

内容中心网络缓存策略研究

王旭辉¹, 张建伟², 蔡增玉¹, 杜春锋¹, 马琳琳¹

(1. 郑州轻工业学院 计算机与通信工程学院, 河南 郑州 450002;
2. 郑州轻工业学院 软件学院, 河南 郑州 450002)

摘要:随着时代的进步与技术的发展,当前以主机为中心的 IP 互联网的主要任务以及互联网用户的关注方向已经发生了翻天覆地的变化,而 IP 互联网由于其固有的特性已经不堪重负。信息中心网络(ICN)作为一种未来网络体系架构,一经提出,便得到了国内外诸多学者的认可。内容中心网络(CCN)作为信息中心网络的一种,被誉为最有发展前景的未来网络体系架构。缓存作为 CCN 的基石,其性能的优劣直接影响着整个网络的性能,良好的缓存策略是 CCN 实现高效内容分发的根本保障。而 CCN 缓存固有的透明化、泛在化以及细粒度化等特性使得传统的缓存理论不再适用于 CCN 的缓存系统。对目前比较有代表性的缓存策略进行了分析,并阐述了其基本原理以及优势与不足,为后续的研究提供参考。

关键词:内容中心网络;缓存策略;透明化;泛在化;细粒度化

中图分类号:TP305

文献标识码:A

文章编号:1673-629X(2018)11-0056-05

doi:10.3969/j.issn.1673-629X.2018.11.013

Research on Caching Strategy for Content Centric Network

WANG Xu-hui¹, ZHANG Jian-wei², CAI Zeng-yu¹, DU Chun-feng¹, MA Lin-lin¹

(1. School of Computer and Communication Engineering, Zhengzhou University of Light Industry,
Zhengzhou 450002, China;
2. School of Software, Zhengzhou University of Light Industry, Zhengzhou 450002, China)

Abstract: With the progress of the times and the development of technology, the main task of the current host-centered IP Internet and the focus direction of Internet users have undergone tremendous changes, and the IP Internet has been overwhelmed by its inherent characteristics. Information-centric network (ICN), a future network architecture, has got the approval of many scholars at home and abroad as soon as it is put forward. Content-centric network (CCN), as a kind of information center network, is praised as the most promising network architecture in the future. As the cornerstone of the CCN, the caching directly affects the performance of the entire network, and well caching strategy is the fundamental guarantee of efficient content distribution in CCN. However, the inherent hyalinization, ubiquitous and fine-grained properties of CCN caching make traditional caching theory no longer applicable to CCN caching system. The current representative caching strategy is analyzed, and its basic principle, advantages and disadvantages are expounded, which gives a reference for subsequent research.

Key words: content-centric network; caching strategy; hyalinization; ubiquitous; fine-grained

0 引言

当前 IP 互联网的设计初衷是为了实现资源(硬件资源)共享,采用的是以主机为中心的端到端的通信模型(host-to-host)^[1]。而随着时代的发展与技术的进步,计算机的硬件资源已经不再匮乏,同时变得非常廉价,所以互联网的任务已经发生了翻天覆地的变化。虽然还是资源的共享,但是已经由硬件资源的共享过

渡为内容资源的共享,即内容的传播;同时,互联网用户的关注方向已经发生了变化,他们不再关注内容存储在哪里(where),而更多的将注意力集中于内容本身,即内容是什么(what)。

以 IP 为中心的互联网关注的是内容的位置而不是内容本身的含义,同时,由于其缓存特有的封闭性,网络中所有的内容都存储在 Source Server 中,使得对

收稿日期:2017-12-10

修回日期:2018-04-18

网络出版时间:2018-06-29

基金项目:国家自然科学基金(61672471,61502436);河南省科技攻关项目(172102210059,172102210060);河南省科技创新人才计划(豫科人组[2017]8号);河南省高校科技创新团队支持计划(18IRTSTHN012);郑州轻工业学院博士基金(2014BSJJ081)

作者简介:王旭辉(1993-),男,硕士,CCF会员(75579G),研究方向为计算机网络、网络安全;张建伟,博士,教授,硕导,研究方向为计算机网络、网络安全、大数据;蔡增玉,硕士,副教授,CCF会员(08965M),研究方向为计算机网络、人工智能。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180629.1700.014.html>

任何内容的请求都需要 Source Server 完成响应,如此不仅导致整个网络的效率不高,同时也造成了巨大的带宽浪费。所以,迫切需要对当前的网络架构进行改进,或者提出全新的网络体系架构^[2]。基于此,国内外学者提出了以信息为中心的新一代网络体系架构——信息中心网络 (information-centric network, ICN),而内容中心网络 (content-centric network, CCN) 作为 ICN 的一个研究分支,被认为是最有发展前景的未来网络体系架构。

与 IP 网不同的是,CCN 支持网内缓存^[3]。在 CCN 中,网络节点(路由器、交换机等)同时具有了转发与缓存的功能。通过网内缓存,CCN 避免了对同一内容的重复传输。当被请求的内容经过某个缓存节点时,该节点可以将此内容存储下来。而且,当这个缓存节点再次收到对该内容的请求时,就可以直接响应该请求,不用再对 Source Server 进行请求。因此,CCN 可以节省网络资源,提升内容的传输效率。

然而,由于 CCN 缓存固有的透明化、泛在化和细粒度化等新的特性^[4],使得传统的缓存理论以及方法等均无法直接应用到 CCN 的缓存系统中。但是,随着对 CCN 研究的深入,许多研究人员在 CCN 缓存理论、模型以及方法等方面都有了创造性的研究成果。

1 CCN 缓存策略

CCN 与 IP 网相比,其主要的特征之一是支持网内缓存。而缓存作为 CCN 的一个重要组成部分,其性能的优劣直接影响 CCN 的整体性能。而在 CCN 中,缓存策略又分为两部分^[4]:其一是缓存决策策略,即在哪个缓存节点缓存什么样的内容;其二是缓存替换策略,即当内容的流行度发生变化或缓存节点的缓存空间被占满,同时又有新的内容需要被缓存时,该如何选择需要被舍弃的内容,释放出空间存储新的内容。

2 缓存决策策略

基于 CCN 的宗旨,其缓存决策策略应该满足以下两方面的要求:

(1) 访问量(流行度)高的内容应该缓存复制到距离互联网用户较近(网络边缘)的缓存节点中,降低用户的请求时延,提升用户使用体验。

(2) 提高整个网络缓存系统的缓存内容的多样性,降低网络缓存内容的冗余度,提升网络的缓存命中率。

为了达到上述要求,降低用户请求时延与内容冗余度,提高缓存内容的多样性,需要网络节点之间能够进行简单而有效的合作缓存机制。而根据合作程度的不同,可以将合作机制大致分为两大类:显式合作机制

与隐式合作机制^[4]。

2.1 显式合作机制

当有内容需要缓存时,显式合作机制要求节点与其周围一定范围内的缓存节点交换彼此的缓存状态信息^[5],进而在此范围内找到最合适的节点缓存该内容。因此,显式合作机制往往伴随着巨大的网络通信开销。以下是几种比较有代表性的显式合作策略。

(1) BCVC (based capacity value caching)。

BCVC 以节点的缓存容量 (cache capacity value, CCV) 作为选择内容缓存节点的的决定性因素^[6-7],每一个缓存节点的 CCV 代表了节点内容存储 (content store, CS) 中的剩余空间,即还能够存储多少的内容。同时,为了记录节点的 CCV,该算法在每一个请求内容的 Interest 包中都添加了 CCV 字段。当请求内容的一个 Interest 包被路由到某一个缓存节点时,该节点首先查找其自身的内容存储 CS,如果在其 CS 中存储有该 Interest 包所请求的内容,该节点就会直接响应 Interest 包的请求,并删除 Interest 包,否则该节点就会将自身的 CCV 写入 Interest 包的 CCV 字段,并继续转发该 Interest 包。当 Interest 包到达另外一个缓存节点时,该节点同样会先查询自身的 CS,如果没有 Interest 包请求的内容,就会将自身的 CCV 与 Interest 包中所携带的 CCV 进行比较,如果小于,就直接转发该 Interest 包,如果大于,就会用自身的 CCV 替代 Interest 包中原来的 CCV 并继续转发该 Interest 包。当 Interest 包到达存储有内容的 Server 时,Server 会根据 Interest 包中所携带的 CCV 选择内容的缓存节点。但是,BCVC 策略没有考虑到内容的流行度,同时,也没有将内容快速地缓存备份到网络边缘的位置。

(2) CCBH (cooperative caching based hash)。

在 CCBH 中,一个内容被分为 n 个大小相同的 chunk^[8-9],而且每一个 chunk 都会携带一个标识,每一个标识就是固定的小于 n 的自然数 $(0, 1, \dots, n-1)$ 。同时,每一个缓存节点及其邻域内的 k 个节点都会有一个编号,分别是 $0, 1, \dots, k-1$ 。当一个 chunk 到达一个缓存节点时,该节点会通过一个 hash 算法决定由其自身还是其邻域内 $k-1$ 个节点中的某个节点缓存该 chunk。具体而言,假设一个 chunk 的标识为 x ,该 chunk 就应该被缓存备份到编号为 $i = x \bmod k (0 \leq i \leq k)$ 的节点,如图 1 所示。

CCBH 策略可以避免同一个 chunk 在邻域内被重复缓存,而且同一个节点不会缓存同一个内容的所有 chunk,提高了网络缓存内容的多样性。但是,因为 CCBH 事先对节点进行了标记,当其中某一个节点被移除或者失效时,CCBH 将无法达到预期目标。同时,CCBH 并没有将内容缓存备份到网络的边缘节点。

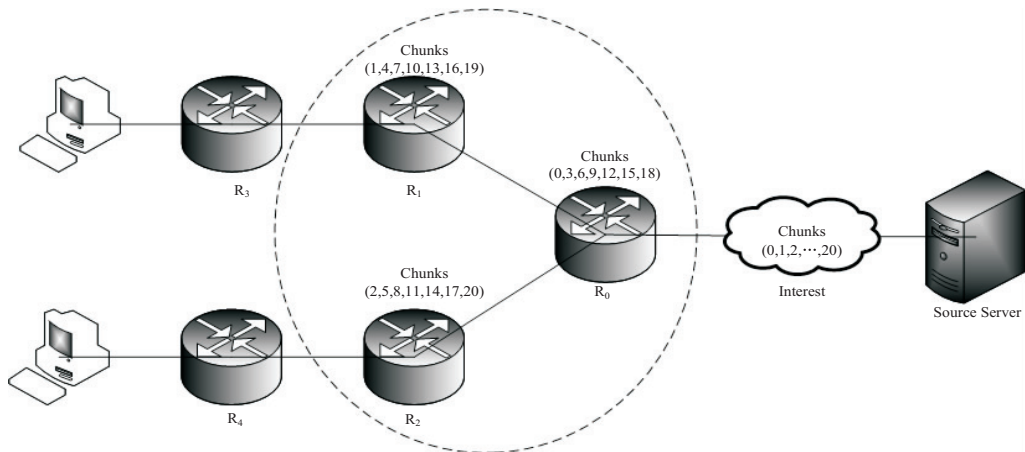


图 1 CCBH 策略

2.2 隐式合作机制

与显式合作机制相比,隐式合作机制不需要如此巨大的网络开销,因为隐式合作不需要缓存节点之间互通状态信息。以下是几种比较有代表性的隐式合作策略:

(1) LCE(leave copy everywhere)。

LCE 也称为 Always Caching^[10],是 CCN 默认的缓存决策策略。该策略的基本思想是:当缓存命中时,在命中节点将用户请求的内容发给内容请求者的传输路径中,沿途上的所有缓存节点都会将该内容缓存下来。

LCE 策略的优势在于可以提高整个网络的缓存命中率,但是会造成巨大的缓存内容冗余,即会在很多的节点中存储有相同内容的副本,使得整个 CCN 网络的缓存内容的多样性不高,而且没有考虑到内容的流行度等因素。

(2) LCD(leave copy down)。

当缓存命中时,仅仅在命中节点下一跳路由器节点对请求的内容进行缓存^[11],但是不会删除该请求内容在命中节点中的缓存副本。

与 LCE 策略相比,避免了对同一内容的大量拷贝,整个 CCN 网络的缓存内容的多样性有所提高,同时,如果对同一内容的访问量较多,该内容将会被复制到靠近用户的地方,这潜在地考虑了内容的流行度。但是,这种策略会在一定程度上降低缓存命中率,另外,如果时间过长,同样会造成巨大的内容冗余。

(3) FixP(fixed-probabilistic caching)。

当缓存命中时,在命中节点将用户所请求的内容发送到用户的传输路径中^[12],沿途上的每一个路由器节点都将会以概率 p 缓存该内容,而以概率 $1-p$ 不缓存该内容。概率 p 可以根据整个网络的缓存情况进行调整。

FixP 策略可以认为是普适化的 LCE,即当缓存概率 $p=1$ 时的 LCE。而且,该策略并没有考虑到内容的流行度等因素数据

(4) ProbCache(probabilistic caching)。

当缓存命中时,在命中节点将用户请求的内容发送到内容请求者的返回路径中^[13],沿途上每一个路由器节点缓存该内容的可能性(概率 p)与其和内容请求者之间的距离成反比。通俗地讲,就是一个缓存节点距离互联网用户越近,该节点能够缓存用户所请求内容的可能性越大。

ProbCache 策略的宗旨就是将内容尽可能地缓存到距离用户较近的边缘节点中,达到降低用户下载内容的平均时延,提高网络资源利用率的目的。但是,ProbCache 策略并没有考虑到内容的访问量等因素,而是一味地将内容尽可能地缓存到距离用户较近的路由器节点中,加大了边缘节点的竞争。

(5) MPC(most popularity caching)。

MPC 是基于内容流行度的缓存策略^[14],其核心思想是:在每一个缓存节点中都有一个流行度表(popularity table,PT),在用户请求一个内容时,沿途的缓存节点会将内容名以及访问量成对地记录下来,当一个内容的流行度达到所设定的流行度阈值之后,持有该内容缓存备份的节点会向其周围的邻居节点(hop=1)发送一条 suggestion,通知其周围的邻居节点缓存该内容,并在收到所有邻居节点已经缓存该内容的反馈信息之后,持有内容备份的节点会重置该内容的流行度以避免重复缓存,如图 2 所示(假设流行度阈值等于 3)。

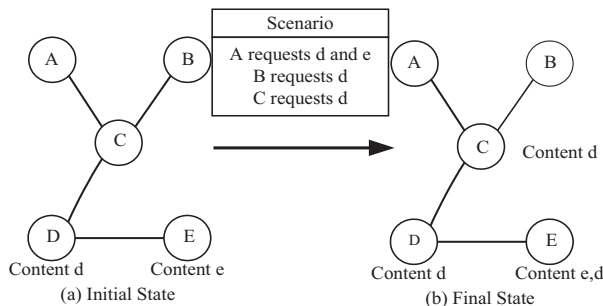


图 2 MPC 策略

MPC 策略的宗旨是尽量多地备份流行度高的内容,这使得用户平均下载时延有了一定程度的降低。然而,正是由于仅仅缓存流行度高的内容,而其他内容并不会得到缓存,使得整个网络资源没有得到充分的利用,致使整个网络系统缓存内容的多样性不高,同时会造成缓存命中率有所下降。

(6) PCBC (popularity and centrality based caching)。

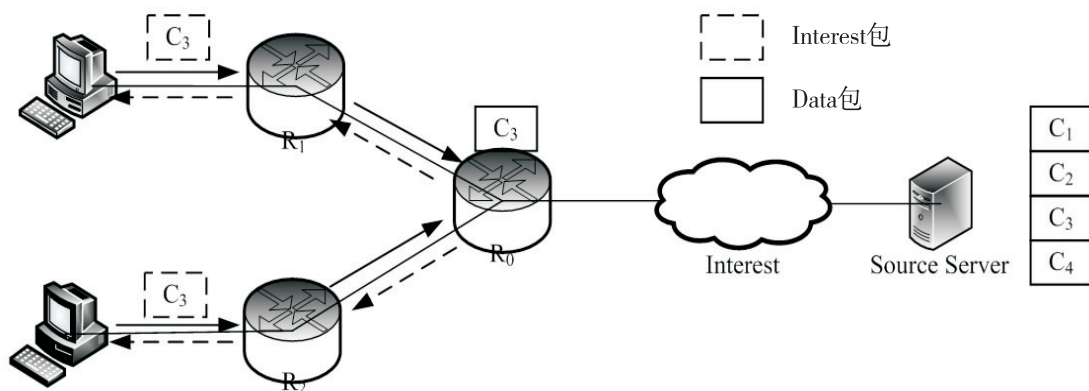


图 3 PCBC 策略

PCBC 策略可以将流行度较高的内容缓存到距离用户较近的边缘节点,达到降低用户下载内容的平均时延,提升用户使用体验的目的。但是,PCBC 策略仅仅将中心度高的节点作为内容的缓存节点,而其他节点并不会缓存任何内容。这样不仅会造成网络资源的大量浪费,同时,PCBC 策略并没有考虑到突发流量的情况,当突发流量到来时,将加大内容缓存节点的通信负载,降低 QoS。

3 缓存替换策略

如前所述,缓存替换策略主要解决的问题是:当内容的流行度发生变化,或者缓存节点的缓存空间被占满,又有新的内容需要缓存时,如何进行内容的替换。

目前,国内外的研究者主要专注于缓存决策策略的研究,而关于缓存替换策略的研究相对较少。以下是几种比较有代表性的替换策略:

(1) LRU (least recently used)。

当有新的内容需要缓存时^[17],缓存节点会将其缓存空间中最近一段时间内访问量最少的内容删除,为新的内容提供缓存空间。

(2) TTL (time to live)。

当一个内容被缓存到节点的缓存空间时,都会生成一个计时器,该计时器会从一个固定的时间开始倒计时,当计时结束时,与其对应的内容就会被删除。

(3) FIFO (first in first out)。

该替换策略的基本思想是:当新的内容需要缓存

PCBC 提出了“中心度”的概念^[15-16],并将节点的中心度作为选择缓存内容节点的决定性因素。中心度用来衡量一个网络节点的“中心程度”,即其在通信链路中的重要性。对一个网络节点来说,与其相关联的通信链路越多,该节点的中心度越高。该缓存策略的基本思想是:将流行度高的内容缓存到中心度较高的缓存节点中。因为中心度越高的节点,越靠近网络边缘,如图 3 所示。

时,缓存节点会删除最先被缓存至该节点的内容,为新的内容腾出空间。

4 结束语

随着网络规模的爆炸式增长,用户兴趣的变化,当前以主机为中心的 IP 互联网已经不能适应时代的发展。国内外学者提出了多种未来网络体系架构,而文中所涉及到的 CCN 就是其中之一。以 CCN 为中心,分析了当前比较有代表性的缓存策略。譬如,LCD 作为 CCN 默认的策略,虽然能够降低请求时延,但是会造成巨大的缓存冗余;PCBC 将流行度高的内容缓存到重要的节点,虽然考虑到了内容的流行度,但是并没有考虑到突发流量的情况,当有突发流量发生时,将加大内容缓存节点的负载。综上所述,在缓存方面,CCN 还有许多问题亟待解决,需要进行更深层次的研究。

参考文献:

- [1] 滕明埏. 命名数据网络中数据缓存策略研究[D]. 重庆:重庆邮电大学,2016.
- [2] 兰巨龙,胡宇翔,张震,等. 未来网络体系与核心技术[M]. 北京:人民邮电出版社,2017.
- [3] 崔现东. 内容中心网络网内缓存策略研究[D]. 北京:北京邮电大学,2014.
- [4] 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.
- [5] 曾宇晶,靳明双,罗洪斌. 基于内容分块流行度分级的信息中心网络缓存策略[J]. 电子学报,2016,44(2):358-364.
- [6] KIM D, LEE S W, KO Y B, et al. Cache capacity-aware

- content centric networking under flash crowds[J]. Journal of Network & Computer Applications, 2015, 50: 101–113.
- [7] TANG Xueyan, CHANSON S T. Coordinated en-route web caching[J]. IEEE Transactions on Computers, 2002, 51(6): 595–607.
- [8] 刘外喜, 余顺争, 蔡 君, 等. ICN 中的一种协作缓存机制[J]. 软件学报, 2013, 24(8): 1947–1962.
- [9] LI Zhe, SIMON G. Time-shifted TV in content centric networks; the case for cooperative in-network caching[C]//IEEE international conference on communications. Kyoto, Japan; IEEE, 2011: 1–6.
- [10] JIANG A, BRUCK J. Optimal content placement for en-route web caching[C]//IEEE international symposium on network computing and applications. Cambridge, MA, USA: IEEE, 2003.
- [11] LAOUTARIS N, CHE Hao, STAVRAKAKIS I. The LCD interconnection of LRU caches and its analysis[J]. Performance Evaluation, 2006, 63(7): 609–634.
- [12] LAOUTARIS N, SYNTILA S, STAVRAKAKIS I. Meta algorithms for hierarchical Web caches[C]//IEEE international conference on performance, computing, and communications. Phoenix, AZ, USA: IEEE, 2005: 445–452.
- [13] PSARAS I, WEI K C, PAVLOU G. Probabilistic in-network caching for information-centric networks[C]//Proceedings of the second edition of the ICN workshop on information-centric networking. Helsinki, Finland; ACM, 2012: 55–60.
- [14] BERNARDINI C, SILVERSTON T, FESTOR O. MP C: popularity-based caching strategy for content centric networks[C]//IEEE international conference on communications. Budapest, Hungary: IEEE, 2013: 3619–3623.
- [15] 芮兰兰, 彭 昊, 黄豪球, 等. 基于内容流行度和节点中心度匹配的信息中心网络缓存策略[J]. 电子与信息学报, 2016, 38(2): 325–331.
- [16] CHAI W K, HE Diliang, PSARAS I, et al. Cache “less for more” in information-centric networks (extended version) ☆☆☆[J]. Computer Communications, 2013, 36(7): 758–770.
- [17] MEGIDDO N, MODHA D S. Outperforming LRU with an adaptive replacement cache[J]. Computer, 2004, 37(4): 58–65.
- +++++
- (上接第 55 页)
- 29(4): 183–190.
- [7] ISCEN C F, EMIROGLU Ö, ILHAN S, et al. Application of multivariate statistical techniques in the assessment of surface water quality in Ulubat Lake, Turkey[J]. Environmental Monitoring & Assessment, 2008, 144(1–3): 269–276.
- [8] 喻 欢, 林 波. 遥感技术在湖泊水质监测中的应用[J]. 环境科学与管理, 2007, 32(7): 152–155.
- [9] GUAN Xian. Monitoring Lake Simcoe water quality using Landsat TM images[J]. Corrosion Science, 2009, 51(4): 703–706.
- [10] 汪小钦, 王钦敏, 刘高焕, 等. 水污染遥感监测[J]. 遥感技术与应用, 2002, 17(2): 74–77.
- [11] 闫友彪, 陈元琰. 机器学习的主要策略综述[J]. 计算机应用研究, 2004, 21(7): 4–10.
- [12] CAO Jian, HU Hongsheng, QIAN Suxiang, et al. Research on water quality assessment method based on multi-class support vector machines[C]//International conference on control, automation, robotics and vision. Hanoi, Vietnam: IEEE, 2008: 1661–1665.
- [13] 宓云翥, 王晓萍, 金 鑫. 基于机器学习的水质 COD 预测方法[J]. 浙江大学学报: 工学版, 2008, 42(5): 790–793.
- [14] LI Shijin, ZHU Haichen, CHEN Deqing, et al. Water quality monitoring based on multiple remote sensing imageries[C]//International workshop on earth observation and remote sensing applications. Guangzhou, China: IEEE, 2016: 112–116.
- [15] HUANG Guangbin, ZHU Qinyu, SIEW C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1–3): 489–501.
- [16] HUANG Guangbin, ZHOU Hongming, DING Xiaojian, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems Man & Cybernetics Part B, 2012, 42(2): 513–529.