Social-Based Cooperative Caching in multi-hop wireless networks

Kothagundla Divya, S. Suresh

Dept. of CSE, Sree Rama Institute of Technology & Science, SKuppenakuntla, Penuballi, Khammam, AP, India

Abstract—Data access is an important issue in Delay Tolerant Networks (DTNs), and a common technique to improve the performance of data access is cooperative caching. However, due to the unpredictable node mobility in DTNs, traditional caching schemes cannot be directly applied. In this paper, we propose DAC, a novel caching protocol adaptive to the challenging environment of DTNs. Specifically, we exploit the social community structure to combat the unstable network topology in DTNs. In such context, we consider a multi-hop wireless network adopting CCN-like cooperative caching, in which each user terminal acts also as a caching node. We propose an interest based insertion policy for the caching, based on the concept of “social-distance” borrowed by online recommendation systems, to improve the performance of the overall network of caches; the main idea is to store only the contents which appear to be of interest for the local user. We show that our proposed scheme outperforms other well-known insertion policies, that are oblivious of such social-distance, in terms of cache hit probability and access delays.

I. Introduction

Mobile networks have been experiencing an impressive growth in the data traffic, mainly due to multimedia application; indeed, 2/3 of mobile traffic is expected to be video by 2016 [1]. Huge amount of this traffic is due to user-generated content and is distributed via popular Internet services, often based on an explicit, user-driven social network (e.g. Facebook and YouTube). This fast growth in traffic has been imposing a significant burden on the current wireless infrastructure, which must be periodically upgraded, increasing the profit gap between the network operators and the Over-The-Top content providers. Indeed, network operators continue to invest into the infrastructure to cope with the increasing traffic, while users spend more money for contents or cloud based services than for the network services. This dichotomy motivates new network paradigms in which the user’s wireless terminals cooperate to decrease the load in the wireless network infrastructure: the Quality of Experience of the user is improved, even if the network infrastructure has not been upgraded. The second factor is the content-reuse, since it is well known [2] that in many contexts a large amount of traffic is caused by few popular contents (e.g., videos, music, apps, software updates, etc.). These popular contents are requested many times after being generated. This fact advocates the use of caching techniques to distribute popular content across the network and to avoid the access at the server storing the original copy of the content. The third factor is the content localization, since the interest for some contents is spatially localized due their specific nature. For example, tourist/event information’s, local news, shop advertisements show a clearly localized region of interest. This factor motivates the cooperation among nodes that are in proximity, since their spatial position increases the content reuse. To enable communications among neighboring nodes, without any infrastructure, layer-2 technologies (like Bluetooth [3] and WiFi Direct [4]) are supporting Device-to-Device
(D2D) communications and enable peer-to-peer capabilities among terminal nodes, thanks also to specific middleware’s (like AllJoyn [5]) to ease the development of applications. Finally, modern mobile devices like smartphones and tablets are equipped with large storage capacity of many gigabytes. The storage capacity can be considered as a free resource nowadays, which the user is willing to share more preferably more than any other resource since it is not affecting directly its Quality of Experience when running applications. All the previous factors advocate the adoption of cooperative caching techniques among neighboring nodes, exploiting D2D communications, multi-hop communications and the available free storage to share among the users. The large storage capacity enables each mobile device to act as a caching node of a wireless CCN; each node stores all the received contents (also the ones to be forwarded in a multi-hop fashion) in its cache. Contents are distributed across the nodes and a user can hopefully access the desired content in its proximity. Thus, cooperative caching can reduce both the delay to access the content (with satisfaction for the user) and the network load (with satisfaction for the wireless operator). The effectiveness of the approach strongly depends.

II. Related Work

We consider a hybrid network scenario which includes Access Points (APs) and mobile nodes. The APs providing data downloading service are located in the Internet infrastructure, which only cover a small fraction of the network area. The connections between mobile nodes and APs are intermittent and opportunistic. A mobile node can directly download data when it has a connection with an AP. When the node has no direct connection with any AP, it can only download data via a multi-hop relay provided by other mobile nodes1. We aim to design a cooperative caching protocol in such a scenario, to let some nodes cache the data and act as the source for future requests. In this way, the performance of data access can be significantly improved. Due to the contact duration limits, a complete data item may not be transmitted during a contact. If the data item is simply fragmented into multiple consecutive native packets, to recover the original data, a node has to collect all the native packets. This suffers from the well-known coupon collector’s problem [5]. Specifically, suppose there are s distinct native packets and each can be collected with an equal probability. Then it has been proven that the expected number of packets a requester needs to collect to accumulate all s distinct packets is on the order of \( \sim (s \log s) \). To mitigate the problem caused by the simple fragmentation, we adopt the random linear network coding [6]. At first, the AP cuts the data item \( P \) into s uniformly sized native packets, \( P = \{p_1; p_2 \ldots \} \), and randomly generates a set of coding coefficient vectors \( \sim j = (j_1; j_2 \ldots \ldots js) \) from a Galois Field to compute the random linear combination of the native packets: \( p_0 j = Ps_{i=1} \_jipi \). Note that all generated coded packets are also uniformly sized. Then the AP sends these coded packets to the data requester via direct contact or by multi-hop relay. Unlike erasure coding which only allows the data source to encode packets, random linear network coding also allows intermediate nodes to perform encoding.

III. CACHING PROTOCOL DESIGN

In this section, we analyze the effects of contact duration on cooperative caching, and propose a distributed contact Duration Aware Caching (DAC) protocol from a social network perspective. Since the nodes in the same community have a higher probability to share their cached data, we exploit this social property to determine the caching solution within each community. Also, it is impractical to maintain the global knowledge of the network at every node, but community based caching only depends on each node’s local knowledge about the community. For simplicity, we assume that
the requester can be any node in the community. A. Main Idea To increase the caching efficiency, the nodes with larger potential to contribute the cached data to others deserve higher priority to be selected as the caching node. When evaluating the potential, traditional caching solutions all assume that the data requester can always get the complete requested data item from the caching node when a connection arises between them. Thus, these solutions take the complete data item as the caching unit; i.e., a node either caches the complete data item or does not cache any of it. However, this assumption does not hold in DTN scenarios. Due to the contact duration limits in DTNs, the amount of data that can be transmitted during a contact is restricted. Under this condition, the caching efficiency is reduced if a node caches too many packets, since the limited contact duration prohibits it from transmitting all of them. shows an example to illustrate the impact of the contact duration limits on caching. Node A is requesting a data item with eight packets, and it will sequentially contact nodes B, C, and D when it moves. In the complete data item is taken as the caching unit. Node B has been selected as the caching node. Under this caching solution, node A can only retrieve five packets from B due to the contact duration limitation. As a result, the remaining three packets cached in node B have not been utilized, and A’s contact opportunities with node C and D have been wasted. can let node B, C, and D to cache five, two, and one packet, as shown in node A can retrieve all eight packets. Under this caching solution both the storage buffer and the contact opportunities are better utilized. As can be seen, in the contact duration limited scenarios, rather than taking the complete data as the caching unit, the problem of how many packets to cache at each node should be carefully determined, to improve the caching efficiency. We model the social relationship among users as “degree of similarity” (i.e., how much the users share common interests) and their interest for a specific content by “degree of interest”. To evaluate quantitatively such values, we adopt a social space model. We start to describe it with a toy example, and then we will generalize the model.

As toy example, assume that all the contents in the catalog are books, which can be classified according to their “history” flavor. Any book is identified by a point on a segment [0; 1] whose position represents the level of “history” present in the book, as shown in Fig.2. The extreme point 1 corresponds to a pure history book, whereas 0 to a book without any history content. Hence, a small distance between two contents implies a large degree of similarity between them, and vice versa. Similarly, the actual interest of a user for history books is represented by her position within the segment. In this way, a small distance between a user and a content implies a large degree of interest for it, and a small distance between two users implies a high degree of similarity in their interests.

For example, regarding the 3 users u1; u2; u3 in we can claim that u2 and, especially, u3 show high interest toward history, whereas u1 is not interested in this genre. To model all the possible categories of contents, it would be necessary to consider a multi-dimensional space, in which each dimension corresponds to a particular category. After properly defining a norm on such a space, a smaller distance between two points represents a larger degree of similarity (between contents) or interest (by a user for a content). Due to the very large (actually, infinite) number of single categories for human interests, a multidimensional space approach is not practically feasible. we adopt a CCN approach for cooperative forwarding and caching. Whenever the user generates a request, the corresponding node sends a request packet with the identification of the requested content towards the server. The request is transmitted between neighboring nodes in a multi-hop fashion, until it reaches the server. Then, the server replies by sending back the content data, which reaches the requester’s node in a multi-hop fashion. Note that in the case of more
generic topologies (not considered in this work), the requests can be propagated through some standard controlled flooding protocols, whereas the data content is sent back to the requester through a single path, discovered during the initial request phase. Each node is equipped with a local cache, denoted by Cache Storage (CS) and a Pending Interest Table (PIT) for storing requests for which the node has no content. When a node receives a content, this is forwarded back to the nodes for which it has stored requests in its PIT. We denote the overall cooperative caching approach we propose as “SOCIAL-CACHE-Rs”, where Rs is a numerical parameter. Referring to the pseudo-code, when a content is received at a node, it is eventually cached according to to store only those contents that would be of possible interest for the corresponding user. Thanks to the social space model, the insertion policy at user u’s node simply consists of storing a content c if their social distance ds(c; s) is below a threshold Rs. Furthermore, if the cache is already full, Least Recently Used (LRU) replacement policy is adopted. If a request packet arrives at a node which does not have the requested content in its cache, a “miss” is experienced; then the request is stored in the PIT and forwarded to the neighboring nodes. Instead, if the requested content is present in the cache, a “hit” is experienced and the content is sent back to the requester in a multi-hop fashion. Our social-aware policy caches only the most requested (i.e., most popular) contents for each user, based on the threshold Rs. The actual number of such contents varies for each user since it reflects the random distances between the user and the different contents. Note that Rs allows a simple implementation since it does not require to know in advance the overall ranking of all the contents, which is a priori unknown for a user, and could be only measured a posteriori. This social threshold is instead based on a different metric, which is the expected interest for the content, deduced with different approaches, as the ones used in recommendation systems. In this work we do not investigate such approaches, and we assume that each node is able to evaluate the social distance between any content and the corresponding user. our novel “social-aware” insertion policy. Caching is an important technique to enhance the performance of the both wired and wireless network. A number of studies have conducted to improve the caching performance in wireless mobile environment [9-12]. Cooperative caching has been studied in the web environment, but little work has been done to efficiently manage the cache in ad hoc networks. Due to mobility and constrained resources (i.e., bandwidth, battery power and computational capacity) in wireless networks, cooperative cache management techniques designed for wired networks may not be applicable to ad hoc networks. In the context of the ad hoc networks, it is beneficial to cache frequently accessed data not only to reduce the average query latency but also to save wireless bandwidth. Hara [5] proposed several replica allocation methods to increases data accessibility.

IV. Conclusion

In this paper, we have considered a CCN paradigm applied to multihop wireless networks exploiting D2D communications. We have proposed a new social-aware insertion policy. We identified the effects of the contact duration limitation on cooperative caching in DTNs. Our theoretical analysis shows that the marginal caching benefit that a caching node can provide diminishes when it caches more data. Based on this observation, we have designed a contact Duration Aware Caching (DAC) protocol, which exploits social network concepts to address the challenge of the unstable network topology in DTNs. Trace-driven simulations show that by adopting DAC, the performance of data access can be significantly improved. To the best of our knowledge, this is the first paper to identify the effects of contact duration on caching in
DTNs. As the initial work, we do not expect to solve all the problems. In this paper, we have addressed the problem of where to cache and how much data to cache, given a fixed number of replicas for a data item. As future work, we will consider how to determine the optimal number of replicas for each data item. This problem is intuitively related to the following four factors: 1) data access pattern, 2) data size, 3) node mobility pattern, and 4) storage limitation. Also, we will deal with new challenges of considering multiple data items: 1) How to determine the priority of the data items when they compete for the limited contact duration and caching space? Intuitively, the priority of a data item is determined by its popularity and redundancy in the community policy for the cache that admits into the cache only contents for which the corresponding user.

References


Ms. Kothagundla Divya is a student of Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, AP, India. Presently she is Pursuing her M.Tech (CSE) from this college.

Her area of interests includes Information Security, Cloud Computing, Data Communication & Networks.

Mr. S. Suresh well known author and excellent teacher. He is working as a H.O.D for the Department of M.Tech., Computer Science & Engineering, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam. He has vast teaching experience in various engineering colleges. To his credit couple of publications both National & International conferences / journals. His area of Interest includes Data Warehouse and Data Mining, information security, Data Communications & Networks, Software Engineering and other advances in Computer Applications. He has guided many projects for Engineering Students.