基于 CA 的流量查询和预警系统

李 伟　丁 伟

(东南大学计算机科学与工程系网络中心，南京 210096)

摘要：针对 CERNET 的计费策略，设计并实现了一个地区中心级的用户国际流量查询和预警系统。该系统采用 CA 证书支持交互以保证信息和系统的安全，并采用柔性模型支持用户个性化的流量自动报警，在最大程度上保证了用户的利益以及网络的安全。

关键字：流量；计费；报警；CA。

中图分类号：TP393　　文献标识码：A

A CA based accounting and over usage warning system

Li Wei　Ding wei

(Network Center, Department of Computer Science and Engineering, Southeast University, Nanjing 210096)

Abstract: Aiming at the accounting policy of CERNET, the paper describes the design and implementation of an accounting and over usage warning system which is used in CERNET regional center. The system uses Digital Certificate during communication to insure security both for system and message, and a traffic warning model is proposed also, which works on user's individual traffic behavior automatically.

Key words: traffic; account; over usage warning; CA.

按国际流入量计费是 CERNET 最重要的计费策略之一。在网络应用不安全因素凸现的今天，由于盗用服务、病毒入侵以及 DOS 攻击等都可能造成异常的流量，对用户的利益和全网的安全造成潜在的威胁。因此，用户希望能够即时地通过计费系统查询自己的流量信息，并且在流量异常的情况下能够及时收到来自计费系统的警告；同时，为了保证用户的隐私和系统的安全，以上的交互应在一个安全的通道中进行。基于以上需求，我们设计并实现了一个基于 CA 的流量查询和预警系统。该系统通过一个安装在地区网边界的流量采集器过滤出国际流量，并根据用户 IP 地址范围分类。可在每天凌晨 1 点后用浏览器以天为单位查询已发生的流量，而当单位时间内的流量超过边界阀值时，系统会通过用户预留的邮箱自动发送一封警告邮件。该系统采用 X509CA 数字证书，实现用户的认证和对浏览资源的访问控制。用户对服务器的访问采用 SSL 协议交互，实现在网络上的密文传输，避免了信息的泄漏。而边界阀值的确定则是通过一个动态可调整的模型进行的。

1　流量采集和数字证书

系统源数据是现有计费系统的流量采集器在高速网上采得的计费信息。同时，系统应用了 CERNET 华东北网络中心的公钥管理体制 CALock 系统，并使用了其数字证书完成用户的认证。

1.1　流量采集

流量采集器通过连接在高速光纤主干网上的分光器从光纤主干上得到网络流的镜像，对经过的报文按照规则集进行匹配。其执行的功能主要有采样、过滤、数据处理等几种。

对分光器取得的镜像，采集系统从以太帧中截取 IP 报文头部开始的 34 字节（包含网络层和传输层的协议头部信息），加上报文时戳 8 个字节一起合成被采样报文的标本，而对报文内容不加处理。然后对 IP 报文中的地址信息进行匹配过滤，符合过滤规则的 IP 报文将按照规则预定义执行接受（ACCEPT）、抛弃(DROP)/抽样(SAMPLING)三种动作。对地址匹配成功的报文，采集系统将采样报文信息，并产生流量信息，送计费服务器进一步分类处理存入

收稿日期：2003-09.　**作者简介**：李 伟（1978－），男，硕士；丁 伟，博士，教授，博士生导师。

数据库。

1.2 CA 数字证书体制

系统采用的证书管理是 CERNET 华东北网络中心现有的公钥管理体制 CALOCK 系统。它基本上采用国际上通用的结构模式，并在此基础上增加了自主性。它采用了 RSA 公司的公钥加密标准 PKCS（Public Key Cryptography Standards）,在基本 PKI 部件，包括数字签名和证书请求格式等上实现了标准化；使用 LDAP 目录为 PKI 系统集成管理数据和控制信息的存储仓库，并实施多层次应用的政策控制管理。同时它使用了证书全生命周期管理，以及基于政策的证书管理体系结构。根据教育网用户诸层管理的特点，在 CA 间的认证方式采用划信任域的层次模型。

CA 的证书实体采用 ITU-T X.509 V3 证书格式，证书类型有 user_nenc、agent_ca、ssl、user_account、user 等五种，分别对应不同的权限范围，证书命名限制。在本流量查询和预警系统内，只用到了 agent_ca 具有签发和管理证书的权限，user_account、user 具有流量查询的权限。

2 流量查询和预警系统结构

流量查询和预警系统的整体框架和运行环境如下图所示。系统从计费系统获得数据源，并从公钥管理系统获得证书管理信息。提供给用户的是基于 SSL 协议的私有 WEB 页面查询和流量超额定所产生的告警信息，并由许可服务器应用 CA 证书策略对用户进行访问控制。

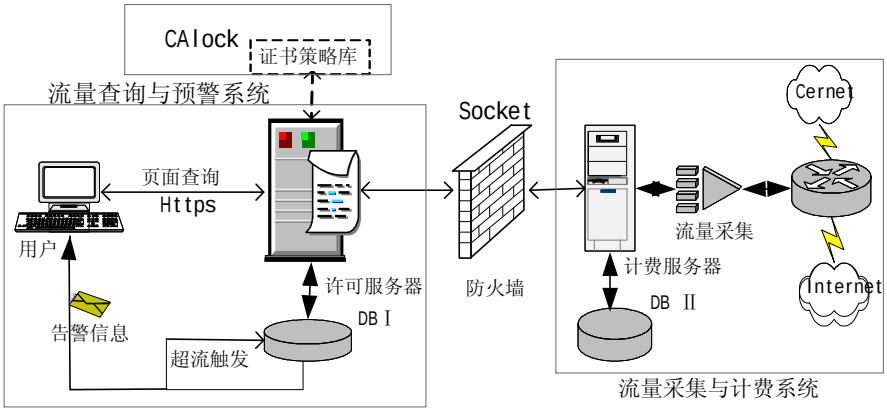

图 1. 系统整体框架图

从功能上说，整个系统分为四个功能模块，分别是：流量采集与传输模块、流量查询模块、流量的监控与预警模块以及系统日志和交互模块。

2.1 流量采集与传输模块

流量采集和传输模块实现了源数据的获取，并将数据存贮在系统数据库里。系统保有的数据库为 DBⅠ，而流量采集器采集的流量数据存放在计费系统的 DBⅡ里，DBⅡ不允许用户直接访问，本系统用户交互的是 DBⅠ。

服务器上的数据库都是采用 SYBASE 数据库。它是基于客户/服务器体系结构的关系数据库，并且有多线索化的特点，本身具有管理用户连接的功能，使在多用户连接时提高了性能。在本系统中 DBⅠ应用了以下几种类型的表：

1） Keytable（用户信息表）：此表的读写权限为 MASTER 级别。表中有以下的表项，Inter_num（序号）、Dn（CERNET 用户编号，对应 CA 的 CertificateSerialNumber）、Netname（用户域名）、Chname（用户中文名称）、Email（管理员邮箱）、Flag（是否接受告警标志）。

2） &Clent&SUM_table 和&Clent&MonthSUM_table(用户的日流量和月流量数据表)：此表为每单位各有一份，&Clent&表示各用户代号，读写权限为 USER 级别。分别记录有 IP 地址信息，时间，以及对应网络流量。

3） trafficvalve（流量阀值表）：读写为 USER 级别。表中有域名，单 IP 流量阀值和子网流量阀值。阀值表是数量上监控流量的依据，其值可以由强制法和自适应法两种方法来确定。

4） abtraffic（超额定流量表）：此表由各个流量表上的触发器动态插入。表项和各单位流量数据表相同。

在每个流量数据表上都建有一个触发器，对应&Clent&MonthSUM_table 为 t_&Clent&MonthSUM_table。触发器有按日和按月两种粒度。

数据库 DBⅠ的数据获取是由系统 CROND 进程每日在凌晨负荷小的时候自动调用 Shell 脚本执行，脚本调用 Sybase 数据库 BCP 函数由 DBⅡ增量转贮到 DBⅠ。从数据完整性出发，转贮将覆盖近三十天的日流量记录和近十二

个月的月流量记录，并且 DBⅡ可作为更久数据的备份。转贮结束后，超额定的流量将使触发器生效，动态插入 abtraffic（超额定流量表），同时写入日志供用户查询，并且用户可以选择是否接受此时的告警。

2.2 流量查询模块

流量查询模块主要完成两方面的功能：用户身份的识别和流量数据的获取。

用户和服务器的交互都是通过 SSL 协议来实现的，通过设置环境变量 SSLVerifyClient 使得服务器端、客户端都要求验证彼此的身份。而许可服务器的证书是在安装 Apache＋SSL 后由管理员手工生成的，客户端用户证书由用户向 CA 服务器申请获得。证书和签名的使用保证了身份的真实性和唯一性，而查询流量是加密传输的，即使被截获也很难破译。另外，许可服务器对外只开放用于 HTTPS 的 443 端口，用户只有通过 HTTPS 这一安全方式才能访问，这更加保证了数据信息的安全。

流量查询模块主要是提供用户流量客观情况的查询。具体分为按月份查询和按天查询两个粒度，并且根据 CA 证书类型不同分为管理员和普通用户两种权限级别，管理员可以查询全网所有的流量信息，普通用户只能查询本用户的具体信息。

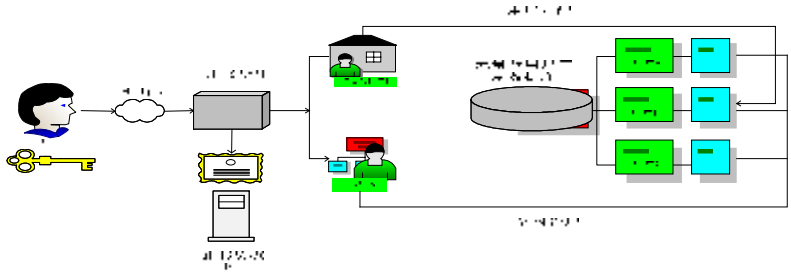

图 2．用户流量查询流程

在按月份查询的粒度下，系统以曲线图的方式表示出历史十二个月的流量变化，并以不同比例显示全网流量变化作为参照。按日查询时，系统提供当月流量的柱状图信息。在超额定流量的情况下，用户可以得到通知并查询具体子网和单机流量以及触发器数值。

2.3 流量的监控与预警模块

流量监控和预警模块检测流量的变化，并从全网和单用户两个角度分析变化特征，在流量异常的情况下分不同的级别产生报警。

从单个用户的角度来看，考虑到网络应用成本，用户希望国际流量能控制在某一个范围内，如果超出额定的范围，用户希望得到提醒。通过对数据库中触发器的设定可以实现这一功能。

而从全网的角度来看，每个用户流量相对于全网应具有相对的稳定性，这种稳定性可以量化为在全网整体流量中所占的比重。这样可以屏蔽由于客观日期（比如节假日）和突发事件对整体的共同影响，而只记录某个用户在全网中的跃迁，并对不同的跃迁程度分别采取不同的报警处理。

综合以上两点，监控与预警模块可以产生异常情况下的报警，报警模型下面有专门讨论。

2.4 系统日志和交互模块

系统日志和交互模块记录系统信息，以公告牌的形式供用户和管理员查询。该模块包含三部分内容：数据库日志、流量告警日志以及管理员信息。其中数据库日志从可读性和有用性来考虑只显示数据库的异常信息，而且管理员信息是用户和管理员有效交互的场所。

3 报警模型

3.1 超额定流量报警

额定流量是指在流量的绝对数值上超出预设值，触发器值即为预设的额定阀值。触发器数值的确定可以有两种办法可供选择：强制法和自适应法。强制法允许管理员根据自己的需要对触发器数值进行更改。自适应法将以触发器数值为基准，根据当前全网的实际流量变化，对触发器进行修正，并将修正后的值作为下一次触发器的基准数值。

触发器的自适应公式如下：

$$TRp = TR * (1 + (\Sigma CLi - \Sigma CLil)/ \Sigma CLi)$$

其中，TR 表示触发器基准，TRp 表示实际应用的触发器数值（修正值），ΣCLi 表示同级用户当前流量值的累计，ΣCLil 表示上一统计时间同级用户流量值的累计（其中，用户相对于单机统计代表同一子网内所有的单机，相对于子网统计代表全网所有的子网）。

当流量值超出触发器所设定的额定流量时，定义为事件 Y ，将引发超额定流量报警。告警事件被载入系统日志，并对管理员 E－Mail 通知。

3.2 流量值突变报警

流量值突变是指用户流量在全网的相对数值的跃迁。将整个系统各用户的流量属性按流量的相对大小分为五类：超低流量、低流量、中等流量、高流量、超高流量，属性值依次升高。分类的算法采用 C 均值聚类算法，流程如下：

设 Vi 为每个类的中心，d（Xj，Vi）为 Xj 与 Vi 的距离，c 为类数（令 c＝5），n 为用户数，Δ 为连续迭代时允许的最小偏差阈值；$E(k) = \sum_{i=1}^{n} \sum_{j=1}^{5} (Xi - Vj)^2$ 为聚类到 k 次时的均差。

1） 数据初始化。任选五个用户数值作为每一个类的中心 Vi，E（k）＝0，k＝0。

2） 对每一个用户 Xj，计算与每个中心点的距离 d（Xj，Vi），然后将 Xj 归入与之差值最小的中心 Vi 代表的聚类中去。

3） 将五类分配完毕，分别计算每一类包含所有数的均值，即为新的中心 Vi，计算新的误差均值 E（k＋1）。

4） 重复步骤 2）和 3）直到| E（k＋1）－E（k）| ＜Δ 或者 k 到达一定数值时为止。

在实际应用中，根据不同的流量情况可能有不同的 c、Δ 和 k 的取值使系统达到最优，目前 c＝5 和应用 k＝20 的循环可以使分类比较理想。算法的结果使用户按流量大小各自归类，这样每日用户流量数据的产生都可以将其规定属性。等第二日数据出现以后，同样采用此聚类算法，将用户重新归类，并记录其类别变化。

当属性值上下浮动为一时，定义为事件 S1；当属性值上下浮动为二三时，定义为事件 S2；当属性值上下浮动为四时，定义为事件 S3。报警级别如下：

1） 事件（$S1 \vee (S2 \wedge \overline{Y})$）为简单事件，载入日志，供管理员登陆时查看；

2） 事件（$S3 \vee (S2 \wedge Y)$）为告警事件，载入日志，并通知用户管理员检查网络（目前为 E-Mail 通知）；

4 结论

本文所介绍的基于 CA 的流量查询和预警系统实现了在现有 CERNET 内应用 CA 证书机制进行流量的查询和监控。提供不同粒度的流量查询，在额定流量的阀值方面实现的自适应方法，并且针对异常的子网流量突变，采取了一种异常检测的模型，提高了用户和全网的安全。应用 CA 数字证书实现了用户的认证和鉴权，采用 SSL 加密传输保证了用户的隐私。当然，为了使系统更具易用性，更简易的客户端平台，更灵活的交互方式等都是可以进一步研究和优化的地方。

参考文献：

[1] 高毓航. PKI 证书管理与政策的研究与实现[硕士论文]. 东南大学 2001.8.

[2] 陶浦洲, 李 强. Sybase 数据库技术大全. 科学出版社 1997.9.

[3] J.P.Marques de Sa(著),吴逸飞(译). 模式识别 －原理、方法和应用. 清华大学出版社 2002.11.