

Internet域间路由系统: 问题与挑战

Internet Inter-Domain Routing System: Issues and Challenges

中图分类号: TP393 文献标识码: A 文章编号: 1009-6868 (2009) 06-0009-04

摘要: 作为Internet的核心基础设施, 基于BGP协议的域间路由系统目前在扩展性、端到端性能、安全性与可信性等方面存在问题。由于部署的广域性、AS的自治性以及ISP之间交互的复杂性、域间路由行为的动态性, 使得人们对域间路由系统的结构和行为规律尚未充分理解, 缺乏全面有效的解决方案, 大规模的系统的网络实验难以开展。理想的域间路由系统及其支撑的Internet在性能上应具有快速恢复和全局优化的能力, 在安全上应具有自主防范和协同控制能力, 在运营上应具有自主配置和协同管理能力。

关键词: 因特网; 域间路由; 扩展能力; 性能; 安全

Abstract: As the critical infrastructure of the Internet, the Inter-Domain Routing (IDR) system based on Border Gateway Protocol (BGP) is suffering from problems in scalability, end-to-end performance, security and trustability. The IDR routing behaviors and structures have not been understood thoroughly due to its wide deployment, autonomy of Autonomous System (AS), complexity of interactions between Internet Service Providers (ISPs) and dynamics of running behaviors. A comprehensive and efficient solution is not available, and it is hard to conduct large scale network experiments to deal with the intricate IDR issues. An ideal IDR and the Internet based on it are expected to have fast recovery and global optimization ability in performance, self-defense and collaborative control power in security, and self-configuration and cooperative management in operations.

Key words: Internet; inter-domain routing; scalability; performance; security

Internet是自治系统(AS)按照各种商业关系互连的集合。AS由互联网服务提供商(ISP)拥有。基于AS的划分, Internet采用层次式路由结构。

1 域间路由系统的结构

Internet结构模型是建立Internet系统模型的基础, 也是域间路由协议设计、域间路由系统开发与部署的重要依据。

1.1 拓扑模型

Internet的模型研究经历了从经验

基金项目: 国家自然科学基金课题 (60873214)

假设到客观分析、从单纯的计算机网络研究到复杂系统特征化研究的过程。根据研究目的和生成方式不同, 分为拓扑生成模型和动态演化模型。前者只要求保证产生的拓扑图符合Internet关键的外在特征, 而动态演化模型力求从内到外、从微观到宏观展现真实的Internet成长过程。

1.1.1 拓扑生成模型

最早的网络拓扑模型是1988年提出的Waxman平面随机模型。此后关注Internet的层次性特征, 例如, Transit-Stub模型将AS域划分为Transit类和Stub类。基于层次关系模型的分

析, 处于顶层的十几个AS绝大部分属于美国的AT&T、Sprint、Verizon、Level3、Global Crossing、Qwest、Savvis以及日本NTT、印度Tata电信等公司, 彼此全互连, 构成了整个Internet的核心。从全球AS互连结构来看, 香港电讯盈科(PCCW)的AS最高处于第二层, 中国电信的AS4134大致处于第三层。1999年Faloutsos兄弟3人发现Internet拓扑结构中存在的幂律关系^[1], 从而将Internet拓扑与生物学、社会学中的复杂网络联系起来, 使其成为无尺度网络的一个实例。值得注意的是, Internet结构的幂律主要体现在AS级互连, 在路由器级由于路由器端口限制以及网络运营部署的约束, 并没有显著的幂律特性。AS级互连还表现出异类相聚、小世界、富人俱乐部和社团特征, 以及一定的自相似性, 例如中国大陆AS级拓扑的宏观特征与全球拓扑的某些类似^[2]。

1.1.2 动态演化模型

AS级的无尺度结构特性是AS局部利益极大化决策的结果, 选择连附上层AS的概率正比于提供商能够免费到达的网络或提供商互连的度数。1999年D. Carlson和J. Doyle提出高度优

朱培栋/ZHU Peidong¹赵金晶/ZHAO Jinjing²邓文平/DENG Wenping¹

(1. 国防科技大学计算机学院, 湖南 长沙 410073;

2. 北京系统工程研究所, 北京 100101)
(1. School of Computer, National University of Defense Technology, Changsha 410073, China;

2. Beijing Institute of System Engineering, Beijing 100094, China)

化的容错模型(HOT)^[3]。建模过程不仅关注统计特性的显式表现,还考虑网络设计中单个ISP需要面对的经济因素和技术约束等。在网络朝最优化方向发展的过程中,幂律等外部特性会自然地表现出来。HOT模型具有高度确定的组织结构,注重提高产出和容错性,关注区域特性,进行网络发展预测和求解域间路由系统的问题更为准确和可信。

1.2 AS商业关系模型

AS关系模型描述AS之间依据商业合同所形成的相互关系。主要包括:客户-提供商,提供商-客户,对等服务以及兄弟关系。L.Gao设计算法通过分析路由表数据可以比较准确地推导各个AS之间的商业关系^[4]。根据商业关系约束可判断路由宣告中AS-Path信息的异常,例如,如果AS把来自一个提供商的路由转发给了其他提供商,即在层次关系图中出现了“谷底”,可判定这条路由违背了商业关系,也可能是伪造的路由。

1.3 域间路由协议的演化

路由系统是随着Internet的发展逐渐成长起来的。在ARPANET初建时只有一个骨干网,后来为了实现多个网络互连设置了各种路由器。随着互联网规模的增长,再让所有路由器保存整个互联网的全部路由信息是不明智的,于是分为域间和域内路由。

边界网关协议(BGP)是目前唯一在用的域间路由协议,最早的协议版本由1989年发布的RFC1105定义,由Cisco和IBM基于EGP协议及其在NSFNET骨干网的使用经验编写。目前广泛使用的是1995年RFC1771定义的BGP-4版本。随着设备制造商对BGP功能的扩展和完善,以及对Internet新技术的支持,IETF的域间路由(IDR)工作组多年来一直非常活跃,先后修订和发布了60多个RFC规范BGP协议及相关功能,最新的BGP-4规范由RFC4271定义,最新的功能规

范是RFC5492对BGP能力宣告参数的重新定义。支持IPv6的BGP4+除了传播的路由信息的地址格式外,与BGP-4没有明显的协议机制差别。BGP-4还扩展了对32位AS号的支持。

2 域间路由的性能问题

虽然基本的BGP只是一种简单的路径向量路由协议,但是域间路由系统却是一个复杂系统。复杂性来源于域间路由系统规模的巨大性、ISP互连关系的丰富性、路由策略交互的动态性和BGP配置的多样性。

2.1 扩展性问题

根据亚太网络信息中心(APNIC)首席科学家G.Huston对核心网络路由信息的统计,2009年7月IPv4路由表中可见的AS号近3.2万,平均每个月新增200多个,表现出超线性趋势;路由表(RIB)有58.7万表项,转发表(FIB)29.9万,并以每两年大致1.2倍和1.3倍的速度增长。Internet体系结构委员会(IAB)为此专门发布RFC4984指出了路由扩展性问题的严重性。对BGP这类基于拓扑的路由协议而言,控制路由表规模实用的主要方法是拓扑聚合,但是多宿主、流量工程、ISP的合并或收购等因素造成大量不可聚合的路由信息。路由表项的不可聚合增加了路由信息宣告的数量和频率,从而加大了路由器处理路由信息和更新转发表的时间。同时,将网络边缘的拓扑状态扩散到整个网络,大量前缀不断修改、产生或撤销造成了路由信息的震荡,这其中也可能存在持续时间比较短的网络前缀劫持。T.Li发现摩尔定律对高端路由器并不适用,主要因为高端路由器采用的低延迟大容量SRAM生产,批量小,其性能的提高和代价的下降跟不上转发表的增长速度。转发引擎的散热问题也制约了路由器性能的进一步提升。可能的对策包括采用MULTI6、SHIM6工作组的多宿主方案,定位器/标志符分离协议(LISP)等。LISP作为互联网边

缘和核心路由器之间的隧道机制,也很好解决了IP地址的重载问题,但是会增加核心路由器的复杂性。隧道技术增加经费并减慢网络流量,甚至会因为传送数据过大而丢失数据。各种方案目前都处于早期的试验阶段。文献[5]指出,Internet结构的无尺度特性是实现基于聚合的层次路由的天然障碍,难以达到与网络规模成对数关系的理想的路由更新数量和路由表大小;LISP这类名址分离的路由结构名址映射表的更新开销也会制约扩展性。因此,如何实现不需要收敛的路由协议,如何像社会网络那样不需了解网络全景视图仍然能够有效选路,以及如何充分利用Internet的各种拓扑特性设计高效紧凑的路由算法,仍然是重要的课题。

2.2 端到端性能问题

收敛是指目标网络的可达性视图在全网达成一致。这种一致是相对的,由于观察点的不同和路由选择策略的差异,允许不同的节点具有到达目标网络的不同路径,但都必须是真的、反映网络拓扑实际互连状态的路径。

2.2.1 不收敛

研究发现BGP路由存在无法收敛的情况,主要表现为某些情况下内部BGP最佳路由的选择无法稳定。由于各个路由器本地的选择策略存在冲突,任何一组BGP路由都无法同时满足它们的需求。统计表明Internet路由中只有25%~35%可用性超过99.99%,10%可用性少于95%。

2.2.2 收敛慢

收敛慢是路由协议的一种病态行为。2000年,C.Labovitz统计发现路由故障平均需要3分钟恢复和重新路由,某些多宿主的故障恢复达15分钟。路由的收敛速度对VoIP、Video Game和商业事务很重要,持续几百毫秒就会中断某些应用。2007年MIT的

N.Kushman等对Skype和Vonage跨域语音电话的性能进行测量,发现BGP路由的慢收敛对通话质量的影响像网络拥塞一样严重,引起呼叫放弃或较长时间的不可用^[6]。D.Pei等发现路由撤销和恢复过程中,路由器最多会遍历 $N!$ 个可能路径(N 是系统中的AS数目),宣告多条暂时的无效路由,经过几次路由扩散才达到收敛状态。另外,其他类型的网络可达性信息的快速变化引起“路由抖动”,也会增加报文丢失率、网络收敛延时,带来路由处理的额外开销。

2.2.3 路径长

ISP互联遵循经济学的规律,主要动力是减少转发代价。各个ISP都是从自身利益的最大化出发,在没有全局规则调和约束的情况下,无法获得全局利益的最大化。例如,具有对等关系的ISP往往采用“热土豆(Hot Potato)”路由,选择最快离开本ISP的出口把流量送到对方网络,而不论流量在对方网络中经历的路径长短。有的ISP承诺使用“冷土豆(Cold Potato)”路由,但是实际的测量发现很可能没有实行这种策略。

从路由器级别考察,同样存在自私路由的情况。E.Tardos等通过理论分析发现,为了快速传递数据,路由器往往选择最少拥塞的路由,宏观来看速度较快的路由很快就会被堵塞;当路由软件再次更换路由后仍然造成堵塞的恶性循环。我们认为目前Internet路由协议还没有这样强的自适应能力和细粒度的调整性能,但是随着流量工程技术的广泛深入应用,在局部优化的同时需要多个ISP共同考虑全局性能。

2.2.4 QoS路由

多年来服务质量(QoS)路由是各种类型网络的重要研究内容。学者们设计的各种QoS路由机制很少得到应用,主要有3个原因:网络规模的日益扩大、新型业务的不断展开和网络安

全的严峻形势,使保持网络的可达性成为网络管理员的核心任务;缺乏简单有效的跨域QoS机制;有些ISP宁愿采用资源过量配置的方法实现高质量服务。但是,测试发现ISP对网络流量做到一定程度的区分服务,例如处理来自不同上游AS的流量时优先考虑来自客户的数据,对BitTorrent和UDP等类型流量的歧视等。但是,在网络资源总量不足的情况下ISP对网络流量区别对待,有人认为违背网络中立性。

目前BGP只是把最佳路由写入转发表并转告给下游路由器,影响了流量工程能力和路由的灵活性,但是如果把学到的多条路径都宣告出去将会使扩展性问题更严重。在全网范围实现跨域的QoS保证还比较遥远。跨域的IP组播由于缺乏有效的商业模式和部署结构的过于复杂,现在的组播往往通过应用层来实现,IP组播的高效率并没有得到彻底的实现。

3 域间路由系统的安全性 与可信性

路由系统的安全性与可信性是彼此交叉密切相关的2个方面,路由系统的安全性又与健壮性密切相关。由于BGP路由信息可以扩散到整个Internet,因此必须从全网角度来考察这3个方面,构建可依赖的路由系统。

3.1 路由系统的安全性

Internet路由系统的安全性多年来为人们所忽视,甚至等同于路由器的安全性。正如域名系统(DNS)的安全性一样,作为网络基础设施的路由系统也面临严峻的安全威胁,具体表现为:对路由系统的破坏;对特定网络的破坏;对网络流量的操纵;基于路由伪造的应用攻击(2005年6月,Google对外服务中断了约半个小时,分析发现是加拿大AS号为174的ISP非法宣告了其前缀)。北美网络运营商协会(NANOG)披露,宣告伪造地址给Email服务器使用,可制造并传播大

量难以溯源的垃圾邮件,由于垃圾邮件服务器需要与合法服务器进行TCP交互,离开路由伪造无法实施这类攻击。网站钓鱼最简单的方法是通过域名伪造实现,如果采用相同的域名可以通过DNS攻击修改名址映射实现,而如果采用和真实网站一样的域名和IP地址,就可以通过路由伪造实现,这种攻击方式更加隐蔽。历史上发生多起路由前缀劫持和路由泄漏的事件。这说明域间路由系统缺乏有效的路由鉴别机制,路由的可信性是威胁路由安全的主要问题。

目前,提高路由安全性的对策主要包括接收方验证机制、路由安全监测和可信性判断系统、新型协议机制的设计等。提出的可信性验证和评估方法主要有IRV和Listen-and-Whisper等。路由安全监测系统主要有UCLA的PHAS、RIPE的MyASN服务、Renesys公司的Gradus服务以及国防科技大学的RouSSeau系统等,主要基于分布采集的路由表和路由报文进行异常判断。为了便于多ISP间异构路由器不同格式路由表的采集和信息发布,IETF开始讨论基于XML的统一格式问题。安全协议机制主要有美国BBN公司的S-BGP、CISCO公司的soBGP、加拿大Carleton大学的psBGP以及国防科技大学的SE-BGP^[7]等,4种方案的安全能力大致相当,采用的信任模型各不相同,分别是以ICANN为根的层次信任模型、Web-of-trust信任模型、基于前缀断言列表的分布式信任模型,以及基于AS联盟的转换者信任模型(TTM)。虽然RFC3779设计了基于X.509的IP地址和AS证书格式,也开展了局部实验,但是S-BGP等离广泛部署还很遥远。

3.2 路由系统的可信性

BGP路由的可信性要求路由资源分配、路由策略、路由信息和数据转发的一致性,即在Internet管理平面、ISP商业平面、路由协议控制平面和路由器转发平面一致。

(1)路由信息的不可信,指ISP传递的路由信息中宣告的前缀不符合资源的分配知识,宣告的AS-Path不是路由信息实际经过的传播路径。

(2)路由行为的不可信,指路由信息与商业合约中声明的路由策略不一致。例如,客户把来自提供商A的路由泄露给提供商B而形成谷底路由;采用Prepend命令使宣告的路由具有较长路径,导致其他ISP不选择自己进行转发,虽然按照商业关系向对方宣告了路由但是实际上并不愿履行数据转发义务。

(3)转发行为的不可信,指ISP的数据转发行为与路由信息中说明的不一致。例如,通过静态路由或者改变下一跳路由器,向没有宣告路由的网络发送流量,以获得免费的出口转发等。

实现全面的可信路由需要在ISP之间建立广泛的信任机制和分布式可信监测体系。

4 域间路由系统的运营

BGP是基于策略的路由(PBR)协议。路由策略综合考虑数据转发的性能、安全、可靠性、经济性等诸多要求。路由器的配置文件是ISP策略的重要体现。随着网络规模的扩展,ISP互连关系的错综复杂,网络协议功能的日益丰富,系统中有大量的故障由人为因素引起,不恰当的配置可能导致路由震荡或引发路由安全事件。MIT开发的路由配置检查器(RCC)已在一些ISP试用。RCC支持一个AS内部多个路由器的配置检查,但是不提供多ISP或多AS协同检查。国防科技大学开发了多ISP路由协同配置系统ISP-Policy,采用多方安全计算方法,在满足ISP之间策略私密性和策略一致性的需求方面取得较好进展。

为了协调多个ISP的协同配置问题,美国自然科学基金会网络(NSFNET)路由仲裁者(RA)项目1995年发起了基于路由注册的方法,现有32个路由注册库(IRR)。与Internet各地区中心的网

络资源数据库(RIR)不同,路由策略蕴含较多的网络运营信息,具有一定私密性,公开到库中的策略信息要么过时,要么与实现的策略不一致,无法保证内容的可信性与完整性。近年来IRR的实施取得较好的进展,例如一些大型ISP强制客户注册,不注册的路由宣告将被过滤;欧洲网络信息中心(RIPE-NCC)在实施路由注册制度方面也取得成效。但是IRR的固有缺陷仍然没有消除,对来自具有对等关系的ISP或者提供商的路由信息仍然无法有效验证和过滤。

5 结束语

域间路由系统的这些问题是密切相关的,有的源于BGP协议和域间路由模式的固有缺陷,有的由于网络运营商的利益约束和交互的复杂性。域间路由系统是以AS为节点构成的自组织系统,单个AS体现出ISP的意志,整个域间路由系统的运行没有统一的管理,因此需要充分考虑ISP的社会属性,结合社会学和经济学的方法,探求域间路由系统问题的求解,例如博弈论、市场理论的应用和信誉体系、协同机制的设计等。域间路由系统是高度动态、快速成长和不断演化的,需要借鉴生态学的原理来引导、借用生物学的机理来设计有效的机制促进其健康发展。域间路由系统具有复杂巨系统的特点,仍然需要物理学的方法探求蕴含的规律,需要数学的方法刻画结构与行为模型,在采用复杂性理论和系统科学方法把握其宏观特征的同时,运用控制科学的理论和方法设计有效的控制与调节机制。

正如Internet一样,尽管存在诸多问题,但是数十年的运行表明,BGP路由协议总的来说是高效、稳定、安全和健壮的。BGP作为最重要的域间路由协议将在相当长的时间内继续存在。通过研究者、制造商和运营商等长期不懈的努力,理想的域间路由系统及其支撑的Internet在性能上应

具有快速恢复和全局优化的能力,在安全上具有自主防范和协同控制能力,在运营上具有自主配置和协同管理能力。

6 参考文献

- [1] FALOUTSOS M, FALOUTSOS C. On power-law relationships of the Internet topology[J]. ACM Computer Communication Review, 1999,29(4): 251-262.
- [2] 张国强, 张国清, 范磊. 中国大陆AS级拓扑的测量与分析[J]. 通信学报, 2007,28(10): 92-101.
- [3] ALDERSON D, DOYLE J, GOVINDAN R, et al. Toward an optimization-driven framework for designing and generating realistic Internet topologies[J]. ACM Computer Communication Review, 2003, 33(1):41-46.
- [4] GAO L. On inferring autonomous system relationships in the Internet[J]. IEEE/ACM Transactions on Networking, 2001,9(6): 733-745.
- [5] KRIOUKOV D, CLAFFY K, FALL K, et al. On compact routing for the Internet[J]. ACM Computer Communication Review, 2007,37(3):41-52.
- [6] KUSHMAN N, KANDULA S, KATABI D. Can you hear me now?! It must be BGP[J]. ACM Computer Communications Review, 2007,37(2): 77-84.
- [7] 胡湘江, 朱培栋, 龚正虎. SE-BGP:一种BGP安全机制[J]. 软件学报, 2008,19(1):167-176.

收稿日期:2009-07-30

作者简介



朱培栋, 国防科技大学计算机学院教授, 从事高性能路由器的研制和路由技术研究, 先后主持10余项国家科研课题。SCI/EI收录论文50余篇, 持有国家发明专利5项。



赵金晶, 博士, 北京系统工程研究所助理研究员, 从事网络路由和信息安全技术研究。发表论文20余篇, 持有国家发明专利1项。



邓文平, 国防科技大学计算机学院在读博士生, 研究方向为路由安全及网络健壮性。