

一种多协议同时进行的网络测量技术

章泰, 李昕

(北京邮电大学计算机学院, 北京 100876)

摘要: 随着云服务的发展, 网络服务质量也越来越受到重视, 评估网络服务质量的重要手段之一就是主动测量获取网络性能参数。在主动测量中, 多数测量工具一次只能使用一种网络协议进行测量, 即在一个时刻只能获取到当前网络针对一种协议的网络服务质量。本文以具备快速探测和对目标网络侵入性小等特点的 FlashRoute 为基础, 在其之上扩展了基于 ICMP 协议和 TCP 协议的测量功能, 并进一步设计和实现了针对同一目标网络同时发送多种协议探测包的功能, 以达到测量目标网络在某一时刻对不同网络协议数据包的服务质量的目的。

关键词: 计算机系统结构; 网络测量; 多协议测量

中图分类号: TP311.1

A multi-protocol simultaneous network measurement technique

Zhang Tai, Li Xin

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract: With the development of cloud services, the quality of network services is becoming increasingly important and one of the most important means of assessing the quality of network services is to obtain network performance parameters through the active measurement. In active measurement, most measurement tools can only use one network protocol at a time, which means that only the network service quality of the current network for one protocol can be obtained at a time. This paper takes FlashRoute, which has the characteristics of fast detection and low invasiveness to the target network, as the basis, and extend its measurement function based on ICMP protocol and TCP protocol, and further design and implement the function of sending multiple protocol detection packets for the same target network at the same time, in order to achieve the purpose of measuring the quality of service of the target network for different network protocol packets at a certain time.

Key words: Computer system structure; Network measurement; Multi-protocol measurement

0 引言

随着互联网、云计算的快速发展, 各行各业为了应对数字化转型, 越来越多的企业正在将服务器转移到数据中心 (Data Center, DC), 使用云服务提供商租赁的云服务器。网络服务质量 (Quality of Service, QoS) 也日益受到重视, 尤其对于云服务提供商而言, 测量当前网络 QoS 对网络运维工作以及提高服务质量具有重要意义。

网络测量是评估网络 QoS 的重要方法, 采用何种协议进行测量也是测量的关键^[1]。例如 Ping 就是常用的测量两个网络设备往返延迟测量的工具, 测量协议通常为 ICMP。ICMP 属于 OSI 网络层, 而在互联网中大多数应用是基于传输层 TCP 协议或者 UDP 进行数据交互的, 网络中路由器的路由转发策略各不相同, 不同协议的数据包在路由器中的转发优先级也有可

作者简介: 1998-, 男, 硕士研究生、计算机科学与技术

通信联系人: 1976-, 男, 副教授、硕导, 软件定义网络. E-mail: cplalx@163.com

能不同,因此使用 ICMP 测得的 QoS 可能并不适用于 TCP 或者 UDP 数据包经过该当前网络时的 QoS。为了解决 ICMP Ping 不能测量 TCP 数据包经过网络时的 QoS 的问题, TCPing^[2] 直接使用了 TCP 协议进行测量。同样地, 还有使用 UDP 协议进行 Ping 测量的工具, 在此不再赘述。但无论 ICMP Ping 工具或者 TCPing, 都不能同时获取到在某个时刻的网络对 ICMP、UDP 和 TCP 三种协议数据包的 QoS。因此, 本文设计并实现了一种能够将上述三个协议探测包同时注入被测网络中的网络测量技术, 以此来采集三种协议数据包经过当前网络时的 QoS, 同时具备针对目标网络降低侵入性等特点。

1 关键技术介绍

1.1 网络测量

网络测量是使用软硬件工具并结合一定的方法和技术来获取表征网络性能指标的一系列活动的总和。网络测量主要包括三个关键点: 测量对象、测量环境和测量方法^{[1][3]}。测量对象, 指的是被测节点以及被测链路的某种或某些网络特性; 测量环境, 主要指的是测量节点的选择、测量设备的类型以及通信链路等; 测量方法主要指的是测量包协议、大小、数量、间隔等参数的确定以及测量的统计方法等。

根据网络测量的方式, 可以将网络测量分为被动测量^[1]和主动测量^[1]。被动测量通过抓取并分析网络中已经存在的流量来获取网络性能参数, 被动测量的优点是可以实时获取网络性能参数, 也可以针对性分析某条流量并且不会对网络中正常的流量造成影响等, 但缺点也较为明显: 依赖于网络中的已有流量。主动测量则是通过主动发送探测包来获取网络性能参数, 不依赖于网络中的现有流量。相比于被动测量, 主动测量可自由控制测量时间、选择被测目标, 以及能够自由选择使用的测量协议, 如 ICMP、UDP 和 TCP 等。主动测量的测量模型如下图:

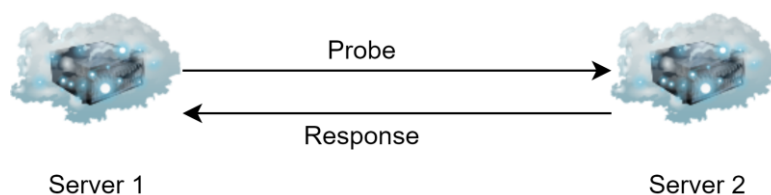


图 1 主动测量模型

Fig. 1 Active measurement model

如上图, Server 1 主动向 Server 2 发送探测包, Server 2 接收到数据包后进行处理并返回给 Server 1, 通过多次发送探测包来获取网络的往返时延 RTT、丢包率等关键的网络质量参数, 而探测包发送的协议有差错报文控制协议、数据报协议等。

1.2 FlashRoute

FlashRoute^{[4][5]}是 Yuchen Huang 等人于 2020 年发表在 SIGCOMM IMC 上的网络拓测量工具, 特点是可以在数十分钟扫描完整个/24 IP 地址, 并且在保证测量结果准确性 (如负载

均衡路径等)的同时又减小对目标网络入侵性。

75

从发布的 FlashRoute 源码^[5]中可以看到，FlashRoute 的测量架构包括 Schedule 调度层、Prober 协议编码与解析层、Monitor 监控层、Network 套接字管理层，每一层的设计和实现都是相互独立、互不影响的，各层交互如下图所示。

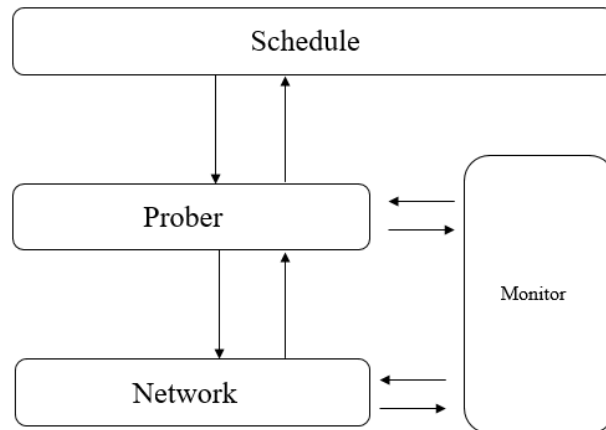


图 2 FlashRoute 架构

80

Fig. 2 FlashRoute's architecture

85

图 2 中, **Schedule** 层的主要功能是进行任务调度以减小对目标网络的入侵性, 这里的任务指的是对某个 IP 发送某个 TTL 的探测包, 具体的编码与发送过程交给下层执行。在具体的实现中, **FlashRoute** 会将所有的目的 IP 进行随机排列, 再使用 **Doubletree** 策略[6]对每个目标 IP 进行探测, 以减轻测量源点附近和目标 IP 附近的网络流量负载, 减小对网络的入侵性。另外, 该层也用来控制程序是执行 **Trace** 的逻辑还是的 **Ping** 逻辑, 即控制每个目标 IP 发送的探测包头部的 **TTL** 的值。

90

图 2 中, Prober 层的主要功能是探测信息的编码和解码。将探测信息编码到协议中的目的是为了

避免记录过多状态, 提高探测速率并节省内存空间。探测信息指的是 IP 头部中的 TTL、发送时间戳等, 编码时也会按照 Paris-traceroute 策略^[7]保证流 ID 的一致性以测量负载均衡路径; 在接收到 Network 层的数据时解析编码在协议中的信息, 用以计算往返时延 (Round-trip Time, RTT), 并将结果反馈给 Schedule 层。在具体的实现中, FlashRoute 会将探测信息编码到 IP 头部和 UDP 头部, 其中 IP 头部使用 IP-ID 字段与 IP 包总长度字段, 这两个字段经过编码后的状态如下两图所示。

IP-ID 字段（16bits）		
10 bits 时间戳低 10bits 的编码	1 bit 标识测量阶段	5 bits TTL 编码

95

图 3 IP-ID 字段的编码

Fig. 3 Encoding of the ID field

IP 包总长度字段（16bits）		
9 bits 填充为 0	1bit 固定为 1	6 bits 时间戳高 6 bits 的编码

图 4 IP 包总长度字段的编码

Fig. 4 The encoding of the total length field

从图 3 和图 4 中可以看出，时间戳只有 16 bits，如果以毫秒为单位，则最多表示 65.535 秒，由此可能导致 RTT 时间计算出错，比如当 RTT 实际为 66 秒时会被错误计算为 0.465 秒即 465 毫秒。

图 2 中，Network 层的主要功能是管理和使用底层 socket。当收到 Prober 层传递下来的数据包时，使用 socket 发送出去；当通过 socket 收到网络数据包时，将数据包传递给 Prober 层进行解析处理。

图 2 中，Monitor 层的主要功能是统计发包和收包信息，例如总共发出多少个探测包、收到多少个响应包、发送速率、差错信息等等。

2 多协议同时测量的设计与实现

本章主要将介绍基于 FlashRoute 设计和实现一种多协议同时进行的网络测量技术。

2.1 多协议同时测量的设计

Flashroute 实现了 UDP 协议的测量，为了同时发送多个协议的探测包进行测量，还需要实现 ICMP 协议与 TCP 协议的测量。

2.1.1 ICMP 测量时的编码设计

在 FlashRoute 的 Prober 层，没有实现 ICMP 协议的测量。新增 ICMP Prober，额外实现 ICMP 协议的测量时候需要考虑的是测量信息应该编码到哪些字段。ICMP Echo Request/Reply 的头部结构如图 5 所示，其中可以用来做有效编码的是 16 位 Identifier 标识字段和 16 位 Sequence Number 序列号字段。可以与 IP 头部中的 32 位一起使用，共 64 位。

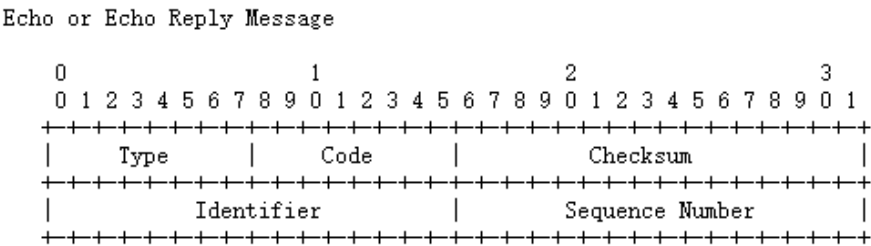


图 5 ICMP Echo Request/Reply 的头部结构

Fig. 5 ICMP Echo Request/Reply Header

但如果采用 FlashRoute 中的 IP 头部编码，当目的 IP 返回 ICMP Echo reply 报文时，该报文不会携带原始探测数据包 IP 头部，所以编码信息全丢失，导致无法定位到该报文是由 TTL 为多少的探测包引发的回应、以及不能确认该探测包的发出时间也就无法计算 RTT。所以，ICMP Echo 报文中能用作编码的只有 ID 和序列号共 4 个字节 32 位。

2.1.2 TCP 测量时的编码设计

在 FlashRoute 的 Prober 层，没有实现 TCP 协议的测量。新增 TCP Prober 额外实现 TCP 协议测量时同样需要考虑测量信息应该编码到哪些字段。TCP 头部结构如下图所示。其中可以用来做有效编码的是 32 位序列号。

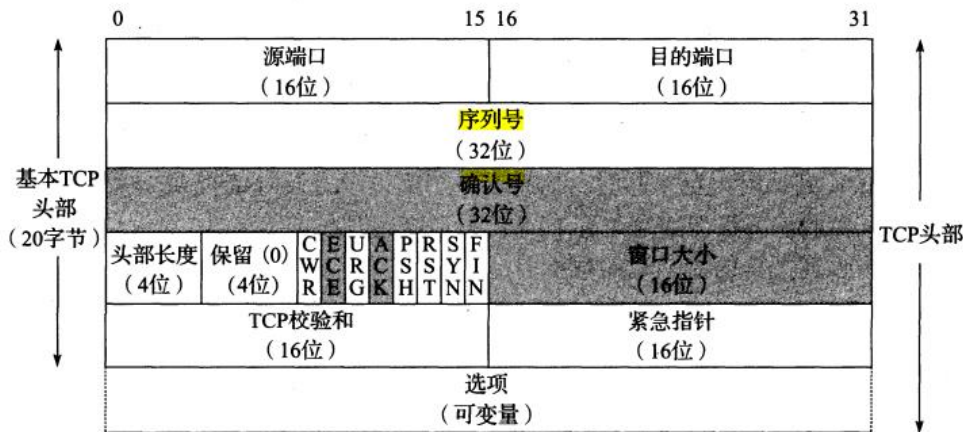


图 6 TCP 头部结构

Fig. 6 TCP header

主动测量时，探测包使用 TCP-SYN 或 TCP-ACK 报文，被探测主机收到后会回复 ACK 报文或 RST 报文，或者回复 ICMP Destination Port Unreachable，这些报文都会附带源探测包的编码信息，不过 RST 报文或 ACK 报文只会附带源探测包中编码在 TCP 确认号或序列号中的信息，不会附带编码在源探测包 IP 头部中的信息，导致探测信息丢失，无法计算出 RTT 等测量结果，而 ICMP Destination Port Unreachable 报文会将这两个信息都附带。所以能用作探测信息编码的字段仅有 TCP 序列号字段共 4 字节 32 位。

2.1.3 同时测量的设计

同时测量指的是探测源以最快速度将不同协议针对同一目的 IP 的探测包发送出去，这里的最快速度指的是 CPU 硬件的指令处理速度，即针对同一目标 IP 的不同协议的探测包的发送速率不受软件层面影响，只受物理硬件影响。

在 FlashRoute 的 Prober 层，没有实现针对一个目标 IP 进行多次编码以及同时发送多个协议探测包的功能。新增 Combine Prober，可根据配置使用多种指定协议进行同时探测。新增的 Combine Prober 是一个包装类，可以将 UDP/ICMP/TCP Prober 都放入其中。当 Prober 层被 Schedule 层调用，需要对某个 IP 执行探测时，Combine Prober 将使用其包装的各 Prober 实例去编码探测包并交给 Network 层发送到网络中；同时，当 Network 层收到数据包并将其传递给 Prober 层的 Combine Prober 实例时，Combine Prober 也会逐个遍历其内的各 Prober 实例去解码数据包，解析成功则直接将结果向上反馈给 Schedule 层，值得一提的是只有对应协议编码的 Prober 才能正确解析收到的响应报文。

2.2 多协议同时测量的实现

本节主要介绍基于 FlashRoute 架构实现多协议同时测量中 Prober 层的实现，该层 UML 如下图所示。

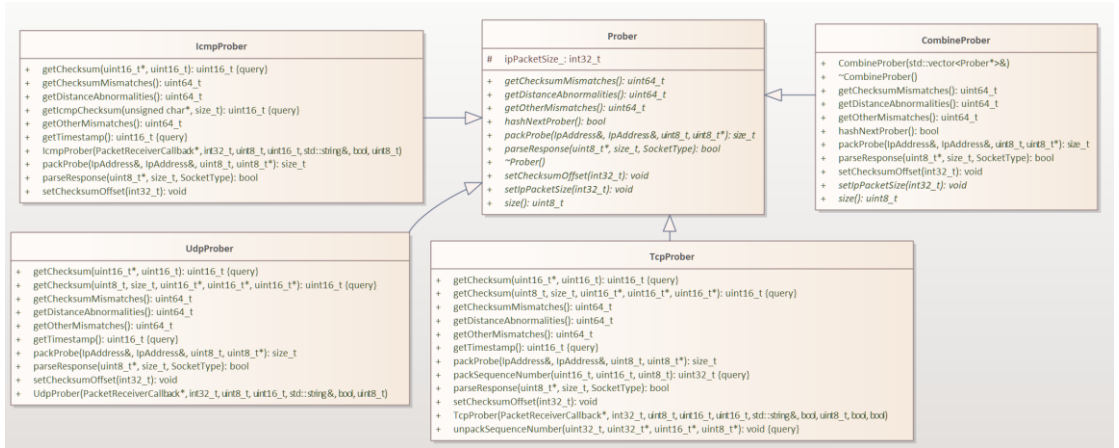


图 7 Prober 层 UML

Fig. 7 Prober layer UML

在上图中可以看到，除了新增的 IcmpProber 和 TcpProber 之外还有一个 CombineProber。CombineProber 的构造参数中传入的是 Prober 数组，当 Prober 层接收到 Schedule 层的探测任务时，将会使用该数组里面的每个 Prober 对象实例进行编码，并将编码后的数据包交给下层 Network 发送到网络中。

3 性能测试

本章主要介绍在 FlashRoute 架构之上实现的多协议同时测量技术的性能测试，包括测试环境、测试说明和测试结果等。

3.1 测试环境

测试环境为局域网内两台 2 核（Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz）4G 配置的 Centos 7 虚拟机，系统功能均已测试正常且两台虚拟机之间 RTT 不超过 1ms。两台虚拟机的其中一台虚拟机 A 用作主动测量发起端，并将多协议同时进行的网络测量工具在虚拟机 A 上提前编译完毕，另一台虚拟机 B 用作被测端。为了评估使用多协议同时探测时，针对同一 IP 地址的不同协议探测包被注入到网络中的时间间隔，虚拟机 A 安装 Tcpdump 进行抓包。

3.2 测试用例

测试用例包含以下几项：

表 1 测试用例

Tab. 1 Test case

测试编号	测试描述	测试目的
1	在虚拟机 A 上使用 ICMP 协议对虚拟机 B 发起测量，测量包大小 64 字节，一分钟内发送 100 个探测包	测试使用 ICMP 协议进行测量时，Prober 层编码与解码的正确性
2	在虚拟机 A 上使用 TCP-ACK 对虚拟机 B 发起测量，测量包大小 64 字节，一分钟内发送 100 个探测包	测试使用 TCP-ACK 进行测量时，Prober 层编码与解码的正确性
3	在虚拟机 A 上使用 ICMP/UDP/TCP-ACK 三种协议同	测试使用多种协议同时进行测量时，

	时对虚拟机 B 发起测量, 测量包大小 64 字节, 测量速率 100K pps, 持续时间 10 分钟	Prober 层编码与解码的正确性
4	在虚拟机 A 上使用 ICMP/UDP/TCP-ACK 三种协议同时对足量且不同 IP 发起测量, 测量包大小 64 字节, 测量速率 100K pps, 使用 tcpdump 持续抓包 10 分钟	评估使用多协议同时探测时, 针对同一 IP 地址的不同协议探测包被准入到网络中的时间间隔

3.3 测试结果

测试 1 的结果: 经过分析测量输出数据, 使用 ICMP 进行测量时, 能够正确解析被测端回包中的数据, 与预期情况相符合; 100 次测量得到 RTT 结果均为 1ms, 与真实情况相符。

测试 2 的结果: 经过分析测量输出数据, 使用 TCP-ACK 进行测量时, 能够正确解析被测端回包中的数据信息, 与预期情况相符合; 100 次测量得到 RTT 结果均为 1ms, 与真实情况相符。

测试 3 的结果: 经过分析测量输出数据, 使用 ICMP/UDP/TCP-ACK 三种协议同时进行测量时, 均能够正确解析被测端回包中的数据信息, 与预期情况相符合; 十分钟内共计约 6 千万次测量得到 RTT 结果均为 1ms, 与真实情况相符。

测试 4 的结果: 经过对 tcpdump 结果的分析, 使用 ICMP/UDP/TCP-ACK 三种协议同时进行测量时, 虚拟机 A 将每一个 IP 的这三种协议探测包发送到网络出口的时间间隔为 164 微秒, 基本满足同时发出的要求, 与预期情况相符合。

4 结论

本文在 FlashRoute 之上给出了一种能够针对同一探测目标同时发出多种协议探测包的
网络测量技术设计与实现方案, 经过测试, 各协议均能正确完成测量任务且多协议同时测量
时满足同时发送的要求。整体上具备 FlashRoute 的快速探测和低侵入性等特点的同时, 扩展
了使用多协议进行网络测量的功能, 使用户能够根据自身需求探测目标网络在某个时刻针对
一种或多种协议的 QoS。

[参考文献] (References)

- [1] 谈杰, 李星. 网络测量综述[J]. 计算机应用研究, 2006, 23(2): 5-8.
- [2] Cloverstd. TCPing 源码[OL]. [2017-7-25]. <https://github.com/cloverstd/tcping>
- [3] 程光, 龚俭, 丁伟. 网络测量及行为学研究综述[J]. 计算机工程与应用, 2004, 40(27): 1-8.
- [4] Yuchen Huang, Michael Rabinovich, and Rami Al-Dalky. 2020. FlashRoute: Efficient Traceroute on a Massive Scale[A]. In Proceedings of the ACM Internet Measurement Conference (IMC '20). Association for Computing Machinery, New York, NY, USA, 443-455.
- [5] Yuchen Huang, Michael Rabinovich, and Rami Al-Dalky. FlashRoute 源码 [OL]. [2020-12-13]. <https://github.com/lambdahuang/FlashRoute/>
- [6] Benoit Donnet, Philippe Raoult, Timur Friedman, and Mark Crovella. 2005. Efficient algorithms for large-scale topology discovery. In Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems. 327-338.
- [7] e Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viger, Timur Friedman, Matthieu Latapy, Cl  nence Magnien, and Renata Teixeira. 2006. Avoiding traceroute anomalies with Paris traceroute. In Proceedings of the 6th ACM SIGCOMM Internet Measurement Conference. 153-158