Future Generation Computer Systems 36 (2014) 170–179

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

# An adaptive per-application storage management scheme based on manifold learning in information centric networks

Yuemei Xu [a,*], Yang Li [a], Tao Lin [a], Zihou Wang [a], Guoqiang Zhang [b], Hui Tang [a], Song Ci [a,c]

[a] *High Performance Network Lab, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*

[b] *School of Computer Science and Technology, Nanjing Normal University, Nanjing 210097, China*

[c] *Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, NE 68182, USA*

## HIGHLIGHTS

- We introduce a manifold learning method to classify the priority of applications.
- We formulate the per-application storage management problem into an optimized resource allocation problem.
- Dynamic programming is adopted to solve the formulated problem.
- An adaptive storage management scheme (ASM) is proposed on the basis of the optimal solutions.
- The proposed ASM scheme is superior to existing static partitioning and shared storage schemes.

## ARTICLE INFO

*Article history:*
Received 15 March 2013
Received in revised form
27 August 2013
Accepted 5 September 2013
Available online 19 September 2013

*Keywords:*
Per-application storage management
Manifold learning
Information centric network

## ABSTRACT

Information Centric Network (ICN) is an emerging network paradigm centered around the named contents rather than the host-to-host connectivity. The common characteristic of ICN leverages *in-network caching* to achieve an efficient and reliable content distribution but also brings challenges. The *in-network caching* technique equips all ICN routers with cache storage. However, no existing works focus on the cache storage sharing mechanism among different applications to satisfy the line speed requirements and diversity of applications in ICN. In this paper, we formulate the per-application storage management problem into an optimized resource allocation problem and introduce a manifold learning method to classify the priority of applications. Dynamic programming is adopted to solve the formulated problem and an adaptive per-application storage management scheme is proposed on the basis of the optimal solutions. Extensive experiments have been performed to evaluate the proposed scheme and show that our approach is superior to static partitioning and shared storage schemes.



## 1. Introduction

The explosively increasing demand for content has motivated a redesign of the Internet architecture, focusing on named contents rather than the host address where the contents are retrieved. This approach is commonly called the information centric network (ICN) architecture of the next-generation Internet [1]. The ongoing proposals of ICN include CCN [2], DONA [3], NetInf [4], PURSUIT [5], etc. The common characteristic of these ICN architectures leverages *in-network caching* to achieve an efficient and reliable content distribution. Every router is equipped with a content store (CS) module and can store contents passed by it so as to satisfy subsequent requests for the same contents. Users' requests routed towards the original server may be satisfied by an intermediate router with a temporary content copy and thus the experience of users can be improved with less content transport delay.

The technique of *in-network caching* makes the ICN architecture very different from traditional cache networks, such as Web cache, P2P cache and content delivery network (CDN). Most of the traditional cache networks are typically designed for certain applications and limited at specific naming systems, where the same contents may be stored using different names and cannot be shared by various applications. For example, a video file can appear either as a HTTP streaming media cached in Web cache or as a BT file stored in P2P cache. The duplicate content placements in Web and P2P cache lead to a waste of resources. ICN exhibits the potential to solve this problem. Multiple applications can share the same content cached in ICN routers due to all contents being identified, routed and cached by global unique names. Such an application-independent caching mechanism is able to cache one content copy for serving multiple applications.

* Corresponding author. Tel.: +86 10 15110099943.
*E-mail addresses:* xuym@hpnl.ac.cn (Y. Xu), liy@hpnl.ac.cn (Y. Li), lint@hpnl.ac.cn (T. Lin), wangzh@hpnl.ac.cn (Z. Wang), guoqiang@ict.ac.cn (G. Zhang), tangh@hpnl.ac.cn (H. Tang), sci@engr.unl.edu (S. Ci).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.09.016

Nevertheless, the *in-network caching* technique of ICN also raises several challenges. First, ICN routers have *line speed requirements* [6], which require content retrieval and replacement to be operated at line speed. Specifically, routers need to search contents in local storage for every arriving request, therefore the available cache sizes of routers in ICN are affected by the memory access speed and should not be as large as that in CDN or P2P cache. It is crucial to orchestrate the limited cache resources for different applications in order to achieve optimal network performance. Second, ICN is naturally designed to support multiple applications in a shared storage space. However, applications are different in their traffic proportions, object and population sizes and popular distributions [7]. For instance, the average content size of Web type and file sharing type applications are 10 kB and 100 MB, respectively. The caching objectives of Web type and P2P type applications are different: the former leverages endeavor in reducing user's access latency while the latter aims at reducing inter-network traffic. There is an urgent need to design an appropriate storage management scheme to satisfy the diversity of applications.

To address the challenges discussed above, this paper focuses on per-application storage management in ICN and solves the problem of managing the cache resource of each router while considering both the limited cache resources of routers and the diversity of applications. Per-application storage management allows routers to allocate more cache size to applications with high performance requirements, to the detriment of other applications and a potential increase of the overall network profit. It has been observed that applications contribute differently to users' experience: users are more sensitive to a delay of web applications than that of file download applications [8]. An intuitive idea is to allocate relative larger cache sizes to some premium applications and to allocate smaller sizes to the remaining regular applications. However, it is challenging to define the importance of applications and existing works simply assume that the priority of applications is known in advance [9]. For the first time, this paper introduces a nonlinear manifold learning method to define the priority class of applications. We collect multidimensional information of applications as raw data to implement data mining and quantify the data mining results by defining *network gain function* for applications. Then the applications which have higher network gain value would be allocated with more cache size accordingly. The major contributions of this paper are summarized as follows:

- A nonlinear manifold learning method is introduced to classify the priority of applications and formulate the per-application storage management problem as an optimized resource allocation problem.
- Dynamic programming is adopted to solve the formulated problem and an adaptive storage management scheme is proposed on the basis of the optimal solutions.
- Our simulations driven from two topologies show that the proposed adaptive per-application storage management scheme exhibits its advantages in managing storage space and is superior to the existing static partitioning and shared storage schemes.

The rest of the paper is organized as follows. Section 2 presents a survey of the related work and background knowledge. Section 3 describes the problem formulation. Section 4 describes how to use the manifold learning method to define network gain functions for classifying applications, while an adaptive per-application storage management scheme is proposed in Section 5. Section 6 is the simulation evaluation. Finally, conclusions are summarized in Section 7.

## 2. Related work and background

### 2.1. Related work

Per-application storage management tries to solve the problem of allocating cache resources of routers to different applications. There are two kinds of per-application storage management strategies: *static* and *dynamic*. In *static* strategies, each application is statically assigned with a fraction of the shared cache resource. This kind of cache resource allocation is fixed but not adaptive to the applications' changing demands. While in *dynamic* strategies, the cache sizes of applications are tuned adaptively for better serving applications, which overcomes the drawback of static strategies but leads to some overhead caused by information maintenance and algorithm computation. In this paper, we focus on the dynamic per-application storage management mechanism, which adaptively adjusts the cache allocation based on data mining results but does not bring too much overhead. We also discuss the complexity of our approach later.

Storage management has already been studied in the context of Web caching. The studies in [10] implemented storage allocation considering only single applications. Lu et al. proposed an adaptive cache allocation scheme for different services [8], which self-tuned the cache resource allocations of services based on approximate linear differential equations. This scheme [8] guaranteed convergence towards a proportional hit rate differentiation among services, but could not guarantee a minimum hit probability to each service. These previous works [8,10] are proposed in the context of Web caching and cannot be applied in ICN architecture, where multiple applications share a common cache space and the per-application storage management is related to many factors, such as the traffic proportion of applications, content sizes of applications, etc.

Carofiglio et al. were the first attempt to explore the storage management in ICN and proposed two *priority* and *weighted fair* storage management algorithms [9]. Carofiglio et al. assumed applications have different priority or weight values. Contents belonging to applications with lower priority or weight values should be removed to make room for incoming contents belonging to applications with higher priority or weight values. Actually, the *priority* and *weighted fair* schemes are cache replacement polices instead of cache allocation schemes and do not discuss how to define priority or weight values for applications. We depart significantly from the above studies [8,10,9] by formulating the per-application storage management into an optimized resource allocation problem. We also differentiate applications and define network gain values for applications by using a data mining method and then we heterogeneously assign cache storage to applications based on the defined network gain values, aiming at maximizing the network performance (e.g., request hit ratio).

A nonlinear manifold learning method is adopted to perform data mining. There are many learning-based data mining algorithms, such as $K$-Means [11], CURE [12], etc. However, most of the learning-based algorithms need to know some a priori information (e.g., locations of nodes) and heavily rely on algorithms' initial values. Furthermore, most of the learning-based algorithms use Euclidean distance for similarity calculation. In this paper, we are unable to get a priori knowledge of cache size allocation and can only collect certain raw data from network measurement variables, which is multidimensional and nonlinear. It is well known that the nonlinear manifold learning method has been widely used for dimensionality reduction without explicitly setting the initial values of algorithms or knowing the network information [13], therefore it is chosen to implement data mining in this paper.

### 2.2. Nonlinear manifold learning

Nonlinear dimensionality reduction (NLDR) is an effective method to discover meaningful low-dimensional structures hidden in high-dimensional data. Often, the high-dimensional data observed in the real world are the consequences of a small number of factors. Manifold learning is one of the NLDR approaches.