

基于节点社团重要度的 ICN 缓存策略

蔡君¹, 余顺争², 刘外喜³

(1. 广东技术师范学院 电子与信息学院, 广东 广州 510665;

2. 中山大学 电子与信息工程系, 广东 广州 510006; 3. 广州大学 电子信息工程系, 广东 广州 510006)

摘要: 全网内置缓存是 ICN(信息中心网络)架构中最重要的特性之一。为使被缓存的内容对象在空间和时间上分布更合理, 提出了一种基于节点社团重要度的缓存策略(CSNIC)。该策略以社团为单位, 不仅把内容缓存到社团内用户容易获取的节点处, 而且使不同流行度的内容对象在各社团内节点处的时间分布上更合理。在多种实验条件下对 CSNIC 策略进行了仿真验证, 结果表明该策略与 CEE-LRU、Betw-LFU、Opportunistic 相比, 能更好地提升包括缓存命中率、跳数减少率、内容差异性及其替换数量等在内的网络缓存性能指标, 而且, CSNIC 策略的额外开销较小。

关键词: 信息中心网络; 缓存; 替换; 节点社团重要度

中图分类号: TN915.9

文献标识码: A

Caching strategy based on node's importance to community in information-centric networks

CAI Jun¹, YU Shun-zheng², LIU Wai-xi³

(1. School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China;

2. Department of Electronics and Communication Engineering, Sun Yat-Sen University, Guangzhou 510006, China;

3. Department of Electronic and Information Engineering, Guangzhou University, Guangzhou 510006, China)

Abstract: In-network caching is one of the most important features of ICN (information-centric networks). A caching strategy based on node's importance to community was proposed to make content objects cached more reasonable in temporal and spatial distribution. Within each community, the strategy not only caches content objects in those nodes where users are easy to access the content, but also makes different popularity of content objects more reasonable in temporal distribution. The CSNIC strategy was implemented under a variety of experimental conditions, and compare it with the previous strategies CEE-LRU, Betw-LFU and Opportunistic. The simulation results show that the strategy can yield a significant performance improvement, such as, cache hit ratio, hop reduction ratio, content diversity ratio and number of replacement. Furthermore, the additional overhead of CSNIC is small.

Key words: information-centric networks; cache; replacement policy; node's importance to community

1 引言

为了适应互联网应用由发送者驱动的端对端通信模式向接收者驱动的海量内容获取模式的转变, 同时增强网络对安全性、业务质量、移动性、可扩展性等方面的支持, 研究者们近年来提出了一类以信息为

中心的新型网络体系架构, 统称为信息中心网络(ICN, information-centric networking), 典型的如: DONA^[1]、CCN/NDN^[2]、PURSUIT^[3]、COMET^[4]和 GreenICN 项目^[5]。在这类 ICN 网络中, 为缓解当前网络流量的快速增长对网络带宽造成的严峻压力, 每个节点都增加了内置缓存功能。虽然缓存机制也是目前互联网中用

收稿日期: 2014-11-03; 修回日期: 2015-04-29

基金项目: 广东省自然科学基金资助项目 (S2012040007184, 2014A030313637); 广东省教育厅特色创新基金资助项目 (2014KTSCX149)

Foundation Items: The Natural Science Foundation of Guangdong (S2012040007184, 2014A030313637); Guangdong Provincial Department of Education Innovation Project (2014KTSCX149)

于提高网络性能的重要手段之一,缓存理论及相关技术也已经在 Web、CDN 和 P2P 中得到了较为广泛的应用,但目前的缓存策略主要针对某一具体应用,并需配置成一个覆盖网络。因此,这些缓存策略与中间节点共享数据时,效率较低。然而,ICN 中的缓存是网络架构内在的一部分(每个节点都具有该功能),与应用无关,具有网络自适应等特性,因而,传统的面向 Web、CDN 和 P2P 缓存算法也就不能直接应用于 ICN 中。近年来,研究者们对 ICN 中的缓存优化方法进行了探索,得出许多研究成果,主要集中在 2 个方面:一是判断内容是否被节点缓存的缓存决策策略^[6-14];二是单节点缓存内容的替换策略^[15]。

在缓存决策策略方面:ICN 原始提案实行的是处处缓存(CEE, cache everything everywhere)策略^[6],即所有的内容对象被要求缓存于去往目的地途中的所有节点。这直接降低了缓存系统所能缓存内容的多样性,导致了严重的缓存浪费,因为请求被任意中间节点响应后而不会再往上游节点转发,其上游某些节点缓存的内容可能根本没有机会响应后来的请求而因缓存空间受限被替换。最近,Chai 等^[7]提出了基于介数(betweenness),即 Betw 策略的选择性缓存机制,只将对象放置于 ICN 中兴趣分组(interest packet)沿途中介数最大的节点,即对象仅缓存于最重要的节点处。与 CEE 策略相比,这种策略具有较高的缓存命中率,并且减少了节点的替换次数,但由于内容请求的社团性质,依然存在局限性,即大部分的内容缓存存在全网介数相对大的部分节点处,导致内容对象在空间分布上不合理,网络性能亦随之降低,主要体现在:1) 大部分的网络流量与介数大的节点相关,从而降低了网络的负载容量;2) 高介数节点的替换次数大幅增加,增加了节点的计算开销,因而不能满足 ICN 节点线速执行的要求;3) 由于网络介数是描述节点网络全局重要度的指标,高介数的节点处并不一定是网络中大部分用户最容易访问的位置。针对以上问题,基于 Internet 网络拓扑结构的社团特性^[8]——社团内部节点之间连接相对紧密,社团之间的节点之间的连接则相对稀疏,所以把内容分别缓存于其经过的各社团内节点社团重要度最大的节点处。因为在同一社团中,社团重要度大的节点不仅越容易被社团内的节点访问,而且也越容易被社团外的节点访问。这样做不仅能提高网络的缓存效率,节约网络的缓存空间,而且能使网络内容在网络空间上的分布变

得更加均匀和合理。

在缓存替换策略方面:为保证 ICN 节点运行于高速的网络环境,目前在 ICN 的诸多缓存策略中^[8,9,17],采用较多的还是简单易运行的 LRU(least recently used)或 LFU(least frequently used)。但是,如果在全网中,所有节点都以内容在该节点处被访问的频率大小或被访问的时间先后为标准,采用相同的替换机制,将直接导致内容对象在时间上分布不合理,即在热门时间里,社团内的每个节点都缓存到相同的对象,这将不利于缓存空间的合理利用。然而,在热门时间过后,该内容对象在各节点上又几乎同时消失,若需访问该内容,又会产生比较大的网络时延,抑制网络的整体性能。为此,本文基于 Internet 网络拓扑结构的社团特性,在每个社团内,每一个节点进行内容替换时,不仅考虑内容在该节点处被访问的情况,还考虑到了该节点在社团中的位置,使同一社团中各节点采用混合的替换机制,即通过节点社团重要度确定替换节点缓存队列中的内容对象位置的概率,如节点社团重要度大的节点,在缓存队列中靠后位置的内容对象被替换的概率大;而节点重要度小的节点,在缓存队列中靠前位置的内容对象被替换的概率大,以达到内容对象在时间上的合理分布。

本文贡献如下:1) 基于复杂网络的节点社团重要度,提出了一种确定 ICN 缓存位置的决策策略,使内容对象在空间上分布更合理;2) 基于复杂网络的节点社团重要度和内容对象在各节点处的访问情况,提出了一种 ICN 节点混合替换策略,使内容对象在时间上分布更合理;3) 在多种实验条件下对 CSNIC 策略进行了仿真验证,结果表明该策略与 CEE+LFU、Betw+LRU 和 Opportunistic 相比,能更好地提升包括缓存命中率、跳数减少率、内容差异性及其替换数量等在内的网络缓存性能指标,并且 CSNIC 的通信开销和计算开销都较小。

2 相关工作

缓存作为 ICN 的一个关键组件,最近受到了很多研究者的广泛关注,而缓存决策策略和缓存替换策略是缓存的 2 个关键点。缓存决策策略用于确定内容对象的缓存位置,在最初的 ICN 原始提案中,实行的是 CEE (cache everything everywhere)策略,即所有的内容对象被要求缓存于去往目的地途中的所有节点,这种处处缓存的机制已被证明导致了

严重的缓存浪费^[10]。为此, Eum 等^[11]提出了一种在对象返回沿途节点中随机选择节点进行缓存的机制, 该机制易于实现, 但未考虑对象的访问频率, 缓存效率不高, 缓存性能的稳定性也得不到保证。Psaras 等^[12]提出了一种基于加权概率的缓存机制, 即对象在返回途中的缓存概率与节点和请求者的距离成反比, 该方法提高了内容在距离请求者更近的节点进行缓存的概率, 但也会增加不同对象在边缘缓存节点的竞争。最近, Chai 等^[7]提出了基于介数 (betweenness) 的选择性缓存机制, 只将对象放置于 Interest 沿途中介数最大的节点, 即: 对象缓存于更重要的节点上从而实现更高的击中率, 并减少替换次数。与此类似, He 等^[13]应用另一种网络全局节点重要度的测度方法确定 ICN 节点缓存对象。而实际上, 这 2 种以节点全局重要度来确定缓存位置的方法并不能使内容对象在空间上分布变得更加均匀和合理, 并导致了许多的问题 (如引言中所述)。Hu 等^[14]提出一种 Opportunistic 缓存策略, 该策略基于内容在每个节点的访问情况和该内容源与节点之间距离确定内容是否被该节点缓存的概率, 由于各节点相互独立, 因而缓存空间的利用率不高。刘外喜等^[15]基于用户的潜在需求, 并结合带宽换缓存的思想, 利用链路的冗余带宽, 将内容分流到相邻节点缓存, 改善缓存效率, 但该方法需要一些开销。在张国强等^[9,17]的综述性论文中, 对目前的信息中心网络的一些缓存策略进行了详细的总结。而本文方法与以上方法的主要区别在于, 本文基于 Internet 网络的社团特性, 从节点社团重要度 (局部重要度) 出发, 把内容缓存于社团内和社团外用户容易访问的节点处, 使内容对象在网络上的空间分布相对均匀, 同时实现缓存资源的充分利用。

缓存替换算法已在传统的 Web 缓存中得到了广泛的研究, Podlipnig 等对其进行较全面的综述^[18]。在目前 ICN 缓存策略研究中, 普遍采用 2 种简单易行的 LRU 或 LFU 替换策略。本文同样采用的是 LRU 策略, 但不同的是, 本文是以节点社团重要度为依据, 在同一社团内, 不同重要度的节点实行概率制的 LRU 策略, 使缓存对象在时间上合理分布。

3 CSNIC 缓存策略

3.1 节点社团重要度定义

本文将基于节点局部重要度——节点社团重要

度确定缓存内容对象的位置和社团内各节点实行 LRU 策略的替换概率。据 Chauhan 等^[19]的研究, 网络邻接矩阵的特征谱能清楚地反映网络中社团的数目, 例如, 由 c 个社团组成的网络, 则该网络的邻接矩阵将有 c 个特征值远大于其他特征值, 这些特征值可以作为量化网络社团结构的重要指标。因而, 网络社团强度定义为: $P = \log \prod_{k=1}^c \lambda_k$, 其中 $\lambda_1, \lambda_2, \dots, \lambda_c$ 表示邻接矩阵特征值中按降序排列的前 c 个特征值。当节点 k 离开网络时, 整个网络的社团结构和邻接矩阵特征值都将随之变化, 即 $P' = \log \prod_{i=1}^c \lambda'_i$ 。所以, 节点 k 对网络社团特性的重要度为 $P_k = P - P'$ 。利用摄动理论可得节点社团重要度的近似解^[20,21], 如式(1)所示。

$$P_k = \sum_{i=1}^c \frac{v_{ik}^2}{v_i^T v_i} \quad (1)$$

其中, c 为网络中社团数目; v_i 表示以网络中的路由器为节点, 路由器之间的物理链路为边构建的邻接矩阵的第 i 个特征向量; v_{ik} 表示特征向量 v_i 中的第 k 个元素。 P_k 值越大, 节点 k 在其所属的社团中越重要, 即社团内外的其他节点访问该节点将越容易。对于 n 个节点, c 个社团的网络, 有 $\sum_{k=1}^n P_k = c$ 。为使测量参数的和为 1, 定义 $I_k = P_k / c$, 满足 $\sum_{k=1}^n I_k = 1$ 。在应用 I 之前, 需预先知道网络中社团数目 c 的值。本文利用网络的频谱特性直接确定网络社团数目^[19]。如果 c 给定, 该方法无需对网络进行社团划分, 避免了复杂的社团划分的计算量, 可以直接描述节点对社团的重要度。

3.2 CSNIC 缓存决策机制

CSNIC 缓存决策机制的原理是: ICN 兴趣分组在经过每个社团时, 都会记录下其所经过的每一个社团中节点重要度最大的节点, 同时在以数据分组的形式返程时将内容对象缓存于这些节点上。下面通过一个简单实例对 CSNIC 缓存决策机制进行分析, 如图 1 所示。如用户 A 需从 S 处获取一个内容, 首先用户 A 通过图 1 实线路径发送兴趣分组, 并记录该路径上所经过的社团以及各自社团内所经历的节点中节点社团重要度的最大值。当搜索到内容对象缓存(或内容服务器)的位置 S 后, 内容对象沿

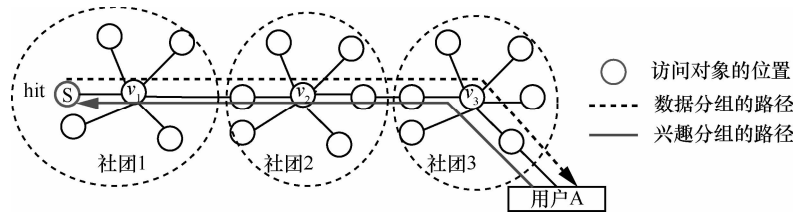


图1 基于节点重要度的决策缓存位置

图1虚线路径(兴趣包的反向路径)返回。在返回过程中,匹配各社团节点重要度的值,若与兴趣包路径保存的节点社团最大值一致,则在该节点处缓存该内容对象,否则,径自通过该节点。如图1所示,数据分组在返回过程中经过了社团1、社团2和社团3,在所经过的每个社团的节点中,节点社团重要度最大的节点分别为 v_1 、 v_2 和 v_3 。因此,将内容对象在 v_1 、 v_2 和 v_3 处缓存。而这些节点正是各自社团内用户容易达到的位置。这样,社团1、社团2、社团3内的用户需再次访问该内容时,直接从缓存该内容的节点处获取即可,与缓存于其他节点相比,该方法实现了内容在网络空间上的合理分布,减少了网络的延时,提高了网络的传输效率。

3.3 CSNIC 替换机制

CSNIC 替换机制的原理是:以网络的社团为一个单位,在同一社团内根据各个节点的节点社团重要度和内容对象在该节点处被访问的时间先后顺序,选择不同的内容对象进行替换,即通过节点社团重要度确定节点缓存队列中的内容对象被替换的概率,如节点社团重要度大的节点,在缓存队列中流行度低的内容对象被替换的概率大;而节点重要度小的节点,在缓存队列中流行度高的内容对象被替换的概率大,以达到内容对象在时间上的合理分布。简而言之,并不是所有节点都使用相同的替换机制。

若ICN节点缓存建模为大小为 C 个对象的队列,以最近最少使用的时间顺序进行排列。假定第 i 个社团内节点 j 的节点社团重要度为 I_{ij} ,在该社团内的平均节点社团重要度为 \bar{I}_{Ci} 。当有一内容对象在该节点需缓存时,将按照概率 $p_{ij}(k)$ 替换该节点缓存队列的第 k 个位置的内容($k \in [1, C]$), $p_{ij}(k)$ 定义如式(2)所示。

$$p_{ij}(k) = \frac{1}{\alpha} \left(\frac{I_{ij}}{\bar{I}_{Ci}} \right)^{\beta k} \quad (2)$$

其中, α 为归一化因子,满足式(3), β 为概率调节系数。

$$\sum_{k=1}^C p_{ij}(k) = 1 \Rightarrow \alpha = \sum_{k=1}^C \left(\frac{I_{ij}}{\bar{I}_{Ci}} \right)^{\beta k} \quad (3)$$

3.4 CSNIC 运行机制

下面对CSNIC缓存策略在ICN架构中的实现过程进行分析。首先,假定每个节点的社团重要度 I 值和每个节点内的每个位置的替换概率是通过离线方式提前计算得到(如通过中心网络管理系统)。当一个用户向某一内容对象发起请求兴趣分组,在兴趣分组中记录下它所经历的社团以及在每个社团中所经历的节点中节点社团重要度的最大值。当兴趣分组达到内容对象所在的服务器或者中间路由器时,这些值将被嵌入数据分组内。在数据分组沿兴趣分组路径逆向传输过程中,每个社团内的路由器匹配自己的 I 值与数据分组的附加值,如果2个值匹配,则该内容被缓存。如果在同一社团内有多个节点的 I 值相同,那么这些节点都将缓存该对象。在确定内容对象需缓存后,再根据该节点缓存队列中各位置的替换概率确定被替换内容对象。从以上分析可知,CSNIC缓存决策策略在执行过程中的开销比较小,每个节点决定是否缓存是相互独立的,完全是基于节点所在社团的重要度的值,既不需要与其他节点交互信息,也无需推断服务器的位置和流量模式。CSNIC替换策略也仅与节点社团重要度和该节点的缓存队列相关。由于在CSNIC策略中,离线方式预先计算的 I 值充分考虑了节点在网络中的位置和与其邻居的关系,因而它又是一种协作式或合作式的缓存策略。在每个社团内其转发请求和内容的伪码如算法1所示。

算法1 (CSNIC 缓存策略)

Content request

- 1) Initialize ($I_{i-\max}=0$); 在兴趣分组中记录第 i 个社团中的节点社团最大值
- 2) foreach (j from 0 to N_i); 在社团 i 中经历 N_i 个节点
- 3) if data in cache
- 4) then send (data)

```

5) else
6)   Get  $I_{ij}$ 
7) if  $I_{ij} > I_{i-max}$ 
8)   then  $I_{i-max} = I_{ij}$ 
9) Forward request to the next hop toward  $N$ 
Content data
...
1) record  $I_{i-max}$  from corresponding content request
2) foreach ( $j$  from 0 to  $N_i$ )
3)   Get  $I_{ij}$ 
4)   if  $I_{ij} == I_{i-max}$ 
5)     then cache (data)
6) forward data packet to the next hop toward  $N_i$ 

```

4 仿真实验与分析

4.1 评价指标

ICN 引入缓存的主要目的为：1)减轻服务器负载，因为一次缓存命中，意味着服务器减少一次用户请求；2)降低用户对内容请求的延迟，借助缓存，用户能快速从邻近缓存位置获取所请求内容，而不是从服务器端；3)减少网络流量和网络堵塞，当缓存命中后，内容请求将经过更少的跳数(hops)到达用户。在达到这些目的的同时，由于 ICN 缓存线速执行的条件，对此提出了以下要求：1)在有限的缓存空间内缓存更多的不同种类的内容；2)减少替换次数。

为量化以上 ICN 缓存的目的和要求，引入缓存命中率(CHR, cache hit ratio)对目的 1)进行评估，其定义为：由缓存而不是由服务器端响应用户请求的概率。定义跳数减少率(HRR, hop reduction ratio)对目的 2)和目的 3)进行评估，其定义如式(4)所示。

$$HRR(t) = \frac{\sum_{r=1}^R h_r(t)}{\sum_{r=1}^R H_r(t)} \quad (4)$$

其中， $H_r(t)$ 和 $h_r(t)$ 分别表示在 $[t-1, t]$ 时间段内，用户从服务器和缓存位置获取内容 f_r 所需要的跳数。若网络中没有缓存，则 $H_r(t) = h_r(t)$ ，此时的HRR恒为1。定义内容差异性(CDR, content diversity ratio)对要求 1)进行评估，其被定义为缓存中所有内容的种类数量与网络中由服务器所产生的内容种类总数的比值，CDR值越大，则在相同缓存空间

内所缓存的内容种类数越多。定义替换数量(NR, number of replacement)对要求 2)进行评估，其定义为一个兴趣分组在一个节点上引起的替换次数，由于在 ICN 中缓存要求线速执行，因此频繁地替换缓存内容并不合适。

4.2 ICN 缓存模型

ICN 缓存模型构建方式如下。首先，假定网络具有发布/订阅的网络架构，即网络中的内容请求和分发机制已经存在。用图 $G=(V, E)$ 表示具有 N 个节点、 M 条边的无向无权网络，其中， $V=v_1, v_2, \dots, v_N$ 表示网络中节点的集合， $E=e_1, e_2, \dots, e_M$ 表示网络中边的集合。 $F=f_1, f_2, \dots, f_R$ 表示网络中的内容的集合， $S=s_1, s_2, \dots, s_p$ 表示网络中内容服务器的集合，并且每个内容服务器与一个节点 $v \in V$ 相对应。内容随机地分布在服务器 S 中，并假定每个内容对象仅保存于一个服务器中。

假定到达网络中的内容请求间隔符合指数分布，对 $r(1 < r < R)$ 内容单元请求的过程服从平均速率为 $\lambda = \sum_{r=1}^R \lambda_r$ 的泊松分布，其中 λ_r 为对内容 f_r 请求的发生率。假定网络内容的需求频率符合 Zipf 分布，即内容需求的频率与内容的排名成反比例关系，内容排名越小，用户对它的需求越高。若 M 表示内容的类别， i 表示内容流行度的排名($1 \leq i \leq M$)，则用户对排名为 i 内容的请求频率为 $P(X=i) = (i^{-\alpha})/C$ ，其中 $C = \sum_{j=1}^M j^{-\alpha}$ 。当内容请求在内容分发的路径上找到了与之匹配的内容时表示一次缓存命中，否则，表示一次缓存未命中。在缓存未命中的事件中，内容请求遍历整个内容分发路径直至内容服务器。假设每个节点的内容单元具有相同的大小。在一特定时间内，每个缓存存储器的一个缓存槽仅可容纳一个内容单元。

4.3 实验结果

实验网络的拓扑由 100 个节点和 386 条链路组成^[22]，其中描述网络的社团特性的模块度 $Q=0.794$ ， $Q>0.3$ 表明网络具有明显的社团特性^[23]。本文采用多个性能参数将 CSNIC 策略与 CEE+LFU 策略、Betw+LRU 策略、Opportunistic 策略进行比较，其中包含系统的缓存命中率、跳数减少率、内容差异性和替换数量变化。下面重点讨论节点缓存空间大小、内容数量、替换概率调节系数(β)和网络社团结构对 CSNIC 策略的缓存性能的影响。

1) 节点缓存空间大小的影响

在该实验过程中, 用户数量设置为 50 000, Zipf 参数(α)的参数设置为 1.2, 替换概率调节系数(β)设置为 0.7, 节点缓存空间大小从 10 MB 至 320 MB 变化。

图 2 为 4 种缓存策略的缓存命中率随网络节点缓存空间的变化情况。从图可以看出, 4 种缓存策略的缓存命中率都随网络中节点缓存空间的增大而增大。这是由于节点缓存空间增大, 缓存的内容增多, 用户在缓存内容的节点处获取所需内容的概率增大, 从而导致缓存命中率增大。但在这个过程中, CSNIC 策略的性能一直优于其他 3 种策略。导致这一结果的主要原因是: CEE+LFU 策略实行把所有的内容对象缓存于去往目的地途中所有节点, 这样直接浪费了节点的缓存空间, 网络中缓存的内容减少导致网络的缓存命中率降低; Betw+LRU 策略把内容缓存到全网介数最大的节点, 间接地闲置了其他节点的缓存空间; Opportunistic 策略提高了流行度高的内容在靠近用户节点处的缓存概率, 与上 2 种策略相比, 提高了网络缓存的命中率。而 CSNIC 策略不仅在空间上把内容合理地分布在网络中不同的社团内, 而且在时间上合理地利用了社团内节点社团重要度较低的节点的缓存空间, 因而更能提高系统的缓存命中率。

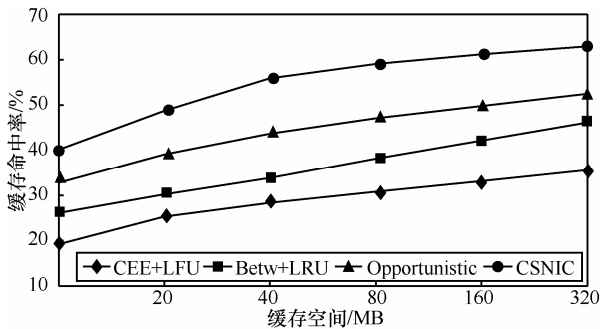


图 2 缓存命中率对缓存空间的影响

图 3 显示了跳数减少率对网络节点缓存空间大小的变化情况, 从图可以看出, 4 种缓存策略的跳数减少率都随网络中节点缓存空间大小的增大而减少, 这是由于节点缓存空间增大, 缓存的内容增多, 用户获取所需内容的跳数减少, 从而跳数减少率降低。但在这个过程中, CSNIC 策略的性能一直优于其他 3 种策略。导致这一结果的主要原因是 CSNIC 策略把内容缓存到社团内节点和社团外节点相对容易访问的位置, 从而降低了跳数减少率。

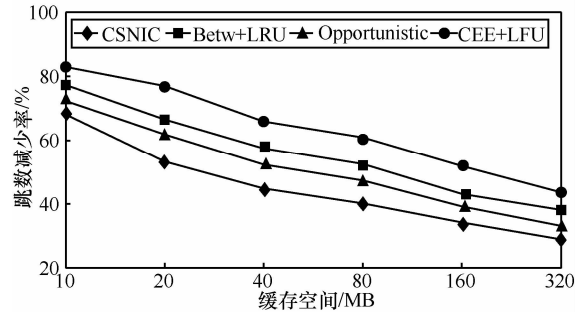


图 3 跳数减少率对缓存空间的影响

图 4 为网络内容差异率对网络节点缓存空间的变化情况, 从图可以看出, 4 种缓存策略的内容差异率都随网络中节点缓存空间的增大而增大, 这是由于节点缓存空间增大, 缓存的内容增多, 节点之间内容差异性增大。但在这个过程中, CSNIC 策略的性能一直优于其他 3 种策略。虽然 Betw+LRU 策略在内容的传输过程中仅缓存介数最大的节点, 在一定的缓存空间条件下, Betw+LRU 策略在网络中缓存内容的差异性通常被认为最好, 但深入的分析发现, 由于缓存空间有限, 其结果是介数大节点的替换次数增加, 并没有大幅增加网络内容的差异性。由于 Opportunistic 策略在内容返回过程不仅考虑了内容在节点处的流行度, 而且通过概率机制考虑了内容的差异性, 因而, Opportunistic 策略性能优于 Betw+LRU 策略。而对 CSNIC 策略, 在同一社团中, 将不同流行度的内容缓存到网络不同的节点处, 使在同一社团内各节点缓存内容具有明显的差异性, 是内容在网络中的时间分布趋于合理, 相当于均匀分配了内容在网络中每个社团内的分布, 直接地提高了缓存内容差异率。同时, 在现实网络中, 不同社团具有不同的喜好, 这样相当于在不同社团之间内容对象再进行一次合理分布。因而, 在这 4 种缓存决策策略中, CSNIC 策略的内容差异率是最佳的。

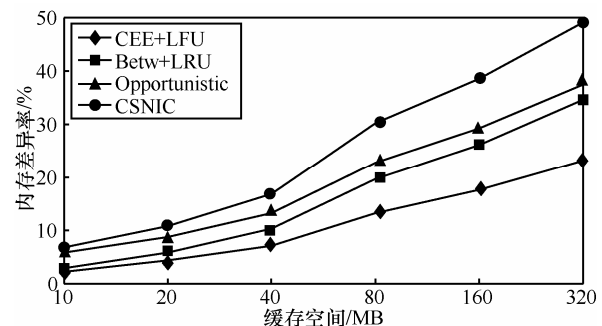


图 4 内容差异率对缓存空间的影响

表 1 为替换数量随缓存大小变化的结果：Betw+LRU 在 4 种机制中的替换数量最小；CEE+LRU 中每一个兴趣分组都会引起一次替换，所以为一常数；而 CSNIC 策略的替换数量略微高于 Betw+LRU，Opportunistic 策略与 CSNIC 策略非常接近。但 Betw+LRU 策略的替换主要集中于高介数的节点处，在单一节点频繁地替换并不满足 ICN 架构中缓存要求线速执行的条件。

表 1 替换数量 VS 缓存大小

Cache	CSNIC	CEE+LFU	Betw+LRU	Opportunistic
10	0.003 28	1	0.002 93	0.003 31
20	0.003 20	1	0.002 81	0.003 26
40	0.002 97	1	0.002 72	0.002 89
80	0.002 90	1	0.002 59	0.002 86
160	0.002 83	1	0.002 51	0.002 78
320	0.002 76	1	0.002 40	0.002 73

2) 内容数量的影响

在该实验过程中，节点缓存空间的大小设置为 20 MB，用户数量设置为 50 000，Zipf 参数(α)的参数设置为 1.2，替换概率调节系数(β)设置为 0.7，其中内容数量在 200 至 6 400 范围内变化。

图 5~图 7 分别为缓存命中率、跳数减少率和内容差异率随网络中的内容数量变化的关系，从图 5~图 7 可以看出，4 种缓存策略的缓存命中率、跳数减少率和内容差异率的性能都随网络中缓存的内容的数量增大而减弱，这是由于随着网络中内容数量增大，需要缓存内容增多，用户在缓存内容节点处获取所需内容的概率降低，从而导致缓存性能减弱。但在这个减弱的过程中，CSNIC 策略的性能一直优于其他 3 种策略。导致这一结果的主要原因与上述节点缓存空间大小情况的讨论类似，网络中内容数量的增加相当于网络中节点缓存空间的减少。

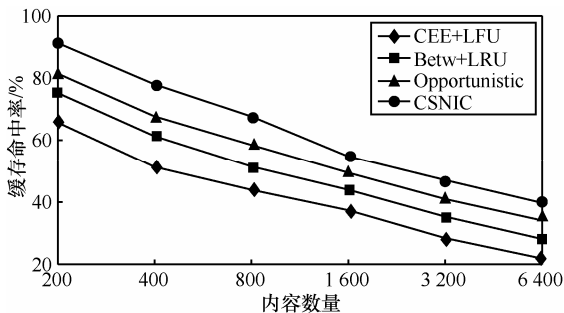


图 5 缓存命中率对内容数量的影响

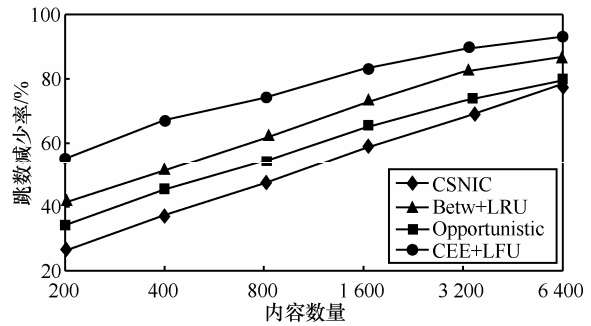


图 6 跳数减少率对内容数量的影响

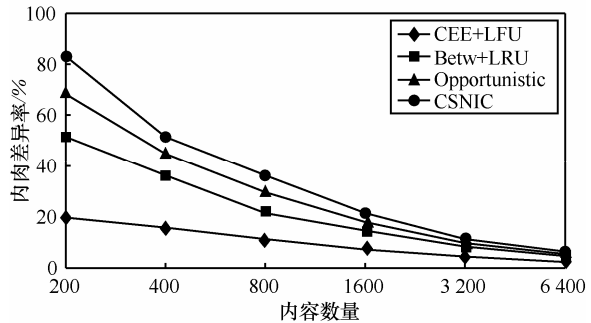


图 7 网络内容差异率对内容数量的影响

4 种缓存策略的替换数量随内容数量变化的结果与缓存空间变化的情况类似：Betw+LRU 在 4 种机制中的替换数量最小；CEE+LRU 中每一个兴趣分组都会引起一次替换，为一常数；而 CSNIC 和 Opportunistic 的替换数量相当，略高于 Betw+LRU。

3) Zipf(α) 参数的影响

在该实验过程中，用户数量设置为 50 000，节点缓存空间的大小设置为 20 MB，替换概率调节系数(β)设置为 0.7，Zipf 参数(α)的参数设置为 0.2 至 1 变化。

现有实证研究显示网络用户对内容的偏好服从 Zipf 分布^[24]，Zipf 参数 α 越大，表示用户的偏好越集中，并且用户对不同网络应用的 α 值也有差异。因此，下面通过变化 α 值测试 CSNIC 策略对不同网络应用的表现。图 8~图 10 分别为缓存命中率、跳数减少率和内容差异率随 Zipf 参数 α 变化的关系。从图 8 可知，4 种策略的缓存命中率都随参数 α 增大而增大，但在这个过程中，CSNIC 策略的性能一直优于其他 3 种策略，导致这一结果的主要原因是 4 种策略都利用了时间的局域性，但除此此外，CSNIC 策略还充分利用空间的局域性。图 9 和图 10 显示，跳数减少率和内容差异率随着参数 α 的增大都有减少，但 CSNIC 策略一直保持较明显的性能优势。

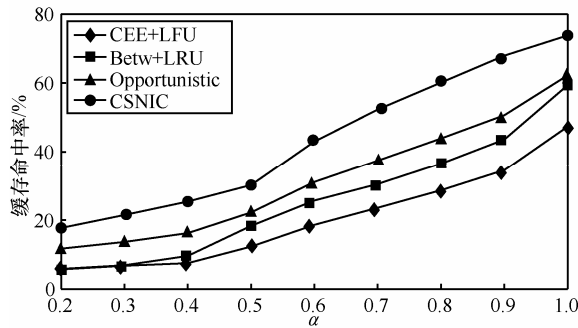


图 8 缓存命中率对 Zipf(α) 的影响

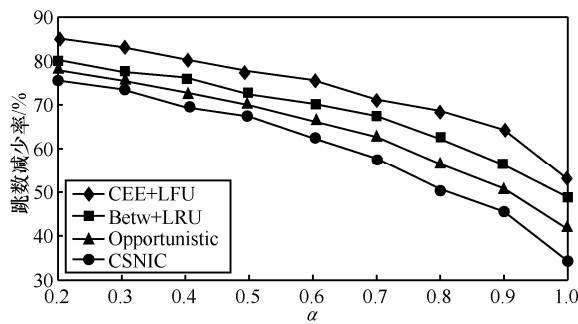


图 9 跳数减少率对 Zipf(α) 的影响

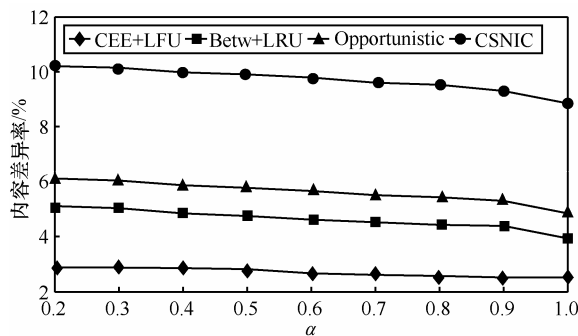


图 10 网络内容差异率对 Zipf(α) 的影响

4) 网络社团结构的影响

现有实证研究显示 Internet 网络具有明显的社团特性^[16], 网络特性的量化指标用模块度 Q 值表示, Q 值越高, 表示网络的社团特性越明显, 现实网络的社团结构在 0.3~0.7 之间。因 CSNIC 策略的实现与节点社团重要度直接相关, 而节点社团重要度与网络的社团结构相联系。为研究网络的社团特性对缓存策略的影响, 本文采用了由 Yan 等提出的 SFC 模型^[25], 该模型生成的网络不仅全网节点的度和社团内节点的度都具有无标度特征, 而且网络社团特性可调, 与现实中 Internet 网络比较接近。选择合适参数生成节点数为 300、模块度 Q 从 0.3 变化到 0.7 的网络。图 11~图 13 为实验结果, 其中横轴为网络模块度的大小。从图可知, 随着网络模块

度的增加, CSNIC 策略的缓存性能(缓存命中率、跳数减少率和内容的差异率)的波动不是很大, 而 CEE+LRU、Betw+LRU 和 Opportunistic 策略随着网络社团模块度的增加, 缓存性能有所下降, 这是由于 CSNIC 策略把内容缓存到社团内节点容易获取的位置, 而其他 3 种策略随着网络社团结构的增强, 若用户想获取的内容不在所属社团内的节点处缓存, 将加大获取的难度。

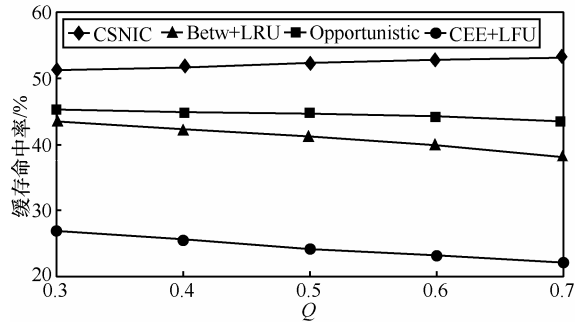


图 11 缓存命中率对网络模块度 (Q) 的影响

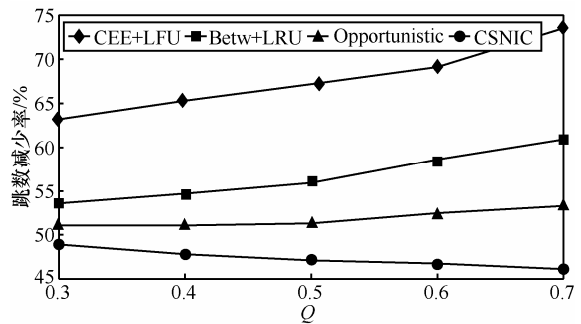


图 12 跳数减少率对网络模块度 (Q) 的影响

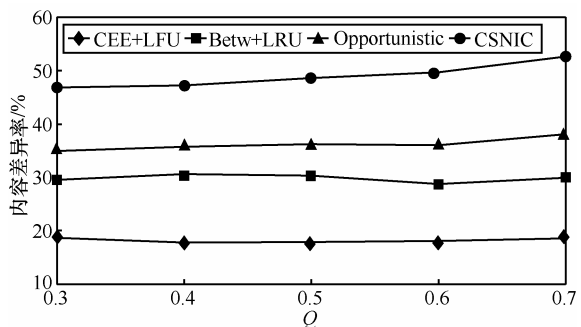


图 13 网络内容差异率对网络模块度 (Q) 的影响

5) 替换调节系数(β)的影响

在该实验过程中, 用户数量设置为 50 000, Zipf 参数(α)的参数设置为 1.2, 节点缓存空间的大小设置为 10 MB 至 320 MB 变化, 替换概率调节系数(β)设置为 0.3 至 1.5 变化。

图 14 为内容差异率随替换概率调节系数(β)和

节点缓存空间的大小变化的曲面图, 从图可以看出, 内容差异率的性能随着 β 系数的变化发生变化。在 $\beta=0.7$ 时, 内容差异率达到最大值, 即在同一社团内, 根据其在社团中的重要度, 通过调节不同节点内内容的替换概率, 以达到每个节点的缓存空间的充分利用。缓存命中率和跳数减少率的变化趋势与内容差异率的变化趋势类似, 限于篇幅, 在文中没有给出它们的变化曲线。进一步实验发现, 为使内容对象在时间的分布上获取一个好的结果, 替换概率调节系数(β)的取值还与网络节点的缓存空间大小分布有关。但由于网络节点缓存空间的大小在网络设计规划时已经固定, 替换概率调节系数(β)对于具体的网络是固定的, 因而在网络中易于布局。

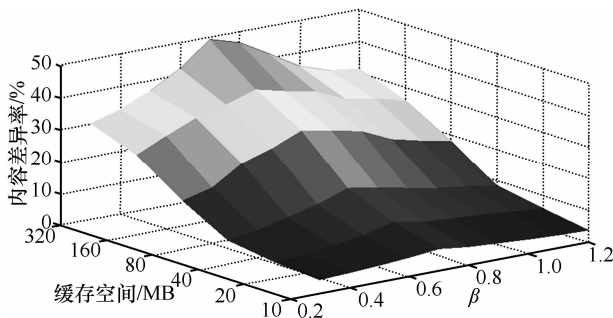


图 14 替换概率调节系数(β)和缓存空间对内容差异率的影响

6) 策略的开销分析

下面从计算开销和通信开销 2 个方面分析 CSNIC 策略。

计算开销: CSNIC 策略缓存决策策略和替换策略中的节点社团重要度的如式(3)所示, 只需求出表示网络节点连接关系的邻接矩阵的所有特征值和特征向量。而现实中的大部分网络为稀疏网络, 利用 Lanczos 算法和 QL 算法, 求稀疏对称矩阵的所有特征值和特征向量的时间复杂度为 $O(nm)^{[26]}$, 其中 n 和 m 分别表示网络的节点数和边数, 其计算复杂度较低, 并且每个节点的社团重要度都是通过离线方式计算存储的。在替换过程中概率的计算也是仅为 $O(k)$, 其中 k 为每个节点的缓存空间大小, 所以 CSNIC 策略满足 ICN 对每个节点缓存处理线速的要求。

通信开销: 在兴趣分组和 Data 分组仅需保存所经过的社团内节点社团重要度最大节点的值, 而每次通信, 找到用户所需内容所经过的社团数量一般在 2 个左右, 显然, CSNIC 策略的通信开销很低。

从上分析可知, CSNIC 策略易于实现, 以较小的开销获得很好的性能改善。

5 结束语

如何高效地利用 ICN 中每个节点的缓存功能, 提高网络传输性能, 成为了 ICN 架构应用于实际环境的关键问题之一。为此, 本文提出了一种应用 Internet 网络拓扑结构的社团特性, 基于每个节点在其所属社团的重要度确定内容缓存位置和缓存替换的缓存策略。通过在不同网络环境进行仿真实验, 实验结果验证了本文方法的有效性。

由于目前实验条件的限制, 本文仅考虑了网络的静态拓扑信息的社团特性, 为更合理地确定内容的缓存位置, Internet 网络的社团特性也应根据网络的实际运行状态动态变化。因此, 下一步工作将基于软件定义网络(SDN)的设计理念, 感知网络的实时状态信息, 从 ICN 架构的实际流量出发, 构建加权有向复杂网络。在 SDN 的控制层, 根据节点的社团特性来确定内容的缓存位置。然后通过 SDN 的数据层执行这一决策, 进一步提高 ICN 架构的缓存性能。

参考文献:

- [1] MOHIT C, BYUNG-GON C, ANDREY E, *et al.* A data-oriented (and beyond) network architecture[A]. Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications[C]. 2007.181-192.
- [2] VAN J, DIANA K, JAMES D. Networking named content[A]. CoNEXT '09 Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies[C]. 2009.1-12.
- [3] Conceptual architecture: principles, patterns and sub-components descriptions[EB/OL]. <http://www.fp7-pursuit.eu/PursuitWeb/>, 2011.
- [4] WEI K C. Curling: content-ubiquitous resolution and delivery infrastructure for next-generation services[J]. IEEE Communications Magazine, 2011, 49(3): 112-120.
- [5] FP7/NICT GreenICN project[EB/OL]. <http://www.greenicn.Org>.
- [6] GHODSI A, SHENKER S, KOPONEN T, *et al.* Information-centric networking: seeing the forest for the trees[A]. HotNets-X Proceedings of the 10th ACM Workshop on Hot Topics in Networks[C]. 2011.1-6.
- [7] CHAI W, HE D, PSARAS I, *et al.* Cache less for more in information-centric networks(extended version)[J]. Computer Communications, 2013, 36(7): 758-770.
- [8] 黄韬, 刘江, 霍如等. 未来网络体系架构研究综述[J]. 通信学报, 2014, 35(8):184-197.
- [9] HUANG T, LIU J, HUO R, *et al.* Survey of research on future network architectures[J]. Journal on Communications, 2014, 35(8):184-197.
- [9] 张国强, 李杨, 林涛等. 信息中心网络中的内置缓存技术研究[J]. 软件学报, 2014, 25(1): 154-175.

- ZHANG G Q, LI Y, LIN T, *et al.* Survey of in-network caching techniques in information-centric networks[J]. *Journal of Software*, 2014, 25(1): 154-175.
- [10] GHODSI A, SHENKER S, KOPONEN T, *et al.* Information-centric networking: seeing the forest for the trees[A]. *HotNets-X Proceedings of the 10th ACM Workshop on Hot Topics in Networks*[C]. 2011.1-6.
- [11] EUM S, NAKAUCHI K, MURATA M, *et al.* CATT: potential based routing with content caching for ICN[A]. *Proceedings of the Second Edition of the ICN Workshop on Information-Centric Networking*[C]. 2012.49-54.
- [12] SAINO L, PSARAS I, PAVLOU G. Hashing routing scheme for information-centric networking[A]. *Pro of the 3rd ACM SIGCOMM Workshop on Information-Centric Networking*[C]. 2013.27-32.
- [13] HE Y, ZHU Y, SHI J, *et al.* A cache strategy in content-centric networks based on node's importance[J]. *Information Technology Journal*, 2014, 13(3): 588-592.
- [14] HU X, GONG J. Opportunistic on-path caching for named data networking[J]. *IEICE Transactions on Communications*, 2014, 97(11): 2360-2367.
- [15] 刘外喜, 余顺争, 胡晓等. CCN 中选择性缓存机制的研究[J]. *计算机学报*, 2014, 37(2): 275-288.
LIU W X, YU S Z, HU X, *et al.* Selective caching in content-centric networking[J]. *Chinese Journal of Computers*, 2014, 37(2): 275-288.
- [16] ERIKSEN K A, SIMONSEN I, MASLOV S, *et al.* Modularity and extreme edges of the Internet [J]. *Physical Review Letters*, 2003, 90(14):148701-148704.
- [17] ZHANG G Q, LI Y, LIN T. Caching in information centric networking: a survey[J]. *Computer Networks*, 2013, 57(16): 3128-3141.
- [18] PODLIPNIG S, BÖSZÖRMENYI L. A survey of Web cache replacement strategies[J]. *ACM Computing Surveys*, 2003, 35(4): 374-398.
- [19] CHAUHAN S, GIRVAN M, OTT E. Spectral properties of networks with community structure [J]. *Physical Review*, 2009, 80(5): 56114.
- [20] WANG Y, DI Z R, FAN Y. Identifying and characterizing nodes important to community structure using the spectrum of the graph[J]. *PLoS ONE*, 2011, 6(11): 1-10.
- [21] 蔡君, 余顺争. 一种有效提高无标度网络负载容量的管理策略[J]. *物理学报*, 2013, 62(5):058901.
CAI J, YU S Z. An efficient management strategy for enhancing traffic capacity in scale-free networks[J]. *Acta Physica Sinica*, 2013, 62(5): 058901.
- [22] CALVERT K I, DOAR M B, ZEGURA E W. Modeling Internet topology[J]. *Communications Magazine, IEEE*, 1997,35(6): 160-163.
- [23] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004,69(2): 26113.
- [24] BRESLAU L, CAO P, FAN L, *et al.* Web caching and zipf-like distributions: evidence and implications[A]. *IEEE INFOCOM*[C]. 1999. 126-134.
- [25] YAN G, FU Z Q, REN J, *et al.* Collective synchronization induced by epidemic dynamics on complex networks with communities[J]. *Physical Review E*, 2007, 75(1): 16108.
- [26] NEWMAN M. *Networks: an Introduction*[M]. Oxford University Press, Inc, 2010.

作者简介:



蔡君 (1981-), 男, 湖南邵阳人, 博士, 广东技术师范学院副教授, 主要研究方向为未来网络、复杂网络。



余顺争 (1958-), 男, 江西南昌人, 博士, 中山大学教授、博士生导师, 主要研究方向为信息安全、信号处理、无线网络。



刘外喜 (1976-), 男, 湖南株洲人, 博士, 广州大学讲师, 主要研究方向为未来网络、网络安全、网络编码。