

# 双结构网络内容共享能力研究

刘 旋<sup>1,2</sup>, 杨 鹏<sup>1,2</sup>, 董永强<sup>1,2</sup>

(1. 东南大学计算机科学与工程学院, 江苏南京 211189;

2. 计算机网络和信息集成教育部重点实验室(东南大学), 江苏南京 211189)

**摘 要:** 内容共享是实现互联网普惠全民(The Internet is for Everyone)的重要方面,为此,中国工程院李幼平院士提出以 Internet 为主结构而以播存网络为次结构的双体系结构内容共享网络,简称双结构网络 DAN (Dual-Architecture Network),本文首先介绍 DAN 的实现模型并进行形式化描述,然后定性分析 DAN 的内容共享机理和内容运营模式,同时采用 2ACT 模型进行定量分析,最后通过原型系统进行原理性验证. 从多个维度综合阐明 DAN 的内容共享能力.

**关键词:** 双结构网络; 广播分发; 内容共享能力; 定性分析; 定量分析

**中图分类号:** TP393 **文献标识码:** A **文章编号:** 0372-2112 (2018)04-0849-07

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2018.04.012

## Research on Content Sharing Capability for Dual-Architecture Network

LIU Xuan<sup>1,2</sup>, YANG Peng<sup>1,2</sup>, DONG Yong-qiang<sup>1,2</sup>

(1. School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 211189, China;

2. Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, Jiangsu 211189, China)

**Abstract:** Content sharing plays the essential role to achieve the vision of "The Internet is for Everyone". To this end, Prof. Li Youping, academician of the Chinese Academy of Engineering, pioneered the concept of DAN (Dual-Architecture Network), which is composed of primary architecture (refers to Internet) and secondary architecture (refers to broadcast-storage network). In this paper, firstly the implementation model of DAN and its formalized description are exhaustively presented, and then qualitative analyses for the content sharing mechanism and content business model are conducted. Furthermore, quantitative analysis is conducted by applying the 2ACT model. Finally, the feasibility is verified through a prototype system. The content sharing capability of DAN is eventually demonstrated in terms of theoretical analysis and system implementation.

**Key words:** dual-architectural network; broadcast distribution; content sharing capability; qualitative analysis; quantitative analysis

## 1 引言

互联网之父 Vint Cerf 提出互联网普惠全民的概念<sup>[1]</sup>, 欧洲一些国家将互联网接入列为最基本的人权<sup>[2]</sup>, IRTF 于 2014 年成立互联网接入工作 GAIA<sup>[3]</sup>. 然而, 据国际互联网协会发布的最新全球互联网调查表明, 受限于人口、地域和经济等因素, 全球仍有近 60% 的人口还未接入互联网<sup>[3]</sup>, 另据国际电信联盟 2016 年

宽带现状报告统计, 全球约 39 亿人还未享受互联网服务, 数字鸿沟已经由语音电话转向互联网内容<sup>[4]</sup>.

而内容共享是实现互联网普惠全民的重要途径. 当前主流内容共享网络是在 Internet 之上构建 CDN (Content Delivery Network)<sup>[6]</sup>. 然而, 这种内容共享网络本质上具备下层由地址驱动的和上层由内容驱动的两体系结构, 且上层结构寄生、依赖于下层结构, 导致在 Internet 基础设施较差地区难以实现内容共享. 因此, 必

收稿日期: 2016-10-23; 修回日期: 2017-03-23; 责任编辑: 李勇锋

基金项目: 国家自然科学基金 (No. 61472080, No. 61672155); 中国工程院咨询研究项目 (No. 2015-XY-04); 国家 863 高技术研究发展计划 (No. 2013AA013503); 软件新技术与产业化协同创新中心资助

须根本打破这种寄生关系来构建新型内容共享网络。

针对上述问题,中国工程院李幼平院士提出以 Internet 为主结构而以播存网络为次结构的双体系结构内容共享网络 DAN<sup>[7]</sup>。然而,作为一种涉及多方利益相关者(包括网络供应商 ISP、内容供应商 ICP 和用户等)的可运营网络,双结构网络能否被真正所采纳及部署实施,还需综合认知其内容共享能力,以便为利益相关者的投资决策提供指导。因此,本文定性分析其内容共享机理和内容运营模式,并采用统一的 2ACT 模型对其进行定量分析,最后通过 DAN 原型系统来进行可行性验证。从定性分析、定量分析和原型系统等多维度综合展示 DAN 的内容共享能力。

## 2 互联网内容共享相关方案

受限于网络设计理念和运营模式,TCP/IP 从根本上难以打破互联网基础设施较差地区的用户接入经济壁垒<sup>[8]</sup>,所以欧盟 Horizon2020 RIFE 项目<sup>[9]</sup>采用以信息为中心的组网理念解耦内容和主机,以发布/订阅访问范型在时空上解耦内容访问,以内置缓存降低访问延迟和增加负载均衡<sup>[10]</sup>。虽然能够减少用户接入成本,但却极大增加了 ISP 的网络重构成本。另外,这种革命式方案研究发展较为缓慢,据 Gartner 预测<sup>[11]</sup>,ICN 组网技术<sup>[12]</sup>成熟至少需要 10 年以上,所以短期内难以部署实施。

因此,从 2015 年起,OneWeb<sup>[13]</sup>、SpaceX<sup>[14]</sup> 和 Leosat<sup>[15]</sup> 等多家卫星互联网巨头纷纷宣布利用低轨小卫星组成巨型星座来提供互联网全球接入。一方面提供蜂窝回程服务,解决“最后 1 公里”问题,另一方面将卫星接收设备作为小区热点,拓展现有地面网络<sup>[16]</sup>。虽然卫星具有广覆盖特点,但是卫星发射和长期维护开销迫

使 ISP 承担巨大经济压力。此外,谷歌公司的 Loon 项目<sup>[16]</sup>采用氦气球来扩大地面互联网覆盖范围,但受雷雨、气压、风流等环境因素影响,这种“浮动基站”的连接稳定性较差。而脸谱公司的 internet.org 项目采用太阳能无人飞行器,为偏远地区提供新闻、医疗卫生、和教育等免费基础信息服务<sup>[18]</sup>,但这种方案饱受破坏网络中立性原则等问题的争议。

为了推动偏远地区的互联网内容共享,一些近实时或非实时共享方案也相继被提出。Outernet<sup>[19]</sup>利用卫星广播信号公益性地向互联网盲区主动推送基本资讯信息,并利用本地终端接收设备自动缓存广播内容,提供近实时互联网内容服务。然而,在这种单向内容广播系统中,一方面 ISP 免费提供互联网内容服务,运营模式不健全,另一方面主动推送内容给用户,难以感知用户兴趣。Gareth Tyson 等人<sup>[20]</sup>提出利用互联网盒子(Internet Box)预先存储维基百科、电子书和音视频等静态内容,离线状态下供本地用户访问,典型代表包含 Internet-in-a-Box<sup>[21]</sup>、LibraryBox<sup>[22]</sup> 和 Library for all<sup>[23]</sup>。但是这种方案本质上是一种非实时的“知识图书馆”,内容更新周期较长,且内容新鲜度较低。

而双结构网络 DAN 能够利用广播单向传输通道提供近实时的多媒体内容服务。和 CDN、RIFE 以及其它高空平台方案等双向通道相比,DAN 采用广播单通道免路由分发,理论上内容共享不限规模。和传统互联网用户主动请求以及 Outernet 和互联网盒子被动接收的方式相比,DAN 不仅能被动接收广播内容,还能基于兴趣适配机制主动预测、推送内容。另外,消费驱动的内容运营模式为可持续的网络生态提供了良好保障。表 1 详细对比了互联网内容共享相关方案。

表 1 互联网内容共享方案对比

	传输通道	内容类型	提供方式	请求方式	内容更新	运营模式
CDN	双向	所有	实时	主动请求	即时	生产驱动
RIFE	双向	所有	实时	主动请求	即时	生产驱动
卫星互联网	双向	所有	实时	主动请求	即时	生产驱动
Loon	双向	未知	实时	主动请求	即时	生产驱动
internet.org	双向	受限内容	实时	主动请求	即时	免费
Outernet	单向	静态内容	近实时	被动接收	周期短	免费
Internet Box	单向	静态内容	非实时	被动接收	周期长	免费
DAN	单向	多媒体内容	近实时	被动接收、主动推荐	周期短	消费驱动

## 3 DAN 实现模型

DAN 主要包括主结构和次结构两个部分,而统一内容标签 UCL(Uniform Content Label)是双结构共栖协同的纽带,用来支持内容信息聚合、传播、管理和个性

化服务等多维度内容资源描述<sup>[24]</sup>。模型主要包括大数据中心、接入网、播发中心、广播通道和边缘盒子 EBox 等五个构件,如图 1 所示。

大数据中心一方面负责采集次结构 EBox 用户历史记录,并结合主结构热门内容来分析、感知和预测偏

远地区个人和群体用户兴趣,并最终生产适配用户兴趣的内容;接入网是当地用户连接传统广域互联网的有线或无线物理通道.在网络基础设施相对不健全的农村等偏远地区,接入网链路带宽低、时延高,甚至具有间歇性连接的不稳定状态;播发中心是广播分发的内容源,为每一条预播发的内容正文进行 UCL 标引,即凝练内容的语用和语义信息并填充 UCL 字段,同时广播分发互联网内容正文和 UCL;广播通道是单向广播数据传输信道;EBox 是主结构和次结构的物理交汇处,既是广播内容边缘存储库,也是广域互联网接入终端.

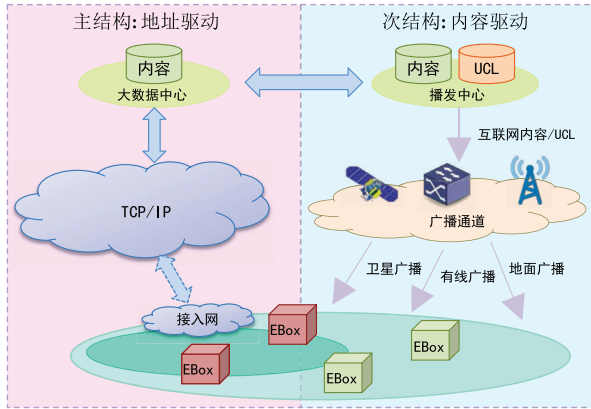


图1 DAN实现模型

下面对 DAN 实现模型进行形式化描述,用四元组表示为  $DAN = (Node, Chan, Func, Serv)$ , 其中,  $Node$  指的是网络实体集合,包括大数据中心  $BDC$ 、边缘盒子  $EBox$  和播发中心  $BRD$  三个子集,且  $Node = BRD \cup EBox \cup BDC$ ;  $Chan$  指的是数据传输通道,包括广播和传统 IP 两类传输通道,分别表示为  $B\_Chan$  和  $T\_Chan$ ,即  $Chan = \{B\_Chan, T\_Chan\}$ ;  $Func$  指的是网络实体所司职功能,主要包括广播分发  $B\_Send$ 、广播接收  $B\_Receive$ 、内容缓存  $C\_Cache$ 、数据处理  $D\_Process$ 、UCL 标引  $UCL\_Label$ 、个体兴趣适配  $Ii\_Adap$ 、群体兴趣适配  $Gi\_Adap$  和内容导航  $C\_Nav$ ,  $Func = \{B\_Send, B\_Receive, C\_Cache, D\_Process, UCL\_Label, Ii\_Adap, Gi\_Adap, C\_Nav\}$ ;  $Serv$  指的是内容服务类型,包括互联网内容全文  $C\_Full$ 、UCL 和传统互联网服务  $T\_Serv$ ,即  $Serv = \{C\_Full, UCL, T\_Serv\}$ . 用  $P(A)$  表示集合  $A$  的幂集,令函数  $t: Node \rightarrow P(Func)$ , 则  $\forall x \in Node$ , 有  $t(x) =$

$$\begin{cases} B\_Send \text{ and } UCL\_Label, x \in BRD \\ B\_Receive \text{ and } C\_Cache \text{ and } Ii\_Adap \text{ and } C\_Nav, x \in EBox \\ D\_Process \text{ and } Gi\_Adap, x \in BDC \end{cases}$$

令函数  $s: Node \times Node \rightarrow P(Chan)$ , 则  $\forall x, y \in Node$ ,

$$\text{有: } s(x, y) = \begin{cases} B\_Chan, & x \in BRD \text{ 且 } y \in EBox \\ T\_Chan, & \text{else} \end{cases}$$

令函数  $u: Node \times Node \rightarrow P(Serv)$ , 则  $\forall x, y \in Node$ ,

$$\text{有: } u(x, y) = \begin{cases} UCL \text{ and } C\_Full, & x \in BRD \text{ 且 } y \in EBox \\ T\_Serv, & \text{else} \end{cases}$$

基于上述定义, DAN 满足以下约束条件:

$$\textcircled{1} |Node| \geq 2 \wedge |EBox| \geq 1 \wedge |Chan| \geq 1;$$

$$\textcircled{2} \text{ 若 } |BDC| \neq 0, \text{ 满足:}$$

$$a. |BRD| \geq 1 \wedge |EBox| \geq 1 |BDC| \geq 1 \wedge |Chan| \geq 3$$

$$b. \forall x, y (s(x, y) = B\_Chan) \wedge (u(x, y) = UCL \text{ and } C\_Full)$$

$$c. \exists x, y (s(x, y) = T\_Chan) \wedge (u(x, y) = T\_Serv)$$

$$\textcircled{3} \text{ 若 } |BDC| = 0, \text{ 满足:}$$

$$a. (BRD \cap EBox = \phi) \wedge (BRD \cup EBox = Node)$$

$$b. |BRD| \geq 1 \wedge |EBox| \geq 1 \wedge |Chan| \geq 1$$

$$c. \forall x, y (s(x, y) = B\_Chan) \wedge (u(x, y) = UCL \text{ and } C\_Full)$$

## 4 DAN 内容共享能力定性分析

由于主、次结构间共栖协同、互补共驱, DAN 能够打破当前内容共享网络结构间的寄生关系.同时,次结构采用物理广播免路由分发,能够使内容直达边缘,且理论上内容共享不限规模,而边缘设备缓存广播内容供用户本地访问,降低请求响应时延;在商业模型上,采用 UCL 驱动的内容运营模式,能够推动 ICP 和用户间进行直接、对等、无中介的内容交易,继而激励 ICP 降低内容接入费用和 ISP 完善网络基础设施.故 DAN 是构建内容共享网络的最佳选择.下面将分小节予以详细阐述.

### 4.1 UCL 驱动的内容共享

主结构是互联网内容的源头,为了向次结构用户提供兴趣适配的内容服务, DAN 采用 UCL 技术,具体格式规范请详见文献[24],主要功能如下:

(1) 内容导航. 为用户提供 UCL 信息门户,通过类型、栏目和关键词等语义信息方便用户快速筛选内容语义,通过内容标题和内容摘要等语用信息方便用户快速筛选内容本身.

(2) 兴趣适配. 为用户提供广播内容的个体和群体用户兴趣适配,而个体适配又包括类别适配和内容适配.类别适配根据用户定制的兴趣类别等信息一阶过滤接收广播内容,内容适配根据个体用户兴趣模型二阶过滤接收广播内容,群体适配则根据区域内群体用户兴趣模型预测待播发的内容.

(3) 内容治理. 为了 ICP 提供内容共享和治理的方法,通过内容指纹等信息来防止内容被篡改和保护知识产权,通过内容出处等信息来有效定位内容来源、实现内容追溯和管理.

DAN 能够实现两种场景下的内容共享.在网络环境相对良好的状态下,次结构首先广播分发内容正文

及其 UCL,并由广播覆盖范围内的 EBox 负责接收和存储,然后用户利用终端设备接入 EBox 并访问 UCL 信息门户及其对应的内容正文.另外,大数据中心通过采集用户的内容访问历史数据,并结合大数据、协同过滤等技术分析、预测和推荐用户感兴趣的内容;而在无广域网或网络断连等状态下,次结构同样广播分发内容正文及其 UCL,但由于大数据中心没有感知用户兴趣的通道,所以只能由 ICP 自主决策播发内容,而 EBox 被动接收内容.

#### 4.2 消费驱动的内容运营模式

互联网主要利益相关者包括 ISP、ICP 和用户,图 2(a)展示了当前互联网内容运营中三者间的关系.

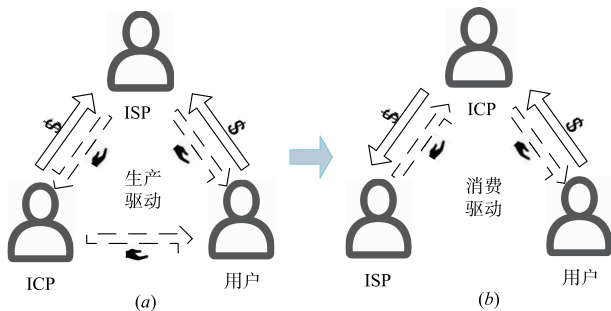


图2 内容运营模式

其中,从资金流角度,ICP 和用户的资金分别流向 ISP,而从服务流角度,ISP 分别为 ICP 和用户提供服务,而 ICP 为用户提供内容服务.因此,当前的内容运营是由 ICP 生产内容驱动的.然而,这种运营模式却存在 ISP 制约 ICP 发展、内容运营复杂化、破坏网络中立性原则以及 ISP 缺少竞争压力等弊端.

而 DAN 采用消费驱动的内容运营模式,如图 2(b)所示.在消费驱动的内容运营模式中,用户起到主导地位,用户对内容的需求能够带动 ICP 在内容类型、形式和质量等方面推陈出新,而其对内容分发性能需求则能够推动 ISP 竞争和发展,用户位于整个生态链上游,也是最终服务对象.具体而言,ICP 不再受 ISP 制约,而 ICP 成为 ISP 赖以生存的上游客户,从而推动 ISP 不断提高网络服务性能;摒除 ISP 服务中介的角色,回归 ICP 和用户间最直接的服务供求关系;差异化传输分级服务的主导权由 ISP 转变为 ICP,ICP 自主根据需求和财力向 ISP 购买相应级别的传输服务,而不再由 ISP 主导,确保网络中立性原则;ICP 不断提高的内容服务传输需求,推动 ISP 竞争和发展,最终促进网络基础设施的完善.

### 5 DAN 内容共享能力定量分析

网络体系结构是否支持适应不断增长和变化的内容需求是判断其内容共享能力一个重要方面,所以本

文采用清华大学徐恪教授等人提出的 2ACT 模型<sup>[25]</sup>从边缘缓存和经济开销两个方面对 DAN 的内容共享能力进行定量分析.

#### 5.1 分析模型

将网络体系结构的内容共享能力表示为<sup>[25]</sup>:

$$U_a = \alpha(\text{Serv})_a + \beta(\text{Cost})_a \quad (1)$$

其中, $a$  指的是待分析的网络体系结构, $U_a$  表示网络体系结构的内容共享能力, $(\text{Serv})_a$  是指网络体系结构为用户提供内容服务的能力, $(\text{Cost})_a$  是指为支持内容服务而部署、维护和管理网络体系结构所产生尽可能少开销的能力. $\alpha$  和  $\beta$  是影响因子,这里假设它们之间存在线性关系. $(\text{Serv})_a$  (衡量数据传输延迟、资源占用)和  $(\text{Cost})_a$  (衡量部署、维护等开销)越高,则  $U_a$  越小,表示其内容共享能力越高.

由于数据传输是网络提供用户服务的基本方式,所以其等同于提供内容服务的能力,则有:

$$(\text{Serv})_a = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n p_{ij} N_{\text{packets}} (\text{Perf})_{ij,a}, \text{ s. t. } \sum_{j=1}^n p_{ij} = 1 \quad (2)$$

其中, $m$  是待分析网络体系结构内容分类方式数量, $n$  是某种分类方式下的内容数据类型种类, $p_{ij}$  是在第  $i$  种分类方式下第  $j$  类内容数据在全网中所占比例, $N_{\text{packets}}$  是网络分组总量, $(\text{Perf})_{ij,a}$  是网络体系结构  $a$  中第  $i$  种分类方式下第  $j$  类内容数据传输性能,为传输延迟、带宽和链路利用率等性能参数的函数.

#### 5.2 内容共享能力定量分析

由于而数据转发跳数能够直观反映数据传输延迟,因此选择利用数据转发跳数和单跳转发性能的乘积来表示  $(\text{Perf})_{ij}$ ,即在网络体系结构  $a$  中第  $i$  种分类方式下第  $j$  类内容数据平均转发跳数为  $H_{ij,a}$ ,单跳转发性能为  $(\text{Phop})_{ij,a}$ ,综合式(1)和(2)得到:

$$U_a = \frac{\alpha}{m} \sum_{i=1}^m \sum_{j=1}^n p_{ij} N_{\text{packets}} (\text{Phop})_{ij,a} H_{ij,a} + \beta C \quad (3)$$

本文将内容数据分为不可缓存数据和可缓存数据两类,例如,实时语音和消息等为不可缓存数据,而非实时多媒体和文件等为可缓存内容.综合文献[26]的统计结果,不可缓存数据平均转发跳数为  $H = 16$ ,假设互联网可缓存数据平均转发跳数为  $H' = 0$ ,而 DAN 采用广播免路由分发,所以其可缓存数据平均转发跳数为  $H' = 1$ ,据文献[27]统计,ICN 可缓存数据平均转发跳数为  $H' = 7$ .由于内容是应用层数据,而 IP 分组是网络层数据,内容转发处理延迟和资源占有率更大,所以通常认为数据转发开销要大于 IP 分组.因此,假设 TCP/IP 网络单跳数据转发性能为  $(\text{Phop})_{\text{IP}}$ ,相应经济开销为  $C_{\text{IP}}$ ,则 DAN 的单跳转发性能为  $r(\text{Phop})_{\text{IP}}$ ,且  $r \geq 1$ ,相应的经济开销为 TCP/IP 的  $t(t \geq 1)$  倍.另外,假

设全网中不可缓存数据流量比例为  $p_1$ , 可缓存数据流量比例为  $p_2$ , 且  $p_1 + p_2 = 1$ . 最终, 根据式 (3) 可分别得到 DAN 和 TCP/IP 的内容共享能力:

$$U_{IP} = 16\alpha(Phop)_{IP} + \beta C_{IP} \quad (4)$$

$$U_{DSN} = \alpha r(Phop)_{IP}(16p_1 + p_2) + t\beta C_{IP} \quad (5)$$

接下来围绕边缘缓存和经济因素两个方面对 DAN 的内容共享能力进行分析.

### (1) 边缘缓存

为了清楚分析边缘缓存对内容共享能力的影响, 假设  $r$  为缓存单跳传输性能与 TCP/IP 数据传输性能的比值. 这里忽略经济因素, 令  $t, C_{IP}$  和  $(Phop)_{IP}$  取值为 1. 取不同  $r$  值, 分析 DAN、TCP/IP 和 ICN 的内容共享能力.

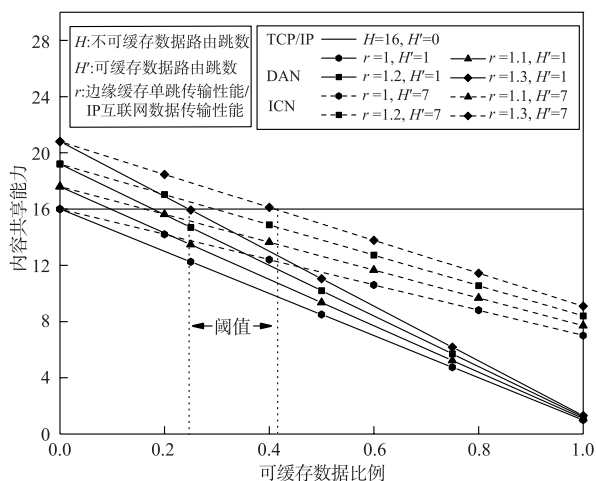


图3 可缓存数据对内容共享能力的影响

从图 3 可以看出, 可缓存数据比例对 TCP/IP 的内容共享能力无影响, 而对 DAN 和 ICN 而言, 内容共享能力随着可缓存数据比例增加而提高. 和 ICN 相比, 无论  $r$  取值, DAN 的内容共享能力都占有明显优势, 而且可缓存数据比例越大, 优势越明显. 然而和 TCP/IP 相比, 当  $r=1$  时, DAN 的内容共享能力占有明显优势, 而随着  $r$  取值变大, 当可缓存数据比例增加到某个阈值时, DAN 的内容共享能力优于 TCP/IP. 而 DAN 中该阈值要低于 ICN, 说明与 ICN 相比, DAN 能够以更少的缓存数据来获取更高的内容共享能力, 同时也说明 DAN 对现有基础设施的改动需求越小, 有利于网络体系结构的演变.

### (2) 经济开销

将经济开销和单位数据转发性能的比值定义为  $n = C_{IP}/(Phop)_{IP}$ , 通过对  $r, t$  和  $n$  取不同的值来分析经济开销对内容共享能力的影响. 如图 4 所示, 实线和虚线分别表示 DAN 和 TCP/IP 的内容共享能力.

显然, DAN 的内容共享能力曲线为单调递减函数, 在相同  $r, t$  和  $n$  取值下, 可缓存数据越多, 其内容共享

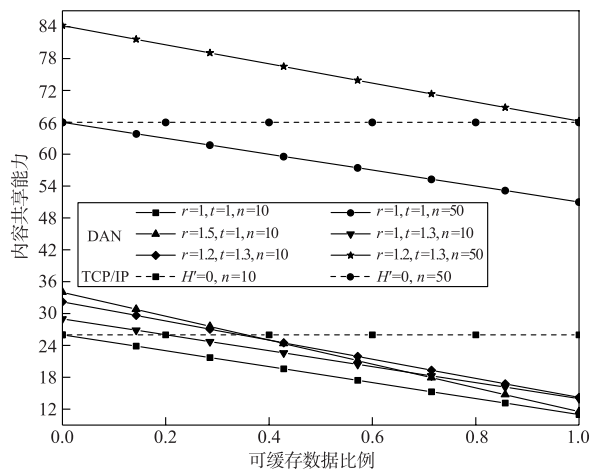


图4 经济开销对内容共享能力的影响

能力越好. 取  $n=10$  时, DAN 的内容共享能力完全优于 TCP/IP, 而取  $n=50$  时, 只有可缓存数据比例为 1 时, DAN 的内容共享能力才和 TCP/IP 持平, 因此,  $n$  值越小, 越适宜采用 DAN. 另外, 在可缓存数据比例过小的情况下,  $r, t$  和  $n$  取值增大时, DAN 的内容共享能力低于 TCP/IP, 而随着可缓存数据比例增加, DAN 的内容共享能力逐渐高于 TCP/IP, 因此, 在部署 DAN 时, 需综合权衡经济开销和可缓存数据比例使其内容共享能力最优.

## 6 DAN 原型系统

为了对 DAN 的内容共享能力进行原理性验证, 课题组研发了 DAN 的内容共享原型系统. 下面简单介绍包括内容聚类与解析、广播内容分发以及个性化兴趣适配等在内的 DAN 主要功能模块.

对于内容聚类与解析, 使用网络爬虫技术对互联网的网页进行爬取, 并采取 URL 去重、内容去重以及信息页存储等技术. 解析本地存储的网页文件并提取网页的关键信息, 生成 UCL. 用户通过 UCL 信息门户界面进行浏览, 界面如图 5 所示.

对于广播内容分发, 广播发送端从内容数据库中选择将要发送的文件, 将文件数据封装新型 TS 流进行广播发送. 广播接收端对 TS 流进行还原和个性化过滤, 还原成文件并存储本地数据库.

对于个性化兴趣适配, 一方面通过 UCL 特征向量和用户兴趣特征向量的相似度计算, 得到 UCL 的兴趣适配度值实现实时推荐, 另一方面基于用户的协同计算综合分析群体用户的历史行为, 通过相似用户进行评分预测实现离线推荐.

## 7 总结

为了实现互联网普惠全民(The Internet is for Eve-



图5 UCL信息门户界面

ryone), 李幼平院士提出了利用双结构网络来解决互联网内容共享的问题. 本文首先详细介绍了 DAN 实现模型, 并对其进行了形式化描述, 然后详细介绍了 UCL 的功能以及 UCL 驱动的内容共享机理, 总结了当前 ICP 生产驱动的内容运营模式的弊端, 并详细阐述了消费驱动内容运营模式特点, 从内容共享机理和内容运营模式两个方面定性分析 DAN 内容共享能力. 接着采用 2ACT 分析模型从边缘缓存和经济开销两个维度对 DAN 内容共享能力进行定量分析, 和 TCP/IP 相比, 随着单跳数据转发性能比值的增大, 当可缓存数据比例增加到某个阈值时, DAN 内容共享能力优于 TCP/IP, 而与 ICN 相比, DAN 能够以更少的缓存数据来获取更高的内容共享能力. 另外, 若使 DAN 内容共享能力最优, 在部署实施时需综合权衡经济开销和可缓存数据比例. 最后通过双结构网络原型系统从基于 UCL 的内容聚类与解析、广播内容分发以及个性化兴趣适配等方面原理性验证 DAN 的内容共享能力.

## 致谢

衷心感谢李幼平院士对本文研究工作的悉心指导, 感谢审稿专家和编辑老师的宝贵意见.

## 参考文献

- [1] VINT Cerf. The Internet is for everyone [OL]. RFC 3271, 2016-10-10.
- [2] SATHIASEELAN A, Crowcroft J. LCD-Net: lowest cost denominator networking [J]. ACM Sigcomm Computer Communication Review, 2013, 43(2): 52-57.
- [3] SATHIASEELAN A. Global Access to the Internet for All Research Group [OL]. <https://irtf.org/gaia>, 2016-10-01.

- [4] KATHY Brown. Global Internet Report 2015 [OL]. <http://www.internetsociety.org/globalinternetreport>, 2016-09-04.
- [5] International Telecommunications Union. The State of Broadband 2016; Broadband Catalyzing Sustainable Development [OL]. <http://www.broadbandcommission.org/publications/Pages/SOB-2016.aspx>, 2016-10-20.
- [6] NYGREN E, SITARAMAN R K, SUN J. The Akamai network: a platform for high-performance internet applications [J]. ACM Sigops Operating Systems Review, 2010, 44(3): 2-19.
- [7] 杨鹏, 李幼平. 播存网络体系结构普适模型及实现模式 [J]. 电子学报, 2015, 43(5): 974-979.  
YANG P, LI Y P. General architecture model of broadcast-storage network and its realization patterns [J]. Acta Electronica Sinica, 2015, 43(5): 974-979. (in Chinese)
- [8] SATHIASEELAN A. Researching the global access to the internet for all (gaia) [J]. IETF Journal, July 2014.
- [9] MARTIN Potts, MARTEL Consulting. Architecture for an Internet for everybody [OL]. <https://rife-project.eu>, 2016-10-09.
- [10] TROSSEN D, SATHIASEELAN A, Ott J. Towards an information centric network architecture for universal internet access [J]. ACM SIGCOMM Computer Communication Review, 2016, 46(1): 44-49.
- [11] DANELLIE Young, BJARNE Munch. Hype cycle for networking and communications [R]. Gartner, Inc, 2015.
- [12] JACOBSON V, SMETTERS D K, THORNTON J D, et al. Networking named content [A]. International Conference on Emerging NETWORKING Experiments and Technologies [C]. ACM, 2009. 117-124.
- [13] OneWeb [OL]. <http://oneweb.world>, 2016-09-11.
- [14] SpaceX [OL]. <http://www.spacex.com>, 2016-08-01.
- [15] LeoSat [OL]. <http://leosat.com>, 2016-10-03.
- [16] ALVAREZ J, WALLS B. Constellations, clusters, and communication technology: Expanding small satellite access to space [A]. IEEE Aerospace Conference [C]. IEEE, 2016. 1-11.
- [17] Loon [OL]. <https://www.solveforx.com/loon>, 2016-08-05.
- [18] Free Basics [OL]. <https://info.internet.org>, 2016-09-05.
- [19] Outernet [OL]. <https://outernet.is>, 2016-09-20.
- [20] TYSON G, SATHIASEELAN A, OTT J. Could we fit the Internet in a box? [A]. Workshop on Embracing Global Computing in Emerging Economies [C]. Springer International Publishing, 2015. 21-30.
- [21] Internet-in-a-Box [OL]. <http://internet-in-a-box.org>, 2016-07-26.

- [22] LibraryBox[OL]. <http://librarybox.us>,2016-06-25.
- [23] Library For All[OL]. <http://www.libraryforall.org>, 2016-05-17.
- [24] 李幼平,杨鹏. 共享文化大数据的新机制[J]. 中国计算机学会通讯,2013,5(9):36-40.  
LI Y P, YANG P, A new mechanism for sharing cultural big data[J]. Communications of the CCF,2013,5(9),36-40. (in Chinese)
- [25] XU K, ZHU M, WANG N, et al. The 2ACT model-based evaluation for in-network caching mechanism [A]. 2013 IEEE Symposium on Computers and Communications (ISCC) [C]. USA:IEEE,2013.000636-000641.
- [26] HUFFAKER B, FOMENKOV M, MOORE D, et al. Measurements of the Internet topology in the Asia-Pacific Region[J]. Microscopy & Microanalysis, 2010, 34(3): 109-113.
- [27] 朱敏,徐恪,林嵩. 面向应用适应能力的互联网体系结构评估方法[J]. 计算机学报, 2013, 36(9): 1785-1798.

ZHU M, XU K, LIN S. The evaluation method towards the application adaptability of Internet architecture [J]. Chinese Journal of Computers, 2014, 36(9): 1785-1798. (in Chinese)

#### 作者简介



刘 旋 男. 1985 年出生, 山东济宁人. 现为东南大学计算机科学与工程学院在读博士研究生, 师从于中国工程院李幼平院士, 主要研究方向包括双结构未来网络、信息中心网络等.

E-mail: yusuf@seu.edu.cn



杨 鹏(通信作者) 男. 1975 年出生, 四川南充人, 副教授, 博士. 现为东南大学未来网络研究中心副主任、国家专业标准化技术委员会委员、新华社统一内容标签(UCL)标准工作组副组长, 主要研究方向包括新一代网络体系结构、双结构未来网络、创新内容标识技术、网络内容大数据治理、分布式计算等.

E-mail: pengyang@seu.edu.cn