doi:10.3969/j.issn.1000-436x.2013.z2.002

基于 IPFIX 的网络流量日志系统

马云龙，张千里，王继龙
（清华大学 信息化技术中心，北京 100084）

摘 要：针对高速网络海量数据采集、存储和管理问题，分析了传统 IPFIX 流量日志系统在高速网络中的性能问题，提出了基于 IPFIX 协议的用户网络流量日志系统体系结构的优化设计，改进了数据聚类和存储算法，包括二元归并方式采集数据以及多层结构的散列算法存储数据。经校园网部署应用证明，可提供万兆链路下用户网络日志详单及准确上网流量计量值。
关键词：IPFIX；散列；流量日志；海量数据处理
中图分类号：TP393 **文献标识码**：A **文章编号**：1000-436X(2013)Z2-0005-04

Network traffic analysis system based on IPFIX

MA Yun-long, ZHANG Qian-li, WANG Ji-long
(Information Technology Center, Tsinghua University, Beijing 100084, China)

Abstract: To deal with the large-scale traffic capture and management in high speed network, the performance of IPFIX system was studied. An optimized architecture was proposed and implemented based on improved binary information integration to collect data and improved multi-layer hash algorithm to storage data. This system was deployed in Tsinghua university campus network and can scale to the 10 GB network traffic collection and accurate pricing.
Key words: IPFIX; hash; traffic log; large-scale data processing

1 引言

随着互联网应用的发展，网络使用模式出现了巨大的变化，P2P 及网络视频等新应用占据了大量的网络带宽，如果不加以有效调控和管理，校园网出口带宽的成本会过高，也会降低校园网用户的上网体验。通过了解校园网用户的上网行为，有助于规范网络管理、合理分配网络带宽、增强信息安全以及确保校园网络环境的平稳，从而提高校园网用户的工作效率。

网络流量分析的主要目的正是为了研究校园网用户的上网行为。网络流量分析可以帮助网络管理员知道什么业务在什么时候占用了多少网络资源？占用过多资源的用户是那些人？用户在过去的某段时间哪些应用增长比较快？网络异常流量发生在什么时间，是哪些子网中的哪些用户产生的？这些问题的分析结果对于优化网络及进行高效的网络管理具有重要意义。

流量数据的收集和分析有多种解决方案，包括 SNMP 协议的接口统计方式、RMON 方式和数据流量探针等[1~3]。这些方式各有特色，在不同网络环境中有各自的应用案例。清华大学校园网是全国最大的校园网络之一，拥有 12.7 万用户，高峰在线用户达到 4.5 万，即使在网络使用的低谷时期（凌晨 4 点左右），在线用户数都超过了 2 万人。为了满足广大用户的上网需求，清华大学高峰时段出口带宽达到 6 Gbit/s，每天产生的网络入流量约为 170 Tbit/s，在如此大规模的网络环境中使用以上的解决方案难以有效地采集、存储和分析出口网络的流量。

为此，本文基于 IPFIX 数据流进行了用户网络流量日志系统的研发，可对网络流量进行记录、存储及分析，既可准确定位网络异常流量，也能明确

收稿日期：2013-09-05
基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（2009CB320505）
Foundation Item: The National Basic Research Program of China(973 Program)(2009CB320505)

查出过去时间段内各种网络应用的带宽占用情况、用户的流量行为比例以及用户的上网日志明细等信息。

2　IPFIX 技术

IPFIX（IP flow information export）是 IETF（Internet engineering task force）基于思科 Netflow v9（RFC3954）开发的最新数据流输出标准。IPFIX 不但继承了 Netflow v9 基于模板的流信息输出格式，而且在此基础上对数据流输出的典型应用进行了输出规范的建议，并对 Netflow 中未涉及到的安全性问题进行了改进和扩充[3]。IPFIX 的部署主要包括流信息采集（exporter）、流信息收集（collector）和流信息分析（analyzer）3 类设备。

Exporter 将收集的网络流的统计信息封装在 UDP 报文中，发送给 Collector。一个 UDP 报文可以携带多条流的统计信息。IPFIX 的报文结构包括 IP、UDP 和 IPFIX 数据三部分。IPFIX 数据按照 NetFlow v9 格式进行组装，由报文头（packet header）以及流组（FlowSet）组成，图 1 为其报文头格式[4~6]。

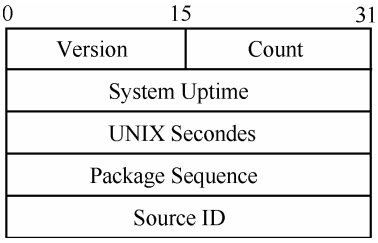


图 1　IPFIX 报文头格式

图 1 中，Version 表示 0x000A；Count 表示报文中携带的记录数量；System Uptime 表示设备运行的时间，以 ms 为单位；UNIX Seconds 表示从 UTC 时间 1 700 的 0 时至现在的秒数；Package Sequence 表示报文序列号，依次累加；Source ID 的取值为 0。

本文通过获取 IPFIX 报文的源 IP、目的 IP 以及 UNIX Secondes 进行相关研究。

选定基于 IPFIX 协议的网络流量处理技术路线后，笔者首先对传统的 CPU 单线程采集及数据库集中存储的 IPFIX 流量日志系统进行研究，发现这些系统在高带宽、高速率的校园网出口上部署效果不理想，海量网络流量数据流会将流量日志系统的 CPU 耗尽，导致 CPU 不响应。为此，笔者对传统算法进行改进，在基于 IPFIX 协议的网络数据流上开发了一套用户流量日志系统，用以提供可溯源的用户网络流量日志详单以及准确的上网流量计量值。

3　基于 IPFIX 的网络流量日志系统实现

当前本校校园网络实际运行中最突出的一个矛盾是用户对高带宽、高速率情况下短时间产生的计费流量不认可，而在用户对网费有异议时校园网计费系统尚无法提供详实的上网流量细节。基于 IPFIX 的网络流量日志系统的研发能很好地解决此问题，其已成功部署于校园网出口，如图 2 所示。

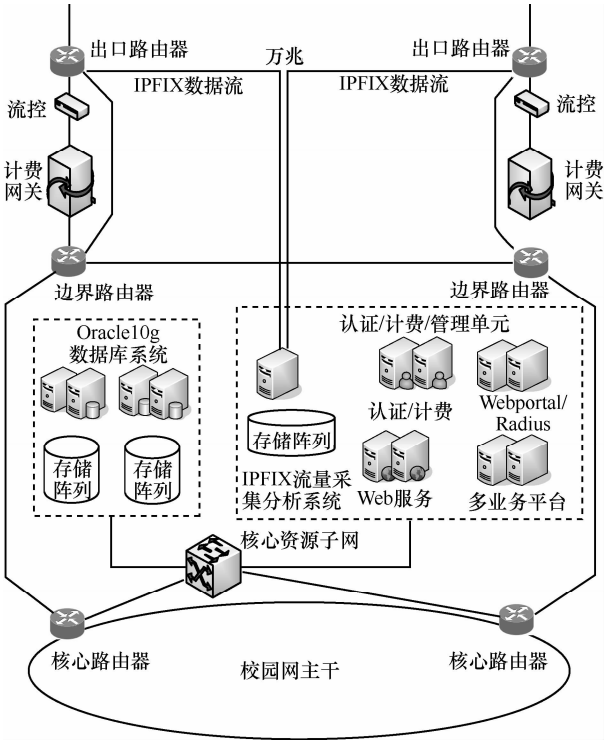


图 2　基于 IPFIX 的网络流量日志系统拓扑图

出口路由器通过 UDP 协议将 IPFIX 数据流发送给 IPFIX 流量服务器，IPFIX 流量服务器执行采集、存储和分析的系统功能，系统提供 Web 接口供管理员查询。

通过实际测试，当开启出口路由器 1:1 的采样比模式时校园网出口每分钟的 IPFIX 数据流为 800 万条左右，因此采用集中式数据库的表结构存储数据显然无法完成。传统的一些设备和应用系统为了将数据流写入数据库会大幅删减数据及降低分析时间粒度，这会导致此类系统一般仅可提供短时间内部分参数的 Top-*N* 信息。为了实现对历史流量的存储和分析，必须采取更高效的数据存储及处理方式。

为此，本系统选用开源嵌入式 SQLITE 文本数据库[7]，并按照系统需求对 IPFIX 数据流进行了合理高效的数据归并：通过对原始数据进行二元归并的方式降低写数据库的负担，并保证数据库内信息的准确。本系统以源地址和目的地址为关键字段进行归并，记录下相同 IPFIX 数据流的分组数和总长度，并记录下相同数据流的到达时间和更新时间。当一条数据流的更新时间超过了 60 s 时，则将这条记录写入数据库进行存储，据此可对海量出口网络流量数据进行精简并将需要的信息存入数据库。

在本应用系统中主要有 2 个模块：报文采集模块和数据库写入模块。为了应对庞大的数据流量，报文采集模块采用多线程的方式进行运作。为了最大化提高模块的处理效率，系统启用的线程数为 CPU 的核数减 1（例如 CPU 为 24 核，则启 23 个线程）。系统架构如图 3 所示。

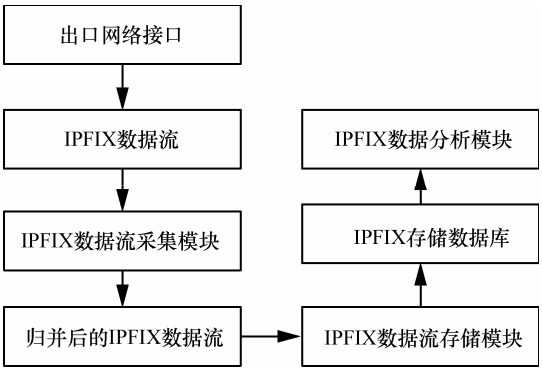


图 3　系统架构

采集模块从 IPFIX 报文中取出源 IP、目的 IP、分组数及分组长字节数并写入一个内存缓存表。缓存表的容量为 1 000 万条记录，格式为源地址、目标地址、写入时间、更新时间、分组数和字节数。缓存表对 IPFIX 数据流做进一步二元组归并。每条记录的写入时间为第一个报文的到达时间，更新时间为最后一个报文的到达时间。

存储模块为单线程，每 10 s 检测一次缓存表中每条记录的更新时间，如果某条记录的更新时间距当前时间已大于 1 min，即表示 1 min 内没有此条记录的相关报文产生，则将此条记录写入 SQLITE 文本数据库中，同时清空缓存表中相应信息，通过如上处理可对持续的数据流（如下载文件）起到很好的归并作用，避免记录数过多影响查询效率。

缓存表的缓存机制是本套系统的关键所在，其采取了多层结构的散列算法[8,9]。散列表采用多层表的方式，每一层表的容量为 16 bit，即 65 536。由散列函数可知，不同源 IP 和目的 IP 的数据流在散列表中位置相同的概率很小，同时由于使用多层的散列表结构，每个相同位置对应的存储空间可能有 $1\times10^7/65\ 536=152$ 个冲突，即每个线程最多寻找 152 次就可以找到对应数据流所在的位置。由上文的论述可知，一个数据流在缓存表中存在的时间约为 1 min，互相冲突的数据流在间隔 1 min 左右都可获得一次重新竞争缓存区的机会，理论上来说获得缓存表的几率是相等的，因此当数据流的冲突超过了散列表的承受范围而导致数据流丢失时并不会造成严重后果。只要数据流存在的时间超过一定时长，则系统就不会丢失该数据流的所有信息；过短的数据流虽然会被丢弃，但由于是较小的数据流，所以对整个用户系统的记录也不会产生较大的影响。图 4 为多层结构的散列算法原理图。

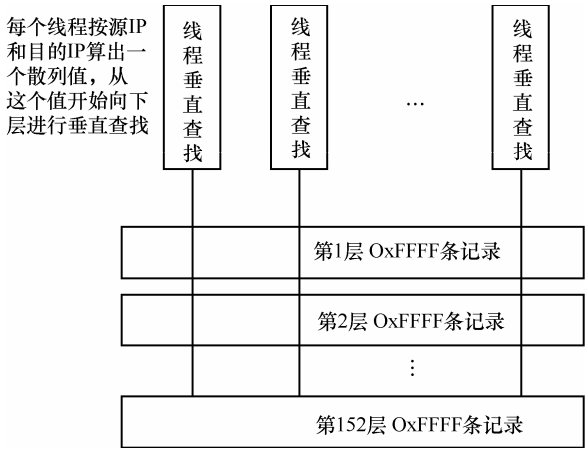


图 4　多层结构的散列算法原理

多层结构的散列表虽然消耗的内存比较大，但是效率高，可实现多线程 0XFFFF 个分支的并行查找。

4　实验结果与分析

IPFIX 流量采集及处理服务器使用 IBM 的 X3650 M4 服务器（配置 2 颗主频为 3.33 GHz 的 6 核 CPU，内存为 24 GB），服务器挂载存储阵列作为数据存储空间。本系统目前已正式部署在清华大学校园网出口，每天产生的流量日志记录达到 130 亿条，但 CPU 的使用率一般仅保持在 1%左右，因而能够很好地适应大型校园网络的实际使用。此外，本系统的 CPU 使用率在网络使用的高峰期和低谷期差别并不大，这是因为两点，首先，IPFIX

系统已经对流量进行了流归并，在高峰和低谷期流的数目差别要小于流量的差别（因为流的数目主要受扫描、DOS 攻击之类的异常行为影响）；其次，架构中所用的多层散列算法中采用了以空间换时间、以物理内存换取 CPU 消耗时间的思路，无论在流量高峰期还是低谷期，CPU 查找数据流所在位置的时间并不会有太大的变化。

目前本系统在校园网实际运行过程中数据库大约每分钟增长 30 MB，每天生成约 45 GB 的文件。按源地址或目的地址单个查询条件查询一天的流量日志，大概费时 20 s 左右即可得到一天的流量日志。

本系统还可作为计费系统的验证系统。以某用户的实际运行案例来看（如图 5 所示），该用户在 1 小时 27 分 29 秒的联网时段内计费系统产生了 13.4 GB 的入流量，从本系统查询可知该时段内用户的详细流量去向，以及该时段内总入流量与计费系统产生的计费数据是基本一致的。

序号	开始时间	结束时间	时长	源IP	目标IP	包数量	包长度
1	2013/7/10 16:20	2013/7/10 16:20	0	61.147.103.98	59.66.161.81	1	40
2	2013/7/10 16:22	2013/7/10 16:22	16	221.130.45.201	59.66.161.81	6	312
3	2013/7/10 16:22	2013/7/10 16:22	0	222.192.185.19	59.66.161.81	4	208
4	2013/7/10 16:22	2013/7/10 16:22	18	110.75.146.112	59.66.161.81	4	208
5	2013/7/10 16:22	2013/7/10 16:22	4	111.91.133.91	59.66.161.81	2	88
6	2013/7/10 16:22	2013/7/10 16:22	0	119.75.217.56	59.66.161.81	1	52
7	2013/7/10 16:22	2013/7/10 16:22	0	175.6.0.123	59.66.161.81	2	104
8	2013/7/10 16:22	2013/7/10 16:23	5	175.6.0.102	59.66.161.81	2	104
9	2013/7/10 16:23	2013/7/10 16:23	0	218.61.20.9	59.66.161.81	1	52
10	2013/7/10 16:22	2013/7/10 16:23	45	208.69.152.105	59.66.161.81	5	260
11	2013/7/10 16:23	2013/7/10 16:23	20	207.46.70.144	59.66.161.81	4	208
12	2013/7/10 16:22	2013/7/10 16:23	51	121.195.178.201	59.66.161.81	9	444
13	2013/7/10 16:22	2013/7/10 16:23	76	117.79.93.51	59.66.161.81	15	768
14	2013/7/10 16:23	2013/7/10 16:23	11	121.195.178.202	59.66.161.81	3	156
15	2013/7/10 16:23	2013/7/10 16:23	0	68.232.44.251	59.66.161.81	3	156
16	2013/7/10 16:23	2013/7/10 16:23	0	192.154.107.190	59.66.161.81	1	40
17	2013/7/10 16:23	2013/7/10 16:24	15	221.130.45.198	59.66.161.81	3	156
18	2013/7/10 16:22	2013/7/10 16:24	90	72.246.103.32	59.66.161.81	6	312

图 5 某用户的流量日志截图

5 结束语

介绍了一种基于 IPFIX 协议的高带宽大规模校园网络的流量日志系统，该系统采用了高效率的数据采集、存储及查询模式，尤其是多层结构的散列表结构模型，可有效解决服务器 CPU 对于高带宽、高速率的海量数据流无法高效响应的问题。该系统目前已作为用户流量日志查询的常规运行系统，实际运行效果良好，可以很好地解释用户的联网行为、联网流量的去向和多少，更深入的数据挖掘研究可依托本系统继续进行。

参考文献：

[1] 王珊, 陈松, 周明天. 网络流量分析系统的设计与实现[J]. 计算机工程与应用, 2009,45(10):86-88.
WANG S, CHEN S, ZHOU M T. Design and implementation of network flow analysis system[J]. Computer Engineering and Applications, 2009,45(10):86-88

[2] 李兴国,费玲玲.基于 NetFlow 的流量分析技术研究[J].微计算机信息,2008,24(3-5):198-200.
LI X G, FEI L L. The research of NetFlow-based flow analyzing technology[J]. Microcomputer Information,2008,24(3-5):198-200.

[3] 刘克难,赵慧娟,魏俊超.IPFIX 网络流量分析技术研究与系统设计[J].微型机与应用, 2011,30(4):11-13.
LIU K N, ZHAO H J, WEI J C, Research and system design of IPFIX flow analysis technology[J]. Microcomputer & Its Applications, 2011, 30(4):11-13.

[4] IPFIX 技术白皮书[EB/OL]. http://www.maipu.cn/ new. aspx?id=1920, 2010.
IPFIX technical white paper[EB/OL]. http://www.maipu.com/new.aspx?id=1920, 2010.

[5] IP flow information export (ipfix)[EB/OL]. http://datatracker.ietf.org/wg/ipfix/charter/, 2010.

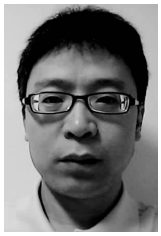
[6] TRAMMELL B, BOSCHI E. An introduction to IP flow information export[J]. Communications Magazine, IEEE, 2011,49(4):89-95.

[7] 沈永增,姚萌萌,周巍.空间数据在嵌入式导航系统中的索引[J].计算机系统应用,2010,19(4):85-88.
SHEN Y Z, YAO M M, ZHOU W. Spatial data index in embedded navigation system[J]. Computer Systems & Applications, 2010,19(4): 85-88.

[8] 陈一骄,卢锡城,孙志刚.面向流管理的哈希算法研究[J].计算机工程与科学,2008,30(4):26-29.
CHEN Y J, LU X C, SUN Z G. Research of the hashing algorithms based on IP flow management[J]. Computer Engineering & Science, 2008,30(4):26-29.

[9] 强士卿,程光.基于流的哈希函数比较分析研究[J]. 南京师范大学学报(工程技术版),2008,8(4):25-28.
QIANG S Q, CHENG G. Comparison and analysis of hash algorithm based on flows[J].Journal of NanJing Normal University (Engineering and Technology Edition), 2008,8(4):25-28.

作者简介：



马云龙（1972-），男，黑龙江嫩江人，硕士，清华大学工程师，主要研究方向为宽带认证、计费、Oracle 数据库、大规模邮件系统等。

张千里（1975-），男，内蒙古包头人，博士，清华大学副教授，主要研究方向为 IPv6 网络真实源地址、互联网网络安全、网络监测等。

王继龙（1973-），男，黑龙江大兴安岭人，博士，清华大学教授，主要研究方向为大规模互联网的规划、建设、运行和研究等。