基于RPC机制的容错计费系统的设计与实现*

杨家海 吴建平

(清华大学计算机科学与技术系 北京 100084)
(清华大学信息网络工程研究中心 北京 100084)

摘要 网络计费管理是商业化计算机网络的重要网络管理功能.随着Internet商业化进程的推进和企业Intranet的广泛发展与应用,人们对网络计费管理的需求越来越迫切.对计费系统的一个基本要求就是高效、可靠.文章提出了一个基于RPC通信机制的容错计费系统FTCharge(fault-tolerant charge).在简述了计费的基本依据和原理以后,着重论述了FTCharge的总体结构和多容错机制的实现.

关键词 计算机网络,Internet,网络管理,SNMP(simple network management protocol),计费管理,RPC(remote procedure call).

中图法分类号 TP393

随着信息社会对网络的依赖程度越来越高,网络的高效、可靠的运行管理也越来越重要[1].和网络互连技术本身一样,网络管理正在向标准化的方向发展[2].在网络管理技术的研究、发展和标准化方面,国际标准化组织ISO和Internet体系结构委员会及其下属的工作组都做了卓有成效的工作.早在20世纪70年代末,国际标准化组织在提出其开放系统互连参考模型OSI/RM(open system interconnection reference model)的同时,就提出了网络管理标准的框架,即开放系统互连管理框架(ISO 7498-4)[3],并制定了相应的协议标准,即公共管理信息服务和公共管理信息协议CMIS/CMIP(common management information service/common management information protocol)[4,5].开放系统互连管理框架ISO7498-4将网络管理划分为5大功能域,即配置管理、失效管理、性能管理、计费管理和安全管理[3].在此框架基础上,Internet体系结构委员会提出了相应的网络管理标准,即简单网络管理协议SNMP(simple network management protocol)[6,7].

网络计费管理是商业化计算机网络的重要管理功能[8].早期Internet网络是一个学术研究网络,网络管理技术的发展相对滞后,尤其是计费管理更未得到应有的重视.随着Internet商业化进程的推进和Intranet的广泛发展与应用,人们对网络计费管理的需求日益突出[9].

为了满足这种需求,一些网络运行主管部门也自行开发了一些计费软件,但这些计费软件基本上都是单机运行的,一旦该机出现故障或者连接该机的局部网络中断,就会丢失计费数据,造成直接的经济损失.因此,如何确保计费系统的高效、可靠运行,就成了一个亟待解决的问题.提高可靠性的一个直接办法是多个计费系统同时运行.但由于大型网络的计费数据量很大,而且计费数据本身是动态变化和不可重复出现的,因此如何保证几个备份的计费系统之间同步协调工作就成为这类计费系统首先要解决的问题.本文提出了一个基于RPC机制的容错计费系统FTCharge(fault-tolerant charge).本文在简述了计费的基本依据以后,着重论述了FTCharge的总体结构和多容错机制的实现.

---

* 本文研究得到国家"九五"科技攻关项目基金资助.作者杨家海,1966年生,副教授,主要研究领域为协议一致性测试,网络互连技术与协议,网络管理.吴建平,1953年生,教授,博士生导师,主要研究领域为协议一致性测试,网络互连技术与协议,网络管理.

本文通讯联系人:杨家海,北京 100084,清华大学CERNET网络中心

本文1998-02-17收到原稿,1998-07-01收到修改稿

1 计费的依据和原理

计费的基本依据是用户使用网络资源的情况，如：信息传输量、占用线路的时间等. 对于专用租线来说，最常用的方法就是根据每个用户通过网络使用的信息传输量(通常称为计费数据)来计算其费用. 计费数据的获取有两种方式，一种是通过标准的 SNMP 操作来获取，另一种是通过总线监听来获取.

Internet 标准网络管理协议 SNMP 在定义了基本的网络管理操作的同时，也定义了一系列支持操作语义的管理信息变量 MIB(management information base)，其中就有和计费相关的 MIB 变量. 只要对被管理对象(通常是连接本网络和外部网络的边界路由器)作适当的配置，被管对象将自动记录所有通过该路由器的进出流量. 以 Cisco 路由器为例，它所记录的计费信息是如下格式的一张表：

Source IP Address	Destination IP Address	*n*Pkts	*n*Bytes
202.120.186.45	206.15.106.131	23	4 080
202.121.0.2	207.68.156.49	19	1 267
……	……	……	……

当每一个数据包由路由器通过时，路由器将搜索表中是否有与之相匹配的 Source IP Address 和 Destination IP Address 对. 如果找到匹配的记录，则将其累加，否则创建一个新记录. 这些记录可通过 SNMP 标准操作而获得.

另一种计费数据获取的方法是利用以太网总线的广播特性，在网络连接的适当地方接入一个只被动接听信息的设备，然后运行一个只截获数据包头的软件. 通过分析数据包头的有关信息可以得到用户使用网络的各种信息，如总的进出流量、流量的类型(如 telnet，ftp，http，mail 等等).

上述两种计费数据获取方法各有利弊，第 1 种方法是标准的方式，通过 SNMP 操作即可实现，而且也无需额外的设备投资. 但这种方法的一个缺点是计费数据粒度较粗，只计到 IP 级，即用户只能知道通信量的总字节数. 第 2 种方式的好处是用户可以根据需要指定计费数据的粒度，不但可以获取总通信量，还可以知道各种业务类型分别占用多少，这将有利于用户对不同流量类型采取不同的收费标准，也有利于用户就近配置适合的资源. 缺点是缺乏标准，且需额外的设备投资，而且依赖于网络拓扑结构. 用户在具体应用中可以根据需要选择其中的一种.

2 系统总体结构

FTCharge 系统是以 RPC 通信为主的分布式计费系统. 系统的总体结构如图 1 所示.

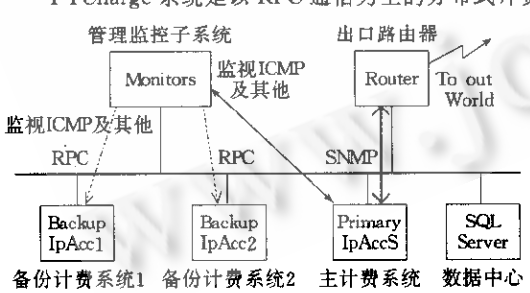

图1 FTCharge系统总体结构

整个 FTCharge 计费系统由 1 个管理监控子系统、1 个主计费系统和 1 个或多个备份计费系统组成. 管理监控子系统 Monitors 的主要功能是监视主计费系统和各备份计费系统的工作情况. 当它发现主计费系统出现故障时，将根据系统预先设定的策略自动地从多个备份计费系统中选出一个正常工作的系统，并启动该系统以接替计费数据的采集工作；当它发现主计费系统的故障已经修复并正常工作时，又会自动停止备份计费系统的工作，以确保任一时刻只有 1 个计费系统在工作. Monitors 以 ICMP 和一个进程监视工具来监视计费系统的网络连接、主机本身和系统各进程是否正常工作.

计费子系统是整个 FTCharge 的核心，它负责所有计费数据的采集、处理和加工工作. 从理论上讲，这些工作是可以在一个系统上实现的，但考虑到计费数据可能存在于多个主机上，为了维护数据的完整性和一致性，将数据的采集和后续的加工处理分开，专门设置一个数据库服务器来加工处理来自各个主机的计费数据. 这样做也有利于数据采集系统集中力量收集数据. 事实上，计费数据采集的实时性是很强的.

由于计费数据在被读取之前需在缓冲区中暂存，且缓冲区一旦被存满之后，后续的数据不再写入，因此，数

据采集系统在读取数据之后要负责清空缓冲区.因此,任一时刻只允许一个数据采集系统进行数据采集,否则将引起数据的不完整性和不一致性,甚至丢失数据.

备份系统本身的功能是与主计费系统完全一样的.只是在正常情况下它处于空闲(idle)状态.一旦主系统出现故障,它将收到来自监控子系统的指令,并进入工作(active)状态.此时,它完成与主计费系统完全一样的工作.

3　容错性的实现

容错性的实现是提高计费系统可靠性的关键.本系统采用基于远程过程调用 RPC(remote procedure call)的分布式监视和控制机制来实现系统的容错性.对计费系统的监视和控制是由监控子系统实现的.

监控子系统从逻辑上分成这样几个步骤:

(1) 监视主计费系统的网络、主机及系统运行情况;

(2) 当主计费系统发生故障时,选择一个可用的备份计费系统,并启动;

(3) 监视当前工作的备份计费系统;

(3.1) 继续监视主计费系统(故障恢复情况);

(4) 在当前运行的备份计费系统也发生故障时,从剩下的备份计费系统中选择一个可用的系统,并启动它;

(4.1) 当主计费系统已经恢复正常时,停止备份计费系统的工作.

为了更好地描述监控系统的工作过程,下面以半形式化的语言来描述它的实现算法.

Monitors 算法.

```
BEGIN
  定义主计费系统:Primary;
  备份计费系统数量:NUM;
  定义备份计费系统集:Backup[1...NUM];
  定义最大重试次数:MaxRetry;
    WHILE (TRUE) {                    /* 不间断运行 */
        Result:=Test(Primary);        /* 测试主计费系统的运行情况 */
        IF (Result="ok")
        THEN                          /* 主计费系统正常 */
        {     sleep a while;          /* 隔一段时间继续测试 */
              continue;
        }
        ELSE
        {                             /* Result="failed":当前测试失败 */
              i:=0;                   /* 一次失败不足以下结论 */
              WHILE (i<MaxRetry) {
                    Result:=Test(Primary);
                    IF (Result="ok")        /* 是瞬时性间断,现在又恢复了! */
                    THEN
                        i:=MaxRetry+1;      /* 跳出循环 */
                    ELSE
                        i:=i+1;     /* 进行下一次测试 */
              }
              IF (i=MaxRetry+1)
              THEN
              {
                  sleep a while;
                  continue;
              }
```

```
            ELSE                    /* Max Retry 次测试都有故障,可以认为该系统确实发生故障 */
            {
                x:=1;
                WHILE (x<=NUM){     /* 从备份系统中选择一个好的 */
                    IF (Backup[x] is OK)
                    THEN
                    {  RPC_start (Backup[x]);  /* 远程过程调用,启动备份系统 X */
                       startmonitorB (Backup[x]);  /* 开始对 X 进行监视 */
                       monitorP (Primary);  /* 继续监视主计费系统,以了解其修复情况 */
                       exit;
                    }
                }
            }
        }
    }  /* END of Outer WHILE */
END
```

对被启动的备份系统的监视流程与此相似,对故障主计费系统进行跟踪监视的过程也与此相似,只不过判断条件和执行的动作正好相反.限于篇幅,这里就不再重复了.

4　计费子系统

真正完成计费数据采集、处理、加工、统计乃至最终生成报表帐单等工作的是计费子系统.我们设计的计费子系统以关系数据库为核心,由数据采集、预处理、计费数据处理和计费信息查询等几个模块组成.其组成结构如图 2 所示.

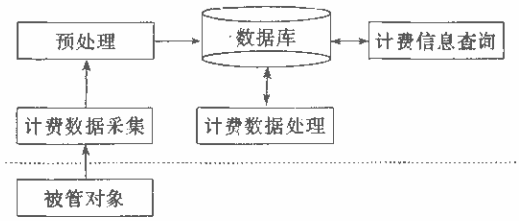

图2　计费子系统的组成结构

经过多方论证,在目前阶段,我们决定先实现第 1 种方式的数据采集功能,但为将来采用第 2 种方式(总线监听)的数据采集留出扩展接口.计费数据以小时为单位,先形成数据文件.预处理模块定期接收并处理这些文件,预处理模块将分析数据文件格式的正确性、IP 地址的合法性,并做初步的统计(累加)工作,最后形成数据库记录,并入库.

计费数据处理模块的功能是对数据库中的数据进行进一步的加工和处理,生成各种适用于专门需求的数据,并把这些数据以报表或帐单的形式打印出来.这些数据或报表可根据用户的不同要求灵活指定,如特定网段的数据汇总表、分项数据汇总表、IP 地址流量明细表、各种统计分类明细表以及计费帐单等.

计费信息查询模块的主要功能是向用户提供和计费相关的各种数据的查询.由于计费系统中保存着连网单位用户的所有计费数据,所以计费系统应该允许对帐单有疑问的连网用户查询或核实本单位个人或 IP 地址对外通信的流量情况.此外,计费系统还保存着连网单位的各种信息,如最新、最精确的 IP 地址分配情况等信息,这些信息可以对网络管理人员开放,同时也可以对具备一定条件和权限的其他用户开放.所有这些都将通过计费信息查询模块来实现.为了最大限度地实现计费信息的分布式查询,计费信息查询模块采用 WEB 技术加以实现,使用户的查询可以直接通过 WEB Browser,如 Netscape 等进行.

5　结束语

我们已经实现了一个基于上述设计的实际系统,并投入中国教育和科研计算机网网络中心的实际管理运行,经过一段时间的试验运行,证明上述设计是可行的,系统运行的可靠性比原先单机运行的时候有显著的提高,而且保证了计费数据的完整性和一致性,基本达到了预定的设计目标.需要指出的是,这样一种容错性的设

计并不局限于计费系统,任何对可靠性有较高要求的系统都可以采用.

这种设计方案的一个不足之处是对设备投资的要求比较高.对于一些中小型的网络运行管理部门来说,可以在经济性和可靠性之间作一个很好的折衷.事实上,FTCharge 系统的很多功能是可以合并到一个主机上来运行的,如数据的处理和数据采集,管理监控子系统和备份计费系统之间等都可以合并.当然,这将在一定程度上影响可靠性.如何在降低成本的情况下仍保持足够的可靠性,是我们下一步的工作目标.

参考文献

1 Aronoff R *et al*. Network management functional requirements. Management of Networks Based on Open Systems Interconnection (OSI): Functional Requirements and Analysis. NIST Special Publication 500-175, Nov. 1989. http://csrc.nist.gov/nistpubs/sp500—175.txt

2 杨家海等. Internet 网络管理. 中国计算机用户,1996,(17),9~11
(Yang Jia-hai *et al*. Internet network management. China Computer User, 1996,(17):9~11)

3 ISO/OSI. Basic Reference Model for Open System Interconnection. IS7498, Oct. 1984

4 ISO/OSI. Management Information Service Definition, Part 2: Common Management Information Service. IS9595-2, ISO, Switzerland, May 1988

5 ISO/OSI. Management Information Protocol Definition, Part 2: Common Management Information Protocol. IS9596-2, ISO, Switzerland, May 1988

6 Jeffrey D Case, Mark Fedor *et al*. A Simple Network Management Protocol (SNMP). RFC1157, Internet Network Working Group, May 1990

7 James M Galvin, Keith McCloghrie. Administrative Model for Version 2 of the Simple Network Management Protocol (SNMPv2). RFC1445, Internet Network Working Group, Apr. 1993

8 徐冈,汤俭. 网络计费管理. 见:张德运编. CERNET 的研究与发展. 西安:西安交通大学出版社,1997
(Xu Gang, Tang Jian. Network accounting management. In: Zhang De-yun ed. The Research and Development of CERNET. Xi'an: Xi'an Jiaotong University Press, 1997)

9 Cisco Systems Inc.. Cisco Enterprise Accounting for Netflow. Version 2.0, Cisco Enterprise Accounting for ISDN, Version 2.0. [on-line], URL http://www.cisco.com

Design and Implementation of an RPC-based Fault-tolerant Accounting Management System

YANG Jia-hai　WU Jian-ping

(*Department of Computer Science and Technology　Tsinghua University　Beijing*　100084)
(*Network Research Center　Tsinghua University　Beijing*　100084)

Abstract　Network accounting management is the important management function of commercialized computer network. With the commercialization of Internet and the occurrence of lots of Intranets, it's urgent to develop network accounting management system. A principal requirement of such systems is efficiency and reliability. A fault-tolerant accounting system based on RPC (remote procedure call) mechanism—— FTCharge (fault-tolerant charge) is proposed in this paper. After the brief discussion of the accounting principle, the design of the architecture and the fault-tolerance mechanism of FTCharge system are discussed.

Key words　Computer network, Internet, network management, SNMP (simple network management protocol), accounting management, RPC (remote procedure call).