开放式网络测量数据分析系统 IP TASCM

朱海婷，丁伟，夏震，曹旭，陈兴其

（东南大学计算机科学与工程学院；江苏省计算机网络技术重点实验室 南京 210096）

摘 要： IP TASCM 是一个以海量 IP Trace 为核心的管理系统。它可以对各种类型的 IP Trace 和分析代码进行有效管理，并基于代码转移的思想，支持各种有关分析程序的运行。其核心目标是在充分利用 IP Trace 和分析代码资源的同时，能在最大程度上减少 Trace 移动对主干带宽带来的压力。

关键词： 网络被动测量 IP Trace 管理系统

中图分类号： TP393

IP TASCM: An open system for network measurement analysis

ZHU Hai-Ting, DING Wei, XIA Zhen, CAO Xu, CHEN Xing-Qi

(College of Computer Science & Engineering, Southeast University;

Key Laboratory of Computer Network Technology in Jiangsu, Nanjing 210096, China)

[Abstract] IP TASCM is a massive IP Trace based management system. Using an idea that codes, not data, are chosen to move through networks, it efficiently manages varieties of IP Traces, analytical codes and supports their computations on our platform. We aim at minimizing the data movements which may cause unnecessary backbone bandwidth pressure, while making full use of present IP Traces and analytical code resources.

[Key words] passive network measurement; IP Trace; management system

1 引言

在高速互联网主干信道上以零或非常低的丢报率获取 IP 报头并将其存储为 IP Trace 需要非常强大的硬件平台的支持，这些 IP Trace 是所有从事网络流量及相关领域研究的人员进行统计分析和验证假设模型的唯一的真实数据来源。

相对于研究小组的数量而言，具备主干信道采集和存储能力的采集存储点的数量要少很多，绝大多数的研究人员都是通过下载网上公开发布的 IP Trace 来进行自己的工作。这也是目前该领域的研究工作中被普遍使用的工作方案和技术路线。然而，即使是在 2.5Gbps 的 OC-48 主干信道上，平均 50%的信道利用率和平均报长为 512 字节的情况下，只采集最基本的 44 字节长度的 IP 报头，每小时的存储量也在 50GB 左右。以这样的强度，连续 24 小时的存储量就将达到 1.2TB。通过网络传递这样海量的数据是非常困难的。所以，大部分公布的主干 IP Trace 都只有几分钟，基本不超过 15 分钟，如此的数据量用于宏观模型的研究和验证是远远不够的。

相对于海量的 IP Trace，分析程序的代码长度几乎可以忽略不计，因此如果采用将分析程序从本地提交到 IP Trace 所在的计算机或集群上运行并获得结果的方法，可以大大减少数据的传输量，从数据的利用率、网络带宽节约等角度给研究工作带来很大方便。具有这样功能的系统还可以方便地支持有效地管理已有的分析结果，实现结果的共享和代码的复用。即使是在同一个研究小组内，这样的系统也有助于研究效率的提高。

相关的工作在国内外均有开展，但大多是按照提供免费下载的 Trace 和分析工具（代码）的传统思路进行的。国外的研究机构有 NLANR、CAIDA、WIDE、LBNL、ACM 等，国内的有清华大学的 SANTT[1]。

项目： 国家支撑计划 2008BAH37B04

作者简介： 朱海婷（1983-），女，博士研究生，主要研究方向：计算机网络测量，网络行为学；丁伟，女，教授、博士生导师。

E-mail： htzhu@njnet.edu.cn

此外，SIMR[2]关注各个数据源的数据格式不统一导致的数据描述的问题，并用数据库来管理数据并且提供各种分析工具，IST-MOME[3]数据库也做了类似的工作。更权威的 CAIDA 的 IMDC[4]在此基础上，增加了对分布在整个互连网上的测量数据的有效跟踪。

作者所在的实验室是 CERNET 的一个省网中心，拥有 2.5Gbps 和 10Gpbs 采集平台 WATCHER[5]和一个海量数据存储中心，它可能在一段连续时间内，以低丢包率对双向 IP 报文头部进行镜像并存储。以此为基础，基于这样的理念，我们设计并实现了一个以数据为核心的开放式网络测量数据分析系统 IP TASCM（IP Trace Analysis System based on Code Moving），首先用于支持本实验室相关的研究工作，在此基础上，我们希望能够将其建成一个开放的公共服务平台，通过互联网为该领域的研究人员提供服务。本文将阐述这个平台的核心设计思路和关键问题的解决方案。

2 系统的功能

在相关领域，传统方式的分析步骤是：

1. 寻找满足需要的 IP Trace；
2. 在本地硬盘上留出足够的空间；
3. 下载这个（些）Trace；
4. 用自己的分析代码对这些 Trace 进行分析，并获得结果。

IP TASCM 是将上述过程中的 Trace 的移动变为代码的移动，而获得分析结果的目标是相同的，因此系统的功能至少要有 2 类：

一、执行类。支持用户提交包括参数文件在内的程序代码，将其在指定的 Trace 上运行并返回结果。这样的运行过程可以通过重新指定参数重复进行；

二、管理类。管理 Trace、用户、代码等系统所涉及的对象，并满足有关查询要求。

3 关键问题研究

这样的设计带来的问题主要体现在 2 个方面。从使用者的角度，分析程序的编写者对执行代码的控制远不如传统方式方便，这会影响调试和运行的过程；从系统的角度，执行代码的移入，不可避免会带来安全隐患。围绕这 2 点进行更具体的分析，我们认为该系统需要解决的关键问题有以下几个：

1. 工作模式的设计

系统本身不支持调试过程，移动的代码是调试完成后的执行码，这使得系统的使用者必须面对一个异步的工作方式，这个工作方式的设计是系统的基础，要考虑包括参数的获取、结果的返回等所有可能涉及到的方面，还要尽可能地做到安全、合理、方便。这个方式还包括重复使用代码时，参数的提交过程的设计。

2. 环境异构问题

由于分析代码的调试和运行是在不同的环境中完成的，IP TASCM 必须解决使用者在本地调试的代码能在系统所在的服务器上正常运行的问题。

3. 安全问题

主要来自 2 个方面，一是合法获得权限的程序由于自身存在问题不能正常结束，导致死锁或者死循环以及类似的问题；二是程序本身带有恶意目的，包括恶意侵占系统资源、恶意攻击、恶意篡改 Trace 数据等。

4. 系统功能和结构的设计

系统需要在能够解决上述问题的基础上，完成具体功能的设计，这个设计还包括对 Trace、代码、结果、用户等角色的管理和查询功能。更进一步的还应包括对并发计算的调度的考虑。在此基础上，设计合理的结构，保证所有关键问题的解决和设计功能的实现。

对上述关键问题的解决我们采取了如下的方案：

- 系统采用 B/S 方式工作，系统的所有功能均通过浏览器实现。
- 相对于传统分析方式，我们为本系统用户设计的工作模式可简单描述为：

1. 代码上传：准备好执行代码并将其和参数文件等一起上传到系统；
2. 任务提交：要求系统将指定代码在指定参数文件上完成运行；
3. 结果返回：通过浏览器将结果文件返回本地。

- 系统提供将参数文件转换成 XML 格式的功能。这样在重复使用代码时，不必重新组织参数文件，可通过浏览器输入完成。
- 系统为每类格式的 Trace 数据提供一个 200M 的测试数据，用户可将其下载到本地用于调试。
- 环境异构的解决采用 2 种手段，一是建议用户使用 JAVA，一是用户直接在系统所在的主机环境下调试程序。前者面向所有用户，而后者只面向内部用户。
- 对安全问题，采用的解决方案包括：
  1. 将系统结构设计成内外两个子系统，之间可设防火墙，所有核心成分均位于内子系统。
  2. Trace 数据存储盘使用 RAID6 进行冗余保护。对外只提供可读权限，防止被恶意篡改。
  3. 对用户先按内部和外部分类，再进行分级，并给以不同权限。外部用户提交执行代码要用 CA 证书加密传输，并验证身份。系统向用户提供隐私信息的保护机制，用户对其上传的代码可定义不同的隐私等级，如可见、可读、可运行等。
  4. 系统的实现采用分阶段的方式，第一阶段以内部用户为主，首先实现核心功能，然后再逐步完善安全功能并向外开放。

4 设计与实现

系统在结构上设计为内和外两个子系统，两者用 Socket 通信。作为面向用户的接口，外子系统又有 2 个实现版本分别是提供给用内部网访问系统用户使用的“普通版”和用 Internet 访问系统用户使用的“安全版”，后者与内子系统在物理上位于不同的设备，且需要在 CA 证书的支持下才能进入内子系统使用代码上传和任务提交等核心功能。具体如图 1 所示。

目前，包括内子系统和“普通版”的外子系统在内的第一阶段的开发已经完成，并可以通过内部网支持用户完成代码上传、任务提交、代码复用等执行功能，和注册、查询等管理功能。系统的模块结构图如图 2 所示。篇幅的原因，进一步的细节不再讨论。

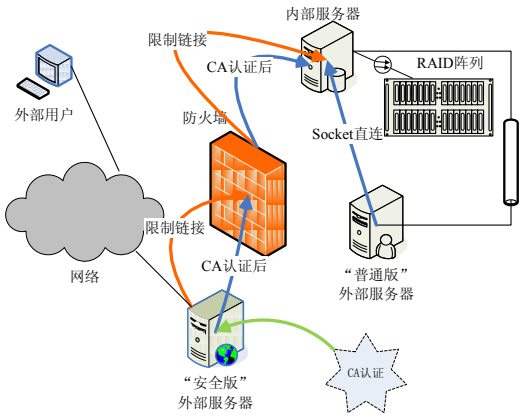

图 1 IP TASCM 结构图

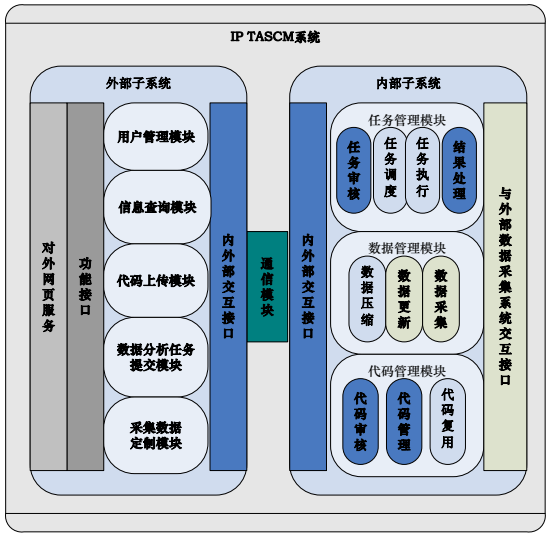

图 2 IP TASCM 模块结构图

5 研究总结

IP TASCM 能够以更有效的方式，支持所有面向 IP Trace 的相关研究工作。目前包括 100 多个条 Trace 数据的第一版的开发已经完成并开始向内部网用户提供服务。

今后的工作将在以下几个方面展开：

1．尽快完成“安全版”的外子系统，并通过 CERNET 提供相关服务。

2．研究并实现高效的 Trace 管理、进程并发管理策略，并将其集成入系统，目前系统在这方面的实现方案比较简单。

3．通过 CERNET2 和“安全版”的外子系统，完成包括实时数据在内的 Trace 定制和发布。

6 参考文献

[1] 李凤华,邵筱昕,孔师今等. SANTT:网络流量共享平台[J].大连理工大学学报, 2005 年 1 期：P 25-28.

[2] Mark Allman, Ethan Blanton and Wesley M. Eddy. A Scalable System for Sharing Internet Measurements[C].In Proceedings of the 2002 Passive and Active Measurement Workshop, Fort Collins, USA,March2002.http://www.icir.org/mallman/papers/simr-pam2002.ps.

[3] The MOME Project Consortium. Information Technologies Society - Cluster of European Projects aimed at Monitoring and Measurement [EB/OL]. http://www.ist-mome.org/.

[4] Colleen Shannon, David Moore, Ken Keys, et al. The Internet Measurement Data Catalog [J]. ACM SIGCOMM Computer Communication Review, Oct, 2005: Volume 35, Number 5.

[5] 程光，丁伟， 龚俭．高速网络流量通用测量平台-WATCHER [C]．见：CSIT 2004 会议论文集, 南京, 2004．122-128．