

# 一种信息中心网络内置缓存空间大小的设计策略<sup>\*</sup>

蔡君<sup>1</sup>, 赵慧民<sup>1</sup>, 魏文国<sup>1</sup>, 刘外喜<sup>2</sup>

(1. 广东技术师范学院电子与信息学院, 广东 广州 510665;  
2. 广州大学电子信息工程系, 广东 广州 510006)

**摘要:** 节点缓存空间大小的设计是信息中心网络(ICN)中缓存策略研究的核心问题之一, 并且缓存空间大小设计策略的优劣直接制约信息中心网络缓存策略的有效实施。基于Internet网络的社团特性, 应用节点社团重要度, 提出了一种设计ICN节点缓存空间大小的策略(简称NIC策略), 并与先前基于度中心设计缓存空间大小的DC策略进行了比较。仿真实验结果显示, 与DC策略相比, NIC策略能更好地提升网络缓存空间的利用率和网络的传输性能。

**关键词:** 信息中心网络; 缓存; 缓存空间; 节点社团重要度

**中图分类号:** TP393 **文献标志码:** A **文章编号:** 0529-6579(2015)06-0072-05

## A Strategy of Designing In-Network Cache Size in Information-Centric Networking

CAI Jun<sup>1</sup>, ZHAO Huimin<sup>1</sup>, WEI Wenguo<sup>1</sup>, LIU Waixi<sup>2</sup>

(1. School of Electronic and Information, Guang Dong Polytechnic Normal University, Guangzhou 510665, China;  
2. Department of Electronic and Information Engineering, Guangzhou University, Guangzhou 510006, China)

**Abstract:** The design of cache size is one of the core issues in the research of caching strategies in Information-centric Networking, and the effective implementation of cache strategies is directly controlled by the pros and cons of the cache size allocation scheme. A novel cache size allocation scheme is proposed based on the node importance to community (NIC), and compared with the previous DC scheme that based on the node degree centrality. The simulation results show that NIC strategy can improve the utilization of network cache space and the network transmission performance.

**Key words:** information-centric networking; cache; cache size; Node's important to community

据思科(Cisco)公司统计, 全球网络流量在过去5年中增长了4倍。2012-2017年期间, 全球网络流量仍将以23%的年均复合增长率高速增长。并且随着新型应用(如物联网模式、云计算

模式)的不断涌现, 流量产生和传输方式也将发生根本变, 其中, 大部分流量都将源自用户驱动的内容获取类应用<sup>[2]</sup>。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

\* 收稿日期: 2015-05-11

**基金项目:** 国家自然科学基金资助项目(61571141); 广东省自然科学基金资助项目(S2013040014339, 2014A030313637, 2014A030310349, 2015A030313672); 广东省教育厅省级重大项目资助项目(2014KZDXM060); 广东省教育厅特色创新项目资助项目(2014KTSCX149); 广东省公益研究与能力建设专项基金资助项目(2014A010103032)

**作者简介:** 蔡君(1981年生), 男; **研究方向:** 未来网络、复杂网络; E-mail: gzhcaijun@126.com

获取模式的转变,同时增强网络对安全性、业务质量、移动性和可扩展性等方面的支持。近年来,国内外针对全新未来互联网的设计启动了各种研究项目,我国对此也启动了 2 个 973 计划项目,各种新架构相继被提出,其中,ICN (Information-Centric Networking,信息中心网络)是各种以信息为中心的架构的统称,其具有以信息为中心的通信模式、基于全网缓存的信息分发模式、内在地支持移动性和安全性等特点,因而正逐渐地被公认为是最有前景的架构之一,典型的如: DONA<sup>[1]</sup>, CCN/NDN<sup>[2]</sup>, PURSUIT 和 COMET<sup>[3-4]</sup>。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,加入了全网内置缓存。虽然缓存机制也是目前互联网中用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<sup>[5]</sup>,但目前的缓存策略主要针对于某一具体应用,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地一部分(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,得出了许多的研究成果,张国强等人<sup>[6-7]</sup>对其进行了全面的综述。其中研究者们重点关注的是缓存位置决策策略、缓存替换策略和节点缓存空间大小的设计三个方面。然而,在这其中,节点缓存空间大小的设计直接制约信息中心网络缓存位置决策策略、缓存替换策略的实施效率。目前,在应用网络拓扑信息构建节点缓存空间大小的策略上,重点考虑了度中心(Degree Centrality, DC ( $n$ ))、压力中心(Stress Centrality, SC ( $n$ ))、介数中心性(Betweenness Centrality, BC ( $n$ ))、接近中心性(Closeness Centrality, CC ( $n$ ))和图中心性(Graph Centrality, GS ( $n$ ))等指标。文献[8]利用这些指标进行了 ICN 架构节点缓存空间大小的设计,仿真结果发现最简单的度中心指标的性能最优。究其原因,如果我们从 ICN 架构增添节点缓存的目的—把内容缓存到尽量靠近用户最容易访问的地方—出发,就会发现度中心指标的性能最优的根本原因是:在这些指标中,只有度中心指标描述的是网络的局部属性。由于一个节点所缓存的内容一般为满足局部用户的内容需求,如果一个节点在某一局部最重要,理应其缓存空间最大,但是网络中反映节点的局部重要性不仅与自

身的信息(度)有一定关系,而且与该节点邻居或邻居的邻居存在一定的关联性。因而,基于度中心指标设计缓存空间大小仍然具有一定的局限性。为此,本文基于 Internet 网络结构具有明显社团这一特性和节点缓存的目的为满足局部用户对内容的需求<sup>[9]</sup>,应用节点在所属社团的重要度的量化指标,提出了一种基于节点社团重要度设计 ICN 架构中节点缓存空间大小的方法(简称 NIC 策略)。仿真结果显示,与先前的 DC 策略相比,NIC 策略能更好地提升网络缓存空间的利用率和网络传输性能。

## 1 NIC 策略

本文将基于节点局部重要度—节点社团重要度确定 ICN 网络中节点缓存空间的大小。据 Chauhan 等<sup>[10]</sup>研究表明,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  $c$  个特征值远大于其它特征值,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<sup>[11-12]</sup>,如式(1):

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

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,即社团内外的其它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<sup>[9]</sup>。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。由公式(1),计算每个节点的社团重要度,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  $O(nm)^{[11]}$ , 其中  $n$  和  $m$  分别表示网络的节点数和边数, 计算量比较低, 适合于实际应用。

假定网络中节点缓存空间的总和为  $C_{tot}$ , 若每个节点的缓存空间相同, 那么  $C_{tot} = |V|C_i$ , 其中  $|V|$  表示网络中的节点数。公式 (2) 表明网络中节点缓存空间的大小完全与该节点的社团重要度的量化值成比例关系, 该式仅考虑节点社团重要度来设计节点的缓存空间大小, 是一种相对理想的方法。而在实际应用中, 由于节点的缓存模块是由多个缓存单元  $c$  组成, 因而, 这种以公式 (2) 分配得到的缓存空间很少能恰好达到缓存单元的倍数关系, 因而它可能在实践中几乎不可行。为此, 引入公式 (3) 进行节点缓存空间大小的设计策略, 其中  $C_i^Q(k)$  为多个缓存单元  $c$  的整数倍, 这样, 节点缓存空间的大小不仅与节点社团重要度相关, 而且与实际的缓存单元相关, 容易在网络中进行实际部署。

$$C_i^P(k) = C_{tot}I_k / \sum_{j \in V} I_j, \quad \forall k \quad (2)$$

$$C_i^Q(k) = \max(c, [C_i^P(k)/c]c) \quad (3)$$

## 2 性能评估

### 2.1 仿真环境

本文采用 ICN 中的内容中心网络 (CCN) 为基础框架来进行测试, 在 CCN 中, 通信的主体是内容而不是主机, 用户只关心内容本身, 不关心内容的位置, 加入全网的内置缓存。在 CCN 的通信过程中, 请求用户发送带名字标识的兴趣包 (Interest Packet), 该包发至靠近用户的 CCN 路由节点, 首先查询该路由器的内容存储器 (Content Store, CS), 若 CS 含有该兴趣包所对应的内容, 则该路由器直接将该内容以数据包 (Data Packet) 的形式发送给请求用户, 完成一次缓存命中; 若 CS 中未查询到该兴趣包所对应的内容, 将查询该路由器的未决请求表 (Pending Information Table, PIT), 若 PIT 中存在与之匹配的信息, 则将该请求端口添加于 PIT 列表中, 否则通过转发库 (Forwarding Information Table, FIB) 查询达到该内容服务器的下一跳接口, 直到找到用户请求的内容, 最后将包含名字和内容的数据包沿着兴趣包的反向路径传送给请求用户, 完成信息传输过程。在这其

中, 缓存位置选择机制采用最近由 WEI 等<sup>[13]</sup>提出的基于介数 (Betweenness) 的放置策略, 该策略只将对象放置于 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命中率, 并减少替换次数。缓存替换策略选择 LRU (Least Recently Used) 策略, 即: 最近最少使用的内容被替换。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,  $i$  ( $1 \leq i \leq M$ ) 表示内容流行度的排名, 则用户对排名为  $i$  的内容请求频率为  $P(X = i) = (1/i^\alpha)/C$ , 其中  $C = \sum_{j=1}^M 1/j^\alpha$ 。因此, 请求内容的百分比直接与前  $k$  个最流行的内容相关, 如:  $\sum_{i=1}^k P(X = i)$  对应于前  $k$  个最流行的内容被请求的百分比。若  $\alpha$  取值比较大, 缓存的对象将变得比较小, 如: 当  $\alpha = 2$  时, 在 108 个文件类别中, 仅 120 MB 的内容需缓存, 前  $k = 12$  个流行度的内容满足了至少 95% 的用户, 即在  $\alpha = 2$  时,  $\sum_{i=1}^{12} P(X = i) > 0.95$ 。现将  $\alpha$  的值分别设置为 1.25 和 1.5, 同样需满足 95% 用户对内容的需求, 此时的  $k$  值发生了显著的变化, 分别变化为 50 000 (大约需 500 TB 的存储空间) 和 234 (大约需 2.4 GB 的存储空间)。因此, Zipf 的参数  $\alpha$  在确定 CCN 的缓存性能过程中比较重要。由于  $\alpha$  值设置为 1.25 和 1.5 对应于缓存性能的变化较大。因此, 在下面的仿真实验中, 仅对  $\alpha \in [1.25, 1.5]$  的边界值进行分析。

### 2.2 评估指标

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, 我们引入缓存命中率 (Cache Hit Ratio, CHR) 对缓存目的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 (Hop Reduction Ratio, HRR) 对缓存目的②和③进行评估, 其定义如公式 (4):

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

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 (Content Diversity Ratio, CDR) 对缓存要求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量的比值, 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 (Number of Replacement, NR) 对缓存要求②进行评估, 其定义为一个 Interest 在一个节点上引起的替换次数, 由于在 ICN 中缓存要求快速执行, 因此频繁地缓存内容替换并不合适。

### 2.3 实验结果与分析

实验网络的拓扑由 100 个节点和 386 条链路组成<sup>[14]</sup>, 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<sup>[15]</sup>。在实验过程中, 采用多个性能参数将 NIC 策略、全网节点缓存空间大小设置相同的 HA 策略和基于度中心指标的 DC 策略进行比较, 其中包含总缓存空间对缓存命中率、跳数减少率、内容差异性和替换数量的影响。其它参数的设计如下: 用户数量设置为 50 000, 内容大小平均为 1 MB, 节点缓存也以 2 MB 为一个单元, Zipf 参数 ( $\alpha$ ) 的参数设置为 1.5。以下所有结果都是经过 10 次试验后得到的平均。

图 1 为系统的缓存命中率对网络总缓存空间大小的变化情况, 其中网络中总的缓存空间大小从 200 ~ 6 400 MB。从图可以看出, 四种机制的节点缓存空间大小设计策略的缓存命中率都随网络中的总缓存空间大小的增大而增大, 但在这个过程中, NIC 策略的性能一直优于其它两种策略。导致这一结果的主要原因是: DC 策略仅考虑了节点的局部属性, 并且这个局部属性只与它的邻居相关, 这样不能很全面的反映整个网络对该节点缓存大小需求。NIC 策略 (NIC-P 表示严格按节点社团重要度进行节点缓存空间分配的策略, NIC-Q 表示以节点社团重要度为依据, 按实际缓存单元大小进行缓存空间分配的策略) 以社团为单位, 而在一个社团内, 节点之间的访问比较频繁, 网络对该节点缓存空间的需求主要来自于社团内的节点。因此本文以节点社团重要度为基础, 就能更好的诠释网络对节点缓存空间大小的需求。

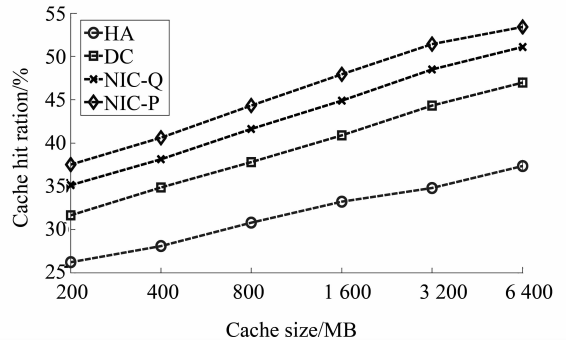


图 1 缓存空间对缓存命中率的影响

Fig. 1 Cache size vs cache hit ratio

图 2 为系统的跳数减少率对网络总缓存空间大小的变化情况, 从图可以看出, 三种缓存空间大小设计策略的跳数减少率都随网络中的总缓存空间大小的增大而减少, 这不能理解, 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, 从而跳数减少率减少。但在这个过程中, NIC 策略的性能一直优于其它两种策略。导致这一结果的主要原因与上述导致缓存命中率情况类似, 也是由于在一个社团内, 最容易达到该社团内其它节点的缓存空间大, 缓存的内容多, 从而跳数减少率降低。

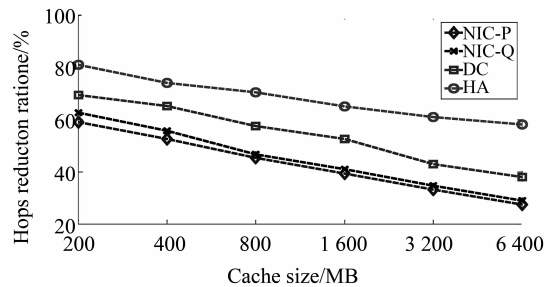


图 2 缓存空间对跳数减少率的影响

Fig. 2 Cache size vs hops reduction ratio

图 3 为系统的内容差异性对网络总缓存空间大小的变化情况, 从图可以看出, 三种机制的缓存空间大小的设计策略的内容差异比率都随网络中的总缓存空间大小的增大而增大, 由于节点缓存空间增大, 不仅缓存的内容增多, 而且缓存内容的种类也增多。在这个过程中, NIC 策略的性能也一直优于其它两种策略。导致这一结果的主要原因是: 一方面虽然网络的总缓存空间大小一定, 如果每个节点的缓存空间大小设置相同, 在不同节点之间所缓存的内容将出现重复的可能性增加, 从而在一定的缓存空间内所缓存的内容的种类将减少; 另一方面在

一个社团内容,所有节点的缓存空间按照节点在该社团的重要度进行设置,使易访问的节点的缓存空间增大,缓存种类增多,其它节点缓存空间依次随节点重要度减少而减少,使网络空间分配实行差异化分配机制,从而网络的内容差异性得到相应地提高。

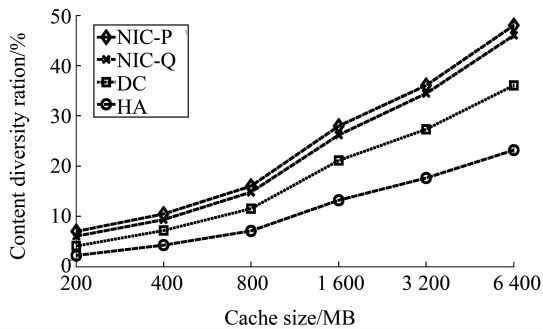


图3 缓存空间对内容差异性的影响

Fig. 3 Cache size vs content diversity ratio

由于在整个实验过程中,缓存位置的选择机制相同,因而,当系统运行稳定后,三种机制的替换数量基本相当。

在仿真过程中,改变用户访问模式  $\alpha$  的值和网络中内容的数量,所得结果与上述结论基本一致。

### 3 结 论

本文提出了一种基于 Internet 拓扑网络的节点社团重要度设计 ICN 架构内节点缓存空间的大小的方法,以 ICN 的 CCN 架构为基础,在具有社团结构的 Internet 网络中,应用介数缓存决策策略和 LRU 缓存替换策略进行了仿真实验。实验结果显示,本文的 NIC 策略在缓存命中率、跳数减少率和内容差异率等性能指标都优于目前相对性能较好的 DC 策略。

虽然本文的方法提高了网络节点缓存空间利用率和缓存性能,但由于所依靠的仍然为网络的静态拓扑信息,因而不能完全的反映网络流量的真实情况,为更合理的规划好节点的缓存空间大小,节点缓存空间的设计也应根据网络的状态进行动态变化,因此,我们的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,基于 SDN 的控制层的缓存空间大小决策策略来指导节点缓存空间大小的分配。

### 参考文献:

- [1] MOHIT C, BYUNG-GON C, ANDREY E, et al. A data-oriented (and beyond) network architecture [C] // Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, 2007: 181 - 192.
- [2] VAN J, DIANA K, JAMES D. Networking named content [C] // CoNEXT'09 Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, 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] 魏文国,赵慧民,庄林凯. 一种基于时钟自适应的改进缓存替换算法[J]. 中山大学学报:自然科学版, 2012, 51(6): 54 - 62.
- [6] ZHANG G Q, LI Y, LIN T. Caching in information centric networking: A survey [J]. Computer Networks, 2013, 57(16): 3128 - 3141.
- [7] 张国强,李杨,林涛. 信息中心网络的内置缓存技术研究[J]. 软件学报, 2014, 25(1): 154 - 175.
- [8] DARIO R, GIUSEPPE R. On sizing CCN content stores by exploiting topological information [C]. INFOCOM WKSHP, 2012: 280 - 285.
- [9] 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.
- [10] CHAUHAN S, GIRVAN M, OTT E. Spectral properties of networks with community structure [J]. Physical Review, 2009, 80(5): 56114.
- [11] 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.
- [12] 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 - 1 - 058901 - 8.
- [13] WEI K C, DILIANG H, IOANNIS P, et al. Cache "Less for More" in information-centric networks (extended version) [J]. Computer Communications, 2013, 36(7): 758 - 770.
- [14] CALVERT K I, DOAR M B, ZEGURA E W. Modeling internet topology [J]. Communications Magazine, IEEE, 1997, 35(6): 160 - 163.
- [15] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69(2): 26113.