УДК 316.776

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

INFORMATION CENTRIC NETWORKING FOR WEB-BASED CONTENT DISTRIBUTION AND MANIPULATION

Modern Internet architecture is based on the model of interaction with the host-center, which was convenient at the initial stage of the Internet development. However, today most users are interested in receiving (often it is huge amounts) information, regardless of its physical location. This shift paradigm in the Internet model, as well as the need to improve security and mobility support, led to a radical revision of the future Internet architecture. In this connection a lot of researches have appeared. In particular, these are some models with an information-centric network (ICN) organization. There are a lot of open questions in ICN. Some of which are discussed in this article. First, these are the basic functional capabilities of ICN architectures. Secondly, it is naming and data routing organization. Thirdly, these are issues of security and caching.

Key words: web-content, Information-Centric networks, architecture, model, naming, routing, caching.

Introduction. At the recent years, the field of Internet networking has replenished with a new paradigm – Content-Centric (CCN) or Information-Centric networks (ICN). While the architecture of packet-switching networks was based on a host-to-host communication path, the architecture of the ICN is based on making a connection between users and contents (or information) [1]. The term "content" means video, audio, web pages, software and the like. In a content-centric information world, what the user wants is becoming more important than where the server is located. In this case, the content becomes the starting point. Information is named, addressed and delivered regardless of location (it can be located anywhere in the network [2–6]). Instead of host specifying, the information name is indicated. As a result of the transformation from the host naming model to the information naming model, the receiver searches for information (in the existing Internet architecture, the sender controls the data exchange). In ICN, after sending a request, the network is responsible for finding the best source that can provide information. So, the information routing request seeks the best source of information based on a location-independent name. For example, communication in ICN is initiated by receivers that request for Named Data Objects (NDOs). The senders, on the other side, provide the required NDOs for receivers by means of publishing the objects. The ICN can identify the users' requests from any resource having a copy of the object. This ability enables caching [2, 7].

So, the main aims of ICN networking architecture are providing better services regarding content distribution, improving mobility of the network, while the networks are more resistant against disruptions and breakdowns [8–9].

Theoretical base. The related projects with the ICN are: Data Oriented Network Architecture (DONA) project; Publish-Subscribe Internet Technology (PURSUIT); Publish-Subscribe Internet Routing Paradigm (PSIRP); Scalable and Adaptive Internet solutions (SAIL); 4WARD architecture (as the ancestor of SAIL); Content Mediator architecture for content-aware networks (COMET); CONVERGENCE; Named Data Networking (NDN); Content Centric Networking (CCN); Mobility; ANR Connect. The above-mentioned projects and architectures are still under development, and they all share a number of major functionalities.

DONA is one of the first complete ICN architectures. Each piece of information (or services) is associated with the "principal". Names consist of a cryptographic hash, is flat, self-certifying (they allow the subscriber to verify that the received data matches the requested name) and uniquely identifying information. Name resolution in DONA is provided by specialized logical servers – Resolution Handlers (RHs) (Fig. 1).

Publisher (principal) sends a "Register" message with the object's name to its RH, who stores a pointer to the principal. The RH then propagates registration to the other RHs, following the established routing policies. Subscriber sends a "Find" message to RH, until a matching registration entry is found. When a "Find" message reaches an appropriate publisher, the "Data" can be directly sent to the subscriber using IP routing and forwarding

(Fig. 1). DONA supports on-path caching via the RH infrastructure (data will be cached there).

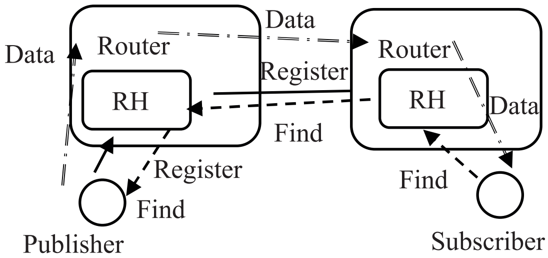

Fig. 1. DONA architecture

NDN provides for the rebuilding of the Internet protocol stack. In NDN a "strategy" layer mediates between the named "data" layer and the underlying network technologies, while a "security" layer applies security functionalities directly on "named" data. Names in NDN are hierarchical and may be similar or not to URLs. In NDN subscribers send "Interest" messages to request information objects which arrive in the form of "Data" messages (Fig. 2).

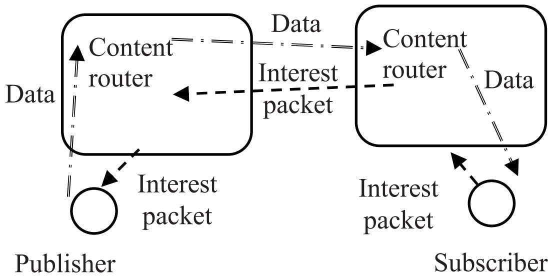

Fig. 2. NDN architecture

All messages are forwarded by Content Routers (CRs). Each "Data" message contains a signature (name and the information included in the message) and information about the public key (to produce the signature). This allows any node and CRs to verify that the information comes from an authorized source. NDN supports on-path caching.

Information objects in PURSUIT are identified by a unique pair of IDs: the "scope" ID (related information object) and the "rendezvous" ID (identity for a particular piece of information). PURSUIT can support both on-path and off-path caching and the Packet Level Authentication (PLA) technique for encrypting and signing individual packets.

The SAIL architecture combines elements present in the NDN and PURSUIT approaches and can even operate in a hybrid mode. Information object names in SAIL provide structure and can be hierarchical, but they do not carry location or organizational information. Hash values are included in names, which allows self-certification. In addition to on-path caching at the Content Routers (CRs) SAIL, has the deployment of large scale information object caching and replication mechanisms.

The core component of the COMET architecture is a Content Mediation Plane (CMP) which mediates between the network providers and the information servers. Information names are provided by a Content Resolution System (CRS) when the information is registered by the publishers. COMET adopts security techniques from other ICN architectures and on-path and off-path caching.

In CONVERGENCE object names consist of a "namespace ID" and a "name part", whose format is determined by the "namespace ID". The default format of CONVERGENCE names is similar to DONA, hierarchical names may also be used as in NDN or URLs. Each "Data" message contains a digital signature.

Functionalities implementation in ICN architectures

Functionalities / ICN	DONA	NDN	PURSUIT	SAIL	COMET	CONVERGENCE
Naming	Flat	Hierarchical	Flat/hierarchical	Flat	Unspecified	Flat/hierarchical
Caching	On-path caching at resolution handlers Off-path caching requires additional registrations		On-path caching Off-path caching requires additional registrations			On-path caching at content routers Offpath caching relies on unspecified name resolution system
Mobility	Subscriber mobility via new requests				Specialized mobility-aware content routers at access network	Subscriber mobility via new requests
Security	Principal is hash of public key	Signatures included in packets	Packet level authentication Names can be self-certifying	Names can be self-certifying	Unspecified	Signatures included in objects

The ICN architectures have similarities and differences, as shown in Table.

Naming. One of the major characteristics of all ICN architectures is the structure of name for a content that makes content communicable over the network. The common point among all the ICN architectures is that the content names are all independent of location. On the other side, the point of difference among the ICN architectures is that the names can be flat or hierarchical; or the names can be human-redeable or not [3].

Name resolution and data routing. The concept of name resolution is process of finding a provider or source that would be able to supply the requested named information. Data routing includes building a path from the provider to the requested information deliver. The major issue here is that whether name resolution and data routing are integrated or decoupled. If these two functions are integrated, the request for information will be routed to the information provider first. Then, the provider sends the information to the host asking for information by going through the reverse path on which the request has been forwarded. On the other hand, the first function that is name resolution does not define or limit the path over which the data will be delivered from the provider to subscriber [3–4].

Caching. There is a difference between on-path caching and off-path caching. If on-path approach for caching is used, the network employs cached information as well as the path through which name resolution request is delivered. On the other hand, the network in off-path caching uses the information that is cached outside the path. If name resolution and data routing are decoupled in the ICN architectures, the name resolution system (that is responsible for managing caches as normal publishers of information) must be able to support off-path caching method. If integrated name resolution and data routing are utilized in the network, off-path caching is needed to be supported by the routing system [4] *Mobility.* Mobility of subscriber is a matter that is naturally supported by all ICN architectures. The reason is that mobile subscribers are only able to send new information subscriptions after a handoff. It is more troublesome to support mobility of publisher because if the integrated name resolution and data routing is used, the name resolution system must be updated; and if the decoupled approach is used, the routing tables are needed to be updated [5–7].

Security. The issue of security is highly in relation with naming structure. If human-readable names are used, an entrusted agent (or a trust association with the name resolution system) is required to ensure that returned content matches the requested name. On the other hand, if flat names are used, self-certification is supported. However, flat names are no human-readable. As a result, an entrusted system is needed to map human-readable names into flat one [5–7].

Conclusion. In modern Internet architecture, there are certain problems that require a fundamental rethinking of how the Internet should work. Several ICN architectures have already been proposed, to address such issues as information delivery and mobility support. Although the field of ICN-related research is still being formed, a set of core ICN functionality has been defined, such as naming, name resolution and data routing, caching, mobility, and security, which compared the functionality of ICS projects. Thus, ICN is a promising research area that requires further research to assess potential benefits and productivity.

References

1. Al-Naday M. F., Thomos N., Reed M. J. Information-centric Multilayer Networking: improving performance through an ICN/WDM architecture. *IEEE/ACM Transactions on Networking*, 2006, no. 25 (1), pp. 83–97.

2. Anand A., Gupta A., Akella A., Sesha S., Shenker S. Packet caches on routers: the implications of universal redundant traffic elimination. *ACM SIGCOMM Computer Communication Review*, 2008, vol. 38, no. 4, 2008, pp. 219–230.

3. ANR Connect project. Available at: http://anr-connect.org/ (accessed 23.03.2017).

4. Arianfar S., Nikander P., Ott J. On content-centric router design and implications. *Proceedings of the Re-Architecting the Internet Workshop*, 2008, p. 5.

5. Barakabitze A. A., Xiaoheng T., Tan G. A Survey on Naming, Name Resolution and Data Routing in Information Centric Networking (ICN). *Int. J. Adv. Res. Comput. Commun. Eng.*, 2008, no. 3 (10), pp. 8322–8330.

6. Bari M. F., Chowdhury S. R., Ahmed R., Boutaba R. A survey of naming and routing in information-centric networks. *IEEE Communications Magazine*, 2012, no. 50 (12), pp. 44–53.

7. Bloom B. H. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 2008, no. 13 (7), pp. 422–426.

8. Carzaniga A., Wolf A. L. Content-based networking: A new communication infrastructure. *Workshop on Infrastruture for Mobile and Wireless Systems*, 2006, Springer Berlin Heidelberg, pp. 59–68.

9. Cerf V. G., Icahn R. E. A protocol for packet network intercommunication. *ACM SIGCOMM Computer Communication Review*, 2005, no. 35 (2), pp. 71–82.

10. Choi J., Han J., Cho E., Kwon T., Choi Y. A survey on content-oriented networking for efficient content delivery. *IEEE Communications Magazine*, 2001, no. 49 (3), pp. 121–127.

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 Natallia Vladimirovna – PhD (Engineering), Associate Professor, Assistant Professor, the Department of Information systems and technologies. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: n.patsei @belstu.by

Received 25.04.2017