and search engine in minimizing the *TD* and *F*.

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. Patsei N., Jaber G. Routing Schema for Information-Centric Networking. *11th International Conference NEET.* Lublin, 2019, p. 29.

3. 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.

4. Patsei N. V., Jaber G. Semantic base addressing strategies for Information-Centric Networking. *Problems of Information Technology: materials of the III All-Ukrainian Science-and-Technology Conference.* Poltava, 2019, p. 103.

5. Jaber G., Patsei N. V., Rahal F. Semantic information-centric networking naming schema. *Trudy BSTU* [Proceedings of BSTU], series 3, Physics and Mathematics. Informatics, 2020, no. 1, pp. 69–73.

6. Vasilakos A. V, Li Z., Simon G., You W. Information centric network: Research challenges and opportunities. *J. Netw. Comput. Appl.*, 2015, vol. 52, pp. 1–10.

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

Received after revision 05.05.2020