Tube Caching: An Effective Caching Scheme in Content-Centric Networking

Hao Wu[†] Bin Liu[†] Yang Li[†] Huichen Dai[†] Weiyi Zhang[‡] Yi Wang [†¶]
[†] Tsinghua University [‡] AT&T Labs Research [¶] Huawei Future Network Theory Lab

I. Introduction

Content-Centric Networking (CCN) is regarded as a promising architecture since it has the built-in caching capability improving content delivery[1][2][3]. However, compared with the exponential increase of Internet content, the storage capacity of a content router is so limited that we need effective caching schemes to decide which contents are qualified to be stored, where to store them, how long should they be hosted and how should the cache objects be evicted when a new object needs to be stored. In this work, we investigate *the effective caching in CCN, focusing on reducing an ISP's energy consumption*. Our motivation is straightforward: if the data chunk can be sent from a local cache rather than the remote content source, the transmission energy can be reduced with some extra energy consumption for storing replicas in cache. As we will show, the caching energy is generally much less than the reduced transmission energy, which implies that CCN built-in caching capability favors energy-efficient content distribution.

II. CCN Caching Model and Energy Consumption

A CCN router is equipped with cache that can store contents to serve the future requests. We study an ISP with multiple gateways and consider the contents stored outside the ISP's network. With a fixed gateway router, a content's requesting paths originated from access routers form a forwarding tree rooted in the gateway.

Our objective is to reduce the energy consumption of an ISP with the help of cache. Let $\mathscr{R}$ be the set of content routers in ISP, $\mathscr{O}$ be the content category. We divide the continuous time into a series of identical time slots, each of which lasts for τ. $t_m (m = 0, 1, 2...)$ represents the m-th time slot. During a single time slot, the request arrival rate remains constant. Define $\lambda_{k,i}(t_m)$ as the request arrival rate for content O_k in router R_i in t_m, and s as the size of the fixed-size chunk in CCN. The binary indicator $x_{k,i}(t_m)$=1 means O_k is cached in R_i in t_m, otherwise $x_{k,i}(t_m)$=0. Energy consumption during $[0, T]$ is comprise of two main components: the transmission energy $E^{tr}(T)$ and the caching energy $E^{ca}(T)$.

This work is sponsored by Huawei Innovation Research Program (HIRP), NSFC (61373143, 61432009), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20131019172), 863 project (2013AA013502), China Postdoctoral Science Foundation (No. 2015T80089), and Jiangsu Future Networks Innovation Institute: Prospective Research Project on Future Networks (No. BY2013095-1-03). Corresponding author: Bin Liu (liub@tsinghua.edu.cn).

We consider CCN over optical network in this work. As for the transmission framework in the optical layer, we utilize the currently most representative <ROADM + DWDM> solution. In this solution, one optical link consists of two ROADMs (Reconfigurable Optical Add-Drop Multiplexer), one DWDM (Dense Wavelength Division Multiplexing) link and several EDFAs (Erbium-Doped Fiber Amplifier). Thus, $E^{tr}(T)$ is expressed as:

$$E^{tr}(T) = \sum_{m=0}^{n} \sum_{O_k \in \mathscr{O}} \sum_{(u,v)\in \mathscr{E}} x_{k,uv}(t_m) \cdot s \cdot e^{tr}_{u,v} \quad (1)$$

$$\begin{aligned} e^{tr}_{u,v} &= 2(P_R + P_{ROADM}) + P_{DWDM} \\ &+ \left(\left\lceil \frac{l_{u,v}}{80} - 1 \right\rceil + 2\right) \cdot P_{EDFA} \end{aligned} \quad (1a)$$

where $x_{k,uv}(t_m)$ is the frequency O_k has been transported on link (u, v) in t_m; $\mathscr{E}$ is the set of links of the topology; $e^{tr}_{u,v}$ is the transmission energy efficiency (Joules/bit) of link (u, v). In Eq. (1a), P_R is the energy efficiency of content routers. $P_{ROADM}, P_{DWDM}, P_{EDFA}$ are the energy efficiency of ROADMs, DWDM links and EDFAs respectively; $l_{u,v}$ is the geographical distance in KM of link (u, v) and the factor of $(\left\lceil \frac{l_{u,v}}{80} - 1 \right\rceil + 2)$ represents the number of EDFAs installed on link (u, v), for the span of a single EDFA is 80 KM. Given transmission energy occupies the main proportion in network energy consumption, the router is supposed to take it as the **primary benefit** gained from storing a certain content replica locally.

On the other hand, the caching energy $E^{ca}(T)$ is described as:

$$E^{ca}(T) = P^{hold}_{ca} \cdot s \cdot \tau \cdot \sum_{m=0}^{n} \sum_{O_k \in \mathscr{O}} \sum_{R_i \in \mathscr{R}} x_{k,i}(t_m) \quad (2)$$

where P^{hold}_{ca} is the power efficiency (Watts/bit) of the cache medium in routers. $E^{ca}(T)$ is roughly proportional to the total occupied cache capacity within the ISP. Without sacrificing the caching performance, we try to evict the inactive replicas from cache medium by diluting their caching benefit to reduce the caching energy.

III. Tube Caching

Guided by the design insights derived in §II, we propose Tube Caching, a heuristic scheme that makes caching decisions based on the weighted benefit in terms of energy consumption.

Let's start from an access router U_j. We define Req_k as the request for O_k, and $POP_{k,j}$ as the amount of Req_k reaches U_j, namely the popularity of O_k in U_j. Let $E^{tr}_{k,U_j} = \sum_{(u,v)\in path_{k,U_j}} e^{tr}_{u,v} \cdot s$ be the transmission energy consumed to deliver $Chunk_k$ from the gateway router connecting to source O_k to U_j, where $path_{k,U_j}$ is the set of links on the delivery path from the gateway to U_j with respect to O_k. From the viewpoint of energy, the benefit of caching O_k in U_j for an expected period T is $E^{tr}_{k,U_j} \cdot POP_{k,j} - T \cdot s \cdot P^{hold}_{ca}$. For simplicity, we consider P^{hold}_{ca} to be the same for all routers. Thus, the energy consumed for holding contents in cache is fixed and we can simply compare the weighted transmission energy for different content when making caching decisions. We define the caching benefit of O_k in U_j as: $b_{k,j} = E^{tr}_{k,U_j} \cdot POP_{k,j}$, which is taken as the quantitative criterion of cache admission and replacement.

Tube Caching treats a content's tube-like forwarding path as a whole cache space and explores the position with the maximal caching benefit to place contents. Every router holds a priority queue recording 3-tuples <content identifier, caching benefit, popularity> for the candidate contents. We add the 3-tuples $< rid, b_{max}, if\_empty >$ into the interest packet. For the routers traversed by Req_k, we use rid and b_{max} to record the ID and the corresponding caching benefit of the maximal beneficial router for O_k. Accordingly, when the chunk returns, it has a field rid_{ca} recording the selected router to store a replica. When R_i receives Req_k, it first checks local cache. If there is a hit, the router drops Req_k and returns the matched chunk. If cache misses, in the case that either R_i has available cache room or there is a cached item that is less beneficial than O_k, we can choose R_i as the new candidate router if $b_{k,i}$ is larger than b_{max}.

When $Chunk_k$ returns to the ISP from external content source, it first reaches the gateway router. The gateway looks up PIT and picks out the ID of the candidate router for caching the chunk. It then writes the candidate caching router ID into rid_{ca} field and forwards $Chunk_k$. If there is not a valid candidate router ID, it writes *NULL*. When $Chunk_k$ reaches R_i and has not been cached on the passed path, the router checks $Chunk_k.rid_{ca}$. If R_i is the indicated caching router, R_i stores a replica of $Chunk_k$ in cache and sets the cached flag ($Chunk_k.if\_cached$), avoiding the chunk being cached again afterwards.

To evict these inactive replicas and accelerate the renewal process for the new-born popular contents, Tube Caching adds timely dilution method for caching benefit. For every content involved in the router's benefit queue, its popularity is decreased at constant rate. If a cached replica's popularity reaches 0, it will be evicted from the cache. In most cases, it is replaced by the incoming content when the caching benefit drops out of the cacheable zone in benefit queue.

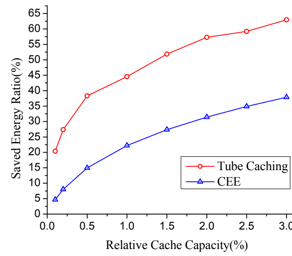

Fig. 1. Saved Energy Ratio

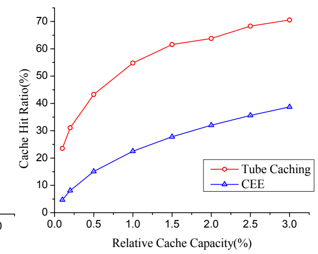

Fig. 2. Cache Hit Ratio

We evaluate the performance of Tube Caching under different router relative cache capacity. We use an AT&T ISP[4] as the network topology that contains 154 nodes and 3 gateways. The size of content category is $M = 100,000$. The overall request arrival rate in every router follows the uniform distribution $U(200,000, 1,000,000)$. Requests in each access router are identically and independently distributed (i.i.d.) within set $\mathscr{O}$, and follow Poisson arrival process. The fixed size of a data chunk is $s = 4$ MB. The dilution rate is dynamically adjusted according to the real-time request arrival rate at every router. Since there is not a generally available prototype of CCN content router, we refer to the energy efficiency of the typical IP-based equipment. The *relative cache capacity* is the ratio of a single router's physical cache capacity to the total content size $M \cdot s$, which is changed from 0.1% to 3.0% in the evaluation.

We compare the Saved Energy Ratio and Cache Hit Ratio of Tube Caching and CEE scheme which is the default caching scheme in CCN/NDN proposal[1]. The results are shown in Fig. 1 and Fig. 2. Clearly, Tube Caching outperforms CEE in both metrics, which show the effectiveness of the proposed scheme.

IV. Conclusion

We investigated the cache allocation and replacement problems in CCN within a single ISP and propose the scheme called Tube Caching that can dynamically distribute contents across the forwarding paths based on the energy-related benefit. Through the preliminary evaluations, Tube Caching has been proven to be effective.

References

[1] V. Jacobson, D. K. Smetters *et al.*, "Networking named content," in *CoNEXT '09*. ACM, 2009, pp. 1–12.

[2] J. Llorca, A. M. Tulino, K. Guan *et al.*, "Dynamic in-network caching for energy efficient content delivery," in *IEEE INFOCOM*, 2013, pp. 245–249.

[3] Y. Wang, B. Xu, D. Tai, J. Lu, T. Zhang, H. Dai, B. Zhang, and B. Liu, "Fast name lookup for named data networking," in *2014 IEEE 22nd International Symposium of Quality of Service (IWQoS)*, May 2014, pp. 198–207.

[4] The at&t isp topology. [Online]. Available: http://totem.run.montefiore.ulg.ac.be/files/examples/att/att.xml