文章编号：2095-2163(2020)03-0330-05 中图分类号：TP393 文献标志码：A

一种基于虚拟化技术的域际路由模拟平台

于 鹏，秦超逸，张 宇
（哈尔滨工业大学 计算机科学与技术学院，哈尔滨 150001）

摘 要：网络模拟在计算机网络研究、协议设计和网络管理等领域具有广泛的用途和意义，但是目前基于数学模型的模拟工具往往只是计算和预测网络行为，或仅仅构建网络协议栈以支持转发。基于对自治域级网络拓扑模拟的需求，提出了基于虚拟化技术的域际路由模拟平台构建方案。方案从数据源采集 BGP 路由拓扑数据，这就从自治域级拓扑向 IP 地址级拓扑的映射方案，使用虚拟交换技术和图划分算法实现多机扩展。最后，对该方案的实现初步评价了功能实现情况和基本性能指标，并介绍了对未来工作的展望。

关键词：域际路由；虚拟化；网络模拟

A border routing emulation platform based on vritualization technology

YU Peng, QIN Chaoyi, ZHANG Yu
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 Network simulation has a wide range of uses and senses in computer network research, protocol design and network management. However, current mathematical model-based emulation tools often only calculate and predict network behavior, or simply build a network protocol stack to support forwarding. Based on the requirements of autonomous system network topology emulation, the paper proposes a construction scheme of inter-autonomous routing emulation platform based on virtualization technology. The solution collects real BGP routing topology data, designs a mapping scheme from autonomous-system-level topology to IP-address-level topology, uses virtual switching technology and graph partitioning algorithm to implement multiple hosts extension. Besides, the paper initially evaluates the function implementation and the basic performance indicators. Finally, the paper introduces the prospects for future work.

【Key words】 inter-autonomous routing; virtualization; network simulation

0 引 言

网络模拟是一种利用软件来实现计算不同网络实体之间的交互，以模拟网络行为。模拟的实体包括路由器、交换机、节点、链路等。通过网络模拟来解决真实互联网环境中的问题更加安全有效，具备更小的风险。本文提出了一种基于 Docker 虚拟化技术的网络模拟方案，利用该方案模拟了真实 BGP 路由网络，并进行了初步的测试和评价。

边界网关协议（Border Gateway Protocol，BGP）是运行在自治域边界路由器上的路由协议。作为一种域际路由协议，通过维护 IP 路由表或网络前缀来实现自治域之间的可达性，能够去中心化地使各个网络自治。在自治域系统中，自治域间的路由不只是选择到达目标的最短路径，还需要符合特定的一些路由策略。一个自治域通常有至少一个 BGP 路由器，但是为了简化网络复杂度，本方案对于每一个自治域部署一台虚拟 BGP 路由器。这在不影响模拟真实性的情况下，使研究工作关于本论文所述的内容更具针对性。

1 相关工作

1.1 域际路由模拟技术

国内外目前有很多基于路由协议模拟的工作，这些网络模拟技术在协议验证和安全事件模拟中具有重要的意义。当前对路由协议的模拟分为静态路由模拟和动态路由模拟，其中动态路由模拟包括对 BGP 和 OSPF 等动态路由协议的模拟[1]。Quoitin 提出了模拟大规模拓扑中 BGP 协议的模拟器 C-BGP[2]，但并未实现真实 TCP 连接和真实数据包交换的模拟。基于 SSFNet 的 BGP 模拟技术[3]模拟了几乎 BGP 协议中的所有细节，具有较高的真实度。

同时，作为用于互联网自治系统之间的路由协议探究，Varadhan 等人[4]的研究说明了路由测量配置将引起路由震荡或者发散。Labovitz 等人[5]对 BGP 的研究证明了路由震荡对路由网络的危害，且

基金项目：国家重点研发计划（2016YFB0801303）。

作者简介：于 鹏（1994-），男，硕士研究生，主要研究方向：计算机网络、下一代互联网；秦超逸（1992-），男，博士研究生，主要研究方向：计算机网络、网络安全、网络协议；张 宇（1979-），男，博士，副教授，主要研究方向：计算机网络、网络安全、下一代互联网。

收稿日期：2019-06-06

文献[5-6]通过对路由变化的测量研究表明 BGP 聚合时间的延迟将很大，这种延迟将会造成拥塞或者网络瘫痪。

1.2　虚拟化技术

虚拟化作为一种资源管理技术，虚拟化[7-8]将计算机及互联网体系中的各种实体进行一定程度的抽象、转换后呈现出来，对计算机领域生产力提升有着重要价值。软件定义网络[9-10]（Software Defined Network，SDN），即通过软件程序的形式定义和控制网络。应用软件定义网络技术，能够对设备建立依照需求的互连，而且能够随时根据业务逻辑的修改来重新定制网络连接方式，更好地实现对网络的控制和管理。

2　设计

2.1　系统概览

本文研究最初是基于对 BGP 路由拓扑模拟的需求，并提出了一个可行的模拟方案。域际路由模拟系统概览如图 1 所示，系统部署在一定数量的宿主机上，宿主机之间网络可达。在每个宿主机上生成虚拟 BGP 路由器并建立虚拟网络的连接关系，对于跨宿主机的链路，使用虚拟交换技术构建。虚拟 BGP 路由器之间运行路由协议，从而构建为路由拓扑网络。

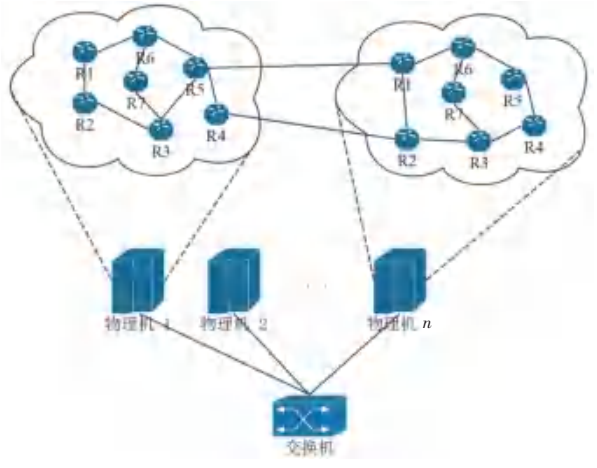

图 1　域际路由模拟平台系统概览

Fig. 1　Inter-autonomous routing simulation platform overview

本方案从 BGP 采集项目下载 BGP 数据并分析得到所需信息（3.2 节）。对这些信息进行预处理：将自治域级的拓扑信息映射到可以直接配置路由器的 IP 级拓扑信息（3.3 节）。模拟系统将这些信息作为配置读入，根据配置生成虚拟 BGP 路由节点，并在虚拟 BGP 路由器之间根据拓扑结构构建对应的虚拟网络连接。若在多宿主机上模拟，则在构建节点前将节点基本均分到不同的宿主机上，且跨宿主机的拓扑链路尽量少（3.4 节）。最后，根据拓扑信息配置每个节点上的路由软件并启动路由服务。

2.2　BGP 数据采集

RIPE 和 RouteViews 等 BGP 路径采集项目每天会更新自己的 BGP 数据。这些采集项目的采集器从其对等点获得 BGP 消息，定期地累计出从其对等点获得的完整路由表和路由更新。程序每天从这些采集项目下载 BGP 数据，经过对原始数据的处理和分析后得到自治域链接关系，前缀起源信息，监测点信息等。同时利用自治域路径推断自治域关系，得到其商业关系。通过收集历史和实时 BGP 数据和自治域信息进行处理，构建 BGP 拓扑结构数据库，其中包含用于模拟所需的 BGP 拓扑结构信息，用于本论文所提出的网络模拟平台的系统内输入，同时这些信息也可以用于分析和应对 BGP 协议自身缺陷引起的网络问题。

2.3　IP 地址分配

IP 地址分配实际上是从已知的自治域级拓扑结构信息向可供直接生成虚拟 BGP 路由器节点的 IP 地址级的信息映射，其目的是为生成节点、生成链路、编辑路由软件配置提供信息。以图 2 描述的自治域级拓扑为例进行阐述，本论文提出了 3 种 IP 地址分配方案。对此可做阐释分述如下。

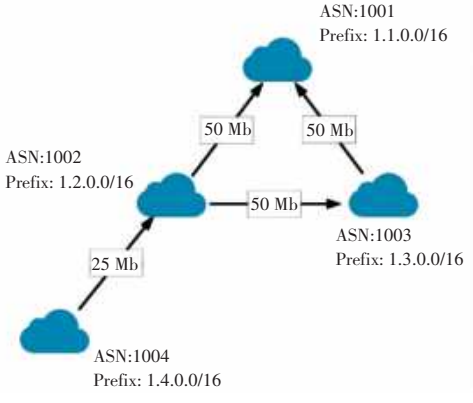

图 2　自治域级拓扑示例图

Fig. 2　An autonomous system level topology example

2.3.1　非子网模式

非子网模式的 IP 地址分配方案从每个自治域管理的网络中为每一条路由器到路由器的链路划分一个子网，其分配步骤详见如下。

（1）遍历每一个自治域，确定其邻接自治域的数量。对与当前遍历到的自治域相关的每一条的链路，判断自己和邻居的 IP 地址是否已经分配。如果已经分配，则跳过该邻接自治域，访问下一个邻居。反之，执行（2）。

(2)从自己的网络前缀中取出一个可用的足够大小的子网(至少为"/30"),为自己的网卡和对方的网卡分配 IP 地址。对下一个邻居重复执行(2),直到当前遍历节点的所有邻居都访问过。

(3)给自己的虚拟网卡 lo:0 分配一个属于自己网络前缀下的 IP 地址,以供连通性测试。

由图 2 自治域级拓扑生成的一种 IP 地址级拓扑信息运行结果如图 3 所示。

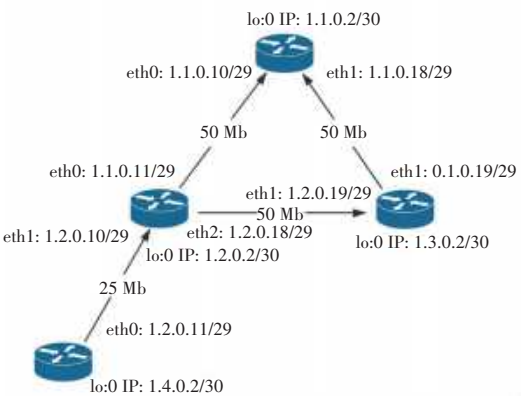

图 3 非子网模式 IP 地址分配结果

Fig. 3 Non-subnet mode result of IP address distribution

2.3.2 强子网模式

强子网模式的 IP 地址分配方案的提出是出于对实际模拟条件的适应。对于当宿主机的实际硬件条件导致不支持创建大量虚拟子网网络时,可以采用强子网模式。这种模式在不影响虚拟路由器之间的邻接关系的情况下,能够有效减少需要创建的子网。该分配方案的思想是对于 A、B、C 三台两两互连的路由器仅创建一个 IP 地址个数大于 3 的子网,这三台路由器通过同一子网下的 IP 地址建立通路,然后利用 BGP 路由配置来构建符合拓扑的路由关系。由图 2 自治域级拓扑生成的一种 IP 地址级拓扑信息图如图 4 所示。

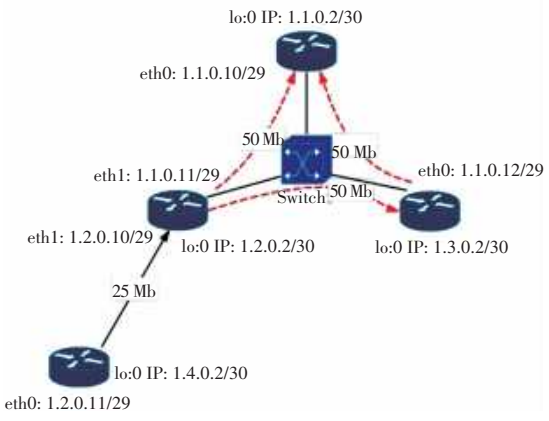

图 4 强子网模式 IP 地址分配结果

Fig. 4 Fine-grained subnet mode result of IP address distribution

2.3.3 弱子网模式

弱子网模式对于宿主机生成网络能力的需求介于强子网模式和非子网模式之间。这种模式中对于两两互连的自治域 A、B、C,若当前遍历到自治域 A,则对于 A-B、A-C 两条链路只创建一个子网,A 自治域的 BGP 路由器使用一个网卡与 B、C 会话。而对于 B-C 的链路,创建一个子网并给 B、C 的这两个网卡分配 IP,且可使用该 IP 会话。由图 2 自治域级拓扑生成的一种 IP 地址级拓扑信息图如图 5 所示。

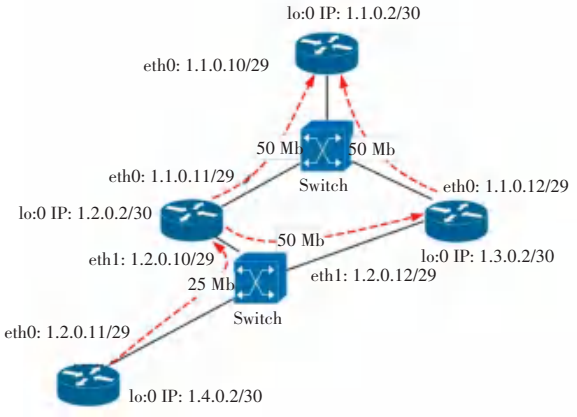

图 5 弱子网模式 IP 地址分配结果

Fig. 5 Coarse-grained subnet mode result of IP address distribution

2.4 节点划分

为了提高平台可模拟的虚拟 BGP 路由器的规模,本方案从提供一台真实物理机来模拟扩展到提供多台。主要涉及的问题是各个物理机上分配的虚拟节点的数量以及如何分配能尽可能减少跨物理机链接的规模。本方案使用节点划分算法。

节点划分目的是最小化不同分区(宿主机)之间的连接数(链路数量),具体地说,链路中有一些将会是跨宿主机的,因此希望能够使跨宿主机的链路尽量少的需求是合理的。节点划分应用一种称为 FM 算法的线性时间的启发式划分算法,旨在减少不同分区之间的连接数,也就是分区之间的沟通成本。

其主要思想是将整个将要模拟的拓扑作为一个图来处理。先将图划分为若干等份(对应模拟时的宿主机数量),每一份有基本相同数量的节点。然后对所有节点估计移动其能带来的收益值(收益值与跨区域的节点之间的边的数量有关),再从中选择收益高的点进行移动,直到所有的点都移动过之后,选取整个移动过程中划分效果最好的情况作为结果输出。

3　实验评价

3.1　实验环境

宿主机：Intel Xeon(R) CPU E5-2603，32 G 内存，2T 硬盘。宿主机操作系统：Ubuntu 16.04。容器操作系统：CentOS 6.9。实验拓扑图如图 6 所示。

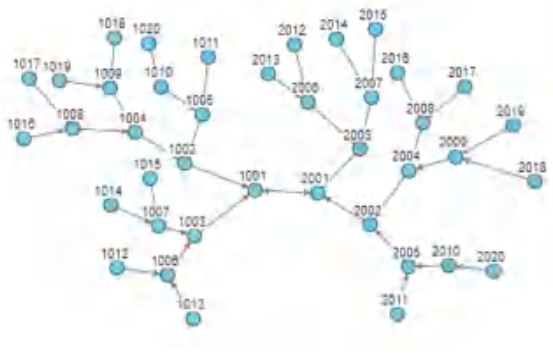

图 6　实验拓扑图

Fig. 6　Test topology

3.2　模拟平台运行性能

(1)连通性和延迟

① 实验方法：任选一个节点(虚拟 BGP 路由器)，向拓扑中所有其它节点做 Ping 测试，观察连通性和网络延迟。多次实验结果去平均值并进行统计。

② 实验结果及统计：根据实验结果反映，同宿主机上的虚拟 BGP 路由器之间互访的延迟较小，而跨宿主机的虚拟 BGP 之间则延迟较大。另外，执行 Ping 命令的源节点和目的节点之间虚拟链路距离越长(跳数越多)，则延迟越大，这与真实网络拓扑环境中的实验结果是一致的。

RTT 测试结果见表 1。通过测试可知，本域际路由模拟平台实现了完整的 BGP 路由拓扑，通过配置模拟使每一个自治域由一个有独立操作系统的软件 BGP 路由器实现，且任意一个虚拟 BGP 路由器都具备全局范围的路由可达性，跨宿主机的虚拟 BGP 路由器由于具有客观物理空间距离的存在而在访问延迟上有体现，未来将对这个问题做进一步的讨论和探究。

表 1　RTT 测试结果

Tab. 1　RTT test result

RTT \ 实验条件/ms	同宿主机	跨宿主机
最大值	0.160	0.344
最小值	0.086	0.272
均 值	0.125	0.313

(2)带宽测试。使用 Iperf3 带宽测量工具对模拟平台多个链路进行了带宽测试，得到了带宽的数据，实验环境同上。带宽测试结果见表 2。

表 2　带宽测试结果

Tab. 2　Bandwidth test result

带宽 \ 实验条件	同宿主机/(Gbits · sec^{-1})	跨宿主机/(Mbits · sec^{-1})
最大值	8.82	91.7
最小值	7.09	86.0
均 值	7.65	90.9

从测试结果来看，Docker 虚拟化网络提供了较大上限的虚拟网络链路带宽，而跨宿主机的虚拟 BGP 路由器之间的链路带宽由于受到实际物理网卡带宽限制而相对较低。如果想要模拟真实网络中的带宽环境，可以通过第三方工具如 TC、Wondershaper 等带宽限制工具根据需求设置虚拟链路带宽。

3.3　拓扑搭建过程性能

在单宿主机上测试了 3 组数据，分别是生成 30 节点、60 节点和 90 节点拓扑的情境。研究可知，CPU 使用情况见图 7。从测试结果来看，30 节点拓扑构建时间约为 37 s，60 节点拓扑构建时间约为 67 s，90 节点拓扑构建时间约为 93 s，随着节点数量的增加，构建时间的花费基本上呈线性增长趋势。从图 7 中可见，在构建过程中 CPU 的使用率基本在 8% ~10%，并且不会由于构建规模的增大而增大，这是一个可以接受的 CPU 使用率和变化情况。

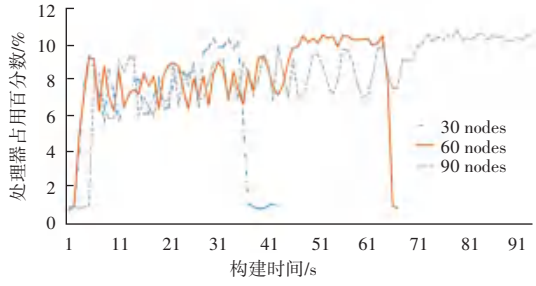

图 7　构建过程 CPU 使用率

Fig. 7　CPU usage during build

拓扑构建过程中内存占用如图 8 所示。内存使用会随着构建规模的增大而增大，而且一旦节点构建完成，模拟平台对内存的使用是持续的。直到这些模拟任务结束后、这些节点被销毁时，内存才会恢复原状态。从图 8 中可知，构建 30 节点规模的拓扑约需要内存 3%(32 G 内存总量)，构建 60 节点的拓扑约需要内存 5%，构建 90 节点的拓扑约需要内存 7%。基本上内存使用情况也是随着节点规模增长而线性增加的。

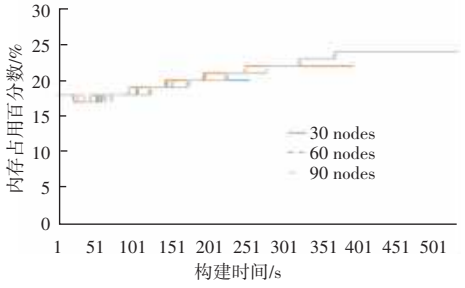

图 8 构建过程内存使用率

Fig. 8 Memory usage during build

4 结束语

本文提出了一种基于虚拟化技术的域际路由模拟方案,并初步模拟了真实的 BGP 路由拓扑。通过对域际路由拓扑的网络模拟,总结了提出的模拟技术的特点和价值。具体内容如下:

(1)这种网络模拟具有高扩展性和可定制性。虚拟化的网络设备实际上是一台轻量级容器,不仅仅可以模拟 BGP 路由器,也可以通过运行其它路由协议从而扮演其它角色,而且作为操作系统在网络层之上也可以安装应用层服务等。

(2)由于这种模拟具有较高的真实度,不仅可以作为一种测量和分析网络运行状态的工具,更能够作为一个进一步构建网络实验的平台,如进行各种基于网络流量的应用设计研究、网络靶场设计探索等。另一方面可将本方案提出的模拟平台中的节点与真实物理设备相连,通过连接真实路由设备来验证模拟平台构建的路由会话与真实路由一致。

参考文献

[1] 丁金科. BGP 协议模拟技术研究[D]. 哈尔滨:哈尔滨工业大学,2009.

[2] QUOITIN B, UHLIG S. Modeling the routing of an autonomous system with C-BGP[J]. IEEE Network,2005,19(6):12.

[3] DARPA. SSFNet 2000 [EB/OL]. [2005-07]. https:// www.ssfnet.org.

[4] VARADHAN K, GOVINDAN R, ESTRIN D. Persistent route oscillations in inter-domain routing[J]. Computer Networks, 2000,32(1):1.

[5] LABOVITZ C, AHUJA A, BOSE A, et al. Delayed internet routing convergence[J]. IEEE/ACM Transactions on Networking, 2001,9(3):293.

[6] TANGMUNARUNKIT H, GOVINDAN R, SHENKER S, et al. The impact of routing policy on Internet paths[C]// Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No. 01CH37213). Anchorage, AK, USA:IEEE,2001:736.

[7] UHLIG R, NEIGER G, RODGERS D. Intel virtualization technology [J]. IEEE Computer Society, 2005, 38(5):48.

[8] SMITH J E, NAIR R. The architecture of virtual machines [J]. Computer, 2005, 38(5):32.

[9] MIT Technology Review. 10 breakthrough technologies, TR10: Software-defined networking [EB/OL]. [2009]. http://www2.technologyreview.com/article/412194/tr10 - software - defined - networking/.

[10] NUNES B A A, MENDONCA M, NGUYEN X N, et al. A survey of software-defined networking: Past, present, and future of programmable networks [J]. IEEE Communications Surveys and Tutorials, 2014,16(3):1617.

(上接第 329 页)

4 结束语

根据 PEST AHP 模型分析结果显示,在中国健康管理信息化模式发展过程中,技术因素是影响其发展的主要方面,主要体现在健康管理信息化发展模式的研究支出和医疗服务模式的信息化程度。当前中国健康管理信息化模式的发展处于起步阶段,亟需互联网新技术的快速融入与有效支持。

根据前文研究结果,本文主要从如下方面做出改进建议:从国家层面,加大对“互联网+”模式发展的扶持力度,创新医疗服务新模式和健康管理信息化模式的实现途径。其次,应当健全健康管理相关的法律保护和监管体系,为健康管理信息化的模式发展提供安全保障和规范的医疗服务市场[7]。从社会层面,推广互联网技术与医疗服务相结合的技术产品的运用,鼓励健康管理信息化模式与各级医院进行数据共享,构建统一的数字健康管理平台,使老年人对于健康管理的新模式接受起来更为便捷与轻松,提高医疗管理效率,全面推进医疗服务管理系统信息化建设。

参考文献

[1] 刘远立, 郑忠伟, 饶克勤,等. 老年健康蓝皮书:中国老年健康研究报告(2018)[R]. 北京:社会科学文献出版社,2019.

[2] 张开金,夏俊杰. 健康管理理论与实践[M]. 南京:东南大学出版社,2011.

[3] 李宁生,承晓梅,仲学萍. 医院健康管理的实践与探讨[D]. 医学研究生学报,2013,26(6): 627.

[4] 杜丹, 路文如. 基于 PEST 分析的中国农业电子商务竞争环境研究[J]. 中国农学通报,2009,25(8): 266.

[5] 万莎. 收入不平等、医疗保险与老年人健康[J]. 山西财经大学学报,2015,37(6):1.

[6] 余冬苹,张玉良,赵彦涛. “互联网+医疗”发展趋势探讨[J]. 移动通信, 2016,40(13): 12.

[7] 栾运波,田珍都. 我国“互联网+医疗”存在问题及对策建议[J]. 行政管理改革,2017(3):59.