

# 无线网络中文件缓存命中率和能耗的联合优化

王悦 邹君妮

(上海大学通信与信息工程学院 上海 200072)

**摘要:** 任何一个缓存问题的重要目标都是减少网络中的流量,但在内容中心网络中能量的消耗是一个不得不考虑的因素,包含缓存能量消耗和传输能量消耗。提出了传统的缓存效率衡量指标,例如命中率,和能量消耗之间存在一定的对立关系。提出一种既可以提高命中率又考虑到减少能量消耗的新的缓存机制,并将这个优化问题表达成亚模的形式,通过贪心算法找到命中率和能量消耗的平衡点,并且找到一种近似最优的缓存机制。最后通过仿真验证了这一结果,在内容中心网络中可以达到命中率和能量消耗的一个均衡的缓存机制。

**关键词:** 内容中心网络;能量消耗;命中率;亚模

**中图分类号:** TN929.5    **文献标识码:** A    **国家标准学科分类代码:** 510.5015

## Joint optimization of cache hit ratio and energy consumption in wireless network

Wang Yue Zou Junni

(Communication & Information Department of Shanghai University, Shanghai 200072, China)

**Abstract:** An important goal of any caching problem is to reduce traffic in the network, but the energy consumption in the content centric network is a factor that needed to be considered, including cache energy consumption and transmission energy consumption. This paper presents a traditional relationship between the cache efficiency measures, such as hit rate, and energy consumption. This paper presents a new caching mechanism that can improve the hit rate and also reduce the energy consumption. Besides this paper express this optimization problem in the form of submodular, find the balance of the hit rate and energy consumption through the greedy algorithm, and find a kind of approximate optimal cache mechanism. Finally, the simulation results show that this result can achieve a balanced cache mechanism of hit rate and energy consumption in the content center network.

**Keywords:** content centric network; energy consumption; hit ratio; submodular

## 0 引言

随着多媒体流量的爆炸式增长,现有的通信网络架构难以承受巨大的用户需求。同时通信系统中网络流量随时间有着明显的变化,在高峰时期网络会十分拥堵,而在非高峰时间网络空闲并不能得到很好的利用。

缓存机制通过在非高峰时间段将内容存储在网络里有存储能力的节点上(路由器、服务器或者终端用户等),在高峰时间段这些缓存可以满足用户的请求。这样的方式可以有效的将网络流量从高峰时间段转移部分到非高峰时间段,从而降低网络拥堵<sup>[1]</sup>。由于网络中流量的巨大一部分源自于对热点内容文件的重复下载,那么网络中的边缘服务器(靠近用户)缓存这些热点内容将会减少网络带宽消耗并且提高内容传输延迟。这样的缓存机制做为应对用户对

内容需求的爆炸性增长的方式,在近年来已经获得越来越多的关注。对于缓存机制来说,最重要的是设计最优内容存储方案:对于一个给定的预期内容需求,确定每个缓存节点中放置哪些文件,以便最大化由缓存服务器满足的请求数量(即命中率)。

目前,以内容为中心的网络已经成为互联网中传播内容最有前景的架构,因此在内容中心网络(content centric networking, CCN)<sup>[2]</sup>中考虑缓存问题十分有必要。这种新型的网络架构有两个主要特征:按名称路由和网内缓存。按名称路由确保内容而不是主机在网络中有最高优先级,所以可以从不同地点多次获取单个内容。网内缓存有助于降低传播时延和网络负载。在以内容为中心的网络中,每个内容路由器具有缓存能力,可以缓存目标文件来减少用户请求的距离。通过使用缓存机制,CCN 可以显著减少网

络负载<sup>[3]</sup>。因为用户请求不再需要通过源服务器,而可以由一个离用户更近的 CCN 内容路由器来满足。虽然在过去有大量优秀的工作是针对于 CCN 网络的,但他们仅仅关注网络资源利用率的提高,而关于能量消耗的部分却被忽略。日益严格的环境标准和迅速增长的能量成本已经引领了互联网中能量效率的趋势。互联网中的能量消耗已经在全球能源消耗比例中占据了高达 10% 并且这个比例在持续增长<sup>[4]</sup>。就能量效率来说,CCN 网络可以通过较短的内容传输距离和去重复传输来降低传输的能量。但同时,CCN 网络通过在每个内容路由器端提供缓存能力而产生额外的能量消耗<sup>[5]</sup>。不同种类的硬件技术可以实现网络中的缓存机制,每种都有不同的能量系数和存储容量。总之,网络中的传输能量和缓存能量存在倒数关系。因此,在 CCN 网络中研究能量消耗是十分有价值的。

目前针对内容中心网络中缓存问题的研究,大多集中于能量消耗部分。本文研究了命中率对于缓存方案设计的影响。

## 1 系统设计概述

### 1.1 系统网络设计

如图 1 所示为内容中心网络中文件缓存的系统模型。图中服务器 S 包含系统中所需要的所有文件内容,为分布在其周围的路由器提供服务。这些边缘路由器均具有缓存能力,每个路由器覆盖需要被服务的用户。在这个模型中有一个服务器 S 和 3 个路由器 R<sub>1</sub>、R<sub>2</sub>、R<sub>3</sub>。用户向路由器请求所需的文件,如果路由器提前缓存该文件,则直接对用户提供服务。在原始网络中服务器 S 拥有所有的文件。如果只考虑命中率,所有路由器均会选择缓存被请求概率较大的文件。考虑功率的情况会变得复杂,如果传输功率的代价比较大,路由器会将文件存储在自己的缓存中;如果缓存功率是主要的,并且由于 R<sub>2</sub> 与 S 距离较远,那么 R<sub>2</sub> 会缓存这个文件,其他路由器也会从他较近的节点获取这个文件。

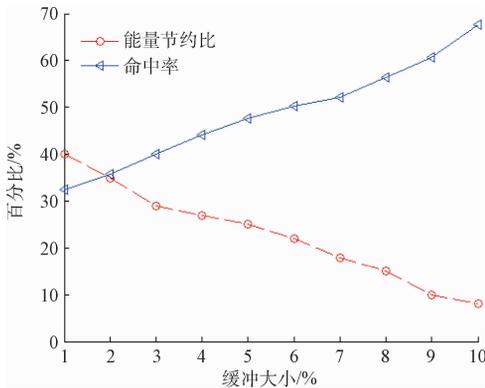


图 1 内容中心网络文件缓存模型

### 1.2 能量消耗模型

对于能量损耗主要考虑两部分,分别为缓存功耗和传

输功耗。缓存功耗  $E^c$  可以使用能量比例模型来定义<sup>[6]</sup>。假设文件  $k$  的大小为  $s^k$ , 节点  $i$  在时间  $t$  内缓存大小为  $s^k$  的文件  $k$  所消耗的能量为  $E_{ik}^c = \omega_{ca} s^k t$ , 在这里  $\omega_{ca}$  是缓存的能量系数,它取决于缓存的硬件条件,例如固态硬盘、动态随机存取存储器和静态随机存取存储器等。传输功耗  $E^r$  指的是当边缘路由器向其他边缘路由器请求内容会产生传输功耗,主要包含路由能量功耗和链路的能量消耗。当节点  $i$  向节点  $j$  请求文件  $k$  时消耗的能量为  $E_{ij}^r = q_i^k s^k [(1 + d_{i,j})(p_d^r + p_d^{\text{roadm}}) + d_{i,j} p_d^{\text{wdm}}]$ 。其中  $q_i^k$  指的是节点  $i$  对文件  $k$  的请求速率,也表示文件的流行性分布。 $d_{ij}$  指两个路由器  $i$  和  $j$  的距离。 $p_d^r$  表示核心路由器的能量密度,  $p_d^{\text{roadm}}$  表示 ROADM 的能量密度,  $p_d^{\text{wdm}}$  表示 WDM 链接的能量密度<sup>[7]</sup>。对于一个传统的没有缓存设备的网络<sup>[8]</sup>,所有文件都需要从服务器进行传输。其总能量消耗为:

$$E^r = q_i^k s^k [(1 + d_{i,-1})(p_d^r + p_d^{\text{roadm}}) + d_{i,-1} p_d^{\text{wdm}}] \quad (1)$$

对于能耗部分,具体定义为:

$$\min \sum_{i=1}^N \sum_{k=1}^M (E_{ik}^c x_{ik} + \sum_{j=1}^N E_{ij}^r y_{ijk} - E_{sr}) \quad (2)$$

式中:  $x_{ik}$  和  $y_{ijk}$  均为指示变量,当  $x_{ik}$  为 1 时表明节点  $i$  缓存了内容  $k$ , 否则为 0; 当  $y_{ijk}$  为 1 时表明节点  $i$  从节点  $j$  下载内容  $k$ , 否则为 0。

### 1.3 命中率模型

所谓缓存命中率指的是节点所请求的文件能在缓存中被满足的概率<sup>[9]</sup>。若想要最大化缓存命中率就是尽量减少从源服务器请求文件的概率。具体定义为:

$$\max: \sum_{i=1}^N \sum_{k=1}^M q_i^k x_{ik} \quad (3)$$

## 2 CCN 缓存机制

本文的目标是在 CCN 网络中路由在有限缓存容量的限制下,如何通过最优的缓存方式达到命中率和功率消耗的均衡。需要在 CCN 网络中找到最小能量消耗和最大命中率之间的均衡点。能量消耗和命中率均衡的问题可以表达成如下的整数线性规划问题:

$$P1: \max_{x,y} \sum_{i=1}^N \sum_{k=1}^M q_i^k x_{ik} +$$

$$\sum_{i=1}^N \sum_{k=1}^M (E_{sr} - (E_{ik}^c x_{ik} + \sum_{j=1}^N E_{ij}^r y_{ijk})) \quad (4)$$

$$\text{s. t. } \sum_{k=1}^M s^k x_{ik} \leq C_i, \forall i \quad (5)$$

$$x_{ik} \in \{0, 1\}, y_{ijk} \in \{0, 1\}, \forall i, j, k \quad (6)$$

$$y_{ijk} \leq x_{ik} \quad (7)$$

在式(4)中,可以写成两部分,能量消耗部分  $\sum_{i=1}^N \sum_{k=1}^M [E_{sr} - (E_{ik}^c x_{ik} + \sum_{j=1}^N E_{ij}^r y_{ijk})]$  和命中率部分  $\sum_{i=1}^N \sum_{k=1}^M q_i^k x_{ik}$ 。其中最小能量消耗用与传统 CDN 相比,缓存所带来的最大节约能量来表征。不等式(5)表示每个节点的缓存内容大小应该小于该缓存容量限制。式(6)表示  $x_{ik}$  和  $y_{ijk}$  均为指示变量,取值只能为 0 或 1。式(7)表示只有内容在节点  $j$  上时,节点  $i$  才能从  $j$  上下载内容  $k$ 。

在过去关于缓存的文献中已经证明了这一类内容放置问题是 NP-hard<sup>[9]</sup>。

### 3 近似算法

#### 3.1 基于拟阵约束的亚模问题

对于这样的 NP-hard 问题,可以将其阐述成一个拟阵约束的单调亚模函数最大值问题。因此需要证明本文的约束条件和目标函数分别是拟阵和亚模函数。拟阵和亚模的定义可以参阅参考文献<sup>[10]</sup>。

首先可以证明约束条件是拟阵。定义一个基集  $S$ 。用  $x_{ik}$  表示把文件  $k$  放置到节点  $i$  的缓存中,则  $S = \{x_{i1}, x_{i2}, \dots, x_{iM}, \dots, x_{NM}\}$ 。这个基集可以被分成  $N$  个独立集合  $S_1, \dots, S_N$ 。其中  $S_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$  是所有可能被放在节点  $i$  缓存中的文件的集合。

**引理 1:** 约束条件是基集  $S$  的分区拟阵<sup>[11]</sup>。

证明:对于优化问题 P1,需要找到一种最优的方法将文件放置到节点的缓存中。每一种内容放置的方法可以看成是  $S$  的子集  $X$  即  $X \subseteq S$ ,  $X$  被称为放置集合。假如  $x_{i1} \in X$ , 节点  $i$  缓存文件  $k$ 。节点  $i$  中的一系列元素是等同于  $X_i = X \cap S_i$ , 在这里  $S_i$  是基集  $S$  关于节点  $i$  的子集。所以关于节点容量的约束可以写成  $X \subseteq I$ , 其中  $I = \{X \subseteq S; |X \cap S_i| \leq M, \forall i = 1, \dots, N\}$ 。对比该表达式和拟阵的定义,可以知道本文的约束  $X$  是拟阵的。

矩阵  $y_{ijk}$  表示节点  $i$  是否从节点  $j$  下载内容  $k$ , 其证明与  $x_{ij}$  的证明类似,不再赘述。

其次需要证明目标函数是单调亚模的。因为亚模函数的和仍然是亚模的。所以可以将式(4)的证明分解成两部分。分别证明命中率部分和能耗部分都是亚模的。

命中率:只需要证明  $h(X) = q_k^i x_{ik}$  是亚模函数即可。如果要证明函数的亚模性其实也就是证明该函数具有边际效用递减的性质,也指在集合中增加新的元素时,函数的值会减小<sup>[12]</sup>。这里需要证明当加入新的文件  $k$  到任意一个节点缓存中时,放置集合增大然而函数的值减小。为了方便问题的解决,我们假设节点中文件的存储是和其流行性相关的,  $q_1^i \geq q_2^i \geq \dots \geq q_M^i$ 。这表明对于所有节点文件 1 是最受欢迎的节点,文件 2 是第二受欢迎的节点。在此考虑两种情况:1)假如文件  $k$  不在节点的缓存中且节点缓存空间还有富余时,那么节点会缓存该文件,其边际效用为  $h(X \cup \{k\}) - h(X) = q_k^i$ ; 2)当其他节点已经缓存了该文件时,节点  $i$  会将该文件与最受欢迎的文件进行交换。把这样的文件记做  $k'$ 。这样边际效用为  $h(X \cup \{k\}) - h(X) = q_{k'}^i$ 。

当把文件  $k$  加入到一个集合  $Y$  中,其中  $Y \subseteq X$ , 分两种情况考虑。1)文件  $k$  不在  $Y$  中,那么  $X$  也不包含文件  $k$ 。边际效用将是  $h(Y \cup \{k\}) - h(Y) = h(X \cup \{k\}) - h(X) = q_k^i$ 。2)如果  $k$  是属于  $Y$  的,边际效用为  $h(X \cup \{k\}) - h(X) = q_{k'}^i$ ,  $k'$  指的是已经存在于节点缓存中的最受欢迎的文件。考虑两种子情况:1) $k$  不在  $X$  中,  $h(X \cup \{k\}) - h(X) = q_k^i \geq$

$q_{k'}^i$ ; 2)  $k$  也存在  $X$  中,那么  $h(X \cup \{k\}) - h(X) = q_{k'}^i \geq q_k^i$ , 这是因为文件  $k'$  是在文件  $k$  挑选后的子集中挑选的。

因此,在集合  $Y$  中增加文件的边际效用是小于等于在集合  $X$  中增加文件。不难理解当缓存中缓存的文件数越多时,那么更多的请求会被缓存服务。这表明  $h(Y) \geq h(X)$ 。可以看出  $h(X) = q_k^i x_{ik}$  是一个单调递增的亚模函数。

能量消耗:根据亚模函数的叠加性质<sup>[13]</sup>,只需要证明

$E_{ik}^\alpha x_{ik} + \sum_{j=1}^N E_{ijk}^r y_{ijk}$  是亚模的即可。将该表达式进一步分解,

可以分别证明  $g(X) = E_{ik}^\alpha x_{ik}$  和  $g(Y) = \sum_{j=1}^N E_{ijk}^r y_{ijk}$  为亚模函数即可。这两部分证明类似,以  $g(Y) = E_{ijk}^r y_{ijk} = q_k^i s^k [(1 + d_{i,j})(p_d^r + p_d^{rdm}) + d_{i,j} p_d^{wdm}] y_{ijk}$  的证明为例。令  $p_{ij} = s^k [(1 + d_{i,j})(p_d^r + p_d^{rdm}) + d_{i,j} p_d^{wdm}]$ , 则  $g(Y) = p_{ij} q_k^i y_{ijk}$ 。

考虑两个放置集合  $Y_1$  和  $Y_2, Y_1 < Y_2$ 。把新文件  $k \notin Y_2$  加入到两个集合中。对于放置集合  $Y_1$  来说,同样分两种情况考虑,(1)假如文件  $k$  不在节点的缓存中且节点缓存空间还有富余时,那么节点会缓存该文件,其边际效用为  $g(Y_1 \cup \{k\}) - g(Y_1) = p_{ij} q_k^i$ 。(2)当其他节点已经缓存了该文件时,节点  $i$  会将该文件与最受欢迎的文件进行交换。把这样的文件记做  $k'$ 。这样边际效用为  $g(Y_1 \cup \{k\}) - g(Y_1) = p_{ij} q_{k'}^i$ 。当把文件  $k$  加入到集合  $Y_2$  中,因为文件  $k$  不属于  $Y_2$ , 其边际效用是  $g(Y_2 \cup \{k\}) - g(Y_2) = p_{ij} q_k^i$ 。因此能耗部分边际效用可总结为  $g_Y(d) = p_{ij} q_k^i$ 。

当把文件  $k$  加入到集合  $Y_2$  中,因为文件  $k$  不属于  $Y_2$ , 其边际效用是  $g_Y(Y_2 \cup k) = p_{ij} q_k^i$ 。又因为  $q_k^i \geq q_{k'}^i$ , 那么  $g_Y(Y_1 \cup k) - g_Y(Y_1) \geq g_Y(Y_2 \cup k) - g_Y(Y_2)$ 。因此  $g_Y(d)$  是个亚模函数。显然  $p_{ij} q_k^i > 0$ , 因此  $g(Y) = \sum_{j=1}^N E_{ijk}^r y_{ijk}$  是一个单调递增的亚模函数。

#### 3.2 贪心算法步骤

贪心算法是一种比较普遍的解决该类型问题的算法<sup>[14]</sup>。贪心算法从空集开始,在每步迭代中,贪心算法选择具有最高边际效用的  $x_{ik}$ 。选择元素  $x_{ik}$  代表节点  $i$  缓存文件  $k$ <sup>[15]</sup>。

$$f(X) = \sum_{i=1}^N \sum_{k=1}^M q_k^i x_{ik} + \sum_{i=1}^N \sum_{k=1}^M (E_{i\alpha} - (E_{ik}^\alpha x_{ik} + \sum_{j=1}^N E_{ijk}^r y_{ijk})) \quad (8)$$

式(8)为亚模函数,随着加入集合中元素的增多,边际效用值会减小。在某一次迭代中如果所选择元素的边际效用值为 0,那么下一个被选择的文件的边际效用值也为 0。因此,本文的迭代将会在边际效用为 0 的时候停止。所有被选择的文件在开始时都根据文件的流行性降序存储在初始集合  $D$  中。初始元素集合  $D$  也就是基集  $S$ 。在每次迭代中,从集合  $D$  中移出元素  $x_{ik}$ 。如果某个节点  $i$  的缓存空间变满,那么将集合  $D$  中和  $i$  相关的元素全部移除,这样节点

$i$  相关元素的边际效用值将不会被考虑在随后的迭代中。放置集合  $X$  和每个被节点  $i$  缓存的文件  $x_{\#}$  在每次迭代的时候更新。最后贪心算法会返回放置集合  $X$ 。综上,贪心算法可以描述为如下。

基于内容中心网络的缓存命中率和能耗的联合优化算法

初始化:令初始元素集合  $D$  为基集  $S$ ,将放置集合  $X$  置为空集。

1) 计算  $k^* = \operatorname{argmax}_{k \in D \setminus X} f(X \cup k) - f(X)$ ;

2) 令  $X = X \cup k^*$ ;

3)  $f(X \cup k) - f(X) = 0$  停止迭代,否则返回步骤 1) 重复上述步骤直到满足条件。

## 4 数值计算与仿真结果

用仿真来评估基于内容中心网络的文件缓存命中率和能耗的联合优化算法的性能。在仿真环境中,总共有 100 个不同的文件。假设每一个文件都有相同的大小,并且文件的概率分布服从齐普夫分布。在仿真网络中一共包含 24 个内容路由器,其中有 20 个路由器与终端用户直连。图 2 所示为在内容中心网络中,根据联合优化算法,在缓存容量为 5 和齐普夫指数为 0.5 的情况下,系统的能耗节约的比率和存储效率的关系,这表明了在 CCN 网络中,我们所提出的缓存方案的效率。从实验结果可以看出,随着存储效率的增加,系统的能耗节约比率是增加的。对于没有缓存的情况,可以认为系统的能耗是没有变化的,所以系统的能耗节约比率接近 0。

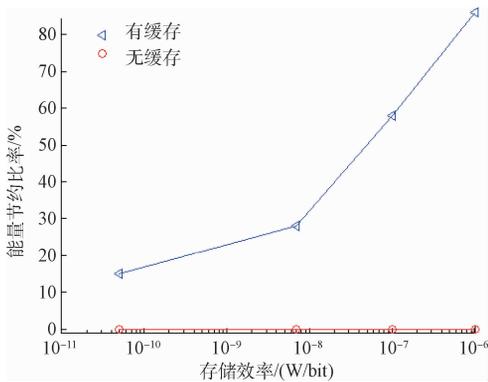


图 2 缓存能耗节约比和存储效率关系

图 3 所示为在内容中心网络中,根据本文提出的文件缓存最优配置方法,在存储效率和齐普夫指数一定的情况下,系统的缓存命中率和缓存容量的关系。从实验结果可以看出,随着缓存大小的增加,每个路由器可缓存的文件数量增加,部分需要通过服务器来提供服务的用户请求,可以通过路由器的缓存来进行服务,那么可服务的用户数量增加,从而命中率也表现出随着缓存容量的增加而增加的性质。

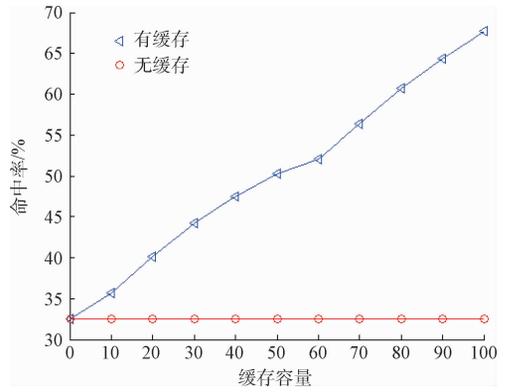


图 3 命中率

图 4 所示在不同的缓存大小的情况下,能耗节约比和缓存命中率的的关系图,可见缓存命中率和能量消耗是一对矛盾的量。随着缓存大小比率的增加,命中率逐渐增大而能量消耗节约比逐渐减小。这从侧面证明了我们所提算法的正确性。

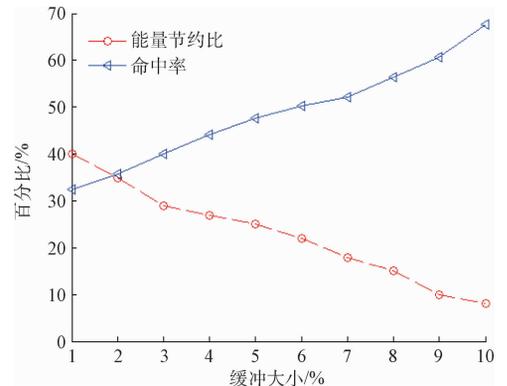


图 4 不同缓存大小下能耗减小比和缓存命中率关系

## 5 结论

本文基于无线网络中流量随时间变化的特点,结合缓存技术的优势,通过在内容中心网络中建立以源服务器为中心,边缘路由器为缓存节点的网络模型,提出了联合缓存命中率和能量消耗的问题模型,证明了该问题是基于拟阵约束的亚模问题,利用亚模函数的性质和贪心算法对该问题进行求解,找到最优缓存放置方案。

## 参考文献

- [1] 李燕君,胡绍利,池凯凯,等. 面向家庭健康监护的传感网中继节点部署方法[J]. 仪器仪表学报, 2015, 36(9): 2153-2160.
- [2] LEE S W, KIM D, KOY B, et al. Cache capacity-aware CCN: Selective caching and cache-aware routing [C]. Global Communications Conference,

- IEEE, 2013; 2114-2119.
- [3] FRICKER C, ROBERT P, ROBERTS J, et al. Impact of traffic mix on caching performance in a content-centric network [C]. IEEE Conference on Computer Communications Workshops, 2012; 310-315.
- [4] POULARAKIS K, TASSIULAS L. On the complexity of optimal content placement in hierarchical caching networks[J]. IEEE Transactions on Communications, 2016, 64(5): 2092-2103.
- [5] 秦勇,梁旭. 基于混合遗传算法的并行测试任务调度研究[J]. 国外电子测量技术, 2016, 35(9): 72-75
- [6] JI M, CAIRE G, MOLISCH A F. Wireless device-to-device caching networks: Basic principles and system performance[J]. IEEE Journal on Selected Areas in Communications, 2013, 34(1): 176 - 189
- [7] CUI X D, HUANG T, LIU J, et al. Design of in-network caching scheme in CCN based on grey relational analysis [J]. The Journal of China Universities of Posts and Telecommunications, 2014(2): 1-8.
- [8] CHOI N, GUAN K, KILPER D C, et al. In-network caching effect on optimal energy consumption in content-centric networking [C]. IEEE International Conference on Communications, 2012; 2889-2894.
- [9] GOLREZAEI N, SHANMUGAM K, DIMAKIS A G, et al. Femtocaching: Wireless video content delivery through distributed caching helpers [C]. INFOCOM, IEEE, 2012, 131(5): 1107-1115.
- [10] POULARAKIS K, IOSIFIDIS G, OURLAS V. Exploiting caching and multicast for 5G wireless networks [J]. IEEE Transactions on Wireless Communications, 2016, 15(4): 2995-3007.
- [11] XU Y, LI Y, WANG Z, et al. Coordinated caching model for minimizing energy consumption in radio access network [C]. International Conference on Communications (ICC), 2014; 2406-2411.
- [12] GREGORI M, GÓMEZ-VILARDEBÓ J, MATAMOROS J, et al. Wireless content caching for small cell and D2D Networks [J]. IEEE Journal on Selected Areas in Communications, 2016, 34(5): 1222-1234.
- [13] LIU D, YANG C. Energy efficiency of downlink networks with caching at base stations [J]. IEEE Journal on Selected Areas in Communications, 2016, 34(4): 907-922.
- [14] LI C, FROSSARD P, XIONG H, et al. Distributed wireless video caching placement for dynamic adaptive streaming [C]. Proceedings of the 26th International Workshop on Network and Operating Systems Support for Digital Audio and Video, 2016.
- [15] 徐琪,刘平,曾禹乔,等. 基于 LabVIEW 的单向阀振动信号采集关键技术设计[J]. 电子测量技术, 2016, 39(11): 118-121.

## 作者简介

王悦, 硕士研究生, 主要研究方向为无线多媒体传输。

E-mail: yuemon@163.com

邹君妮, 教授, 博士生导师, 研究方向为视频编码、认知无线电等。

E-mail: zoujn@shu.edu.cn