Software Engineering and Applications 软件工程与应用, 2013, 2, 131-137
http://dx.doi.org/10.12677/sea.2013.26023 Published Online December 2013 (http://www.hanspub.org/journal/sea.html)

Design and Implementation of a Dataflow Monitoring System Based on the Genetic Algorithm Thought

Ming Zhu, Weiping Yang, Houqin Su

Donghua University, Shanghai
Email: zhuming@dhu.edu.cn, dongganyangweiping@126.com, suhq@dhu.edu.cn

Received: Nov. 11th, 2013; revised: Nov. 30th, 2013; accepted: Dec. 6th, 2013

Abstract: Traffic monitoring plays an extremely important role in network management, which can understand the current network running status, judge its performance and network failures, and also can reflect if it's on the safety time or not, then help to strengthen the network security. Dataflow monitoring is an effective method to know the network running status, performance, some troubleshooting and enhance the network security. In this paper, the monitoring system is designed with data acquisition layer, pre-processing layer, and interaction analysis layer. To upload the local database information to the central database efficiently, the system uses multi-thread to upload the record in the pre-processing program of the pre-processing layer and also uses the idea of the genetic algorithm to assign almost the same data records for each service thread to ensure the time consistency of data collection. The system achieves functions of real-time monitoring, statistics analyzing of the flow information for office network in one large city, and shows flow data results in report form as a basis for strategy formulation of network management and security.

Keywords: Metropolitan Area Network; Flow Monitoring; Flow Analysis; Genetic Algorithm

一种基于遗传算法思想的流量监测系统的设计与实现

朱　明，杨蔚萍，苏厚勤

东华大学，上海
Email: zhuming@dhu.edu.cn, dongganyangweiping@126.com, suhq@dhu.edu.cn

收稿日期：2013 年 11 月 11 日；修回日期：2013 年 11 月 30 日；录用日期：2013 年 12 月 6 日

摘　要：流量监测在网络管理中的承担着极其重要的角色，既可以了解当前网络的运行状态，判断其性能和网络故障，并能够反映安全时间发生与否，帮助网络安全的加强。本文所设计与实现的流量监测系统采用数据采集层、前置处理层、交互分析层的三层系统架构，并且为了高效地上传前置采集数据库信息到基础数据库中，在前置处理层中的程序中采用了多线程并行上传，并运用了遗传算法的思想为每条服务线程分配近乎相同的数据记录，从而保证数据上传在时间上的一致性。该系统实现了对大型城市办公网络的流量信息的实时监控、统计分析功能，并能以报表形式展示流量数据结果以作为网络管理和网络安全策略制定的依据。

关键词：城域网；流量监测；流量分析；遗传算法

1. 引言

网络流量信息的获取是影响一个网络管理是否行之有效的不可或缺的途径，就大型城市的办公网络而言，更是迫切需要对其网络进行流量监控以保证工

作的高效性和数据传输的安全性。

目前针对国内外各类网络流量的监测系统的研究及应用较为广泛[1-6]，而专门针对大型城市的办公网络的流量监测系统的设计与实现为数不多，且由于大型城市的办公网络结构相对复杂，流量采集难度较大，流量信息的储存、分析也更为困难；市场上网络监测软件尽管管理功能相对完善，但价格比较昂贵，而且为充分发挥功能需要二次开发[7]。本文正是针对现有网络监测系统存在此类问题，设计并实现了一种针对某大型城市办公网络的流量监测系统，通过在网络汇聚点设置数据采集点，进行分布式的流量数据采集，进而在本地数据库存储该流量信息，而后通过将各本地数据库中的流量信息上传至中心数据库，继而统一地进行流量信息分析，最终提供给用户最为直观报表，从而实现较为全面的网络流量监测，在加强监管的同时提高了网络的安全性。

2. 问题分析

本系统的网络结构是以办公外网管理中心为核心枢纽，且该管理中心亦是连接市办公外网和国家办公外网、公共互联网等外联网络的枢纽；外网结构上使用的是 MPLS/VPN 网络作为承载网。所有区县二级办公网络、委办局、直属单位都是接入在 MPLS/VPN 承载网上。

当市办公网用户单位和管理中心、外部网络有数据交换时，经由 MPLS/VPN 承载网到达管理中心，由管理中心进行路由选择；当市办公外网用户单位和市办公外网内部用户单位有数据交换时，直接在 MPLS/VPN 承载网上完成数据交换。这种架构有效缓解了是管理中心的网络在数据交换带来的压力。但同时也带来了问题，因为是办公外网用户单位之间有数据交换时，流量不通过外网管理中心，因此该管理中心不能有效的监控、管理市办公外网用户单位之间的数据交换。

为了使办公外网管理中心监视到接入在承载网上单位之间的数据交换信息，需要在各个接入点配以数据采集设备(需带有数据库以储存流量信息)通过程序将 IP 数据包中的信息，存储与本地数据库中，而后将各个接入点下的数据库内的数据记录上传至管理中心的中心数据库，并将流量数据进行分析处理，将分析结果通过交互页面向用户展示，即符合数据采集，前置处理，分析交互的三层结构。

在实际的承接网上存在几十个接入点，且在办公网络环境下，每个数据的数据采集设备(采集设备的数据库)内都储存着大量的流量信息记录，问题的难点在于：首先如何高效的将本地数据库中的纪录上传至中心数据库中，在实际项目中可以通过开设多条服务线程并行作业来解决；其次，承接上一个解决方法，接下来难点即需要确定每条服务线程承载多少数据量(流量数据记录)。诚然，为了使各服务线程可以在几乎相同的时间内完成数据上传的任务，以保证同一周期内数据在采集时间上的一致性，理应为每条服务线程分配尽量相同数量的数据条数(即记录一定时间内各个数据库的记录条数，按照数量的多少，力求平均地分配到各个服务线程中)。这也是本文需要处理的最大的难点所在。

3. 数学模型与遗传算法

3.1. 数学模型的建立

首先，承接上文的难点描述为本地存放了一定数量的数据记录，需要通过指定数目的服务线程上传，并且每条服务线程需要分配尽量相同的数据库记录条数。

这是一个典型的集合划分问题(Set Partitioning Problem, SPP)。所谓集合划分问题，是指将一给定集合划分为指定个数互不相交的子集，并使每个子集含有的元素大小之和尽可能一致。它的判定问题严格叙述为：

(SPP)实例：有穷集合 $A=\{a_1,a_2,a_3,\cdots,a_n\}$，以及每一个 $a\in A$ 的“大小” $w(a)\in R^+$；正整数 $m\in Z^+$。问：是否存在一个关于 A 的划分 $\sigma=\{A_1,A_2,\cdots,A_j\}$，使得对 $\forall j=\{1,2,\cdots,m\}$，有

$$\sum_{a\in A_j} w(a)=\frac{1}{m}\sum_{a\in A} w(a)\ ?$$

将一给定的信息主题集合 $A=\{a_1,a_2,\cdots,a_n\}$ (a_i 即对应本项目中的某个本地数据库中的数据记录条数)，划分为 m 个互不相交的子集(即分配至 m 个服务线程)，并使每个子集(对应每条服务线程)含有的元素大小(各个服务线程上承载的数据库记录数目)之和尽可能一致。

在数学上，称使各子集中元素大小之和尽可能一

致的 SPP 为 SPP 优化问题。对于它，人们通常从两个方面来考虑。其一是使最大的子集最小化，另一是使最小的子集最大化。

针对使最大子集最小化的数学模型可进行如下描述

$$\min \max_{1\le j\le m} S_j \tag{1}$$

约束条件：

$$\sum_{j=1}^{n} x_{ij} = 1 \ \left(i = 1,2,\cdots,n\right) \tag{2}$$

指定元素 i 只能分配到一个子集，$x_{ij} = \{0,1\}$。其中 $S_j = \sum_{i=1}^{n} a_i x_{ij}$，表示分划分后的各个子集的元素之和(即对应各个服务线程中的数据记录条数的和)。a_i 仍表示对应本项目中的某个本地数据库中的数据记录条数。x_{ij} 表示元素 i 是否分配到子集 j 中，是则取值为 1，否则为 0。

为了进一步简简化模型(1)，将模型转化为

$$\frac{\min\left(\sum_{j=1}^{m}\left|s_j - s\right|\right)}{m} \tag{3}$$

经证明[8]，两个数学模型等价，其中：$S = \left(\sum_{j=1}^{m} s_j\right)\Big/ m$ 为全部子集的元素之和的平均值，该式中的目标函数表示各个子集和到平均值的平均距离。

3.2. 数学模型的解决方法——遗传算法

解决式(3)中的模型的方法，即是求解 $\min f\left(x_1, x_2, \cdots, x_n\right)$ 的最优化问题，而遗传算法是一种在计算机科学人工智能领域中用于解决最优化的一种搜索启发式算法，它可以以用来生成有用的解决方案来优化和搜索问题，所以遗传算法十分适合求解如下类型的最优化问题：$\min f\left(x_1, x_2, \cdots, x_n\right)$ 或 $\max f\left(x_1, x_2, \cdots, x_n\right)$，因此该模型可以直接遗传算法用来解决问题。并且为了提高算法的收敛速度，本文采用精英保留策略。在第一代进化完成后，其最优秀的个体被复制到“当前最优”个体。在以后的每代进化完成，计算出各个体的适应度，分别找出最优和最差的个体。再把最优个体与“当前最优”个体进行比较。如果前者比后者还优秀，则用前者覆盖后者，作为新“当前最优”个体。总是把“当前最优”个体替换该代的最差个体。

从式(3)描述的数学模型可见，在基因中用 1 个二进制位来表示一个 x_{ij} 是最简单直观的。但是，这种普通的表示方法，无法保证一个元素必须并只能分配到一个子集，从而会导致产生很多违反模型约束条件的基因。因此本文采用基因表示[8]，如图 1 所示。A 中第 i 个元素，若分配到第 j 个子集，则对应的 $d_i = j\left(1 \le j \le m\right)$。基因 d_1 到 d_m 组成。这样的基因表示法，保证了一个元素必须并只能分配到一个子集，从而不会导致产生任何违反约束条件的基因。如图 1 所示。

算法思路：

1) 在初始种群中计算各个个体的适应度，在此种群大小定为 50；

2) 采取精英保留策略，将适应度最好的个体保留；

3) 计算出个体的生存概率，通过轮盘赌选择算子进行个体选择；

4) 以固定概率进行杂交；

5) 为每个个体的基因片段随机生成随机数[0,1]之间，若该随机数小于给出的固定变异率则该基因片断变异；

6) 第二代个体生成；

7) 计算第二代个体的适应度；

8) 继续采取精英保留策略，挑选出最好和最差，若此代的最好比之前的最好更优，则替换之前的最优解，否则，用当前最优解替换最差的个体；

9) 若满足结束条件，则停止，不然，跳转第 3)步，找到所有符合条件的规则。

4. 监测系统体系框架设计

本文设计针对大型城市的办公网的流量监测系统在体系结构上划分为 3 个层次，由底至上依次为数据采集层、前置处理层、交互分析层，如图 2 所示。

数据采集层：用于基础数据的采集和预分析。

前置处理层(分析)：其基本功能为将采集层收集的数据进行分析统计，将前置数据库中的数据读出，并转换成中间基础数据保存到基本数据库中，并将数据传给界面层。

交互分析层(界面)：本层采用 BS 结构，用户通过浏览器来访问 WEB 操作界面。该层包含两个功能：

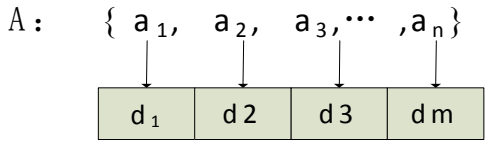

Figure 1. Gene expression
图 1. 基因表示

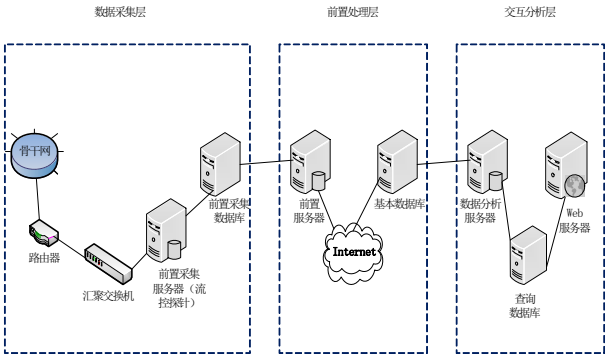

Figure 2. System architecture diagram
图 2. 系统架构图

提供与用户之间进行交互的界面和分析统计功能。

数据处理流程即为首先通过监听数据采集设备的两个以太网端口，分别捕捉监控镜像端口的进站、出站流量和出站的镜像数据包。数据采集设备将捕获的 IP 数据包进行基本梳理，获取 IP 数据包中的 socket 地址(源地址、源端口、目的地址、目的端口)、通讯协议、包的大小并加上时间戳，存储在数据采集设备的前置数据库中。数据采集设备按照后台监控服务器的指令要求(兼职采集程序)，将记录的数据通过网络发送给后台监控服务器的基础数据库中，而后后台服务器上数据分析统计程序的进行各种分析统计业务，然后将客户需要的结果保存到查询数据库中，而后 Web 程序会从查询数据库中获取数据，然后以图形或表格的方式呈现给用户。

5. 系统实现

考虑到目前外网网络已经建设成型，并且已经有了很多接入单位和应用。选择在不改变现有网络环境，不影响现有网络的基础上进行系统的实施和部署。

5.1. 数据采集层

本层主要负责利用数据采集设备采集流量信息并进行预处理。

目前广泛采用的交换网络中监听所有流量有相当大的困难，因此需要通过配置交换机来把一个或多个端口(VLAN)的数据转发到某一个端口来实现对网络的监听在本系统中，为了有效监控该网络的用户，我们在其二级网络的汇聚层设备上，利用 SPAN 复制上联 MPLS/VPN 端口数据到一台新增设的前置采集机(流控探针)上，端口的数据中包括端口发送的所有帧和端口接收的所有帧。

5.1.1. 原始抓包

原始抓包采用 Raw Socket 技术，接收本机网卡上的数据帧或者数据包，该程序是以发送接收 ip 数据包的方式创建这种 socket。首先使用经过初始化各个参数值后，预先定义一个 IP 头结构，而后与数据库(Mysql)进行连接，将网卡的工作模式设置为混杂模式以获取网络设备的所有数据包，同时设置一个过滤器，以确定只抓取 ICMP，TCP，UDP 的数据包，并对接收到的信息进行拦截处理，丢弃非法的信息，而后通过获取以太网的包头信息获取传输层的协议类型，根据不同的协议内容(ICMP, TCP, UDP)，分别烈性初始化相应的数据结构(入库的记录结构)，而后设置一个缓冲区数组，来存放流量数据，并将缓冲区的最大容量设置为一分钟内可以允许最多的抓包的记录数，达到上限时不再像缓冲数组插入数据，未达到上限的情况下，要遍历记录，若存在源地址和目标地址均相同的要予以合并，随后程序设置了存盘的时间间隔十分钟，若达到该时间，将缓冲区数组内的数据存入本地数据库中，而后清空缓冲区数组中的数据，未达到时间要求则继续像缓冲区数组内插入流量数据，而后一直循环从拦截消息之后的全部过程。

5.1.2. 数据包信息预处理

通过抓包程序已将 IP 数据包获取并以数据库的记录形式存于前置采集数据库中，为了方便之后的分析处理，所以需要将流量信息记录中的 IP 地址替换为对应的单位名称，即通过在遍历前置采集数据库中的流量信息数据表，获取其 IP 地址，一次同时遍历存储各个单位 IP 地址范围的数据表，若该 IP 地址在某单位的允许的 IP 地址范围内，则将该 IP 地址替换为单位名称，以更新原来存储于前置采集数据库中相应记录。

5.2. 前置采集层

5.2.1. 前置采集程序

主要负责下载数据采集设备上采集到的流量数据，将数据下载到前置服务器中的基本数据库后，删除前置数据库上的原有下载数据。

程序从相关数据表中获取各采集设备的数据端口地址，而后通过 IP 网络从前置数据库中获取数据，每 10 分钟获取一次。将获取的数据保存到基本数据库中的相关数据表中(此处需要通过多线程实现各个前置数据库中的数据记录上传到基础数据库中，下小节详述)，然后在前置数据库中的对应表中将相应的记录置为已获取状态。程序每 10 分钟检查一次基本数据库中的该数据表，将表中已经统计过的记录删除。具体流程详见图 3。

5.2.2. 多线程并行上传

在本文的实际项目中的承接网上存在 35 个接入点，每个采集点下连接一个数据采集设备(设备中自带数据库)内都储存着流经该接入点中的流量信息记录。正如通过开设多条服务线程并行作业的方式以高效地将前置数据库中的纪录上传至基础数据库中；而为实现各服务线程可以在几乎相同的时间内完成数据上传的任务保证同一周期内数据在采集时间上的一致性，即目标是为每条服务线程分配尽量相同数量的数据条数。诚如本文第二部分所讲，此处才有集合划分的思想，利用公式(3)中的数据模型和遗传算法的思想解决模型中的最小值问题。因此需要确定服务线程的数目以及各个前置数据库中的记录条数。

本文中，实际上在服务器上开启 5 条服务线程(经测试，前置数据库每分钟平均处理 1000~2000 条数据，35 个采集设备中可储存近 50,000 条数据，而根据服务器的行配置，一条服务线程一分钟内最多可以承担 12,000 条数据记录，所以理应开设五条服务线程。而通过对各个采集设备中的记录一个月的观察，以统计的方法求得，服务器每天处理记录约 3000 万条，而以下数据是各个采集设备中在一天之内存储的记录条数的，以集合形式表示：A = {26 285 17 8 70 167 27 61 72 15 24 220 54 154 72 28 137 132 51 179 1 42 215 18 208 64 53 191 41 23 35 116 18 23 153}。

按照集合划分问题里的遗传算法的思想中对应

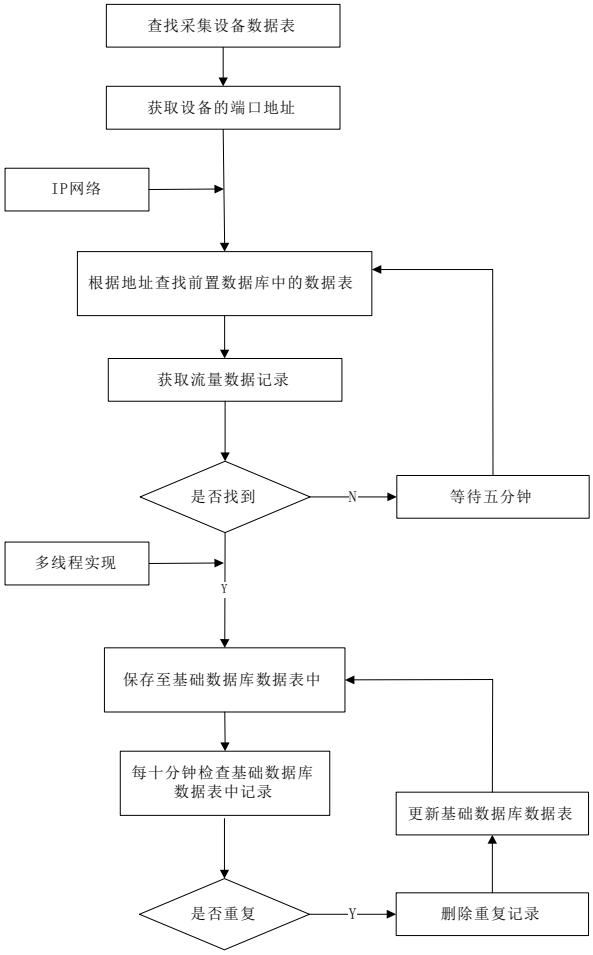

Figure 3. Front acquisition program flow chart
图 3. 前置采集程序流程图

的基因表示方法，即 A 中第 i 个元素，若分配到第 j 个子集，则对应的 $d_i = j\ (1 \le j \le n)$。基因 d_1 到 d_m 组成。，就本文的例子来说，d_i = {1,2,3,4,5}，m = 35，$1 \le j \le 5$。

5.3. 交互分析层

负责对采集到的数据分析处理，以形成流量统计报表，并将结果存入 web 服务器的数据库中供用户查询，可以查询到流量总计、起始地址和目标地址等信息以及提供与用户交互的界面。

在基础数据库中的数据表中，根据用户的各查询需求，创建相应的存储过程，利用存储过程在基本数据库中提取相应的数据集，一方面将该数据集存入结果数据库中，供给用户查询；另一方面利用该数据集生成需要的报表(日、周、月报表)。

生成报表方面，对于单位到单位(端口)的流量数

据分析，由于日报表以 10 分钟为统计单位，所以基本数据库中的带有端口的数据可以直接生成日报表，在此基础上，在源端口和目标端口一致的条件下，利用存储过程对以十分钟为单位的数据库记录进行累加，形成以一小时单位的数据记录，同理，月报表是在周报表的基础上形成 24 小时为单位的，数据记录。

对于不带端口的流量数据流量的分析处理，与带端口的大致相同，得到带端口的流量数据后，将公司 A 到 B 的所有端口的数据记录累加，形成日报表数据记录，而后过程同上。

而后分析处理后的数据记录存储在用于与用户交互的 WEB 服务器的结果数据库中，待用户按需求进行信息和相关报表的查看。

6. 运行测试

由于篇幅限制，此处重点展示数据采集，采用遗传算法下实现多线程分配任务量均等分配的运行结果，以及流量查询时生成的报表。

首先测试系统中数据采集部分，在对交换机进行正确的配置后，将经过端口镜像，程序抓包，初步解析，以十分钟为单位进行累计的流量信息存储在 Linux 操作系统下的本地数据库(MysQL)中。查询数据中的前 20 条记录的呈现结果，如图 4 所示。

其次，前置抓取程序中用多线程的形式实现从 35 个本地数据库到中心数据库的上传，开设了 5 条服务线程，为每条服务线程尽量分配数量相同的记录条数，运用了遗传算法的思想来解决这一集合划分问题。实际的分配结果如表 1 所示。

在实际运行的程序中，由于目标函数中需要将划分后的该集合元素与所有元素的平局值的差值作为分母，所以考虑的零的出现，无法作为分母，因此将出现零的时候赋予了一个 10^{-10} 的一个数字，因此 best fitness 会出现 10^{10} 这样的数字，实际证明当遗传到 1771 代是结果已经稳定，结果表示将原集合划分为各个元素相加相等(各子集中元素和为600，sumbest (x) = 600)的五个子集，5 组分别{167,220,137,53,23}，{26,208,191,41,116,18}.{17,61,15,24,72,28,132,1,215,35}，{8,70,27,72,154,51,42,23,153}，{285,54,179,18,64}，结果表明一号采集记录被分配到服务线程五中，二号被分配到服务线程四中，依此类推。

```
mysql> select  * from Area_To_Area  limit 0,20;
+----------+---------+---------+-------+---------------------+---------------------+------+-------+----------+
| ATA_ID   | SAreaID | DAreaID | DPort | BeginTime           | EndTime             | Size | SPort | DeviceID |
+----------+---------+---------+-------+---------------------+---------------------+------+-------+----------+
| 32550631 |     500 |    1871 |  2061 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  312 |     0 |        7 |
| 32550630 |     500 |    1871 |  2062 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  786 |     0 |        7 |
| 32550629 |     500 |    1871 |  3727 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  186 |     0 |        7 |
| 32550628 |     500 |    1871 |  3724 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 | 13506 |     0 |        7 |
| 32550627 |     500 |    1871 |  3726 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  2503 |     0 |        7 |
| 32550626 |     500 |    1871 |  3725 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 | 26864 |     0 |        7 |
| 32550625 |    1849 |     500 | 10079 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |    60 |     0 |        7 |
| 32550624 |     500 |    1849 |  4427 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |   712 |     0 |        7 |
| 32550623 |     500 |    1849 | 64354 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |    60 |     0 |        7 |
| 32550622 |    1017 |    1841 | 53323 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  4309 |     0 |        7 |
| 32550621 |     500 |    1849 | 64362 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |   557 |     0 |        7 |
| 32550620 |     500 |    1747 |  4685 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |    60 |     0 |        7 |
| 32550619 |     500 |    1871 |  4133 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |   120 |     0 |        7 |
| 32550618 |     500 |    1849 |  4428 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  5612 |     0 |        7 |
| 32550617 |     500 |    1849 |  2394 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 | 40431 |     0 |        7 |
| 32550616 |     500 |    1849 | 50600 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |    62 |     0 |        7 |
| 32550615 |     500 |    1849 | 53874 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |   449 |     0 |        7 |
| 32550614 |     500 |    1849 | 55968 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |   391 |     0 |        7 |
| 32550613 |     500 |    1849 |  2793 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |  3100 |     0 |        7 |
| 32550612 |     500 |    1747 |  4699 | 2013-07-11 15:20:00 | 2013-07-11 15:29:59 |   186 |     0 |        7 |
```

Figure 4. The flow information record of the database
图 4. 数据库中的流量信息记录

Table 1. The allocation result chart of the local database by the thread
表 1. 线程分配的本地数据的结果图

var(1) = 5	var(2) = 4	var(3) = 2	var(4) = 3	var(5) = 3
var(6) = 1	var(7) = 3	var(8) = 2	var(9) = 3	var(10) = 2
var(11) = 2	var(12) = 1	var(13) = 4	var(14) = 3	var(15) = 2
var(16) = 2	var(17) = 1	var(18) = 2	var(19) = 3	var(20) = 4
var(21) = 2	var(22) = 3	var(23) = 2	var(24) = 4	var(25) = 3
var(26) = 4	var(27) = 1	var(28) = 5	var(29) = 5	var(30) = 5
var(31) = 2	var(32) = 5	var(33) = 5	var(34) = 1	var(35) = 3
sumbest (1) = 600	sumbest (2) = 600	sumbest r(3) = 600	sumbest(4) = 600	sumbest (5) = 600
Best fitness:10000000000				

最后测试流量统计与分析，以报表形式展现，本系统可生成单位流量统计报表、单位到单位流量统计报表、单位到单位端口流量统计报表，并且每种类型的报表都包含日、周、月三种时段，其中日报表精确到 10 分钟，周报表精确到 1 小时，月报表精确到 1 天，此处给出通过查询单位到单位之间日流量统计结果而生成的饼状日报表以及双击饼状图的对应区域后形成的点对点的报表，如图 5 所示。

通过点击报表选项卡后生成的饼状日报表，可以直观的看出流向各处流量的比例大小。以及单击对应区域后，可以得知从市区所监站流向 unknown 的平均流量值，以及在各个时刻的流量大小便于网站管理人员的查看。

7. 结语

本文设计和实现的一种针对某大型城市的办公网络的流量监测系统，完成了在多级网络结构下的流量采集任务，通过分布式的部署流量采集点以及流量数据本地存储的方式确保了流量信息的完整性；而后通过多线程并行上传各本地流量信息至中心数据库

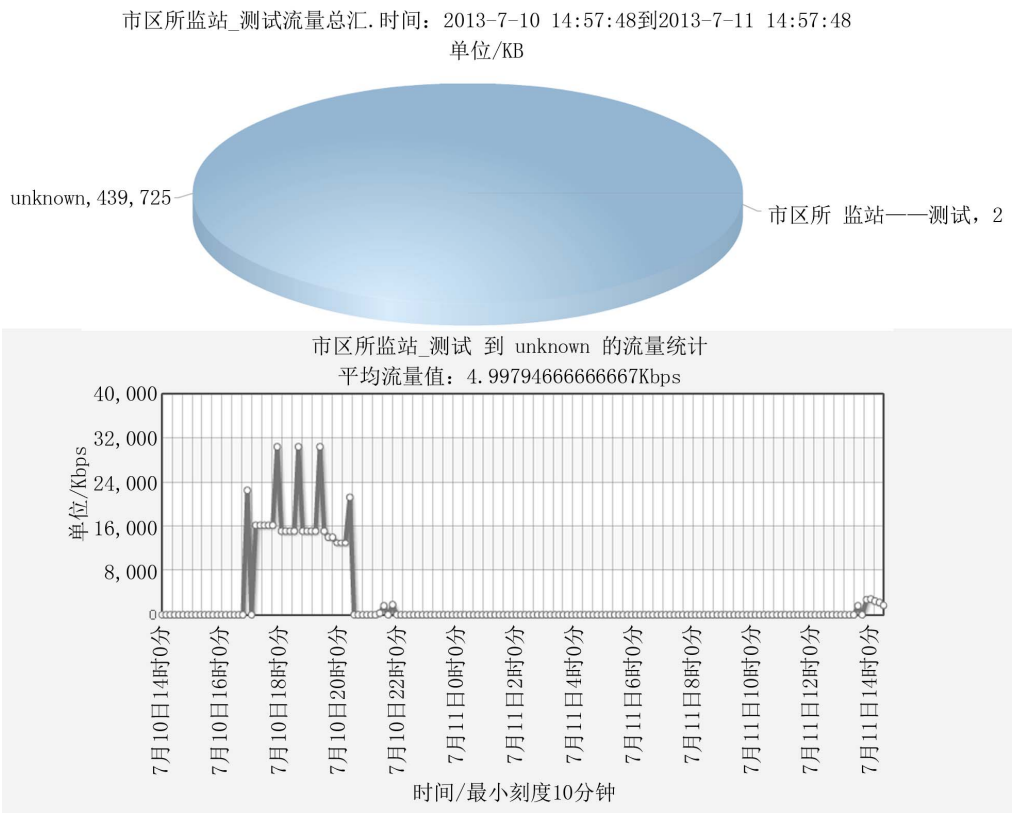

Figure 5. Daily flow data statistics pie report between different companies daily and daily point-to-point report
图 5. 单位与单位间的日流量统计饼状日报表和点对点日报表

(其中在为各服务线程分配任务量是，更是运用了遗传算法的思想，实现了“平均”分配原则，统一了数据上传的周期时间)，进行统一流量信息处理的方式保证了流量信息分析的全面性。最后，相关网络管理人员可以在页面上按需求查询经过细化统计分析后的流量数据并可以针对查询的流量数据生成直观的报表，保证了流量信息分析结果展示的直观性。该系统的展示结果为流量控制提供了有价值的参考信息，减轻了网络管理工作负的同时网络的安全性。该系统经过大量的用例测试后，仍然可以正常运行。

参考文献 (References)

[1] 赵新元, 王能 (2007) 基于 Web 的网络流量监测系统的设计. *计算机工程*, **33**, 237.

[2] 鲍江宏, 李炯城 (2008) 基于遗传算法的集合划分问题求解. *计算机工程与设计*, **11**, 2280-2281.

[3] Lim, K.S. and Stadler, R. (2005) Real time views of network trafficusing decentralized management. 9*th IFIP/IEEE International Symposium on Integrated Network Management* (*IM*2005), **5**, 16-19.

[4] 郑晓霞 (2012) 校园网异常流量分析系统设计与实现. 硕士论文, 中国海洋大学, 青岛, 22-24.

[5] 左靖, 王海龙, 杨奔全 (2009) 基于 WSDM 的校园网流量监测系统设计与实现. *电子技术应用*, **6**, 148-151.

[6] 刺婷婷 (2012) 网络流量监测系统的设计与实现. 硕士论文, 陕西师范大学, 西安, 41-43.

[7] 欧亮, 陈迅, 沈晨, 黄晓莹, 吕屹 (2013) IP 网络流量流向分析与预测技术研究. *电信科学*, **7**, 24-25.

[8] 王宏 (2008) 网络综合流量管理关键技术研究. 硕士论文, 国防科技大学, 长沙, 21-24.