

称号及び氏名 博士（工学） **Kevin Pognart**

学位授与の日付 平成 29 年 3 月 31 日

論 文 名 「 **Design and Evaluation of the Keyword-Based
Information-Centric Networks** 」

論文審査委員 主査 戸出 英樹

副査 本多 克宏

副査 吉岡 理文

論文要旨

Nowadays, the Internet is mainly use to share contents via social networks and files sharing websites such as YouTube, Spotify or web storage solutions. However, the current Internet has limited performances due to its architecture. The current Internet is based on a host-to-host architecture which is very efficient for the communication between two machines, but not optimal for today users' needs. In this context, new network architectures are studied to provide better performances, and Information-Centric Networking (ICN) is a very hot topic among those.

The two main ideas in ICN are to focus on content identifiers instead of network addresses to route requests, and to perform content copy caching at some nodes. The inspiration comes from peer-to-peer system such as BitTorrent where all users contribute to the content sharing efficiency. In ICN, content copy shares the same identifier as the original content. Hence, when a request is send into the network, the network will search content corresponding to the identifier requested, even if it is a copy. The goal is to reduce server traffic and to speed the download up, by the caches and the identifier uniqueness between content and its copies. In such environment, another issue appears from a user point of view: the difficulty to know content identifiers. To perform the routing, content identifiers must be unique. This issue is similar to the URL issue in

the current Internet. To know a website address, users use web search engine by using some keywords to retrieve the wanted website address.

Some solutions regarding the keyword-based search feature for CON have been proposed. It is important to add this feature in ICN because the current web search engines use centralized data centers, and they cannot access caches. Hence, some advantages due to basic concepts of ICN are not used. One approach to provide such a feature is to implement a system similar to typical multimedia search engines into ICN. This system searches the contents similar to the content the user includes inside its request. When a search by content name is performed, the search interest is flooded over the network. Each node sharing searchable content performs a feature extraction task: a feature vector containing information about the content characteristics is extracted for each content, and an index is formed for each content type. When a request for similar contents is received by a node, it performs similarity search by comparing the request descriptors (the feature vector of the content requested) against the index to find a set of the most similar contents. When a node has similar contents, a new content is created: it is a collection of corresponding links constituted by a label name for the similar object descriptors and a target name for the content name. An interest is sent to the requester to inform him about its availability. He requests the collection objects from each interest received. Then, data packets carrying the collections of names of similar objects are sent to the requester. This approach seems to be extendable to a keyword-based search feature but it does not seem feasible for a network such as the current Internet because of the flooding of messages all over the network. Another approach consists in giving each keyword (or group of keywords) a numerical keyword ID as in Independent Search and Merge and Integrated Keywords Search. Using these keyword IDs, the system can retrieve the answers for a keyword-based request by finding contents related to the keyword. This information is stored in router tables.

We propose in this thesis some keyword-based search features for Information-Centric Networks (ICN). These propositions are based on Breadcrumbs, an ICN solution, which has attractive features. In Breadcrumbs system, content creates some metadata about itself in nodes on its path. Metadata contains information about content identifier, previous node on the content path, next node, and time information to manage metadata. It is a very interesting ICN solution because content caching and routing systems are simple. In addition, Breadcrumbs can be used as a hybrid solution between the current Internet and ICN by using IP routing at user node or server node, and by using Breadcrumbs system on message path if it is available. In Breadcrumbs (BC), we assume that users and possibly routers have a content cache. Routers

have also a BC table used to route requests. When content passes through a router, a BC entry (metadata) is created in its BC table to indicate the direction of the cached content. If the content goes through a node having a content cache, the content is cached. Requests are firstly sent to a server to download contents by using IP routing. When a request arrives at a router where a BC entry for the same content identifier exists, the request is redirected to follow the direction shown in the BC entry. Each next node will redirect the request according to the direction in BC entries until finding the content in a content cache. If an issue occurs during the redirection, the BC entries are invalidated and the request is forwarded again to the server by IP routing

Hence with our Keyword-based Breadcrumbs (KBC), we adapted and improved Breadcrumbs system to perform keyword-based searches too. Our goal is to add an intrinsic keyword-based search feature to BC system while preserving the BC advantages in terms of simplicity, scalability, feasibility and working. For this purpose, we add elements to BC system to allow two ways of working: the standard working using content name-based request and sending back of content, and a new one using keyword-based request, where KBC entries are used to find other contents in other location than server, and where answers are information about content and not the content itself. The basic idea is to use KBC to find closest corresponding contents. In the initial state, there are no cached contents and no guidance information. When content is downloaded, KBCs are created on-path like in the BC system. The difference appears for a keywords-based request. For the KBC request, the first destination remains a server. If the request reaches a node with one or more KBC entries whose keywords correspond to the requested ones, the request will be replicated. Replicated requests follow their KBC trail while the original request continues its path to the server. Then, when a right content is found, an answer containing the content identifier, its list of keywords and its location is sent back to the requester. By this method, the requester can get a large number of answers with information for choosing the one he wants and if there are several identical contents, he can select the closest one. Also, IP-routing is used for downloading a content found by such a request because the answer gives the content ID and the location, and so performing another BC request for this content ID is unnecessary.

With KBC, we want to take advantage of Breadcrumbs behavior for making our keyword-based search feature similar in its principle to ICN concept: using at its maximum on-path discovery to route request, and avoid relying too much on addresses. Hence, we investigate the possibility of such a system to perform or enhance keyword-based searches for ICN. We also discuss the possibility of combining KBC with other existing keyword-based search

system to enhance the searching performances.

As for the content of the thesis, it is organized as follows.

In Chapter 1, we essentially contextualize our work. This Chapter has two main parts. The first one introduces the main ICN solutions and compares them. We detail the shared ideas to define the common environment in ICN. In the other main part, we focus on Breadcrumbs to explain the reasons why we use it as the basis of our proposition. Also, we present some keyword-based search systems for ICN.

In Chapter 2, we describe the main concepts regarding keyword-based search feature. We focus on Keyword-based Breadcrumbs. This Chapter is organized as follows. We first introduce the different KBC solutions we implemented. We present and explain all the settings used. We explain the differences and the motivations behind those different KBC versions. Then, we discuss about the adaptations made to KBC for avoiding issues, because some fundamental behaviors are different from Breadcrumbs. We finally analyze their performances to conclude on the strong aspects and on the weak aspects of KBC.

In Chapter 3, we introduce the Active Breadcrumbs function into the KBC framework. We describe how to extend the ABC feature into the KBC system, and design the extended system in detail. Through the extensive performance evaluation, we clarify the performance enhancement by combining two different features.

In Chapter 4, we combine smartly KBC with other keyword-based search systems to improve the performances and to give additional results information to the requester. When KBC is used as the only keyword-based search system, results are greatly dependent on the requester neighborhood, but the global efficiency is poor. To counterbalance this situation, we propose to associate cleverly KBC with another keyword-based search system which is configured to have perfect global retrieval efficiency. We present in details the association implemented, with the interaction created between the two keyword-based systems. We discuss on the pros and cons of such an association by analyzing simulation results.

In Chapter 5, we finally conclude this thesis.

審査結果の要旨

本論文は、コンテンツ取得の効率性を追求した将来の情報指向ネットワークにおいて、ユーザが直接所望のコンテンツ名を指定する入力のみならず、キーワードを入力することによりネットワー

ク内のキャッシュに存在する当該コンテンツ集合の情報を取得し、最終的に所望のコンテンツを取得するシステムを実現するための研究であり、以下の成果を確認した。

- (1) 将来の情報指向ネットワークにキーワード検索機能を付与するアーキテクチャを実現した。具体的には、コンテンツ転送経路上に **Breadcrumbs** と呼ばれるログ情報に加えて、経由コンテンツのキーワード情報をも登録し、その後のキーワード検索メッセージの到来に対して適切にそのメッセージを複製・誘導することにより近隣のキャッシュ上に存在する多数の関連コンテンツ情報を取得する **Keyword-Based Breadcrumbs (KBC)** 方式を提案し、詳細に設計した。
- (2) **KBC** 方式を適用することにより、キーワードにマッチする関連コンテンツ情報をサーバ以外の複数の網内キャッシュから迅速に取得することが可能であり、リクエスト送付先サーバ以外のサーバに存在する関連コンテンツの情報も最大 **10** 数%の割合で獲得することが可能であることを明らかにした。
- (3) コンテンツの転送経路上にログ情報を記録するだけでなく、キャッシュをもつノードが拠点となって、キーワード情報を周辺に積極的に配布する機能を加えた **Keyword-Based Active Breadcrumbs (KABC)** 方式を提案し、性能評価の結果、最終的な関連コンテンツ情報の取得性能は **KBC** 方式と変わらない一方で、サーバからではなく、網内に分散されたキャッシュノードから迅速に関連情報を取得できるコンテンツ情報の割合を平均で **60%**増加させることを明らかにした。
- (4) ネットワークワイドなオーバーレイを構築してキーワードに関連するコンテンツ情報を管理し、キーワード検索リクエストに対する網羅性の高い応答を実現する **Independent Search and Merge (ISM)** 方式と、近隣キャッシュからの迅速なレスポンス性能を提供する **KABC** 方式とを融合したキーワードベースのコンテンツ検索・取得システムを提案し、**KBC/KABC** 方式の特徴である、高人気もしくは新規のコンテンツに対するキーワード検索並びに通常のコテンツ名検索に対する迅速な所望コンテンツ取得、及び **ISM** 方式の特徴である、上記以外のコンテンツに対する確実なコンテンツ取得の双方を実現するシステムが構築可能であることを明らかにした。

以上の諸成果は、将来のキーワードに基づくコンテンツ取得を指向するネットワークの設計指針に大いに貢献すると考えられ、本分野の学術的・産業的な発展に貢献するところ大である。また、申請者が自立して研究活動を行うのに必要な能力と学識を有することを証したものである。