内容中心网络中面向隐私保护的协作缓存策略

葛国栋* 郭云飞 刘彩霞 兰巨龙
(国家数字交换系统工程技术研究中心 郑州 450002)

摘 要：针对内容中心网络节点普遍缓存带来的隐私泄露问题，在兼顾内容分发性能的基础上，该文提出一种面向隐私保护的协作缓存策略。该策略从信息熵的角度提出隐私度量指标，以增大攻击者的不确定度为目标，首先对于缓存策略的合理性给予证明；其次，通过构建空间匿名区域，扩大用户匿名集合，增大缓存内容的归属不确定性。缓存决策时，针对垂直请求路径和水平匿名区域，分别提出沿途热点缓存和局域hash协同的存储策略，减小缓存冗余和隐私信息泄露。仿真结果表明，该策略可减小内容请求时延，提高缓存命中率，在提升内容分发效率的同时增强了用户隐私保护水平。

关键词：内容中心网络；协作缓存；隐私保护；内容路由
中图分类号： TP393 **文献标识码**： A **文章编号**：1009-5896(2015)05-1220-07
DOI: 10.11999/JEIT140874

A Collaborative Caching Strategy for Privacy Protection in Content Centric Networking

Ge Guo-dong Guo Yun-fei Liu Cai-xia Lan Ju-long
(*National Digital Switching System Engineering & Technological Research Center*, *Zhengzhou* 450002, *China*)

Abstract: How to mitigate the privacy attacks related to the ubiquitous presence of caching poses challenges to the content delivery in Content Centric Networking. On the basis of trade-off between content distribution performance and users' privacy, a collaborative caching strategy for privacy protection is proposed. First, the privacy metrics are designed and the rationality of the proposed strategy is demonstrated by applying the concept of information entropy. And then, the anonymity domain is constructed to increase the uncertainty of which nearby consumer recently requested certain cached content. When making the caching decision, in order to eliminate the cache redundancy and privacy leaks, the hottest on-path caching and the collaborative hash caching are proposed for the vertical requesting path and horizontal anonymity domains, respectively. The simulation results show that the strategy can decrease the request latency, increase the cache hit ratio, and enhance the protection of users' privacy while improving the efficiency of content distribution.

Key words: Content Centric Networking (CCN); Collaborative caching; Privacy protection; Content-based routing

1 引言

随着互联网技术与应用的飞速发展，“宽带化”、“内容化”与“个性化”已成为未来网络发展的主旋律，人们对于数据内容的需求日益强烈，网络应用的主体逐步向内容请求和信息服务演进[1]。据Cisco VNI的统计预测，2012年~2017年期间，IP流量将以23%的年均复合增长率增加。其中，大部分流量都源自内容获取类应用，预计到2016年，仅视频类流量将占据超过 86%的份额[2]。为了适应不断增长的内容访问需求，内容中心网络(Content Centric Networking, CCN)[3]作为一种革命式(clean-slate)的未来互联网设计思路，让内容本身成为网络通信的主体单元，将网络通信模式从关注“在哪”(地址、服务器)转变为关注“是什么”，成为未来互联网设计的重要模式。CCN在中间层用命名数据取代IP，数据传输采用“发布-请求-响应”模式，直接以内容名字进行路由。当沿途节点接收到兴趣包请求后，依据内容名字依次在内容存储器(Content Store, CS)、未决请求表(Pending Interest Table, PIT)和转发信息库(Forwarding Information Base, FIB)中进行匹配查询[4]，采用网络内在普遍缓存(In-network caching)的方式，在兴趣包沿途转发路径的所有节点上存储应答内容，使得网络不仅是一个传输体，更是一个内容存储、服务平台。

内容的泛在存储在提升内容分发性能的同时，

2014-07-02收到，2014-10-29改回
国家973计划项目(2012CB315901, 2013CB329104)，国家自然科学基金 (61372121)和国家863计划项目(2011AA01A103)资助课题
*通信作者：葛国栋 ggd@mail.ndsc.com.cn

却增大了用户隐私的攻击平面和探测范围，给用户的隐私安全带来严重威胁[5]。在 CCN 中，由于内容命名语义与数据本身紧密相关，节点缓存内容会泄露大量的用户通信痕迹和请求行为信息，攻击者只要获取内容名字，即可请求相应的数据内容，导致严重的隐私信息泄露。

文献[6, 7]对于用户隐私威胁和信息泄露问题进行了全面的分析，指出在 CCN 中，必须综合考虑内容分发性能和用户隐私安全之间的关系，将隐私信息保护融入到 CCN 内在的缓存机制设计中。文献[5]提出了 3 种隐私攻击模式，并分别分析了攻击执行的条件和具体流程；文献[8]提出了使用随机延迟的方式，通过对就近缓存内容的响应时间附加额外时延，以使攻击者不能依据数据响应时间执行缓存内容探测，防止信息泄露。但是该方案却增大了用户请求时延，导致 CCN 网络缓存就近响应带来的低时延优势无法发挥；文献[9]采用洋葱路由思想对数据包进行多重隧道加密，实现信息中心网络(Information-Centric Networking, ICN)中的隐私性保护。但是，由于需要执行多次隧道的加解密操作，将会引入较大的内容传输延迟和额外开销；文献[10]提出了一种隐藏内容名字和数据信息的思想，将请求目标内容和掩护内容名字进行混合，以增加攻击者解析难度和探测成本，增强用户隐私保护；文献[7]指出，在缓存策略设计时，可以通过局部节点的协作缓存，增大请求者的匿名集合，实现用户隐私保护，但是文中并没有给出具体的实现机制。现有方案的不足之处主要体现在：(1)以牺牲内容分发效率来换取用户匿名性和隐私保护的提高；(2)隐私防护策略需要网络增加额外的功能和报文处理，引入大量的计算和代价开销，没有从 CCN 网络内在缓存设计的角度来解决隐私泄露问题。

本文认为缓存策略的设计必须综合考虑内容分发效率和用户隐私安全，将隐私信息保护融入到 CCN 内在的缓存策略设计中。为此，本文提出一种面向隐私保护的协作缓存策略(Collaborative Caching Strategy for Privacy Protection, CCSPP)，从增大攻击者判断的不确定性出发，以扩大用户匿名范围来增加缓存内容的归属不确定度，通过请求内容的热点存储和混淆放置来减小隐私信息泄露，提升用户隐私保护水平。

2　隐私度量与合理性证明

2.1 隐私度量

为了度量用户的隐私保护程度，采用信息熵理论，来量化攻击者对于内容请求者推测的不确定程度。信息熵用于表示某种特定信息的出现概率，一个系统越是规则有序，信息熵就越低。对于攻击者而言，其主要目标是推测缓存内容对应的真正请求者，获知用户的请求行为。攻击者推测内容实际请求用户的不确定度越大，用户隐私保护水平越高。

定义 1　匿名区域(Anonymity Domain, AD)：表示内容路由器(Content Router, CR)与其接入用户构成的局部空间区域。当攻击者与合法用户共同接入 CR 时，AD 代表了攻击者可以探测的邻居用户集合，也构成了内容请求者可以进行匿名混淆的空间范围。

定义 2　隐私匿名熵(Privacy Anonymity Entropy, PAE)：表示对于缓存内容，攻击者推测实际请求者的平均不确定度。例如，当 CR 的接入用户数量为 k，在攻击者没有任何先验背景知识下，内容请求者被识别的概率不超过 $1/k$，从而实现了一种空间 k-匿名。对于 CR 构成的匿名区域 AD，对应的 PAE 大小为

$$\mathrm{PAE}=-\sum_{k} p(u)\log_2 p(u)=-\sum_{k}\frac{1}{k}\log_2\frac{1}{k}=\log_2 k \quad (1)$$

其中，$p(u)$ 表示用户被识别的概率。对于节点的缓存内容，其包含的接入用户数量越多，对应的匿名集合越大，攻击者探测的不确定度越大，用户隐私保护程度越高。

对于攻击者而言，主要目标是判断内容的实际请求者并窃取用户的敏感信息。对于那些流行程度大，众多用户同时请求的热门内容，攻击者难以进行单独的区分定位，大幅降低了攻击者推断目标个体的成功率[5]。内容请求概率越大，不确定度减小量越小，其所能泄露的隐私信息越少。为此，为了度量缓存内容对于用户隐私的危害程度，提出隐私泄露度的概念。

定义 3　隐私泄露度(Privacy Leak Degree, PLD)：定义为单位存储空间缓存内容泄露的信息量大小，表示缓存内容对于用户隐私的危害程度。对于请求概率为 $p(c_i)$ 的缓存内容 c_i，其包含的信息量 $\mathrm{I}(c_i)$ 为

$$\mathrm{I}(c_i)=-\log_2 p(c_i) \quad (2)$$

内容的请求概率越小，攻击者获得的不确定度的减小量越大，泄露的信息量越多。对于缓存空间为 U 的 CR，单位存储空间对应的隐私泄露度 PLD 大小为

$$\mathrm{PLD(CR)}=\frac{\sum\limits_{c_i\in C(\mathrm{CR})}\mathrm{I}(c_i)}{U}=\frac{-\sum\limits_{p(c_i)\in P(\mathrm{CR})}\log_2 p(c_i)}{U} \quad (3)$$

其中，$C(\mathrm{CR})$ 为节点的缓存内容集合，$P(\mathrm{CR})$ 为对

应的请求概率分布。在缓存决策时，节点存储内容的请求概率越大，PLD 取值越小，对用户的隐私危害程度越低。

2.2 合理性证明

定理 1 隐私匿名熵 PAE 与匿名区域 AD 包含的接入用户数量 k 成正比。

证明 对于任意给定的两个匿名区域 AD_1 和 AD_2，包含的请求者数量分别为 k_1 和 k_2，不妨设 $k_2 > k_1$，且有 $k_1 \in \mathrm{N}^+$，$k_2 \in \mathrm{N}^+$。按照式(1)定义，其对应的隐私匿名熵分别为：$\mathrm{PAE}_1 = \log_2 k_1$ 和 $\mathrm{PAE}_2 = \log_2 k_2$。因为 k_1，k_2 分别为大于 1 的整数，$\log_2$ 为单调递增函数，且 $k_2 > k_1$，则有：$\mathrm{PAE}_2 > \mathrm{PAE}_1$。所以，对于隐私匿名熵 PAE，匿名区域 AD 包含的接入用户数量越多，对应的 PAE 取值越高，攻击者推测的不确定程度越大。 证毕

结论 1 为了提高内容请求者的隐私保护水平，增大隐私匿名熵 PAE，应扩大接入用户对应的匿名区域范围，增加攻击者推测的不确定性。

定理 2 隐私泄露度 PLD 与 CR 缓存内容的请求概率成反比。

证明 假设 CR 的缓存空间大小 U，初始缓存内容集合为：$C(\mathrm{CR}) = \{c_1, c_2, \cdots, c_u\}$，请求概率分布为：$P(\mathrm{CR}) = \{p(c_1), p(c_2), \cdots, p(c_u)\}$，对应的隐私泄露度：

$$\mathrm{PLD(CR)} = \sum_{c_i \in C(\mathrm{CR})} I(c_i)/U = -\sum_{P(c_i) \in P(\mathrm{CR})} \log_2 p(c_i)/U$$

不妨设缓存内容 c_i（$c_i \in C(\mathrm{CR})$）的请求概率由原有的 $p(c_i)$ 增大为 $p'(c_i)$，且 $0 < p(c_i) < p'(c_i) < 1$，其对应的隐私泄露度分别为 PLD 和 PLD′。为了证明 PLD 与内容请求概率成反比，即要证明：$\mathrm{PLD}' < \mathrm{PLD}$。依据式 (3) 可知，$\mathrm{PLD}' = -[\log_2 p(c_1) + \log_2 p(c_2) + \cdots + \log_2 p'(c_i) + \cdots + \log_2 p(c_u)]$，$\mathrm{PLD} = -[\log_2 p(c_1) + \log_2 p(c_2) + \cdots + \log_2 p(c_i) + \cdots + \log_2 p(c_u)]$，对比两式不同部分可知，要证明 $\mathrm{PLD}' < \mathrm{PLD}$，只需证明 $-\log_2 p'(c_i) < -\log_2 p(c_i)$ 即可。因为 $0 < p(c_i) < p'(c_i) < 1$，$\log_2(\cdot)$ 为单调递增函数，则有：$\log_2 p(c_i) < \log_2 p'(c_i)$，即：$-\log_2 p'(c_i) < -\log_2 p(c_i)$。所以，CR 缓存内容的请求概率越大，对应的隐私泄露度 PLD 取值越小。 证毕

结论 2 为了减小缓存内容的隐私泄露度 PLD，在执行缓存决策时，应尽量避免对于流行程度低，请求概率小的敏感信息存储。

3 协作缓存策略

CCSPP 主要设计思想包括：(1)扩大用户匿名区域范围，从而增大攻击者推测的不确定性，提高 PAE 取值；(2)沿途热点缓存。将应答内容存储在沿途最大的热点请求区域，避免对于请求概率低的冷门资源的缓存，降低 PLD; (3)域内协同存储。AD 所属节点基于一致性 hash 实现请求内容的协同存储，内容的协同混淆存储，使得攻击者推测的难度和不确定度成倍增加，隐私匿名熵 PAE 取值明显提升。

3.1 匿名区域构建

对于包含 n 个节点和 m 条边的任意 CCN 网络拓扑，用无向无权图 $G = (V, E)$ 来进行表示，其中 $V = \{v_1, v_2, \cdots, v_n\}$ 为节点集合，$E = \{e_1, e_2, \cdots, e_n\}$ 为对应边集。对于节点 v_i，d_i 表示节点度(degree)的大小，$D = \{d_1, d_2, \cdots, d_n\}$ 为节点集 V 对应的度分布。初始时，网络不包含任何匿名区域 AD，节点间无管理和隶属关系。当执行匿名区域构建后，整个网络将被划分为大小范围不同的匿名区域集合，$\mathrm{AD}(G) = (\mathrm{AD}_1, \mathrm{AD}_2, \cdots, \mathrm{AD}_j)$。每个 AD 由一个管理节点和若干隶属节点共同组成，$\mathrm{AD} = (M, S)$。其中，M 代表管理节点，S 为隶属节点集。下面给出 AD 构建算法的具体步骤：

步骤 1 依据节点度分布 D，选择 V 中度数最大节点作为管理节点，将其一跳邻居节点作为隶属节点，建立匿名区域 AD。若出现度数相同节点，随机进行选择。

步骤 2 添加 AD 到 $\mathrm{AD}(G)$ 中，更新管理和隶属节点集合。

步骤 3 更新剩余节点集：$V = V - \mathrm{AD}(G)$。

步骤 4 判断 V 中节点是否全部包含于 $\mathrm{AD}(G)$。若 $V \neq \varnothing$，则说明还有剩余节点没有进行匿名区域划分，重复执行步骤 1～步骤 3。

步骤 5 若 $V = \varnothing$，说明网络所有节点已划分完毕，输出匿名区域集合：$\mathrm{AD}(G) = (\mathrm{AD}_1, \mathrm{AD}_2, \cdots, \mathrm{AD}_j)$。

3.2 沿途热点缓存

定义 4 区域内容活跃度(Domain Content Activity, DCA)：表示在整个匿名区域 AD 内，特定内容被请求的频度。单位时间 T 内，内容 c 在 AD 中对应的 DCA 为

$$\mathrm{DCA}(c) = \sum_{v \in \mathrm{AD}} n(v_1) + n(v_2) + \cdots + n(v_i) \Big/ T \qquad (4)$$

其中，$n(v_i)$ 表示时间 T 内，AD 中节点 v_i 对于内容 c 的请求次数。$\mathrm{DCA}(c)$ 是内容 c 在 AD 所有节点上的请求频度之和，衡量了整个匿名区域对于该内容的整体需求程度。

沿途热点缓存用来确定应答内容存储的目标

AD，请求内容只存储在沿途内容活跃度最大的热点请求区域。在兴趣包和数据包中添加 DCA 字段，用于记录和匹配沿途 AD 的内容活跃度信息，下面给出缓存策略的具体步骤：

步骤 1　内容请求者 v_r 发送兴趣包请求内容 c，节点接收到请求报文后，若 CS 已经缓存内容，直接进行响应；若 CS 和 PIT 中没有对应的请求内容和接口信息，将兴趣包转发到所属 AD 对应的管理节点 v_M。

步骤 2　v_M 查询其 CS 是否存储请求内容，如果包含该内容，则直接进行响应。否则，v_M 执行一致性 hash 运算，判断是否执行 AD 内的局域缓存查找(见 3.4 节)，若其它隶属节点含有该请求内容，将兴趣包转发至目标隶属节点，实现局部就近应答。

步骤 3　如果不执行局域缓存查找，管理节点 v_M 在兴趣包中添加该内容对应的区域活跃度信息 DCA(c)，并依据 FIB 表项执行下一跳路由转发。

步骤 4　通过兴趣包逐跳的上行传输，依次记录和添加沿途 AD 对于内容 c 的活跃度信息。每当兴趣包转发至下一个 AD，管理节点 v_M 将相邻 AD 对应的 DCA(c) 取值大小进行比较，并将 DCA(c) 更新为最大值。

步骤 5　最终，当兴趣包到达内容提供者 v_p 后，DCA(c) 记录的就是对于内容对象 c，沿途匿名区域中最大的内容活跃度取值 $\mathrm{DCA}^{\max}(c)$，其反映了垂直请求路径上，内容 c 对应的最大热点请求区域。

步骤 6　当 v_p 发送数据包进行应答时，将上行兴趣包中携带的 $\mathrm{DCA}^{\max}(c)$ 添加到数据包的 DCA(c) 选项中，并依次比对反向路径对应的活跃度信息，匹配目标 AD，实现应答内容的热点区域存储。

3.3 区域协同存储

当数据包到达目标 AD 后，v_M 执行以下两种操作：(1)正常数据转发。节点依据 PIT 表项的逐跳记录，向请求者传送应答数据内容；(2)区域内容存储。管理节点 v_M 依据 DCA 取值，判断是否缓存该应答内容。如果 v_M 中已有缓存内容的活跃度均大于内容 c 的活跃度 $\mathrm{DCA}^{\max}(c)$，v_M 不执行内容缓存，将 c 直接发送到下一级隶属节点进行存储。否则，v_M 将其添加到 CS 存储单元最顶层，将最底层缓存单元内容替换淘汰，并发送到下一级隶属节点进行存储。目标缓存节点 v_t 的确定通过 v_M 执行一致性 hash 来进行计算：

$$\text{hash}\ (C_{\text{name}}) \to v_t\ ,\quad v_t \in S \tag{5}$$

其中，输入为内容名字(C_{name})，目标空间为 AD 包含的隶属节点集合 S，输出为目标缓存节点 v_t。管理节点确定目标存储节点 v_t 后，从对应的转发接口将内容发送至隶属节点 v_t 进行存储。

3.4 缓存查找与路由转发

当管理节点 v_M 缓存了最新到达的内容后，其 CS 替换的内容将会被发送到下级隶属节点 v_t 进行存储。为了维持目标节点的存储状态，v_M 将建立缓存信息表(Cache Information Table, CIT)用于记录相应的存储信息。每当替换内容被发送到 v_t 后，对应的内容名字将被添加到 CIT 表项中，并进行动态更新。节点接收到后续请求兴趣包时，首先查找 CS 中是否已存储了该请求内容 c，若匹配成功，直接返回数据包进行响应。否则，查找 PIT 表项，若已包含该内容的请求条目，直接添加到达接口到已有的请求列表中；如果为最新请求内容，在 PIT 中新增内容对应的请求条目，并将兴趣包转发至所属 AD 的管理节点 v_M。如果 v_M 已包含该内容，则直接进行响应，否则执行以下两种策略：(1) v_M 查询 CIT 表项，如果包含该内容名字，说明该请求内容在 AD 内其他隶属节点已经缓存。v_M 执行一致性 hash 运算，确定目标缓存节点 v_t，将兴趣包转发至节点 v_t；(2)不必执行局域的缓存查找，直接依据 FIB 表项执行下一跳路由转发。

4　仿真与性能分析

4.1 仿真环境与参数设置

采用 ndnSIM[11] 进行仿真与性能分析，在 GT-ITM 下采用 Locality 模型生成 50 个路由节点的平面随机网络拓扑，仿真拓扑在 NS-3 python 下可视化显示。网络中内容对象总数 N 为 10000 个，内容序号以 1~10000 依次排序，大小设为 10 kBytes。节点缓存容量一致，CS 设为 100，链路带宽 100 Mbps。在网络中设置 2 个内容服务器，负责内容对象的存储和发布，各服务器随机存储 5000 个内容对象，并在网络边缘选取节点 0 和节点 49 与内容服务器直接相连。各节点包含的接入请求用户数量 k 服从均匀分布，$k \sim U(1, 10)$。用户发送的内容请求服从 λ=10 个/s 的泊松过程[12]，请求概率服从 Zipf 分布[13]，第 i 个内容的请求概率为：$p(i) = \dfrac{C}{i^{\alpha}}$，$C = \left(\sum\limits_{i=1}^{10000} \dfrac{1}{i^{\alpha}}\right)^{-1}$。为了构造用户请求的局域分布特征，在内容请求对应 Zipf 序列中随机选取一定比例的扰动内容，进行随机化重新排序[14]，扰动比例选取 1%，即 100 个内容对象。仿真时间为 500 s，采样周期 T=5 s。初始节点缓存状态为空，缓存替换策略为最近最少使用策略(Least Recently Used, LRU)。

4.2 性能分析

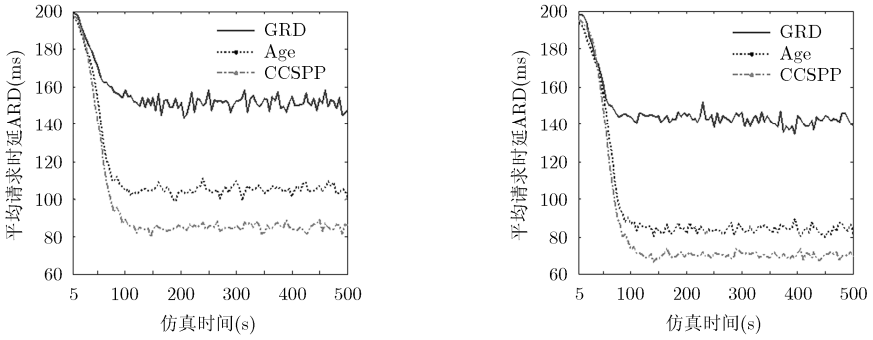
将 CCSPP 与基于随机化延迟的方案(Generate Random Delay, GRD)[8]和基于缓存年龄(Age)的存储方案[15]进行对比分析，性能评价指标包括：平均请求时延(Average Request Delay, ARD)，缓存命中率(Cache Hit Ratio, CHR)，隐私匿名熵 PAE 和隐私泄露度 PLD。

(1)平均请求时延： 图 1 给出 α=0.8 和 α=1.0 时，各方案 ARD 对比，采样时间间隔 T=5 s。仿真初始阶段，由于网络节点存储状态为空，兴趣包都需要转发至内容服务器获取响应，ARD 较大。但随着内容不断存储，缓存内容的响应率逐步增加，ARD 随之减小。

在 GRD 中，采用随机延迟的方式，对就近缓存内容的响应时间附加额外时延，防止攻击探测和信息泄露。但是该方案增大了内容响应时间，对应的 ARD 最大；Age 依据内容流行等级和节点距离数据源的路由跳数，设置缓存时间大小，但该方案仅仅考虑了垂直请求路径方向上的缓存放置，对于沿途路径以外的局部缓存资源无法进行加以利用，ARD 高于 CCSPP; CCSPP 在应答内容存储时，依据 AD 对请求内容的整体需求程度，将应答内容存储在沿途活跃度最高的热点请求区域。内容请求时，不仅可以利用传输路径上缓存资源，对于沿途 AD 隶属节点上的内容副本也可加以利用，减小了内容请求时延，ARD 是各方案中最小的。

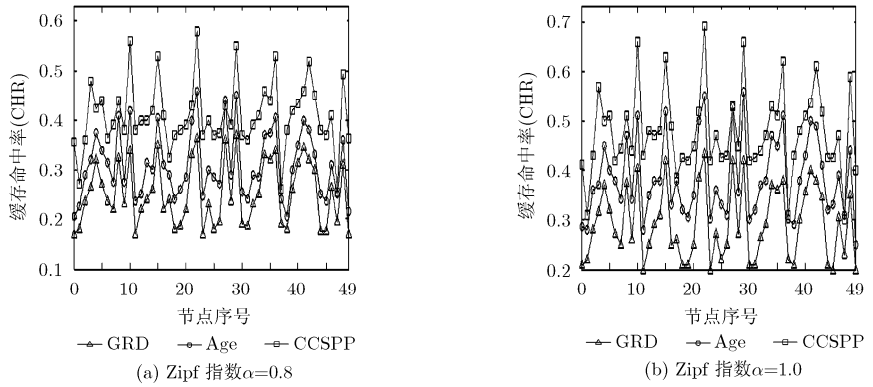
(2)缓存命中率：图 2 给出 α=0.8 和 α=1.0 时，各节点的 CHR 对比。对于 GRD，在缓存决策时，采用的依旧是 CCN 泛滥式的沿途全部缓存方式，大量的缓存冗余和高频率的内容替换更新，导致缓存缺失概率增大，CHR 明显小于 CCSPP; Age 在路由查找时，缓存内容的可用性只局限于沿途传输路径，内容请求对于沿途附近存在的大量缓存资源无法加以利用；对于 CCSPP，整个 AD 最活跃的请求内容存储在度数最大的管理节上，隶属节点缓存缺失内容都将转发至管理节点，有效增大其缓存内容后续利用率，各管理节点对应的 CHR 取值明显高于其它节点。同时，局域缓存查找增加了隶属节点缓存利用率，各节点对应的 CHR 都得到不同程度提高。

(3)隐私匿名熵： 图 3 给出了 $k \sim U(1, 10)$，各方案的 PAE 对比。在 Age 中，节点间缺乏相互协同，内容请求者的匿名范围只局限于接入节点，隐私匿名熵的大小取决于接入节点包含的用户数量，PAE 明显小于 CCSPP; GRD 通过附加额外时延，以使攻击者错误地认为请求内容缓存在一定跳数之外的路由器上，从而增大了合法请求用户的隐私匿名范围，提升了 PAE 取值；对于 CCSPP，整个网络被划分为不同范围的匿名区域，除了少数由单独



(a) Zipf指数α=0.8 (b) Zipf指数α=1.0

图 1 平均请求时延 ARD 对比



(a) Zipf 指数α=0.8 (b) Zipf 指数α=1.0

图 2 节点缓存命中率对比

节点构成，其余都是由多个不同的接入节点协同组成，匿名区域用户数量的成倍增加，使得攻击者推测的成功概率大幅减小。相比 Age 和 GRD，匿名区域的扩大和请求内容的协同混淆存储，使得攻击者推测的难度和不确定度成倍增加，隐私匿名熵 PAE 取值明显提升。

(4)隐私泄露度： 图 4 给出了 α=1.0 时，各方案节点 PLD 的对比。GRD 在执行内容缓存时，没有考虑不同请求内容流行程度的差异性，无法实现应答内容的差异化存储。特别是对于请求频度低的敏感内容，节点不加区分地盲目式存储，将会带来严重的隐私威胁。攻击者可以借助少量背景知识，准确定位内容的实际请求者，各节点对应的 PLD 取值明显高于 CCSPP 方案；Age 依据内容流行等级来设置缓存时间大小，加快冷门资源的沿途替换更新，但是非流行内容依然会存储在沿途多个节点上，直到其缓存时间结束；在 CCSPP 中，AD 内只存储整体需求程度最大的热点请求内容，避免对冷门敏感信息的缓存，有效降低了各节点 PLD 取值。对于 CCSPP，管理节点负责存储整个匿名区域内最活跃的请求内容，对应的 PLD 取值明显小于隶属节点。

(5)隐私侵犯率：由于不同方案在 PLD 和 PAE 方面可以取得不同的性能，为此，提出隐私侵犯率(Privacy Attack Ratio, PAR)来度量各方案总体的隐私保护程度，PAR 定义为归一化后的 PLD 取值与 PAE 取值之比：

$$\mathrm{PAR}=\begin{cases}\dfrac{\mathrm{PLD}}{\mathrm{PAE}}, & \mathrm{PAE}>0\\ 1.0, & \mathrm{PAE}=0\end{cases} \tag{6}$$

其中， $0\le \mathrm{PAR}<1.0$, PAR 取值越小，表示攻击者带来的隐私侵犯率越低，防护策略对应的隐私保护程度越好。当 PAE=0 时，表示局域合法用户的接入数量为 1，攻击者以概率 1 可直接判断特定内容是由该用户请求的，对应的隐私侵犯率 PAR 为 1.0。图 5 给出了 α=1.0, $k\sim U(1, 10)$时，各方案 PAR 的对比分析。GRD 通过附加额外时延，以使攻击者错误的认为请求内容缓存在一定跳数之外的路由器上，提升 PAE 取值。但是，在内容存储时，采用的依旧是 CCN 沿途普遍缓存的方式，并没有考虑不同请求内容流行程度的差异性，PLD 取值是 3 种方案中最大的，导致总体的隐私侵犯率 PAR 大于 CCSPP; Age 在缓存决策时，虽然考虑了不同请求内流行度的差异性，降低了隐私泄露度 PLD，但是内容请求者的匿名范围只局限于接入节点，对于 PAE 取值没有任何提升；对于 CCSPP，通过构建局部匿名区域 AD 和基于 hash 协同的内容混淆放置，增大了用户的匿名范围，提高了 PAE 取值。同时，AD 内只存储局域整体需求程度最大的热点请求内容，避免对冷门敏感信息存储，降低了各节点 PLD 取值，对应的隐私侵犯率 PAR 取值最小。

5　结束语

本文从 CCN 内在缓存策略的设计入手，在兼顾内容分发效率的基础上，提出了一种面向隐私保护的协作缓存策略。CCSPP 借助少量额外代价的付出，在实现内容高效缓存的同时，增强了用户隐私保护水平，仿真结果和性能分析验证了其有效性。后续研究工作主要是针对 CCN 中缓存污染，兴趣泛洪攻击等其它安全威胁，如何设计相应的安全存储机制和防护策略。

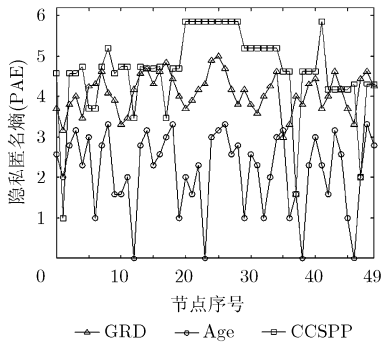


图 3 节点隐私匿名熵 PAE 对比

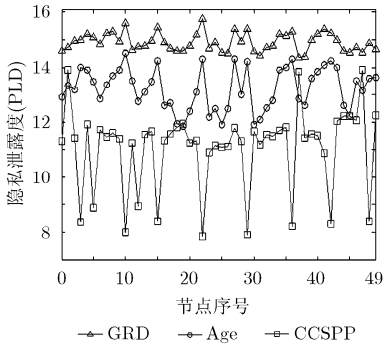


图 4 节点隐私泄露度 PLD 对比

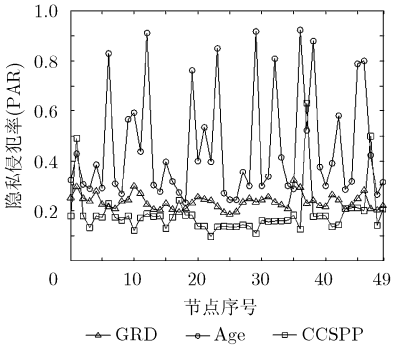


图 5 隐私侵犯率 PAR 对比

参 考 文 献

[1] Xylomenos G, Ververidis C, Siris V, *et al.*. A survey of information-centric networking research[J]. *IEEE Communications Surveys & Tutorials*, 2014, 16(2): 1024-1049.

[2] 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 Guo-qiang, Li Yang, Lin Tao, *et al.*. Survey of

in-network caching techniques in information-centric networks[J]. *Journal of Software*, 2014, 25(1): 154–175.

[3] Jacobson V, Smetters D K, Thornton J D, *et al.*. Networking named content[J]. *Communications of the ACM*, 2012, 55(1): 117–124.

[4] 崔现东, 刘江, 黄韬, 等. 基于节点介数和替换率的内容中心网络网内缓存策略[J]. 电子与信息学报, 2014, 36(1): 1–7.
Cui Xian-dong, Liu Jiang, Huang Tao, *et al.*. A novel in-network caching scheme based on betweenness and replacement rate in content centric networking[J]. *Journal of Electronics & Information Technology*, 2014, 36(1): 1–7.

[5] Lauinger T, Laoutaris N, and Rodriguez P. Privacy implications of ubiquitous caching in named data networking architectures[R]. Technical Report TR-iSecLab-0812-001, 2012.

[6] Lauinger T, Laoutaris N, Rodriguez P, *et al.*. Privacy risks in named data networking: what is the cost of performance ?[J]. *ACM SIGCOMM Computer Communication Review*, 2012, 42(5): 54–57.

[7] Chaabane A, Cristofaro E D, Kaafar M A, *et al.*. Privacy in content-oriented networking: threats and countermeasures[J]. *ACM SIGCOMM Computer Communication Review*, 2013, 43(3): 25–33.

[8] Mohaisen A, Zhang X W, Schuchard M, *et al.*. Protecting access privacy of cached contents in information centric networks[C]. Proceedings of the ACM SIGSAC Symposium on Information, Computer and Communications Security, Hangzhou, China, 2013: 173–178.

[9] DiBenedetto S, Gasti P, Tsudik G, *et al.*. ANDaNA: anonymous named data networking application[C]. Proceedings of the Network and Distributed System Security Symposium, San Diego, USA, 2012: 1–18.

[10] Arianfar S, Koponen T, Raghavan B, *et al.*. On preserving privacy in content-oriented networks[C]. Proceedings of the ACM SIGCOMM Workshop on Information-Centric Networking, Toronto, Canada, 2011: 19–24.

[11] Afanasyev A, Moiseenko I, and Zhang L X. ndnSIM: NDN simulator for NS-3[R]. NDN, Technical Report NDN-0005, 2012.

[12] Chai W K, He D, Psaras I, *et al.*. Cache “less for more” in information-centric networks[C]. Proceedings of IFIP Networking, Prague, Czech Republic, 2012: 27–40.

[13] Kim Y and Yeom I. Performance analysis of in-network caching for content-centric networking[J]. *Computer Networks*, 2013, 57(13): 2465–2482.

[14] Guo S, Xie H Y, and Shi G. Collaborative forwarding and caching in content centric networks[C]. Proceedings of IFIP Networking, Prague, Czech Republic, 2012: 41–55.

[15] Ming Z X, Xu M W, and Wang D. Age-based cooperative caching in information-centric networks[C]. IEEE INFOCOM Workshop on Emerging Design Choices in Name-Oriented Networking, Orlando, USA, 2012: 268–273.

葛国栋： 男，1985年生，博士生，研究方向为新型网络体系结构设计、内容中心网络.

郭云飞： 男，1963年生，硕士，教授，博士生导师，研究方向为新型网络体系结构设计、移动互联网.

兰巨龙： 男，1962年生，博士，教授、博士生导师，研究方向为可重构柔性网络和高性能路由.

刘彩霞： 女，1974年生，博士，副教授，硕士生导师，研究方向为内容中心网络、移动互联网.