软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(3):736−759 [doi: 10.13328/j.cnki.jos.004947]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

绿色网络理论与技术研究*

张国强[1], 许自取[1], 刘 真[2]

[1](南京师范大学 计算机科学与技术学院,江苏 南京 210046)

[2](北京交通大学 计算机与信息技术学院,北京 100044)

通讯作者: 张国强, E-mail: guoqiang@ict.ac.cn

摘 要: 互联网的飞速发展,也使网络能耗急剧增长.但目前网络设备能效低下,未实现能耗比例计算的理念.而网络却为峰值负载而设计,在众多时间处于低负载,存在巨大的节能契机.首先介绍网络设备的能耗模型,继而从两方面阐述网络能耗优化的理论与技术:一方面,在假设网络总流量无法改变的前提下,为网络设备增加能源和性能状态,并优化本地控制策略,从而使单个网络设备实现能耗比例计算,或者在不提高现有网络设备能效的前提下,通过网络范围的协同和流量工程,使网络整体实现能耗比例计算的理念;另一方面,为网络提供缓存能力可以降低或缓解网络流量,从而减少网络的传输能耗或缓解其增长速度,智能的缓存部署、内容存储和请求路由能够进一步优化网络的能耗.在上述基础上,比较了各种技术的优劣,并分析了未来的研究方向.

关键词: 绿色网络;能耗;能耗比例计算;能耗感知路由;泛在缓存

中图法分类号: TP393

中文引用格式: 张国强,许自取,刘真.绿色网络理论与技术研究.软件学报,2016,27(3):736−759. http://www.jos.org.cn/1000-9825/4947.htm

英文引用格式: Zhang GQ, Xu ZQ, Liu Z. Research on green network theory and technology. Ruan Jian Xue Bao/Journal of Software, 2016,27(3):736−759 (in Chinese). http://www.jos.org.cn/1000-9825/4947.htm

Research on Green Network Theory and Technology

ZHANG Guo-Qiang[1], XU Zi-Qu[1], LIU Zhen[2]

[1](School of Computer Science and Technology, Nanjing Normal University, Nanjing 210046, China)

[2](School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: The fast development of the Internet also results in sharp rise in network energy consumption. The energy efficiency of present network devices is very low, which is an obstacle in realizing energy-proportional computing. However, while networks are designed for peak hour load, actual network load can be quite low in a large fraction of time, leaving great opportunity for energy saving. After an introduction of the energy consumption model of network devices, theories and technologies for optimizing network energy consumption are studied from two aspects. First, when the total amount of network traffic cannot be changed, it is possible to add power and performance states to individual network devices and optimize local control policies to realize energy proportionality; or, without sophisticated power management capabilities, it is also possible for a network as an ensemble to realize energy proportionality through

* 基金项目: 国家自然科学基金(61572256, 61174152, 61303243, 61370227); 江苏省自然科学基金(BK20141454); 江苏省未来网络创新研究院未来网络前瞻性研究项目(BY2013095-2-11); 中央高校基本科研业务费专项资金(2015JBM042); 赛尔网络下一代互联网技术创新项目(NGII20150102)

Foundation item: National Natural Science Foundation of China (61572256, 61174152, 61303243, 61370227); National Science Foundation of Jiangsu Province (BK20141454); Jiangsu Future Networks Innovation Institute Perspective Research Project on Future Networks (BY2013095-2-11); Fundamental Research Funds for the Central Universities (2015JBM042); CERNET Next Generation Internet Technology Innovation Project (NGII20150102)

收稿时间: 2015-08-01; 修改时间: 2015-09-14, 2015-10-17; 采用时间: 2015-12-02; jos 在线出版时间: 2015-12-30

CNKI 网络优先出版: 2015-12-30 17:12:36, http://www.cnki.net/kcms/detail/11.2560.TP.20151230.1712.001.html

© 中国科学院软件研究所 http://www.jos.org.cn

coordination and traffic engineering. Second, it is possible to reduce or alleviate the traffic by providing cache capability to the network, which in turn reduces or alleviates the growth of network transmission energy. Intelligent cache deployment, content cache and request routing policies can further optimize network energy consumption. Based on the above discussion, a thorough comparison between different technologies is provided and potential future research directions are analyzed.

Key words: green network; energy consumption; energy-proportional computing; energy-aware routing; pervasive caching

1 研究背景

虽然互联网有助于提高传统行业的生产率和能效,但随着互联网的膨胀,其自身的能耗变得不容忽视.能源技术咨询和投资顾问公司数字能源小组(digital power group)的研究表明,不包括数据中心网络(data center network,简称 DCN)在内的运营商网络基础设施的能耗保守估计为 250TWh/年~400TWh/年[1].DCN 的能耗占数据中心能耗的 5%~10%左右[2],而由于采用了虚拟化等技术,数据中心服务器的能耗增长速度在放缓,DCN 的能耗占比逐渐上升.移动互联网的普及和用户对高质量多媒体数据的需求等进一步加剧了网络流量的增长,加速了网络能耗的增长速度.统计表明:从 2007 年~2012 年,网络能耗的年均复合增长率达到 10.4%[3],超过了 PC 和数据中心的能耗增长速度.全球电子可持续发展推进组织(global e-sustainability initiative,简称 GeSI)的数据显示:按目前的趋势,欧洲电信业到 2020 年将消耗 35.8TWh 的能源,10 年将增长 67%.2007 年,意大利电信便已经成为仅次于国家铁路系统的第二大耗能实体,每年的能耗超过 2TWh[4].

为了提高网络的能效,研究界进行了大量的研究.现有研究大致可分为两大类:一类是在假定数据流量增长趋势不可逆转的前提下,提高网络的传输能效;另一类是试图缓解网络流量的爆炸性增长,实现高能效的网络.前者可以进一步分为:(1) 通过提供硬件支持和优化本地策略实现单个网络设备的能耗比例计算(即,能耗与工作负载成正比)能力,如高能效以太网(energy efficient ethernet,简称 EEE)[5,6]、自适应链路速率(adaptive link rate,简称 ALR)[7]、自适应缓存(adaptive buffer)[8]等技术;(2) 通过网络设备之间的协同使得网络作为一个整体实现能耗比例计算,典型的是能耗感知路由(energy aware routing,简称 EAR)[4,9−27].

无论是单个节点还是网络整体,均可以看成一个系统.如何提高网络资源利用率、降低网络能耗,同时保证网络服务质量,这本身是一个高能效计算问题.文献[28]概括了高能效计算问题的 3 要素.

1) 能耗模型.详细规定了每个能源状态的功耗、状态转换的延迟以及状态转换的能耗;
2) 性能约束集.给出计算问题的性能约束条件,如完成时间、端对端传输延迟、最大链路利用率等;
3) 能耗优化器.基于能耗模型和现有资源的工作负载调度器,使得系统满足性能约束条件,同时最小化能耗.

本文第 2 节介绍网络设备的能耗模型.第 3 节和第 4 节分别从单个网络设备和网络设备协同的角度介绍相应的节能技术.第 5 节从降低/缓解流量爆炸的角度介绍相关的网络节能技术.第 6 节对各种技术予以分析和比较.第 7 节探讨未来的研究方向.第 8 节总结全文.

2 网络设备的能耗模型

2.1 抽象能耗模型

网络设备的能耗可分为与网络负载无关的基准能耗和依赖于网络负载的动态能耗.精确度量网络设备各个部件的能耗极具挑战性.目前,市场上的路由器或交换机设备空载时的基准能耗占其总能耗的比例过大,未实现能耗比例计算的理念[29],即便网络的负载为 0,系统设备的基准能耗也超过了 50%[12,13,30,31].

抽象地,若将网络设备看成一个整体,现有研究大都认为设备的基准功耗 EG_0 较大,而随着负载增加到最大值,达到最大功耗 $EG_{\max}$.一般认为:在$[EG_0,EG_{\max}]$之间,功耗随负载线性增加[17,31].文献[17]给出了 3 种能耗模型(如图 1 所示),可以抽象概括现有的研究.第一种 IdleEnergy 模型具有较高的基准功耗,与大部分现网设备比较接近.而另两种为极端情况:一种是功耗与负载完全线性相关的模型,另一种则是功耗与负载完全无关的 Energy-

© 中国科学院软件研究所 http://www.jos.org.cn

agnostic 模型.

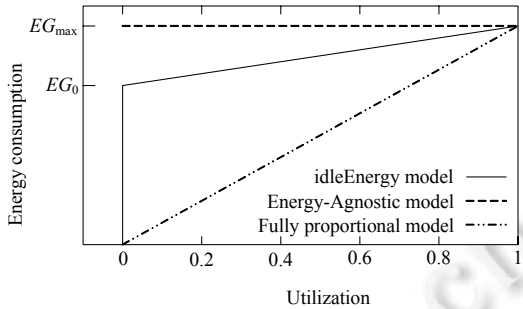


Fig.1 Three energy models for network devices[19]

图 1 网络设备的 3 种能耗模型[19]

2.2 具体网络设备能耗模型

深入到路由器内部,如图 2 所示,现代的高端路由器由不同的部件构成[13,31−34].不同部件的功耗与负载可能呈现不同的函数关系:(1) 亚线性增长关系,例如以太网链路[13,14,19];(2) 线性关系,例如 Batcher,Crossbar 等交换架构[19,35];(3) 超线性关系,例如采用 DVS 和 DVFS 等技术的处理器[14,19,36].

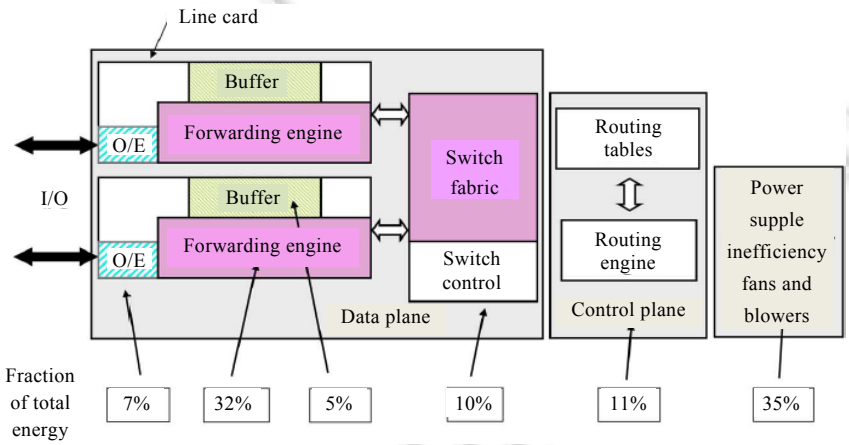


Fig.2 Components of a high-end router and its energy decomposition[32]

图 2 高端路由器的部件构成及其能耗分解[32]

路由器的能耗大致可分为 3 部分[32,34]:(1) 与路由协议相关的控制层能耗;(2) 与数据转发相关的数据层能耗;(3) 与供电和降温相关的能耗.图 2 同时给出了路由器各个部分能耗分解的大致比例**.通常,包括底盘和交换背板在内的基准能耗占路由器能耗的 25%~50%[13,32,37],缓冲存储占路由器能耗的 5%~10%[8,32],所有线卡(包括端口)占整个路由器能耗的 40%左右[13,32,37].而传统路由器由负载引起的动态能耗通常较小[13,17,31,38,39],在 5%~15%.据此,根据路由器或交换机的配置,其功耗可以建模为[13,38]

$$P^{switch}=P^{chassis}+\sum_{i=1}^{L}P^{LineCard}(i)+\sum_{i=1}^{L}\sum_{j=1}^{L_i}(P^{portStatic}(i,j)+P^{portDynamic}(i,j)),$$

其中,L 为线卡数,L_i 表示第 i 个线卡的端口数量,$P^{portDynamic}(i,j)$是与第 i 个线卡的第 j 个端口的负载相关的动态

** 该比例会随着路由器线卡的数量而变化,当线卡数量增多时,线卡占整个路由器的能耗比例会随之增加,而其他部分的能耗占比会随之下降.此外,随着单个设备逐渐引入能耗比例计算的理念,路由器与动态负载相关的能耗占比也将加大.

© 中国科学院软件研究所 http://www.jos.org.cn

功耗.

对于由负载引起的动态能耗,可以进一步将其分为两部分[34,40]:(1) 报文处理的能耗,如更新 TTL、最长前缀匹配等;(2) 存储和传输报文数据的能耗.前者与报文长度无关,仅与报文数量相关,因此,比特到达率相同时,报文长度减小会增加路由器的处理能耗[39];而后者与报文的字节数直接相关.

对于相同类型的网络物理链路,其工作速率基本决定了其功耗.而处于相同工作频率时,负载为 0 和满负载时链路的功耗基本相同,也就是上述的 $P^{portDynamic}(i,j)$基本为 0[7].

而对于路由器 i,j 之间的逻辑连接,可以采用 IEEE 802.1AX[41]定义的 Bundled Link 实现,即,由一组物理链路提供聚合的逻辑带宽.例如,一条 40Gbps 的逻辑链路可以由 4 根带宽为 10Gbps 的 OC-192 电缆组成.根据是否支持 Trunking,可以将这类链路分为两类.

- 第 1 类为 Trunk 链路,其具备两个特征:一方面,逻辑链路中的物理链路可以被单独关闭,从而使得整个逻辑链路的能耗随负载变化呈阶梯变化;另一方面,物理链路一端的线卡可以随着链路的流量呈现多个离散的能耗状态;
- 第 2 类非 Trunk 链路则不具备上述两个特征.

文献[40]给出了上述两类逻辑链路的功耗模型.假设逻辑链路 l 有 n_l 条物理链路构成,链路当前逻辑链路的负载为 x_l,则对于非 Trunk 链路,每条物理链路的负载为 $\frac{x_l}{n_l}$,逻辑链路的总功耗为

$$2n_l\left(\delta_l+\rho_l\frac{x_l}{n_l}+u_l\left(\frac{x_l}{n_l}\right)^{\alpha_l}\right),$$

其中,δ_l 表示物理链路的一端在空闲时的基准功耗;ρ_l 为常数, $\rho_l\frac{x_l}{n_l}$ 表示物理链路的一端存储和传输负载报文所需的功耗;而 $u_l\left(\frac{x_l}{n_l}\right)^{\alpha_l}$ 则表示链路的一端处理负载报文所需的功耗,u_l 为常数.

而对于 Trunk 链路,其功耗为

$$2(n_l-n_c)\left(\delta_l-\delta_c+\rho_l\frac{x_l}{n_l-n_c}+u_l\left(\frac{x_l}{n_l-n_c}\right)^{\alpha_l}\right),$$

其中,n_c 表示在当前负载下可以关闭的物理链路数量.因此,负载将被均摊到剩下的 n_l-n_c 条物理链路上.δ_c 则表示物理链路一端的线卡为满足当前负载所需工作的能耗状态能够在δ_l基础上减少的功耗.

3 单个网络设备的节能技术

单个网络设备应从两方面为节能提供支持:

(1) 硬件:提供类似于 ACPI(advanced configuration and power interface)[42]的能源和性能状态,并能自适应地根据负载调整其状态,或允许控制程序控制其状态;

(2) 本地控制策略:根据流量特征对设备的能源或性能状态进行主动控制,在能耗与性能之间进行有效的折衷.

3.1 网络设备的能源状态

ACPI 为通用计算机系统及各个组成部件的能源和性能状态定义了通用的接口,不同的设备或部件可以支持其子集.ACPI 定义了对用户可见的系统全局能源状态,其中最主要的是工作状态(G_0)和睡眠状态(G_1).其中,G_1 可以进一步细分为多个睡眠子状态 $S_1,S_2,...,S_5$.

- 当系统处于 G_0 时,系统的某些部件可以处于睡眠或关闭状态;
- 而当系统处于 G_0 状态时,系统中最重要的子部件处理器可以处于活跃状态(C_0)或不同程度的睡眠状态($C_1,C_2,...,C_n$).

© 中国科学院软件研究所 http://www.jos.org.cn

此外,ACPI 也为活跃部件定义最多不超过 16 个性能级别 $P_0,P_1,\ldots,P_n$.

但目前,市场上的网络设备极少支持 ACPI 标准.IETF 的 eman 工作组[43]为网络设备定义了能源和管理接口标准,通过定义与能源对象的能源状态相关的 MIB 对象,用于管控网络实体的能源状态[44,45].ETSI 的绿色抽象层(green abstract layer,简称 GAL)标准[46,47]使能源管理程序可以用一致的方式获取网络设备的能源管理能力、它们潜在的能耗和对网络性能的影响以及如何与之交互等信息,从而网络管理员可以有效对网络设备的能源进行管理,在能耗和网络性能之间进行折衷.GAL 通过层次化方式组织一个网络设备的能耗感知实体(energy aware entity,简称 EAE).类似于 ACPI,GAL 也支持能源状态和性能状态,如图 3 所示.

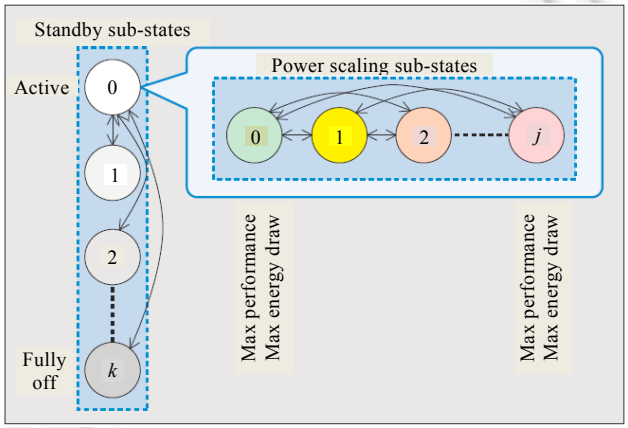


Fig.3 An abstract representation of energy and performance states for network devices based on GAL[49]

图 3 基于 GAL 的网络设备能源状态和性能状态抽象表示[49]

目前,以太网卡的能耗比例计算能力得到了深入的研究.IEEE 802.3az 标准[5,6]定义了网卡在活跃/休眠状态之间进行快速切换的机制.实现了该标准的以太网卡相连时,如果链路没有流量,网卡可以自动进入睡眠状态.支持 802.3az 的以太网卡可以和不支持的以太网卡并存,其效果就和不实现 802.3az 一样,因此可以前向兼容.

当前的众多绿色网络研究都假设网络接口具备下述两种能力之一:一种仅支持 On/Off,另一种支持自适应频率调节[48].但上述能力一般都作为理论研究的前提,认为可以通过统一的控制或自动感知实现 On/Off 状态的切换或不同工作频率的切换.但是目前只有 EEE 实现了基于网络负载的活跃/休眠状态的切换.ALR 虽然也得到了研究[7],但并未进入标准化.其中的一个原因在于不同速率的切换时间较长,为秒级[49].

线卡包含独立的处理器、存储以及多个网络接口.现有研究基本假设线卡具有 On/Off 两种状态,一般默认其处于 On 状态,仅当它的所有接口都处于 Off 状态时,才能将线卡整个置于 Off 状态.目前,对于线卡的切换时间并没有统一认识.文献[30]提到:若假设线卡在休眠时其检测线路处于活跃状态,则一旦检测到输入端口有数据到达就能唤醒线卡,文章假设唤醒时间为 10us.文献[48]则假设,部件在不同的能源状态之间切换的时延为 ms 级.但上述假设对目前的路由器产品可能都过于乐观.文献[33]认为,一个具有 200K 路由条目的线卡需要 5m 才能从关闭状态切换到工作状态.文章通过重新设计路由器架构,可以使线卡的唤醒时间缩小至 127ms.复杂的线卡不仅包含网络接口,还包含处理器.处理器可以采用 DVS 和 DVFS 等方式节能[30].现代的高端路由器转发引擎通常使用 TCAM(ternary content addressable memory)进行基于硬件的报文分类和匹配,但 TCAM 以高能耗著称.近年来,TCAM 厂商提供了将其分成多个部分且仅检索其中某部分的能力.基于此,可以通过软件算法大幅降低 TCAM 的能耗[50,51],从而也能部分降低线卡的能耗.

Chassis 的能耗占整个路由器能耗的近一半.目前的研究也基本假设 Chassis 具有 On/Off 两种状态.仅当所有的线卡都处理 Off 状态且没有流量经过路由器时,Chassis 才能被显式置于 Off 状态,On/Off 状态之间的切换时间也达到数分钟之久[9,15].因此,未来的路由器或交换机应支持多个睡眠状态,使其能快速从睡眠状态切换到

© 中国科学院软件研究所 http://www.jos.org.cn

活跃状态.

文献[30]提到:在线卡睡眠时可以将Crossbar也置于睡眠状态,唤醒线卡时同时唤醒Crossbar.如果线卡不睡眠但 Crossbar 睡眠,则可能会丢包.因此,只有当所有的线卡都睡眠时才能将 Crossbar 置于睡眠状态.

Buffer(包括 SRAM,DRAM)消耗了路由器约 5%~10%的能源.文献[8]将 Buffer 分为片上缓存、SRAM 缓存和 DRAM 缓存这 3 种,其中,片上缓存总是处于活跃状态.而 DRAM 具有 3 个能源状态:Active(正在进行高频率的读/写操作),Idle(基本没有读/写操作),Sleep(无法进行读/写操作).DRAM 可以按行处理,即,每一行处于不同的状态.而 SRAM 的能耗与负载基本无关,因此仅包括 Active 和 Sleep 两种能源状态.从 Sleep 状态切换到 Active 状态需要额外的时间,在此期间到达的报文可能由于空间不足而被丢弃.

3.2 本地策略

虽然硬件可以提供能耗比例计算的能力,但一方面,在不同的能源状态之间切换需要时延;另一方面,切换可能付出额外的能耗[9,37,48].因此,当网络流量呈现高度动态时,任由硬件在不同的状态间切换可能无法达到节能目的.因此,需要本地系统通过策略优化设备在能源状态之间的切换,避免无谓的耗能.

文献[30]提出了单个网络设备非协同睡眠的概念.非协同睡眠需考虑两个问题:

(1) 何时进入/退出睡眠状态?

(2) 网络部件能睡眠多久?

第(1)问题正是本地策略可以控制的.对于何时进入睡眠,一种做法是,当网卡空闲时进入睡眠;另一种是如文献[37]提出的基于窗口预测的策略,即:如果在一个时间窗口内观测到的报文数量低于某个阈值,则将端口切换到睡眠状态一定的时间.对于何时唤醒网络端口,也有两种策略:

- 定时器控制的睡眠,定时器的值可以依据报文到达间隔估算而得.该策略的问题在于对下一个报文到达的预测准确性直接影响了效率.定时器设置过小会过早退出睡眠,而定时器值设置过大则会产生丢包;
- Wake-on-Arrival(WOA),即,当新报文到达时唤醒端口.但由于唤醒需要时间,因此会产生丢包.为解决这个问题,文献[30]提出了一种避免丢包的方法,即:当一个端口希望向一个处于睡眠状态的接口发送报文时,首先发送一个 dummy 报文以唤醒处于睡眠状态的端口,在等待必要的时间后再发送真正的报文.但是,该方法会增加端对端延迟.

对于第(2)问题,网络部件能够睡眠的时间长度取决于报文到达间隔(对于路由器而言,一个上限是路由控制协议如 OSPF 的定时 HELLO 消息),若报文到达过于频繁,则网络部件会频繁在 active/sleep 状态间切换.频繁的切换不利于节能,也不利于设备的使用寿命[52].虽然机会性睡眠对局域网设备可能会奏效[53],但核心网的负载远高于局域网,机会性睡眠效果不佳.为此,如图 4 所示,文献[48]提出了一种 Burst-and-Buffer 机制,即:在入口路由器 S 处对流量进行缓存,人为制造突发流量;同时,S 对报文进行重新排序,使去往同一出口路由器的报文在缓存队列中彼此相邻,从而中间路由器 R 在被唤醒后可以处理一批报文.通过 R 的各个端口转发的报文也呈现突发状态,因此,R 及其下游路由器可以睡眠更长时间.

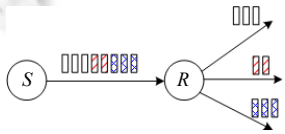


Fig.4 Burst-and-Buffer scheme[51]

图 4 Burst-and-Buffer 机制[51]

文献[48]也提出了在支持速率调节的硬件能力基础上调整端口处理速度的策略,其所提出的 practRA 策略根据预测的未来报文到达速率和当前队列的长度调整端口的速率.研究结果还表明:离散速率值的分布为均匀分布时,节能效果比指数分布要好.

© 中国科学院软件研究所 http://www.jos.org.cn

文献[7]提出了基于 Ethernet 握手 MAC 帧进行 Ethernet 速率快速切换的机制,并提出了若干根据网络的利用率调整链路速率的策略,包括基于输出缓冲区的双门限策略、基于链路利用率门限的策略和基于定时器门限的策略.当报文的到达率满足某个特殊条件时(即:报文到达速率使链路工作在低速率时的队列长度超过 *qHigh*,但又不能使链路在高速率工作时维持队列长度超过 *qLow*),双门限策略可能会导致链路在两个速率之间震荡.基于链路利用率的策略可以有效避免这一震荡,但需要实时统计链路传输的字节数,因此需要额外的计数器和寄存器,增加了网卡的复杂度.

文献[8]对本地 Buffer 进行按需动态分配.实际上,路由器的大部分 Buffer 空间许多时候都处于空闲状态,因此可以依据网络负载按需调节 Buffer 的大小,即:当负载小时,将多余的 Buffer 置于睡眠状态,但必须保证有足够的 Buffer 处于工作状态,以应对唤醒时延内可能到达的报文;而当负载逐渐增大时,按需唤醒所需的 Buffer.研究结果表明,该方法可将 Buffer 的能耗降低 97%左右.

4 网络范围节能技术

由于网络为峰值负载和故障而设计,一般存在大量的冗余带宽和链路.因此,当单个网络设备还无法完全支持能耗比例计算时,可以通过网络设备间的协同,让网络作为一个整体实现能耗比例计算.当网络负载较低时,通过能耗感知的路由/流量工程将网络流量聚合到少量的网络设备和链路上,从而可以将其他的网络设备和链路置于低能耗状态.特别地,对于数据中心网络,还可以利用其拓扑结构的特殊性、应用模式的可预测性和网络资源分配的灵活性,为其设计更优化和简单的能耗感知路由方案[54,55].

能耗感知路由/流量工程是一个典型的高能效计算问题,通常使用链路利用率或网络吞吐力作为性能约束.在能耗模型方面,需要分别建立路由器和链路的能耗模型.能耗感知路由/流量工程的核心在于能源优化器,即,如何实现最优的决策以最大化节能效果.这类决策策略通常都是依据网络拓扑结构知识和流量需求在路由协议层面对流量进行汇聚或重新分配,允许无负载(低负载)的节点/链路进入低能耗状态(如关闭、睡眠或低性能)以达到节能目的.

基于网络设备的硬件能力,这一问题通常分为两类:第一类硬件支持基于 Powerdown 的能耗模型[15];另一类硬件支持多个性能状态,即,支持基于 Speed Scaling 的能耗模型[14].此外,还可以将问题分为全局优化和局部优化:前者通常需要集中式的控制节点进行决策,而后者通常可以采用分布式方法进行决策.无论是哪种硬件模型,全局优化一般都建模成混合整数规划问题,其输入包括:

(1) 网络结构、链路带宽、链路的权值等网络拓扑信息;

(2) 流量需求矩阵和/或流量的特征;

(3) 网络设备(包括节点和链路)的能耗模型.

4.1 基于Powerdown模型的全局优化

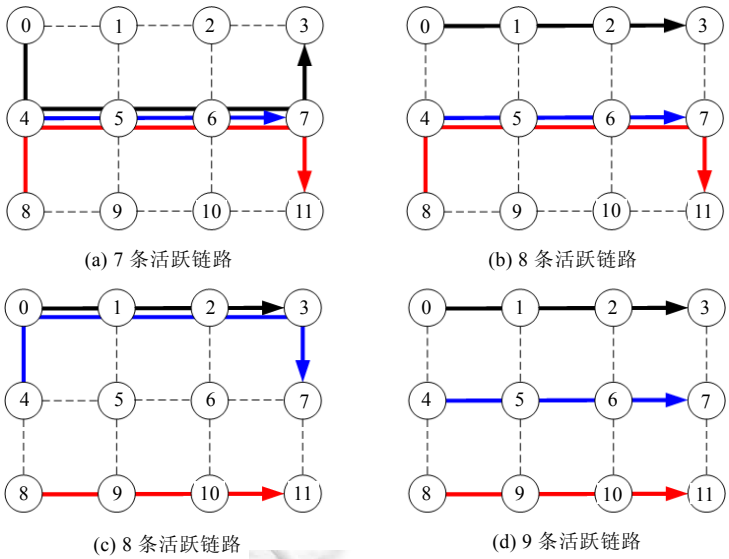
4.1.1 理论模型

在这一模型中,网络端口、线卡或节点只能处于 On/Off 两种状态.其输出包括:(1) 每个端口、线卡和节点的 On/Off 二元状态;(2) 每个流量需求的路由状态.

图 5 给出了 EAR 的例子.假设需求为 0-3,4-7 和 8-11 这 3 个流,图 5(d)采用最短路径路由,需使用 9 条活跃链路和所有节点;而若采用图 5(a)的方式聚合流量,则仅需要 7 条活跃链路和 8 个节点.当然,如果 4-5-6-7 之间的带宽无法同时满足 3 个流的聚合需求,则可能会出现图 5(b)和图 5(c)的路由配置方案,需要 8 条活跃链路和 10 个节点.

目前,针对 EAR 已有众多的建模研究,这些研究的主要区别在于它们对网络设备的能耗模型假设.抽象地,可以将网络看成是抽象的节点和逻辑链路的集合,只有节点或链路会产生能耗[4,11,12,17].深入到路由器内部,每个路由器节点则进一步是 Chassis、线卡和网络端口等构成的组合体,在研究中可以只考虑关闭线卡和网络端口,而不关闭 Chassis[10];或者可以同时关闭 Chassis 和线卡[13].对于由多条物理链路构成的 Bundled Link,文献[27,56]还考虑了如何在满足流量需求矩阵的前提下关闭最大数量的物理链路.

© 中国科学院软件研究所 http://www.jos.org.cn



(a) 7 条活跃链路 (b) 8 条活跃链路

(c) 8 条活跃链路 (d) 9 条活跃链路

Fig.5 Illustration of energy-aware routing (EAR)

图 5 能耗感知路由(EAR)示意图

下面以将网络看成是抽象的节点和链路的集合为例,给出 EAR 优化问题的建模,其他情况基本大同小异,整个模型是容量受限的多物品流问题(capacitated multi-commodity flow,简称 CMCF).

假设网络拓扑 $G(V,E)$,表 1 给出了文中所用到的大部分变量及其含义.路由的目标是对在满足流平衡和链路带宽限制的前提下,使得网络的总能耗最低,整个优化问题可以形式化描述为:

minimize:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij}PL_{ij} + \sum_{i=1}^{N} y_i PN_i \tag{1}$$

subject to:

$$\sum_{j\in A(i)} f_{ij}^{sd} - \sum_{j\in A(i)} f_{ji}^{sd} = \begin{cases} t^{sd}, & \forall s,d,i=s \\ -t^{sd}, & \forall s,d,i=d \\ 0, & \forall s,d,i\neq s,d \end{cases} \tag{2}$$

$$f_{ij} = \sum_{s=1}^{N}\sum_{d=1}^{N} f_{ij}^{sd} \leqslant \alpha c_{ij} x_{ij}, \forall ij \in E \tag{3}$$

$$x_{ij} \leqslant y_i, \forall i, \forall j \in A(i) \tag{4}$$

$$f_{ij}^{sd} \geqslant 0, \forall i,j,\forall s,d \tag{5}$$

$$x_{ij}=0,1,\forall i,j \tag{6}$$

$$y_i=0,1,\forall i \tag{7}$$

其中,式(1)为目标函数,即,整个网络的功耗;公式(2)表示 CMCF 中的流守恒约束;公式(3)表示链路的总流量不能超过链路利用率约束;公式(4)表示仅当节点相连的所有链路都关闭时,才能关闭节点.

上述模型为最基本的模型,不同的场景可能会增加额外约束.例如:上述模型允许流可分割,在某些场景下,需要增加流不可分的约束.此时,可以假设 y_{ij}^{sd} 为 0-1 决策变量,表示流 s-d 是否经过链路 i-j[11].则上述模型的流守恒约束(2)和链路利用率约束(3)变为如下的公式(8)和公式(9):

© 中国科学院软件研究所 http://www.jos.org.cn

$$\sum_{j\in A(i)} y_{ij}^{sd} - \sum_{j\in A(i)} y_{ji}^{sd} = \begin{cases} 1, & \forall s,d,i=s \\ -1, & \forall s,d,i=d \\ 0, & \forall s,d,i\neq s,d \end{cases} \tag{8}$$

$$f_{ij} = \sum_{s=1}^{N}\sum_{d=1}^{N} y_{ij}^{sd} t^{sd} \leqslant \alpha c_{ij} x_{ij}, \forall ij \in E \tag{9}$$

Table 1 Variables and their meanings in the optimization problem formulation

表 1 优化问题描述中的变量及其含义

变量	含义
V	节点集
E	边集
N	表示节点数$\|V\|$
L	表示边数$\|E\|$
c_{ij}	链路 i-j 的带宽
α	链路利用率的上限
$A(i)$	节点 i 的邻居集合
t^{sd}	源节点 s 到目的节点 d 的流量需求
PN_i	节点 i 在活跃状态的能耗
PL_{ij}	链路 i-j 在活跃状态的能耗
x_{ij}	0-1 决策变量,表示链路 i-j 是否处于活跃状态,0 表示不活跃,1 表示活跃
y_i	0-1 决策变量,表示节点 i 是否处于活跃状态,0 表示不活跃,1 表示活跃
f_{ij}^{sd}	取值为$[0,t^{sd}]$,表示 s-d 的流量需求经过链路 i-j 的流量值
f_{ij}	经过链路 i-j 的总流量
y_{ij}^{sd}	0-1 决策变量,假设流不可分,表示 s-d 的流量需求是否经过链路 i-j

4.1.2 启发式算法

由于上述模型属于混合整数规划或整数规划问题,一般属于 NP-Hard 问题.因此对于较大规模网络,一般都需要寻找启发式算法.

文献[10]给出了一种启发式算法,对于任意源和目的节点对(s,d),仅搜索长度最短的k条路径,从而大幅降低了搜索空间.在此基础上,考虑了两种更为严格的限制,即:候选路径为上述 k 条路径中路径长度不超过网络直径的子集(basic+nd)以及候选路径集合为上述 k 条路径中路径长度不超过(s,d)之间最短路径长度两倍的路径子集(basic+2e).同时,还考虑了流平衡的需求,即:在求解上述优化问题得到的活跃链路基础上,进行第二阶段以最小化最大链路利用率为目标的线性规划问题求解.

CARPO[38]利用了网络中不同流之间的相关性.两个流相关是指它们会同时达到峰值,此时不应聚合其流量;反之,则可以聚合流量.提出了一种基于贪婪装箱的启发式算法.对于任意流 f 和可行路径 RT,若 RT 的可用带宽能满足待检测流的需求,且待检测流与路径中的其他流均不相关,将流 f 的路由路径设为 RT.进一步还考虑了速率调节,即,将活跃链路的速率调整为最适应流量需求的速率.

文献[12]提出了包括基于贪婪装箱和拓扑感知的两种启发式算法.前者总是选择最左边的能满足流需求的路径分配给该流,但与 CARPO 不同的是,没有考虑流的相关性.拓扑感知的启发式算法仅适用于 FatTree[57]的规则拓扑,它不需要完整的流量矩阵信息,而仅需要实现了 OpenFlow 的端口计数器信息.

文献[4,16]提出了一种迭代式关闭网络节点或链路的启发式算法.初始时,所有节点和链路都处于活跃状态,然后通过多种策略不断地迭代尝试,每次尝试都试图关闭一个网络设备(节点或链路).由于关闭节点比关闭链路获得的能耗收益更大,因此每次都优先关闭节点.提出了包括随机、最少链路、最小流、opt-edge、最大能耗等多种尝试策略.

由于以满足网络流为出发点的优化问题求解复杂度通常较高,也有一些研究仅通过网络拓扑结构特征选择处于活跃状态的节点/链路集合,或计算潜在可能被关闭的链路或节点集合[58−60],以降低计算复杂度.例如,文献[58]基于节点度来选择所使用的最短路径树,而文献[60]则根据网络拓扑的代数连通度和链路的介数

© 中国科学院软件研究所 http://www.jos.org.cn

(betweenness)来确定可以被关闭的链路集合.

4.2 基于Speed Scaling的全局优化模型

假设每个网络设备 e 都具有自适应的处理速度缩放(speed scaling)能力,且处理速度为 s 时的能耗为 $h_e(s)$,则对于给定的流量需求矩阵,可以通过路由优化降低网络的整体能耗.路由的选择与函数 $h_e(s)$的性质紧密相关,具体地,$h_e(s)$与 s 的关系可能是下面 3 种[14]:

(1) $h_e(s)=us$,即,能耗是处理速度的线性函数.此时,最短路径路由最节省能源;

(2) $h_e(s_1)+h_e(s_2)\geqslant h_e(s_1+s_2)$,即,能耗是处理速度的亚线性函数,如以太网链路、边缘路由器、光纤链路和设备等.例如,以太网卡从 1Gbps 变为 10Gbps,能耗从 4W 变为 20W[18].这一假设下的路由优化归结为 Buy-at-Bulk 网络设计问题[61];

(3) $h_e(s_1)+h_e(s_2)\leqslant h_e(s_1+s_2)$,即,能耗是处理速度的超线性函数,如微处理器.通常,以一个多项式的函数来表示能耗与处理速度的关系,$h_e(s)=u_e s^{\alpha}(\alpha>1)$.在这一假设下,如果流可分割,则该问题是凸线性规划,存在多项式时间解;而如果流不可分,则该问题变成整数规划,一般不存在多项式时间的解.

假设处理 s 单位的数据量引起的能耗是 $h_e(s)$,且流不可分割,f_e 表示链路 e 上的总流量,则文献[14]将该路由优化问题建模为:

minimize:

$$\sum_{e\in E} h_e(f_e) \tag{10}$$

subject to:

$$\sum_{j\in A(i)} y_{ij}^{sd} - \sum_{j\in A(i)} y_{ji}^{sd} = \begin{cases} 1, & \forall s,d,i=s \\ -1, & \forall s,d,i=d \\ 0, & \forall s,d,i\neq s,d \end{cases} \tag{11}$$

$$f_e = \sum_{s=1}^{N}\sum_{d=1}^{N} y_e^{sd} t^{sd}, \forall e\in E \tag{12}$$

$$y_{ij}^{sd} \in \{0,1\}, \forall (i,j)\in E, \forall s,d\in V \tag{13}$$

但这一建模并未考虑网络的链路容量,如果考虑链路容量限制,则应加上如公式(3)或公式(9)的链路利用率约束.文献[19]对这一问题进行了建模,允许流可分割,同时包括了链路和节点的最大负载约束条件.

4.3 基于Speed Scaling的局部优化模型

文献[18]提出了一种实时动态调整链路能耗状态的方法,该方法假设链路工作在有限数量的离散频率 $r_1<r_2<...<r_n$,如 10Mbps,100Mbps,1Gbps.假设链路 l 的当前负载为 y_l,当前工作频率为 r_i,那么如果 y_l 仅略高于频率 r_{i-1}(例如,当前负载为 11Mbps,那么链路需要工作在 100Mbps),则可以将负载适当分配到其他路径(不引起其他路径的链路切换到更高的工作频率),将当前链路的负载降低到 r_{i-1} 以下,从而允许其工作频率切换到 r_{i-1}.量化地,定义满足 $w_l=\dfrac{y_l-r_{i-1}}{y_l}<dm$ (dm 为自定义的豁度阈值)的链路为可以进行负载转移的候选链路集.

与前述的全局优化不同,该方法可以分布式实现,每个入口路由器实时收集源自该节点的所有路径的链路流量并独立进行优化,根据优化结果调节负载在不同路径的分配.在报文从源发送至目的地的过程中,中间路由器在每一跳计算 w_l 的值.如果当前链路计算出的 w_l 值比当前报文携带的值小,则用当前计算出的值替换报文中的值.当报文抵达目的地后,目的地向入口路由器报告该 w_l 的值.入口路由器接收到所有目的地的信息后,会按比例地将需要移动的额外流量均匀移动到无需提升性能状态而能吸收额外负载的路径上.

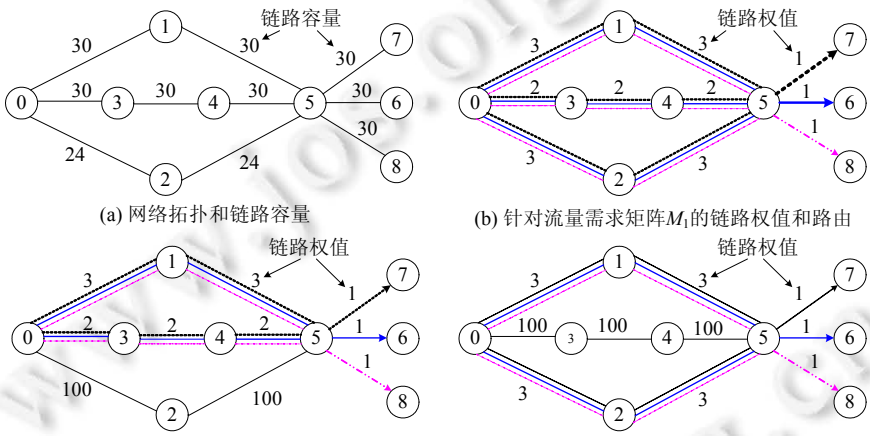
4.4 基于链路权值设置的全局优化

虽然基于流的路由是核心网的一项主要技术(如 MPLS),但基于链路权值的 OSPF 路由也是常用的路由技术.在 OSPF 中,报文会选择最短路径路由,并且支持基于 ECMP(equal cost multiple path)的流量均衡.

© 中国科学院软件研究所 http://www.jos.org.cn

通过为一些链路设置高权值,可以避免最短路径经过这些链路,从而可以将这些链路切换到睡眠状态.进一步地,没有任何流量经过的节点也可以被切换到睡眠状态.如何确定网络链路的权值,满足网络服务质量的约束,并最小化网络能耗,被称为基于最短路径路由的能耗感知流量工程(energy-aware traffic engineering with shortest path routing,简称 E-TESP)问题[62],最近得到了研究界的重视[58,59,62−65].E-TESP 问题的输入和约束条件与基于 Powerdown 模型的 EAR 问题基本相同,而输出则是每条链路的 OSPF 权值以及由此而导出的每个节点/链路的活跃状态.

图 6 给出了这一问题的示例.图 6(a)标注了链路的带宽,表 2 给出了不同时刻的 3 个流量需求矩阵 M_1,M_2,M_3,其中,整个网络的流量需求 M_1 最大,M_3 最小.当负载为 M_1 时,一个可行的路由方案和权值设置如图 6(b)所示,此时,为了满足链路容量约束,所有的链路都必须置于活跃状态;而当负载为 M_2 时,图 6(c)给出了优化的路由方案和对应的权值设置,此时,链路 0-2,2-5 和节点 2 都可以切换到睡眠状态;最后,当负载为 M_3 时,则优化的方案如图 6(d)所示,此时,可以关闭 3 条链路和两个节点.



(a) 网络拓扑和链路容量 (b) 针对流量需求矩阵M_1的链路权值和路由

(c) 针对流量需求矩阵M_2的链路权值和路由 (d) 针对流量需求矩阵M_3的链路权值和路由

Fig.6 An example of energy aware routing based on OSPF link weight

图 6 基于 OSPF 链路权值的能耗感知路由示意图

Table 2 Three flow demands in Fig.6

表 2 图 6 中不同的 3 个流量需求

流量需求矩阵	流量需求		
	(0,6)	(0,7)	(0,8)
M_1	30	30	10
M_2	20	20	10
M_3	20	10	10

E-TESP 也可以建模为混合整数规划问题.除了表 1 的变量外,E-TESP 模型还引入了几类新变量.设 w_{ij} 为决策变量,表示链路 i-j 的权值;u_{ij}^d 为 0-1 变量,表示链路 i-j 是否位于从 i 到 d 的最短路径上;z_i^{sd} 表示按照 ECMP 路由规则,节点 s 经过 i 的 s-d 流量分配到等价路径上的流量与 t^{sd} 的比例;r_i^d 表示从节点 i 到 d 的最短路径长度.当链路的权值确定以后,u_{ij}^d 和 z_i^{sd} 也随之确定.从而,整个优化问题可以表示为

minimize:

$$\sum_{i\in V} PN_i y_i + \sum_{(i,j)\in E} PL_{ij} x_{ij} \tag{14}$$

subject to:

$$x_{ij} \leqslant y_i, \forall (i,j)\in E \tag{15}$$

© 中国科学院软件研究所 http://www.jos.org.cn

$$x_{ij} \leqslant y_j, \forall (i,j) \in E \tag{16}$$

$$\sum_{j \in A(i)} (f_{ij}^{sd} - f_{ji}^{sd}) = \begin{cases} t^{sd}, & i = s \\ -t^{sd}, & i = d \\ 0, & i \neq s \neq d \end{cases}, \forall i \in V, (s,d) \in D \tag{17}$$

$$\sum_{(s,d) \in D} f_{ij}^{sd} \leqslant \alpha c_{ij} x_{ij}, \forall (i,j) \in E \tag{18}$$

$$0 \leqslant z_i^{sd} - f_{ij}^{sd} / t^{sd} \leqslant 1 - u_{ij}^d, \forall (s,d) \in D, \forall (i,j) \in E \tag{19}$$

$$f_{ij}^{sd} / t^{sd} - u_{ij}^d \leqslant 0, \forall (s,d) \in D, \forall (i,j) \in E \tag{20}$$

$$1 - u_{ij}^d \leqslant r_j^d + w_{ij} - r_i^d \leqslant (1 - u_{ij}^d) M, \forall d \in D_d, \forall (i,j) \in E \tag{21}$$

$$u_{ij}^d - x_{ij} \leqslant 0, \forall d \in D_d, \forall (i,j) \in E \tag{22}$$

$$w_{ij} \geqslant (1 - x_{ij}) w_{\max}, \forall (i,j) \in E \tag{23}$$

$$x_{ij} + w_{ij} \leqslant w_{\max}, \forall (i,j) \in E \tag{24}$$

$$1 \leqslant w_{ij} \leqslant w_{\max}, \forall (i,j) \in E \tag{25}$$

$$x_{ij}, y_i \in \{0,1\} \tag{26}$$

$$z_i^{sd} \in [0,\ 1] \tag{27}$$

$$r_i^d \geqslant 0 \tag{28}$$

其中,M 是一个很大的常数,可以设为 $M=2w_{\max}$,$w_{\max}$ 为可设置的最大权值.

在上述建模中:

- 公式(14)为目标函数;
- 公式(15)、公式(16)表示仅当所有邻接链路都被关闭时才可以关闭节点;而当节点关闭时,所有邻接链路都必须关闭;
- 公式(17)和公式(18)分别是流守恒约束和带宽利用率约束;
- 公式(19)为 ECMP 约束,如果 $u_{ij}^d = 1$,即,*i-j* 位于 *i-d* 的最短路径上,则 *i-j* 承载的 *s-d* 的流量占 t^{sd} 的比例是 z_i^{sd};
- 公式(20)表示如果 $u_{ij}^d = 0$,即,*i-j* 不位于 *i-d* 的最短路径上,则 *i-j* 承载的 *s-d* 流量为 0;
- 公式(21)和公式(22)保证路由行为和最短路径的一致性,其中:公式(21)表示如果 *i-j* 在 *i-d* 的最短路径上,则 *i-j* 的权值 w_{ij} 为 $r_i^d - r_j^d$,否则至少为 $r_i^d - r_j^d + 1$;公式(22)表示当链路 *i-j* 关闭时,*i-j* 肯定不在 *i-d* 的最短路径上;
- 公式(23)~公式(25)表示如果 *i-j* 关闭,即 $x_{ij}=0$,则 *i-j* 的权值 $w_{ij}=w_{\max}$;反之,如果 $w_{ij}=w_{\max}$,则链路 *i-j* 处于关闭状态.

4.5 面向受网络限制的流的全局优化

上述优化均基于下述假设,即,不允许改变流的时间传输特性.也即,流应满足稳定的传输速率.

文献[23,24,66]则给出了针对另一类受网络限制(network-limited)的流的优化方法,这类流的特征是应用程序可以按足够快的速度产生流量,从而耗尽可用带宽.这类流的完成时间,取决于路径上由于不同流竞争而存在的瓶颈链路.为了降低这类流引起的网络能耗,不仅要考虑减少参与流传输的活跃网络设备的数量,还要考虑减少网络设备参与传输的时长.图 7 给出了针对受网络限制的流的不同调度算法对网络能耗的影响.其中,网络中有两个受网络限制的流 f_1(服务器 1 到服务器 3)和 f_2(服务器 2 到服务器 4).图 7(c)、图 7(d)的拓扑与图 7(a)、图 7(b)基本相同,唯一的区别在于少了 *e-i* 和 *h-g* 的连接.图 7(a)、图 7(c)中两个流使用边不重叠的路径以最大化吞吐量,而图 7(b)、图 7(d)将两个流聚合到同一路径以最小化活跃设备的数量.假设每个流的大小都为 *Z*,所有

© 中国科学院软件研究所 http://www.jos.org.cn

链路的带宽是 B,每个交换机处于活跃状态的功耗是 P(与负载大小无关),则图 7(a)所需的能耗是 $\frac{4\times Z\times P}{B}$,图 7(b)中,每个流的完成时间变成 $\frac{2\times Z}{B}$, 所需能耗为 $\frac{6\times Z\times P}{B}$.

类似地,图 7(c)和图 7(d)所需的能耗分别为 $\frac{8\times Z\times P}{B}$ 和 $\frac{6\times Z\times P}{B}$.

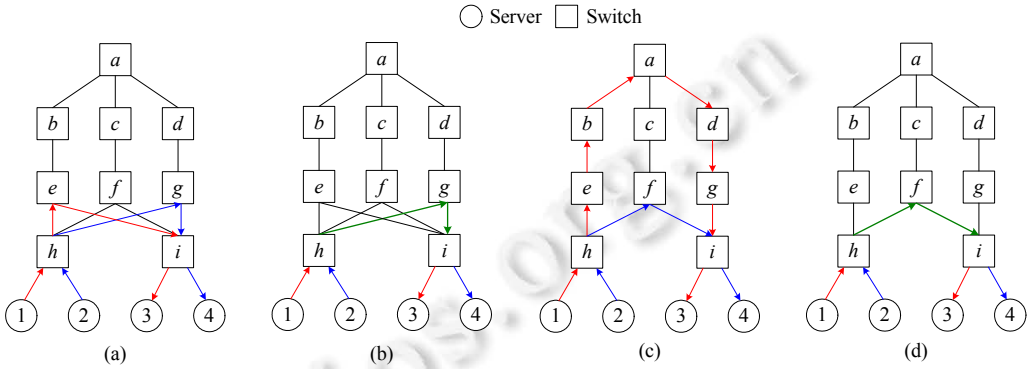


Fig.7 An illustration of flow scheduling for network-limited flows

图 7 受网络限制的流的调度示意图

可见,无论是最大化吞吐量还是最小化活跃设备数量,都可能无法实现最佳的节能效果.为此,文献[66]提出一种流调度方法,使流能够独占链路,从而提高流的吞吐率,减少网络设备参与传输的时长.文献[23]进一步提出了一种贪婪算法,为每个新的流选择一条路径,使得现有流的交换机工作时长总和最小.

4.6 能耗感知路由/流量工程的实现与部署

图 8 给出了为实现 EAR 所需的逻辑实体,包括:

(1) 信息采集模块.负责网络拓扑、流量矩阵、流量特征、能耗描述等输入信息的采集;

(2) 优化器.依输入信息求解最优化问题,输出每个流的路径信息/链路权值和每个设备的能源/性能状态;

(3) 能源控制模块.用于根据优化器的输出结果控制/更新网络设备的能源状态和性能状态;

(4) 路由控制模块.用于根据优化器的输出结果控制/更新网络设备的路由表或链路权值等.

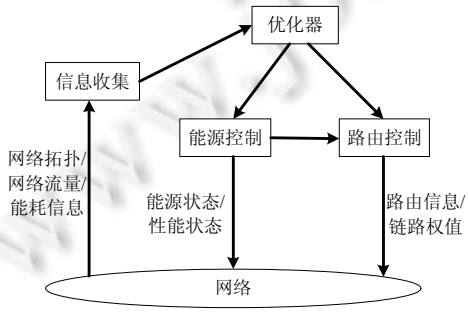


Fig.8 Deployment of energy-aware routing

图 8 能耗感知路由的部署图

- 网络拓扑收集

OSPF 协议通过 LSA 广播链路状态信息,每个节点可以实时地获取全局的拓扑信息.

- 网络流量收集

© 中国科学院软件研究所 http://www.jos.org.cn

一方面,如果流量本身具有高度的规律性,则可以按照历史信息分时段给出静态的流量矩阵预测;另一方面,由于实时收集网络的流量矩阵非常困难,为此,可以通过实时链路利用率计算出网络的流量矩阵[10].链路利用率是 RFC 3630[67]定义的 TE-LSA 的一部分.作为 OSPF LSA 的扩展,TE-LSA 也通过洪泛的方式在网络中广播.路由器在带宽使用率发生较大变化时,会发送 TE-LSA 报文.当控制节点获取了所有链路的利用率后,可以使用 tomogravity 方法[68]来计算网络流量矩阵.该方法的问题在于:路由器通告链路利用率的时刻并不统一,因此计算出的流量矩阵可能不具备实时性.除了 TE-LSA,文献[12]提出:传统上,可以使用 NetFlow,SNMP 和采样的方式来获得流量数据,但 OpenFlow 提供了一种获取流量的新途径.OpenFlow[69]支持基于流的路由,并提供多个级别的与流转发相关的计数器,用于提供流量矩阵数据.文献[70]也基于 Openflow 的计数器来获取链路的利用率信息.另外,可以采用混合的方式,即,基于先验的流量矩阵计算路由信息,同时,实时地收集网络的链路利用率,并据此确定使用哪个路由配置.例如文献[11]中,每隔 T 秒,边缘路由器会向中间路由器发送链路利用率查询消息,边缘路由器只需要向自己产生的路径所经过的中间路由器发送查询消息即可.文献[18]采用分布式优化,每个报文在传输过程中会纪录下路径上的 w_l 的最小值,并由目的节点向入口节点反馈,从而入口节点可以进行流量转移的决策.

- 能耗信息收集

如果预知域内的硬件配置,则可以有先验的设备能耗描述;否则,则需要能通过控制协议收集网络设备的能源信息.虽然目前许多网络设备还不支持统一的能源状态接口,但业界已经开始了标准化的研究,希望为未来的网络设备增加灵活的能源控制接口.ETSI 最近提出的 GAL 标准[46,47]旨在采用层次化的方式来定义设备的组件,并在物理设备和控制器所面向的逻辑设备之间提供抽象的能源接口.IETF 的 Energy Management(eman)工作组[43]也定义了相应的能源管理框架和 MIB 信息库[44,45].

- 最优化问题求解

(1) 集中式 vs 分布式:通常,最优化问题的求解由集中的控制节点完成.求解的关键因素在于求解时间和求解频率.路由更新太快会导致路由震荡,且频繁改变网络设备的能源状态会引起额外的开销和缩减设备的使用寿命[4,52].相反地,分布式方法[18,25,26]通常具有更好的扩展性.文献[18]所提出的方法由边缘节点进行局部优化.文献[25]提出的方法则可以在现有 IP 架构下予以实现,除了最短路径下一跳,每个路由节点还可以根据节能目标选择节能下一跳,当与最短路径下一跳链路的负载较低时,可以选择将流量切换到某个节能下一跳.文献[26]提出的分布式 GRiDA 方法由每个节点根据当前链路负载和对过往决策的学习做出链路关闭决定,在此过程中,节点之间仅通过 OSPF 的 LSA 建立有限的协同.

(2) 在线 vs 离线(offline):在线实时求解可能花费的时间较长,即便是启发式算法,也可能花费若干分钟时间[12,38].ElasticTree[12]提出的基于拓扑结构的启发式算法复杂度较低,但由于其仅适用于 FatTree 结构的网络,无法移植到其他拓扑结构.当流量变化较快时,在线求解会导致求解结果无法满足实时需求;离线求解则是预先计算好路由表,根据实时的流量选择使用哪个路由表或路由权值配置[11,62].

- 路由更新

对于支持 OSPF 的网络,可以通过路由器提供的 OSPF 配置接口对链路权值进行配置,或通过 OSPFv2 的 MIB[71]管理接口对其进行配置,之后,由 OSPF 路由协议根据新权值进行路由的重新计算;而如果是为路由器增加下一跳选项,则可以在现有的 IP 快速重路由(IP fast rerouting,简称 IP-FRR)架构 LFA[72]下予以实现.

对于支持 MPLS 流路由的网络,则需要建立 MPLS 的隧道.文献[10]对此进行了探讨,可以使用受约束路由标签分发协议协议或 RSVP-TE(resource reservation protocol-traffic engineering)建立标签交换路径,从而承载非最短路径的流量.文献[11]认为:在核心网中采用 MPLS 建立流级别的路径,即便对于具有 200 个~300 个出口路由器规模的网络也是可行的.

如果支持 SDN,则可以由控制器为交换机添加流表项来更新路由[11,12,38,66,73].在基于 OpenFlow 的转发控制中,由于流表需要支持并行查询,通常使用 TCAM 来实现,一般仅支持几百到几千条流表项.因此,当流的数量较大时,流表会超出 TCAM 的容量,这会对 EAR 的性能造成较大的影响.文献[74]对具有流表大小约束条件的 EAR

© 中国科学院软件研究所 http://www.jos.org.cn

问题进行了建模,并给出了相应的启发式算法.

路由更新需注意两方面的问题:

➢ 一是为了降低切换过程中的丢包率,切换过程需要遵循一定的顺序.

应当在备用路径建立好并等待一段时间后再关闭节点或链路,否则,正在途中的报文会丢失,当需要唤醒节点或链路时,应立刻将网络设备唤醒,但并不立刻在新路径上传输报文,而是等待链路的两端都就绪后再向新唤醒的链路发送报文[10];

➢ 二是为了避免路由震荡,需要控制路由更新的频率.

OSPF 协议中,关闭或唤醒链路都会引起路由的重新计算.路由更新快,能快速适应网络流量的动态变化,但计算开销大,同时,路由更新过程中也会产生路由循环和黑洞,引起不必要的丢包,影响网络可用性[20,75],为了在关闭或唤醒链路时触发的路由收敛过程中仍然保持网络的可用性,可以使用快速重路由技术[75];路由更新过慢,则难以快速适应网络流量的变化,从而降低节能效果.GreenTE,CARPO 和 ElasticTree 提倡的路由更新频率为 10 分钟左右[10,12,38].而文献[11,20,62]则认为,一天内仅应用少量的路由配置也能逼近理论上的节能上限.例如,文献[62]将一天根据流量的稳定性分为多个宏观时段,并在线下为每个时段计算出最优能耗的 OSPF 权值.

但上述研究都没有考虑网络流量在时间维度的动态性与路由的关系.频繁地切换网络设备的能源状态会造成不利情形有:(1) 路由更新过于频繁,控制层信息额外开销过大;(2) 设备从 Off 状态切换到 On 状态需要额外的能耗;(3) 设备切换需要额外的时间,可能会造成服务质量下降.因此,从时间维度考虑不同时段的流量特性及其关联性,跨时间段进行联合优化可能会具有更佳的效果[9,63].

- 能源状态更新

最后,需要依据最优化问题的求解结果更新/调节网络设备的能源状态,例如开启/关闭设备或链路、改变设备/链路的工作频率.目前的网络设备还不完全支持能源状态的调节.文献[10]提出了一种用 OSPF 流量工程扩展(OSPF traffic engineering extension)中的 TE-Metric 来传递链路的 On/Off 状态的方法.

如前所述,目前,ETSI 和 IETF 的 eman 工作组[43]分别为网络设备定义了能源管理的标准接口,未来可以用于控制和更新网络设备的能源状态和性能状态.

在基于 SDN 的网络控制领域,也有研究试图扩展当前的 OpenFlow 协议,为其增加交换机的能耗状态控制能力.文献[76,77]增加了 3 类 OpenFlow 的指令,用于:(1) 开启/关闭交换机或线卡;(2) 使能/禁用链路或端口,以及调节链路或端口的速率;(3) 改变整个系统的能耗水平.文献[78]则试图在 OpenFlow 中加入 GAL 的能力,即扩展 OpenFlow 协议,使得其增加 GAL 标准中的发现、供给和监控能力.

5 基于泛在缓存的网络节能技术

近年来,研究人员提出了一类以内容/信息为中心的网络,如 DONA[79],CCN[80],NetInf[81],以适应互联网应用模式从点对点通信为主向点对面的海量内容获取为主的转变.这类网络架构的重要特征是:将内容提供者和接收者解耦,利用泛在、透明的网络内置缓存和灵活的请求转发机制来满足海量的、具有重尾和异步特征的用户请求,从网络体系架构层面为可扩展的高效内容分发提供原生支持.

泛在缓存通过降低网络流量节约了传输能耗,但同时也增加了存储能耗.因此,泛在缓存能否节约能耗,与网络设备的传输/存储能耗特征、对象的请求频率、缓存的位置以及缓存内容的分配和替换等紧密相关.

5.1 基于缓存的网络设备能效分析

文献[82]首次提出利用 CCN 优化互联网能耗的想法.不同网络设备传输单个比特所需的能耗(以焦耳/比特衡量)差异很大[82−84],表 3 给出了不同类型设备的传输能效.例如传输相同的数据量,家庭网关和 PC 的耗能是核心路由器的 100 倍~1000 倍.因此,类似于 Nano Data Center[85]或 P2P 的内容分发方式虽然能降低核心网络的流量,但从能耗角度可能并非最佳选择.

为网络设备增加 CCN 能力的关键在于存储,存储设备包括 RLDRAM,DRAM,SSD 和 Disk 等多种功耗、访问速度、代价都不同的类型[84,86,87].表 4 给出了不同的存储介质的技术参数.

© 中国科学院软件研究所 http://www.jos.org.cn

Table 3 Transmission energy efficiency for different device types[82,83]

表 3 各种不同设备类型的传输能效[82,83]

设备类型	所处位置	能效
Cisco CRS 1 系列路由器	核心网	15W/Gbps
Cisco GRS 12000	核心网	28.6W/Gbps
Cisco 7507	边缘路由器	80W/Gbps
Zyxel IES-500M	接入设备(服务提供商端)DSLAM	208.3W/Gbps
D-Link DSL 2320B	接入设备(用户端)DSL Modem	1536W/Gbps
Fujitsu FA2232U	接入设备(服务提供商端)GPON OLT	16W/Gbps,25W/Gbps
Allied Data's GPON Gateway	接入设备(用户端)GPON OLT	245.8W/Gbps
具有 Xeon 处理器的刀片服务器	数据中心	360W/Gbps
具有双核的 PC	P2P	对 DSL 为 15KW/Gbps 或 10KW/Gbps 对 GPON 为 3KW/Gbps 或 2KW/Gbps

Table 4 Technical features for different types of storage technologies[84,87]

表 4 各种存储技术的技术参数[84,87]

存储技术	能效	最大规模	访问时间	成本
TCAM	15	~20MB	4	200
SRAM	0.12	~210MB	0.45	27
RLDRAM	0.027	~2GB	15	0.27
DRAM	0.023	~10GB	55	0.016
高速 SSD	0.000 05	~10TB	1 000	0.03
SSD	0.000 01	~1TB	10 000	0.003

为支持内容缓存所需的存储硬件受缓存容量和线速处理需求的制约:

- 一方面,缓存容量大小决定了能够采用的硬件技术,同时也决定了缓存内容索引表的大小,进而决定了索引表能采用的硬件技术和存储能耗.例如:当缓存空间为 10GB 时能采用 DRAM 技术,根据不同的索引技术,索引表大小约在 30MB~100MB 之间,可以采用 SRAM 或 RLDRAM 技术;而当缓存空间超过 1TB 时,则需要采用 SSD 技术,索引表也将达到若干个 GB,从而需要采用 DRAM 技术;
- 另一方面,由于不同的硬件存储技术具有不同的访问速率,为了支持线速查找,Interest 报文的到达速率决定了需要采用的硬件技术.例如:采用 SRAM 实现索引表可以支持较高的 Interest 报文到达率,而基于 RLDRAM 或 DRAM 的索引表能支持的 Interest 报文到达率将大为降低.

为传统路由器增加 CCN 能力,需付出额外的能耗和代价.目前的研究表明,这些额外代价是可控的[82,87].

5.2 基于缓存的内容传输能耗模型

增加缓存后,内容分发的能耗由两部分构成:(1) 传输能耗;(2) 存储能耗.现有的众多研究假设所有网络设备具有相同的存储/传输功耗[84,88−91].对其做适当扩展,假设每个节点和链路都有不同的功耗,在时长为 t 的待考察时段内,设 q_u^k 表示该时段内节点 u 发出的对内容 O_k 的请求次数,s^k 表示内容 O_k 的大小,x_u^k 为 0-1 变量,表示节点 u 是否缓存了内容 O_k,$P_{cache}(u)$表示节点 u 存储单位数据量的功耗,$P_r(u)$和 $P_l(i,j)$分别表示节点 u 和链路 i-j 传输单位数据量的功耗,则满足所有请求的总能耗 $EG_{tot}=EG_{trans}+EG_{cache}$,其中,

- 存储能耗为

$$EG_{cache}=t\sum_{u\in V}\sum_{O_k\in O}x_u^k s^k P_{cache}(u).$$

- 传输能耗表示为

$$EG_{trans}=\sum_{O_k\in O}\sum_{u\in V}\sum_{v\in VPath(u,DES_u^k)}P_r(v)q_u^k s^k+\sum_{O_k\in O}\sum_{u\in V}\sum_{(i,j)\in EPath(u,DES_u^k)}P_l(i,j)q_u^k s^k.$$

文献[92]则以最小的时间粒度(同样大小的数据块在一条链路的传输时间Δt)为基础,考察在某个时间段$\Gamma=|T|\Delta t$ 内,网络在每个时刻的最优状态配置.网络的状态可用 $x_u^k(t)$ 和 $x_{ij}^k(t)$ 予以表示:前者表示在时刻 t 节点 u 是否缓存内容 O_k,后者表示在时刻 t 边 i-j 是否传输内容 O_k.则在时间段Γ内的总能耗为

© 中国科学院软件研究所 http://www.jos.org.cn

$$EG(X(t))=\sum_{t\in\Gamma}\sum_{u\in V}P_{cache}(u)\sum_{O_k\in O}x_u^k(t)\Delta t+\sum_{t\in\Gamma}\sum_{(i,j)\in E}P_l(i,j)\sum_{O_k\in O}x_{ij}^k(t).$$

该能耗模型并未考虑节点传输内容的能耗.若考虑加节点的传输能耗,设 $y_u^k(t)$ 表示在时刻 t 节点 u 是否传输了内容 O_k,则有:

$$EG(X(t))=\sum_{t\in T}\sum_{u\in V}P_{cache}(u)\sum_{O_k\in O}x_u^k(t)\Delta t+\sum_{t\in T}\sum_{(i,j)\in E}P_l(i,j)\sum_{O_k\in O}x_{ij}^k(t)+\sum_{t\in T}\sum_{u\in V}P_r(u)\sum_{O_k\in O}y_u^k.$$

5.3 能耗最优化建模

文献[93]将 CCN 中的高能效缓存技术分为缓存节点放置、内容放置和请求路由这 3 个方面.

5.3.1 缓存节点的放置

文献[86]对层次化结构中单缓存的部署位置与能耗收益之间的关系进行了研究.缓存节点越靠近内容源,所需要存储的副本越少,因此存储能耗越小,但所需要的传输能耗更高.该文量化分析缓存的位置、内容访问的频率、存储能耗和传输能耗对 CCN 有效性的影响.为使采用缓存比不采用缓存更节能,访问频率需要大于某个阈值,该阈值与单位数据的存储功耗成正比,而与缓存节点与内容源的距离(N-M)以及每条链路传输单位数据的功耗成反比.除了能耗,性能在系统评估时也占重要地位.为此,可以使用能耗与延迟的乘积($energy\times delay$)衡量系统在能耗与性能之间的平衡能力.

对非层次化的泛在缓存,文献[82]的实验结果表明,部署 CCN 能够节约能耗,且在边缘部署 CCN 节点的有效性要高于在核心部署.该结论的问题在于:虽然都是 20%的部署率,但由于边缘节点数量远大于核心节点数量,因此在边缘部署 20%的 CCN 节点所付出的代价要比在核心部署 20%的 CCN 节点高的多.

5.3.2 内容放置与请求路由

内容放置和请求路由通常是紧密耦合的,否则难以最大化缓存的效用.文献[88]将内容的平均请求跳数作为重要的指标,研究表明:当传输功耗与存储功耗的比值越大或内容的请求频率越高时,应当在网络中缓存更多的副本以减少能耗.文献[84,89,90]假设在时间 t 内网络的缓存状态不变化,而在这一时间段内,每个路由节点 i 对内容 k 的访问频率为 R_{ik},则可以将内容放置和请求路由联合建模成一个整数规划问题:

minimize:

$$\sum_{k=1}^{F}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}^k x_{ij}^k+t\sum_{j=1}^{N}P_{cache}(j)\sum_{k=1}^{F}s^k y_j^k \tag{29}$$

subject to:

$$\sum_{j=1}^{N}x_{ij}^k=1,\forall(i,k) \tag{30}$$

$$x_{ij}^k\leqslant y_j^k,\forall(i,j,k) \tag{31}$$

$$\sum_{k=1}^{F}s^k y_j^k\leqslant S_j,\forall j \tag{32}$$

$$x_{ij}^k\in\{0,1\},y_j^k\in\{0,1\} \tag{33}$$

其中,$a_{ij}^k=R_{ik}s^k\left[\sum_{u\in VPath(i,j)}P_r(u)+\sum_{(u,v)\in EPath(i,j)}P_l(u,v)\right]$表示 i 从 j 获取 R_{ik} 次大小为 s^k 的内容 O_k 所需的传输能耗,目标函数的第 2 项表示时间 t 内所有节点的存储能耗.

约束(30)表示节点 i 仅从某一个节点获取内容 O_k(即,不采用多路请求转发);约束(31)表示仅当节点 j 缓存了内容 O_k 时,节点 i 才有可能从节点 j 获取内容 O_k;约束(32)表示在节点 j 缓存的所有内容的大小小于节点 j 的缓存大小 S_j.

但是,在时间 t 内多次访问内容而缓存的状态不发生变化是不符合实际的.文献[92]以最小时间粒度为基础,考察网络在某个时间段 $\Gamma=|T|\Delta t$ 内每个时刻的最优状态配置.同时,提出了基于整数规划的最优化决策问题表述,决策结果不仅给出了按照时间间隔的网络状态配置,还给出了请求被满足的路径.

© 中国科学院软件研究所 http://www.jos.org.cn

这些整数规划问题求解的复杂度都较高,文献[84,90−92]给出了一些基于局部搜索和遗传算法的启发式算法.

6 比较与分析

表 5 对前文讨论的各种网络节能技术进行了总结、分类与比较.

Table 5 Comparison and analysis of different network energy saving techniques

表 5 各种节能方案的分析与比较

技术方案分类				适用场景	方法	优点	缺点
网络流量不变化	实现单个网络设备的能耗比例计算	硬件的能源状态支持		任意单个网络设备	从硬件层面为网络设备/部件增加能源状态和性能状态	硬件支持,简单	可能会在不同的能源状态间频繁切换
		网络设备的局部控制策略		任意单个网络设备	基于自身流量状态自适应地通过本地控制策略调节硬件的能源状态	优化设备的能源状态调节;无需全局协同	无法达到全局最优
	能耗感知路由	基于流的路由	Powerdown 模型	硬件具有 On/Off 能力;应用受限的流	根据流量需求矩阵、网络拓扑、节点的能耗模型等信息,将流量聚合到少数网络设备/部件,允许关闭其他网络设备	大部分网络设备都支持 On/Off 的能耗模型	在 On/Off 状态间的切换时延可能较长;计算复杂度高
				硬件具有 On/Off 能力;网络受限的流	根据流的大小、网络拓扑和带宽、截止时间要求对流进行调度,既减少活跃设备的数量,又降低设备活跃的时长	大部分网络设备都支持 On/Off 的能耗模型;兼顾了活跃设备的数量和设备活跃的时长	仅适用特定的流
			Speed Scaling 模型	硬件支持不同的工作频率;应用受限的流	根据流量需求、网络拓扑、节点的能耗模型等信息,将流量分配到不同的网络设备上,使网络设备以最佳工作频率工作	能更细粒度地控制网络的能耗	在现有的网络设备能耗模型下,节能效果有限;计算复杂度高
		基于最短路径的路由	Powerdown 模型	硬件具有 On/Off 能力;应用受限的流	依据流量需求矩阵、网络拓扑、节点能耗模型等信息,优化网络链路的权值,使得在给定权值设置下可以关闭按最短路径路由没有流量经过的网络设备	适用于 OSPF 网络	权值调整会引起 OSPF 路由的重新计算;最优权重的计算复杂度高
					除了最短路径下一跳,每个路由节点可以选择节能下一跳,当与最短路径下一跳链路的负载较低时,可以选择将流量切换到某个节能下一跳	适用于 OSPF 网络,纯分布式实现	路由器需要维护更多的转发选项
改变网络流量	泛在网络缓存			以内容获取为主的网络	在网络内部部署缓存,减少网络流量,从而降低网络的传输能耗	从根本上解决网络能耗问题;可与其他技术并存	增加了存储能耗;需要新的网络架构支持

7 未来研究展望

虽然目前网络节能技术已经有一些研究结果,但还未获得有效应用和大规模部署.展望未来,仍有大量需要进一步研究的关键问题,包括:

(1) 路由设备节能技术标准化.

为了对不同厂商路由设备的能耗进行统一管理,目前,ETSI 和 IETF 都展开了网络设备能耗框架、交互接口

© 中国科学院软件研究所 http://www.jos.org.cn

和协议等方面的标准化工作.如何协调各个组织的研究,使其能够具备统一规范,是互联网节能技术能够在异构网络部署的关键.

(2) 同时考虑 Powerdown 和 Speed Scaling 模型的节能技术.

CARPO[38]虽然考虑了 powerdown 和 Speed scaling,但并非协同优化,而是将优化问题分为两个阶段,第二阶段的链路速率调整是基于单个设备的调整,并未进行全局协同,难以达到整体能耗的最优化.当硬件同时支持 Powerdown 和 Speed Scaling 能力时,基于两种能力的协同流量工程优化能够最优化网络的能耗[94].文献[94]建立了基于随机图的理论模型.结论表明:当静态能耗与动态能耗之比高于一定比例时,Powerdown 模型能够与 Speed Scaling 模型有效结合.但目前,研究界还没有针对此的具体优化算法,有待进一步深入探索.

(3) 基于多核化转发引擎的节能技术.

目前,支持多核多线程处理器的路由器开始在边缘接入网络替代不灵活的基于 ASIC 或网络处理器的传统转发引擎,例如 HP A6600 和 H3C SR6600 系列的路由器.多核使得多种报文处理功能在不同的核上灵活分配[95],从而允许一个或多个核在适当时候进入低能耗状态.如何组织报文的处理功能、如何调度多个核完成报文处理、最小化多核转发引擎的能耗,值得进一步研究.

(4) 利用可再生能源的网络和路由设计.

目前,大部分网络节能研究都着眼于降低网络的整体能耗.而随着风能、太阳能等可再生能源的出现,降低直接的碳排放变得更为重要.例如,Apple 公司最新的数据中心都配有大功率的太阳能发电场,其最终目标是让其数据中心完全依赖可再生能源,实现零碳排放.为此,如何设计数据中心网络的结构和路由,提高整个数据传输中可再生能源的利用率,同时保证服务可靠性和可用性,是新的研究热点[96,97].目前,这方面的研究还刚起步,许多问题都有待进一步探索.

(5) 新型网络体系架构的节能效果.

目前,未来网络体系架构是网络领域的一个研究重点,除了解决网络传输的性能、移动性、安全性、可扩展性等问题,能否克服现有互联网快速增长的能耗问题,也是衡量未来网络架构可行性的重要依据之一.目前的节能研究集中于 SDN 和 NDN 这两种未来网络架构,而对其他新型的网络架构(例如以解决移动性为主的 MobilityFirst[98]、以提供多种主体为目标的 XIA 架构[99]、以提供发布/订阅模式为主的 PSIRP[100]等)的节能研究还很少,需要进一步深入探索.

8 总 结

互联网的流量和能耗在可预计的未来仍将处于急剧增长的通道.提高网络设备的能效、缓解互联网的能耗增长,是一项非常紧迫的任务.网络节能涉及到硬件设备、传输协议、流量工程、网络架构和上层应用等多个层面.目前,围绕绿色互联网,研究界已经在高能效的网络设备、能耗感知路由、泛在网络缓存等方面展开了广泛的研究.但是,这些研究结果距离现实部署还有漫长的路要走.随着国民对环境问题重视程度的提高,网络节能仍将是未来几年的热点研究领域,在网络设备能耗模型的标准化与实现、网络节能策略对现有网络协议有效性的影响、克服网络节能潜力与网络性能之间的矛盾、多方面抑制网络流量的爆炸式增长等方面还有着巨大的研究空间.

References:

[1] Mills M. The cloud begins with coal—big data, big networks, big infrastructure, and big power: An overview of the electricity used by the global digital ecosystem. 2013. http://www.tech-pundit.com/wp-content/uploads/2013/07/Cloud_Begins_With_Coal. pdf?c761ac

[2] Barroso LA, Hölzle U. The Datacenter as a Computer: An Introduction to the Design of Warehouse-scale Machines. Morgan & Claypool, 2009.

[3] Lannoo B. Energy consumption of ICT networks. http://www.fp7-trend.eu/system/files/content-public/502-final-trend-workshop-brussels-24-october-2013-presentations/energyconsumptionincentives-energy-efficient-networks.pdf

© 中国科学院软件研究所 http://www.jos.org.cn

[4] Chiaraviglio L, Mellia M, Neri F. Minimizing ISP network energy cost: Formulation and solutions. IEEE/ACM Trans. on Networking, 2012,20(2):463−476. [doi: 10.1109/TNET.2011.2161487]

[5] IEEE P802.3az energy efficient Ethernet task force. http://www.ieee802.org/3/az

[6] Christensen K, Reviriego P, Nordman B, Bennett M, Mostowfi M, Maestro JA. IEEE 802.3az: The road to energy efficient Ethernet. IEEE Communications Magazine, 2010,48(11):50−56. [doi: 10.1109/MCOM.2010.5621967]

[7] Gunaratne C, Christensen K, Nordman B, Suen S. Reducing the energy consumption of Ethernet with adaptive link rate (ALR). IEEE Trans. on Computers, 2008,57(4):448−461. [doi: 10.1109/TC.2007.70836]

[8] Vishwanath A, Sivaraman V, Zhao Z, Russell C, Thottan M. Adapting router buffers for energy efficiency. In: Proc. of the 7th Conf. on Emerging Networking Experiments and Technologies. New York: ACM Press, 2011. [doi: 10.1145/2079296.2079315]

[9] Addis B, Capone A, Carello G, Gianoli L, Sanso B. Energy management through optimized routing and device powering for greener communication networks. IEEE/ACM Trans. on Networking, 2014,22(1):313−325. [doi: 10.1109/TNET.2013.2249667]

[10] Zhang MG, Yi C, Liu B, Zhang BC. GreenTE: Power-Aware traffic engineering. In: Proc. of the 18th IEEE Int'l Conf. on Network Protocols. Washington: IEEE Computer Society, 2010. 21−30. [doi: 10.1109/ICNP.2010.5762751]

[11] Vasić N, Novaković D, Shekhar S, Bhurat P, Canini M, Kostić D. Identifying and using energy-critical paths. In: Proc. of the 7th Conf. on Emerging Networking Experiments and Technologies. New York: ACM Press, 2011. [doi: 10.1145/2079296.2079314]

[12] Heller B, Seetharaman S, Mahadevan P, Yiakoumis Y, Sharma P, Banerjee S, McKeown N. ElasticTree: Saving energy in data center networks. In: Proc. of the 7th USENIX Conf. on Networked Systems Design and Implementation. Berkeley: USENIX Association, 2010.

[13] Chabarek J, Sommers J, Barford P, Estan C, Tsiang D, Wright S. Power awareness in network design and routing. In: Proc. of the 27th IEEE Conf. on Computer Communications. IEEE, 2008. [doi: 10.1109/INFOCOM.2008.93]

[14] Andrews M, Anta AF, Zhang L, Zhao WB. Routing for energy minimization in the speed scaling model. In: Proc. of the IEEE INFOCOM 2010. IEEE, 2010. [doi: 10.1109/INFCOM.2010.5462071]

[15] Andrews M, Anta AF, Zhang L, Zhao WB. Routing and scheduling for energy and delay minimization in the powerdown model. In: Proc. of the IEEE INFOCOM 2010. IEEE, 2010. [doi: 10.1109/INFCOM.2010.5462279]

[16] Chiaraviglio L, Mellia M, Neri F. Reducing power consumption in backbone networks. In: Proc. of the IEEE Int'l Conf. on Communications 2009 (ICC 2009). IEEE, 2009. [doi: 10.1109/ICC.2009.5199404]

[17] Bianzino AP, Chaudet C, Larroca F, Rossi D, Rougier JL. Energy-Aware routing: A reality check. In: Proc. of the 2010 IEEE GLOBECOM Workshops. IEEE, 2010. 1422−1427. [doi: 10.1109/GLOCOMW.2010.5700172]

[18] Vasić N, Kostić D. Energy-Aware traffic engineering. In: Proc. of the 1st Int'l Conf. on Energy-Efficient Computing and Networking (e-Energy 2010). New York: ACM Press, 2010. 169−178. [doi: 10.1145/1791314.1791341]

[19] Restrepo JCC, Gruber C, Machuca CM. Energy profile aware routing. In: Proc. of the IEEE Int'l Conf. on Communications Workshops 2009. IEEE, 2009. [doi: 10.1109/ICCW.2009.5208041]

[20] Chiaraviglio L, Cianfrani A, Rouzic EL, Polverini M. Sleep modes effectiveness in backbone networks with limited configurations. Computer Networks, 2013,57:2931−2948. [doi: 10.1016/j.comnet.2013.06.016]

[21] Avallone S, Ventre G. Energy efficient online routing of flows with additive constraints. Computer Networks, 2012,56:2368−2382. [doi: 10.1016/j.comnet.2012.03.011]

[22] Kist A, Aldraho A. Dynamic topologies for sustainable and energy efficient traffic routing. Computer Networks, 2011,55: 2271−2288. [doi: 10.1016/j.comnet.2011.03.008]

[23] Li D, Yu YR, He W, Zheng K, He BS. Willow: Saving data center network energy for network-limited flows. IEEE Trans. on Parallel and Distributed Systems, 2015,26(9):2610−2620. [doi: 10.1109/TPDS.2014.2350990]

[24] Xu MW, Shang YF, Li D, Wang X. Greening data center networks with throughput-guaranteed power-aware routing. Computer Networks, 2013,57:2880−2899. [doi: 10.1016/j.comnet.2012.12.012]

[25] Li Q, Xu MW, Yang Y, Gao LX, Cui Y, Wu JP. Safe and practical energy-efficient detour routing in IP networks. IEEE/ACM Trans. on Networking, 2014,22(6):1925−1937. [doi: 10.1109/TNET.2013.2288790]

[26] Bianzino AP, Chiaraviglio L, Mellia M, Rougier JL. GRiDA: Green distributed algorithm for energy-efficient IP backbone networks. Computer Networks, 2012,56:3219−3232. [doi: 10.1016/j.comnet.2012.06.011]

© 中国科学院软件研究所 http://www.jos.org.cn

[27] Lin CQ, Soh S, Chin KW, Lazarescu M. Efficient heuristics for energy-aware routing in networks with bundled links. Computer Networks, 2013,57:1774−1788. [doi: 10.1016/j.comnet.2013.03.006]

[28] Brown DJ, Reams C. Toward energy efficient computing. Communications of the ACM, 2010,53(3):50−58. [doi: 10.1145/1666420.1666438]

[29] Barroso LA, Hölzle U. The case for energy-proportional computing. IEEE Computer, 2007,40(12):33−37. [doi: 10.1109/MC.2007.443]

[30] Gupta M, Singh S. Greening of the Internet. In: Proc. of the 2003 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2003). New York: ACM Press, 2003. 19−26. [doi: 10.1145/863955.863959]

[31] Mahadevan P, Sharma P, Banerjee S, Ranganathan P. A power benchmarking framework for network devices. In: Proc. of the 8th Int'l IFIP-TC 6 Networking Conf. (NETWORKING 2009). Berlin: Springer-Verlag, 2009. 795−808. [doi: 10.1007/978-3-642-01399-7_62]

[32] Tucker RS, Parthiban R, Baliga J, Hinton K, Ayre RW, Sorin W. Evolution of WDM optical IP networks: A cost and energy perspective. Journal of Lightwave Technology, 2009,27(3):243−252. [doi: 10.1109/JLT.2008.2005424]

[33] Pan T, Zhang T, Shi JX, Li Y, Jin LX, Li FL, Yang JH, Zhang BC, Liu B. Towards zero-time wakeup of line cards in power-aware routers. In: Proc. of the IEEE INFOCOM 2014. IEEE, 2014. 190−198. [doi: 10.1109/INFOCOM.2014.6847939]

[34] Vishwanath A, Hinton K, Ayre RW, Tucker RS. Modeling energy consumption in high-capacity routers and switches. IEEE Journal on Selected Areas in Communications, 2014,32(8):1524−1532. [doi: 10.1109/JSAC.2014.2335312]

[35] Ye TT, Benini L, Micheli GD. Analysis of power consumption on switch fabrics in network routers. In: Proc. of the 39th Design Automation Conf. IEEE, 2002. 524−529. [doi: 10.1109/DAC.2002.1012681]

[36] Wierman A, Andrew LLH, Tang A. Power-Aware speed scaling in processor sharing systems. In: Proc. of the IEEE INFOCOM 2009. IEEE, 2009. 2007−2015. [doi: 10.1109/INFCOM.2009.5062123]

[37] Ananthanarayanan G, Katz RH. Greening the switch. In: Proc. of the 2008 Conf. on Power Aware Computing and Systems (HotPower 2008). Berkely: USENIX Association, 2008.

[38] Wang XD, Yao YJ, Wang XR, Lu KF, Cao Q. CARPO: Correlation aware power optimization in data center networks. In: Proc. of the IEEE INFOCOM 2012. IEEE, 2012. 1125−1133. [doi: 10.1109/INFCOM.2012.6195471]

[39] Lin C, Tian Y, Yao M. Green network and green evaluation: Mechanism, modeling and evaluation. Chinese Journal of Computers, 2011,34(4):593−612 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2011.00593]

[40] Yang Y, Xu MW, Wang D, Li SG. A hop-by-hop routing mechanism for green Internet. IEEE Trans. on Parallel and Distributed Systems, 2015. [doi: 10.1109/TPDS.2015.2394794]

[41] IEEE standard 802.1AX: Link aggregation. IEEE computer society. 2008. https://standards.ieee.org/findstds/standard/802.1AX-2008.html

[42] Advanced configuration & power interface (ACPI). 2014. http://www.acpi.info

[43] IETF energy management (eman) workgroup. 2013. https://datatracker.ietf.org/wg/eman/charter/

[44] Parello J, Claise B, Schoening B, Quittek J. Energy management framework. Request for Comments (RFC) 7326, IETF, 2014.

[45] Chandramouli M, Claise M, Schoening B, Quittek J, Dietz T. Monitoring and control MIB for power and energy. Request for Comments (RFC) 7460, IETF, 2015.

[46] Bolla R, Bruschi R, Davoli F, Donadio P, Fialho L, Collier M, Lombardo A, Reforgiato D, Riccobene V, Szemethy T. A northbound interface for power management in next generation network devices. IEEE Communications Magazine, 2014,52(1):149−157. [doi: 10.1109/MCOM.2014.6710077]

[47] Bolla R, Bruschi R, Davoli F, Gregorio LD, Donadio P, Fialho L, Collier LM, Lombardo A, Recupero DR, Szemethy T. The green abstraction layer: A standard power-management interface for next-generation network devices. IEEE Internet Computing, 2013, 17(2):82−86. [doi: 10.1109/MIC.2013.39]

[48] Nedevschi S, Popa L, Iannaccone G, Ratnasamy S, Wetherall D. Reducing network energy consumption via sleeping and rate-adaptation. In: Proc. of the 5th USENIX Symp. on Networked Systems Design and Implementation (NSDI 2008). Berkeley: USENIX Association, 2008. 323−336.

© 中国科学院软件研究所 http://www.jos.org.cn

[49] Mahadevan P, Sharma P, Banerjee S, Ranganathan P. Energy aware network operations. In: Proc. of the IEEE INFOCOM Workshops 2009. IEEE, 2009. [doi: 10.1109/INFCOMW.2009.5072138]

[50] Zane F, Narlikar G, Basu A. CoolCAMs: Power-efficient TCAMs for forwarding engines. In: Proc. of the IEEE INFOCOM 2003. IEEE, 2003. 42−52. [doi: 10.1109/INFCOM.2003.1208657]

[51] Ma Y, Banerjee S. A smart pre-classifier to reduce power consumption of TCAMs for multi-dimensional packet classification. ACM SIGCOMM Computer Communication Review, 2012,42(4):335−346. [doi: 10.1145/2377677.2377749]

[52] Chiaraviglio L, Wiatr P, Monti P, Chen J, Lorincz J, Idzikowski F, Listanti M, Wosinska L. Is green networking beneficial in terms of device lifetime? IEEE Communications Magazine, 2013,53(5):232−240. [doi: 10.1109/MCOM.2015.7105670]

[53] Gupta M, Singh S. Dynamic Ethernet link shutdown for energy conservation on Ethernet links. In: Proc. of the IEEE Int'l Conf. on Communications 2007. IEEE, 2007. 6156-6161. [doi: 10.1109/ICC.2007.1019]

[54] Wang L, Zhang F, Aroca JA, Vasilakos A, Zheng K, Hou CY, Li D, Liu ZY. Green DCN: A general framework for achieving energy efficiency in data center networks. IEEE Journal on Selected Areas in Communications, 2014,32(1):4−15. [doi: 10.1109/JSAC.2014.140102]

[55] Shang YF, Xu MW, Li D. Research on energy-saving routing devices and protocols in the Internet. ACTA Electronica, 2012,40(11):2290−2297 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2012.11.023]

[56] Fisher W, Suchara M, Rexford J. Greening backbone networks: Reducing energy consumption by shutting off cables in bundled links. In: Proc. of the 1st ACM SIGCOMM Workshop on Green Networking (Green Networking 2010). New York: ACM Press, 2010. 29−34. [doi: 10.1145/1851290.1851297]

[57] Al-Fares M, Loukissas A, Vahdat A. A scalable, commodity data center network architecture. ACM SIGCOMM Computer Communication Review, 2008,38(4):63−74. [doi: 10.1145/1402946.1402967]

[58] Cianfrani A, Eramo V, Listanti M, Marazza M, Vittorni E. An energy saving routing algorithm for a green OSPF protocol. In: Proc. of the IEEE INFOCOM Workshops 2010. IEEE, 2010. [doi: 10.1109/INFCOMW.2010.5466646]

[59] Cianfrani A, Eramo V, Listanti M, Polverini M. An OSPF enhancement for energy saving in IP networks. In: Proc. of the IEEE Conf. on Computer Communications Workshops 2011. IEEE, 2011. 325−330. [doi: 10.1109/INFCOMW.2011.5928832]

[60] Cuomo F, Cianfrani A, Polverini M, Mangione D. Network pruning for energy saving in the Internet. Computer Networks, 2012,56: 2355−2367. [doi: 10.1016/j.comnet.2012.03.009]

[61] Awerbuch B, Azar Y. Buy-at-Bulk network design. In: Proc. of the 38th Annual Symp. on Foundations of Computer Science. IEEE, 1997. 542−547. [doi: 10.1109/SFCS.1997.646143]

[62] Amaldi E, Capone A, Gianoli LG. Energy-Aware IP traffic engineering with shortest path routing. Computer Networks, 2013,57: 1503−1517. [doi: 10.1016/j.comnet.2013.02.006]

[63] Moulierac J, Phan TK. Optimizing IGP link weights for energy-efficiency in multi-period traffic matrices. Computer Communications, 2015,61:79−89. [doi: 10.1016/j.comcom.2015.01.004]

[64] Francois F, Wang N, Moessner K, Georgoulas S, Xu K. On IGP link weight optimization for joint energy efficiency and load balancing improvement. Computer Communications, 2014,50(1):130−141. [doi: 10.1016/j.comcom.2014.05.003]

[65] Lee SSW, Tseng PK, Chen A. Link weight assignment and loop-free routing table update for link state routing protocols in energy-aware internet. Future Generation Computer Systems, 2012,28(2):437−445. [doi: 10.1016/j.future.2011.05.003]

[66] Li D, YF, Chen CJ. Software defined green data center network with exclusive routing. In: Proc. of the IEEE INFOCOM 2014. IEEE, 2014. 1743−1751. [doi: 10.1109/INFOCOM.2014.6848112]

[67] Katz D, Kompella K, Yeung D. Traffic engineering (TE) extensions to OSPF Version 2. Request for Comments (RFC) 3630, IETF, 2003.

[68] Zhang Y, Roughan M, Duffield N, Greenberg A. Fast accurate computation of large-scale IP traffic matrices from link loads. ACM SIGMETRICS Performance Evaluation Review, 2003,31(1):206−217. [doi: 10.1145/885651.781053]

[69] OpenFlow switch specification v1.5.0. Open Networking Foundation, 2014.

[70] Kakadia D, Varma V. Energy efficient data center networks—A SDN based approach. Technical Report, No. IIIT/TR/2012/-1, Bangalore: Centre for Search and Information Extraction Lab International Institute of Information Technology, 2012.

© 中国科学院软件研究所 http://www.jos.org.cn

[71] Joyal D, Galecki P, Giacalone S, Coltun R, Baker F. OSPF Version 2 management information base. Request for Comments (RFC) 4750, IETF, 2006.

[72] Atlas A, Zinin A. Basic specification for IP fast-reroute: Loop-Free alternates. Request for Comments (RFC) 5286, IETF, 2008.

[73] Oda S, Nobayashi D, Fukuda Y, Ikenaga T. Flow-Based routing schemes for minimizing network energy consumption using OpenFlow. In: Proc. of the 4th Int'l Conf. on Smart Grids, Green Communications and IT Energy-Aware Technologies (Energy 2014). 2014. 69−72.

[74] Giroire F, Moulierac J, Phan TK. Optimizing rule placement in software-defined networks for energy-aware routing. In: Proc. of the IEEE Global Communications Conf. 2014. IEEE, 2014. 2523−2529. [doi: 10.1109/GLOCOM.2014.7037187]

[75] Yang Y, Xu MW, Li Q. Towards fast rerouting-based energy efficient routing. Computer Networks, 2014,70:1−15. [doi: 10.1016/j.comnet.2014.04.014]

[76] Vu TH, Thanh T, Trong VQ, Thanh NH, Nam PN. Energy saving for openflow switch on the NetFPGA platform using multi-frequency. Int'l Journal of Computing and Network Technology, 2014,2(1):9−15. [doi: 10.1186/s40064-014-0775-8]

[77] Vu TH, Nam PN, Thanh T, Hung LT, Van LA, Linh ND, Thien TD, Thanh NH. Power aware openflow switch extension for energy saving in data centers. In: Proc. of the 2012 Int'l Conf. on Advanced Technologies for Communications. IEEE, 2012. 309−313. [doi: 10.1109/ATC.2012.6404282]

[78] Bruschi R, Lombardo A, Morabito G, Riccobene V, Bolla R, Davoli F, Lombardo C. Green extension of OpenFlow. In: Proc. of the 2014 26th Int'l Teletraffic Congress (ITC). IEEE, 2014. [doi: 10.1109/ITC.2014.6932957]

[79] Koponen T, Chawla M, Chun BG, Ermolinskiy A, Kim KH, Shenker S, Stoica I. A data-oriented (and beyond) network architecture. ACM SIGCOMM Computer Communication Review, 2007,34(4):181−192. [doi: 10.1145/1282427.1282402]

[80] Jacobson V, Smetters DK, Thornton JD, Plass MF, Briggs NH, Braynard RL. Networking named content. In: Proc. of the 5th Int'l Conf. on Emerging Networking Experiments and Technologies (CoNEXT 2009). New York: ACM Press, 2009. 1−12. [doi: 10.1145/1658939.1658941]

[81] Kutscher D, Farrell S, Davies E. The NetInf protocol-draft-kutscher-icnrg-netinf-proto-01. Internet Draft, IETF, 2013. http://tools.ietf.org/id/draft-kutscher-icnrg-netinf-proto-01.txt

[82] Lee U, Rimac I, Hilt V. Greening the Internet with content-centric networking. In: Proc. of the 1st Int'l Conf. on Energy-Efficient Computing and Networking. New York: ACM Press, 2010. 179−182. [doi: 10.1145/1791314.1791342]

[83] Lee U, Rimac I, Hilt V. Toward energy-efficient content dissemination. IEEE Network, 2011,25(2):14−19. [doi: 10.1109/MNET.2011.5730523]

[84] Choi N, Guan K, Kilper DC, Atkinson G. In-Network caching effect on optimal energy consumption in content-centric networking. In: Proc. of the IEEE Int'l Conf. on Communications (ICC) 2012. IEEE, 2012. 2889−2894. [doi: 10.1109/ICC.2012.6364320]

[85] Valancius V, Laoutaris N, Massoulié L, Diot C, Rodriguez P. Greening the Internet with Nano data centers. In: Proc. of the 5th Int'l Conf. on Emerging Networking Experiments and Technologies (CoNEXT 2009). New York: ACM Press, 2009. 37−48. [doi: 10.1145/1658939.1658944]

[86] Braun T, Trinh TA. Energy efficiency issues in information-centric networking. LNCS, 2013,8046:271−278. [doi: 10.1007/978-3-642-40517-4_22]

[87] Perino D, Varvello M. A reality check for content centric networking. In: Proc. of the ACM SIGCOMM Workshop on Information-Centric Networking (ICN 2011). New York: ACM Press, 2011. 44−49. [doi: 10.1145/2018584.2018596]

[88] Guan K, Atkinson G, Kilper DC, Gulsen E. On the energy efficiency of content delivery architectures. In: Proc. of the IEEE Int'l Conf. on Communications Workshops. IEEE, 2011. [doi: 10.1109/iccw.2011.5963557]

[89] Li J, Liu B, Wu H. Energy-Efficient in-network caching for content-centric networking. IEEE Communications Letters, 2013,17(4):797−800. [doi: 10.1109/LCOMM.2013.022213.122741]

[90] Fang C, Yu FR, Huang T, Liu J, Liu YJ. An energy-efficient distributed in-network caching scheme for green content-centric networks. Computer Networks, 2015,78:119−129. [doi: 10.1016/j.comnet.2014.09.017]

[91] Fang C, Yu FR, Huang T, Liu J, Liu YJ. Energy-Efficient distributed in-network caching for content-centric networks. In: Proc. of the IEEE Conf. on Computer Communications Workshops 2014. IEEE, 2014. 91−96. [doi: 10.1109/INFCOMW.2014.6849174]

© 中国科学院软件研究所 http://www.jos.org.cn

[92] Llorca J, Tulino AM, Guan K, Esteban J, Varvello M, Choi N, Kilper DC. Dynamic in-network caching for energy efficient content delivery. In: Proc. of the IEEE INFOCOM 2013. IEEE, 2013. 245−249. [doi: 10.1109/INFCOM.2013.6566772]

[93] Fang C, Yu FR, Huang T, Liu J, Liu YJ. A survey of energy-efficient caching in Information-centric networking. IEEE Communications Magazine, 2014,52(11):122−129. [doi: 10.1109/MCOM.2014.6957152]

[94] Chiaraviglio L, Ciullo D, Mellia M, Meo M. Modeling sleep mode gains in energy-aware networks. Computer Networks, 2013,57: 3501−3066. [doi: 10.1016/j.comnet.2013.07.011]

[95] Dobrescu M, Argyraki K, Iannaccone G, Manesh M, Ratnasamy S. Controlling parallelism in a multicore software router. In: Proc. of the Workshop on Programmable Routers for Extensible Services of Tomorrow (PRESTO 2010). New York: ACM Press, 2010. [doi: 10.1145/1921151.1921154]

[96] Gattulli M, Tornatore M, Fiandra R, Pattavina A. Low-Emissions routing for cloud computing in IP-over-WDM networks with data centers. IEEE Journal on Selected Areas in Communications, 2014,32(1):28−38. [doi: 10.1109/JSAC.2014.140104]

[97] Singh R, Irwin D, Shenoy P, Ramakrishnan KK. Yank: Enabling green data centers to pull the plug. In: Proc. of the 10th USENIX Conf. on Networked Systems Design and Implementation (NSDI 2013). Berkeley: USENIX Association, 2013. 143−156.

[98] MobilityFirst. http://mobilityfirst.winlab.rutgers.edu/

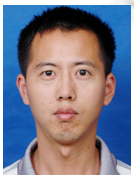
[99] XIA. https://www.cs.cmu.edu/~xia/

[100] PSIRP. http://www.psirp.org/

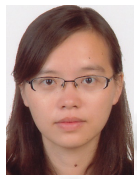
附中文参考文献:

[39] 林闯,田源,姚敏.绿色网络和绿色评价:节能机制、模型和评价.计算机学报,2011,34(4):593−612. [doi: 10.3724/SP.J.1016.2011.00593]

[55] 商云飞,徐明伟,李丹.互联网路由设备与协议节能研究综述.电子学报,2012,40(11):2290−2297. [doi: 10.3969/j.issn.0372-2112.2012.11.023]



张国强(1980－),男,江苏常州人,博士,副教授,CCF 高级会员,主要研究领域为未来网络,网络科学,网络流量优化.



刘真(1977－),女,博士,副教授,CCF 会员,主要研究领域为移动社交网络,云计算与虚拟化,分布式计算.

许自取(1990－),男,硕士,主要研究领域为未来网络.

© 中国科学院软件研究所 http://www.jos.org.cn