电信级以太网虚拟硬件方法的研究

戴锦友[1,2] 余少华[2] 汪学舜[1,2] 朱国胜[1,2]
(华中科技大学计算机学院 武汉 430074)[1] (武汉邮电科学研究院 武汉 430074)[2]

摘 要 网络的不断发展对包括电信级以太网系统在内的网络节点设备的硬件资源容量提出了更高的要求。交换芯片通常是电信级以太网产品的主要功能单元,因不能编程导致可扩展性较差。在硬件结构不改变的情况下,电信级以太网设备提供的二层转发表、三层路由表、访问控制列表等重要硬件资源的容量就基本确定。但在实际应用中,经常会遇到这些资源不能满足应用需要的难题。当前的以太网产品,除了升级硬件,未采取措施来解决或弥补这个缺陷。另一方面,由帕累托原则可知,实际硬件转发表中的少数表项处于相对重要的地位,对网络流量的影响较大,而其它表项则处于重要程度相对较低的地位,对网络流量的贡献较小。基于这个事实,利用某种方法来管理和维护硬件资源,使其充分利用是可行的。提出了一种面向应用的解决硬件资源不足问题的方法,该方法借鉴操作系统的内存管理技术,采用特定的算法,在应用意义上等价扩充二层转发表、三层路由表、访问控制列表等硬件资源从而提高系统转发性能。该方法在内存中建立软二层转发表、软三层路由表以及软访问控制列表等,实际的硬件二层转发表、三层路由表以及访问控制列表可看作它们对应的高速缓存,通过生成访问频度等级信息,采用LFU算法来决定软表和硬表之间的数据交换,使得硬表中尽可能存放重要程度更高的条目,从而改善系统的转发性能。实验表明,该方法能达到预期的目标。

关键词 电信级以太网,虚拟技术,交换芯片,硬件资源
中图法分类号 TP302.8 **文献标识码** A

Research on Virtual Hardware Method for Carrier Ethernet

DAI Jin-you[1,2] YU Shao-hua[2] WANG Xue-shun[1,2] ZHU Guo-sheng[1,2]
(School of Computer, Huazhong University of Science and Technology, Wuhan 430074, China)[1]
(Wuhan Research Institute of Post and Telecommunication, Wuhan 430074, China)[2]

Abstract Continually development of the network brings about higher requests to hardware resources of network node devices including carrier Ethernet systems. Switch chips are usually important components of carrier Ethernet products and they are less expandable because that they can't be programmed. The contents of important hardware resources such as layer 2 table, layer 3 table and access control list are unchangeable without upgrade of the hardware. However, those hardware resources can't often meet needs of applications. As to current Ethernet products, no special measure except upgrading hardware is adopted to solve the problem. On the other hand, Pareto principle indicates that only the minority of items in hardware tables plays important role in practical network and the majority of the items has less influence on network behavior. So, it is feasible to adopt a special method to manage and make full use of the hardware resources. A method was presented to solve the problem of hardware resource inadequacy. Based on the virtual memory technology of operation system, a special algorithm was used to expand the hardware resources in order to improve system forwarding performance. Software tables are set up in the system memory and the hardware tables can be thought to be caches of their corresponding software tables. The LFU algorithm is adopted to exchange data between hardware tables and software tables in order that the hardware tables always store more important items, then the performance of carrier Ethernet system can be improved. The experiments show that the method can reach its expected aim.

Keywords Carrier ethernet, Virtual technology, Switch chip, Hardware resource

1 引言

艾瑞咨询(iResearch)的相关报告揭示,2007年美国社会化网络用户规模达到7200万,较2006年的6030万增长19.4%。预计到2011年,美国社会化网络用户将增长至1.05亿。在经历2007年和2008年两年的高速增长之后,2009年至2011

到稿日期:2009-03-05 返修日期:2009-06-18 本文受国家高技术研究发展计划"八六三"项目(2003AA121110,2005AA121410),电子基金项目("城域网多业务环MSR技术和面向比特的EOS方法")资助。

戴锦友(1968—),博士生,高级工程师,主要研究方向为电信级以太网,E-mail:djy96@fhn.com.cn;余少华(1962—),博士,教授级高工,博士生导师,主要研究方向为IP互联网络和城域网;汪学舜(1970—),博士生,主要研究方向为IP互联网络;朱国胜(1972—),博士生,主要研究方向为IP互联网络。

年的增长率将放缓，但依然保持在5%以上[1]。而在中国，随着互联网的日益普及，中国互联网用户人数也在不断增长，2006年中国互联网用户数达到1.32亿人，增长率为19%。iResearch预计，2009年中国的互联网用户超过2亿，到2010年时将达到2.46亿。2010年的互联网用户将比2005年增长1倍[2]。中美两国网络用户的增长，意味着网络用户的增长是世界性的。网络用户的飞速增长带来了网络的繁荣，也给现有网络系统和设备带来了更大的挑战，这体现在以下几个方面：

(1)网络用户的急速增长，意味着在网络中标识用户终端具有唯一性的以太网地址数量急剧上升，也意味着工作在二层的网络节点设备需要维护和处理的地址表急剧上升。

(2)网络用户的急速增长，意味着网络节点设备和用户终端设备数量同步增长，这意味着网络拓扑会变得更加庞大和复杂，网络中从任一源节点到目的节点的路由也会不断增加，也意味着工作在三层的网络节点设备需要维护和处理的路由表急剧上升。

(3)网络用户的急速增长，带来了整个网络的安全问题，也使得运营商管理和控制用户的行为变得更加困难，这意味着网络节点设备赖以实现上述功能的访问控制列表容量需要有更大的提升。

在网络中已经大量应用的电信级以太网系统同样面临网络飞速发展带来的挑战，也会面临二层地址表、三层路由表以及访问控制列表等硬件资源不能适应网络不断增长的需求难题。

另一方面，随着网络的不断增长，各大运营商和专网也对网络系统和设备的硬件资源数量给予高度重视。某运营商(由于涉及保密因素，这里隐去运营商名称，下同)已在2008年选型测试要求中明确指出：A类电信级以太网设备的二层转发表中条目数量不小于64k，单卡访问控制列表条目数量不小于1k。某行业网络在其选型测试要求中亦明确规定，核心交换机的访问控制列表数量不小于10k，核心交换机和电信级以太网设备的三层路由表容量不小于10k，这些技术指标相对于两年前有1倍或以上的增长。

作为网络中主要节点设备之一的电信以太网设备，面临这些硬件资源要求的冲击是不可避免的。换句话说，电信以太网系统的研究和设计需要妥善解决硬件资源需求不断增长带来的难题。电信级以太网设备的交换子系统的主要功能单元通常由交换芯片组成。与网络处理器相比，交换芯片因不能编程，导致可扩展性较差。在硬件结构不改变的情况下，电信级以太网设备提供的二层转发表、三层路由表、访问控制列表等重要硬件资源的容量就基本确定，但在实际应用中，经常会遇到这些资源不能满足应用需要的难题。当前的以太网产品，除了升级硬件，未采取措施来解决或弥补这个缺陷。升级硬件包括更换具有更高性能的硬件设备、升级当前网络节点设备的硬件配置等。这些方法当然可以解决当前网络设备的硬件资源不能满足应用需要的矛盾，但付出的代价也是昂贵的。更换硬件设备意味着现有网络设备要彻底淘汰，以前为购买、配置和开通这些设备的巨额投入就白白损失，同时也意味着巨大浪费。升级当前网络节点设备的硬件配置同样意味着大量的投入。

另一方面，由帕累托原则[3]可知，实际硬件转发表中的少数表项处于相对重要的地位，对网络流量的影响较大，而其它表项则处于重要程度相对较低的地位，对网络流量的贡献较小。基于这个事实，利用某种方法来管理和维护硬件资源，使其充分利用，从而改善系统转发性能是可行的。

基于以上信息和分析，为提高电信级以太网系统[4]的转发性能，本文提出了一种虚拟硬件[5-7]资源方法，以弥补硬件资源的不足和改善系统的性能。此方法的基本思路是：借鉴虚拟存储器[8,9]的相关技术和方法，以电信级以太网系统固有的二层转发表、三层路由表以及访问控制列表[10]等硬件资源为“主存”，在系统的内存中开辟相应的空间，分别作为二层转发表、三层路由表以及访问控制列表的“辅存”，借助于软件方法实现对这些“主存”和“辅存”的统一管理，并利用算法使硬件表总是存储对系统性能影响较大的表项。这样，逻辑上等价于电信级以太网系统的二层转发表、三层路由表以及访问控制列表等硬件资源等价扩充到“辅存”的大小，同时系统的转发性能也得到很大的改善。

2 评价虚拟硬件资源方法的性能

本文提出的虚拟硬件方法，其主要目标是逻辑上扩充电信级以太网系统的二层转发表、三层路由表和访问控制列表等硬件资源，主要用于电信级以太网系统的二层转发表、三层路由表和访问控制列表等硬件资源不足以满足要求时。

首先看看当这些硬件资源已经用尽的情况下，电信级以太网系统的转发行为将发生哪些变化。

对于二层转发表来说，如果表项已经用完，那么进入设备的数据包若是按照二层转发且具有一个新的源地址A，则A将不能够写入到硬二层转发表中，那么进入系统的目的地址为A的所有数据包将查不到目的端口。对于当前的系统，这些数据包一般只有两种操作：丢弃或在对应的虚拟局域网(VLAN，Virtual Local Area Network)中洪泛。丢弃意味着信息的丢失，而洪泛意味着信息可能到达一个或多个不期望的目的端口，既浪费了目的端口的带宽，也降低了信息的安全性，所以这两种操作带来的结果都不理想。

如果三层路由表已满，那么系统新学习的路由信息不能够进入系统的硬件三层路由表中。进入设备的数据包如果走三层转发流程，且在硬三层路由表中不能成功查找到目的路由，这些数据包一般也只有两种操作：丢弃或按照缺省路由转发。丢弃的后果就不用赘述了，而按照缺省路由也意味着信息常常到达不期望的目的地，在很大程度上也相当于数据包丢弃。

如果访问控制列表已满，那么系统需要增加新的利用访问控制列表实现的功能就不可能，这时候的系统就丧失了部分功能扩展性。尤其在需要用访问控制列表实现基于流的安全功能、QoS、灵活QinQ[11]以及快速重路由等功能时，在很多应用场合，访问控制列表的不足通常使得这些功能难以达到预期效果。

从上面的描述可知，评价虚拟二层转发表(这里把本方法针对二层转发表部分的功能简称为虚拟二层转发表)机制的性能，主要是参照按二层转发的单播数据包的转发行为，可以使用目的地址查不到的比率 P_{l2dlf} 来表示。设在一定时期内，进入系统的总单播数据包个数为 S_2，能成功查找到目的地址而正常转发的数据包个数为 S_{2f}，查不到目的地址丢弃或洪

泛的数据包个数为 S_{2l}，则：

$$\begin{cases} S_2 = S_{2f} + S_{2l} \\ P_{l2dlf} = S_{2l}/S * 100\% \end{cases} \tag{1}$$

还可以采用最优比 R_{l2opt} 来评价性能的好坏。在一定时期内，对于给定的网络数据流集和给定的转发表容量，如果在理想化的情况下，S_l 的值记为 S_{l2opt}，也就是说 S_{l2opt} 是在给定硬件配置和网络流量情况下 S_l 的最小值，则最优比 R_{l2opt} 可以用式(2)得到：

$$R_{l2opt} = S_{2l}/S_{l2opt} \tag{2}$$

例如，一个只有 4 个转发表项的系统，16 个数据包(这里以目的地址-源地址标记一个数据包，如数据包的目的地址是 A，源地址是 B，则该数据包表示为 A-B)进入系统的顺序分别为 A1-A2，A3-A4，A2-A1，A4-A3，A5-A3，A3-A5，A5-A2，A2-A5，A5-A4，A4-A5，A5-A3，A5-A3，A5-A2，A3-A5，A5-A2，A2-A5，则按照现有网络设备的转发行为，可得到：

$$\begin{cases} S = 16 \\ S_l = 9 \\ S_f = 7 \\ S_{l2opt} = 2 \\ P_{l2dlf} = 56.25\% \\ R_{l2opt} = 4.5 \end{cases} \tag{3}$$

显然，针对本例所述的情况，普通方法的转发效能非常低。

对于三层路由表，同样可以使用类似二层转发表的评价方法。以 P_{l3dlf} 来表示相应路由查不到的比率，以 R_{l3opt} 表示三层最优比。设在一定时期内，进入系统的总三层数据包个数为 S_3，能成功查找到相应路由而正常转发的数据包个数为 S_{3f}，查不到路由丢弃或走缺省路由的数据包个数为 S_{3l}，S_{l3opt} 表示 S_{3l} 的最小值，则：

$$\begin{cases} S_3 = S_{3f} + S_{3l} \\ P_{l3dlf} = S_{3l}/S * 100\% \\ R_{l3opt} = S_{3l}/S_{l3opt} \end{cases} \tag{4}$$

对于虚拟访问控制列表而言，情形要复杂得多。因为，对于二层转发表来说，所有经过二层转发的数据包都会查表；同样对于三层路由表而言，所有经过三层转发的数据包也必须查表，而且查找失败会影响转发行为。但访问控制列表不同，首先，仅仅从配置了访问控制列表项且使访问控制功能的端口进入到设备的数据包才需要查找访问控制列表，而且查找失败对转发行为不产生影响。基于访问控制列表的这些特性，上述两种方法用来评价虚拟访问控制列表都不太合适。这里定义期望匹配率 R_{em} 来衡量虚拟访问控制列表的性能。令在一定时期内，期望被访问控制列表匹配的数据包个数为 S_e，而实际匹配的数据包个数为 S_r，则：

$$R_{em} = S_r/S_e * 100\% \tag{5}$$

3 本方法描述

3.1 本方法概述

本方法的基本思路是：借鉴虚拟存储器的相关技术和方法，以电信级以太网系统固有的二层转发表、三层路由表以及访问控制列表等硬件资源为“主存”，在系统的内存中开辟相应的空间分别作为二层转发表、三层路由表以及访问控制列表的“辅存”，借助于软件方法实现对这些“主存”和“辅存”的统一管理，逻辑上等价于电信级以太网系统的二层转发表、三层路由表以及访问控制列表等硬件资源等价扩充到“辅存”的大小。

硬件上的二层转发表、三层路由表以及访问控制列表分别用 Hii，$Hiii$ 和 $Hacl$ 来表示，它们的条目分别用($Hii[1]$，$Hii[2]$，…，$Hii[m1]$)，($Hiii[1]$，$Hi[i[2]$，…，$Hiii[m2]$)和($Hacl[1]$，$Hacl[2]$，…，$Hacl[m3]$)表示；软件上的二层转发表、三层路由表以及访问控制列表分别用 Sii，$Siii$ 和 $Sacl$ 来表示，它们的条目分别用($Sii[1]$，$Sii[2]$，…，$Sii[n1]$)，($Siii[1]$，$Siii[2]$，…，$Siii[n2]$)和($Sacl[1]$，$Sacl[2]$，…，$Sacl[n3]$)来表示；$m1$，$m2$ 和 $m3$ 分别表示硬二层表、三层表和访问控制列表的表项数目，$n1$，$n2$ 和 $n3$ 分别表示软二层表、三层表和访问控制列表的表项数目。

本方法的功能结构如图 1 所示。其中，左边 3 张表分别表示硬件中的二层转发表、三层路由表和访问控制列表，是本方法中涉及的“主存”；而中间的存储块分别对应软二层转发表、三层路由表和访问控制列表，是本方法中涉及的“辅存”；最右边边界虚线的方块表示的是来自电信级以太网系统外部的信息，也就是网络运行中可能存在的二层转发信息、三层路由信息和访问控制信息等。

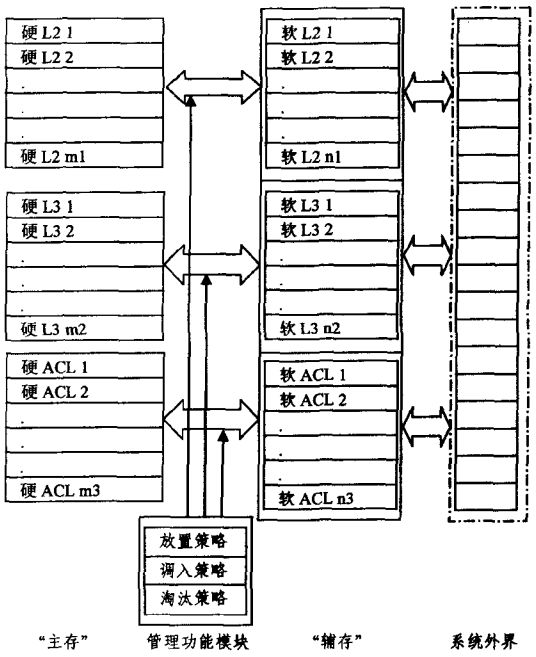

图 1 本方法功能结构示意图

系统的放置策略、调入策略和淘汰策略不仅涉及“主存”信息的管理以及“主存”和“辅存”之间的信息交换，同样包括“辅存”信息的管理以及“辅存”和系统外界之间的信息交换，两者采样的方法基本类似。但本方法的重点是“主存”和“辅存”之间的信息交换。

3.2 虚拟硬件资源方法的管理模块采取的策略

3.2.1 放置策略

放置策略决定信息放置的位置。由于硬二层转发表和硬三层路由表都采取哈希(Hash)算法决定条目在硬件空间的位置，那么在“辅存”也就是软表中决定信息条目放置的位置也需要采用同样的哈希算法。为了信息放置的位置不与硬件的行为发生冲突，也利于信息放置和管理，软二层转发表和软三层路由表条目数分别设为硬二层转发表和硬三层路由表条

目数的 2^l(l 为整数)倍,且采取分页方式,每页的条目数与对应的硬表的条目数相等。而对于硬访问控制列表来说,转发引擎查询该表多采用 TCAM 算法,但管理和维护该表一般采用顺序方式,因此对于软访问控制列表的放置策略也采用按顺序存储。

3.2.2 调入策略

调入策略是指决定信息从辅存调入到主存的时机。主要的调入策略有请调方式和预调方式两种。请调方式是在需要使用时调入,而预调方式则是在需要使用之前就调入。本方法涉及的信息并不是程序执行需要的代码和数据,而是转发引擎需要使用的转发信息表表项,因此调入策略既不同于请调策略,也不同于预调策略。概括起来,本方法的调入策略是时间驱动调入。也就是说,在一定时间内,如果系统发现硬表中存在某条目重要性小于不在硬表中的某条目,就执行调入操作。

3.2.3 淘汰策略

在虚拟存储器技术中,淘汰策略使用得比较多的是近似 LRU[5] 算法。主要原因是最优算法不具有实用性,先进先出算法效能比较低,而 LFU[5] 算法实现代价相对较高。但在本虚拟硬件方法中,目标是提高硬件资源使用的效率,因此应尽可能将最该被淘汰的信息移出硬件表,所以选用 LFU 算法。本方法使用的 LFU 算法在下一节描述。

3.3 本方法的淘汰策略——LFU 算法

实现 LFU(Less Frequently Used)算法最关键的方面是如何确定信息的访问频度,也就是如何确定该信息最近一段时间访问的次数。这包括在如下的 3 个方面:首先,选择多长时间作为"最近",就是个棘手的问题。第二,如何紧凑地记录访问次数也需要仔细斟酌。第三,如何用最小的代价获得访问次数信息同样需要合理设计。下面对这 3 个方面分别进行说明。

(1)获取信息条目的访问次数:交换芯片通常具有将设备收到的数据包可控地送往处理器的机制,比如 ACL 或 PCL 引擎等。中、高端交换芯片还有 SFLOW[12] 机制,这些机制能保证处理器能监测网络上的流量,而且送往处理器的数据包比例(采样率,以 Sr 表示)可以设置得比较小,这样接收数据包占用处理器的开销就比较小。处理器基于收到的数据包结合系统的各种软表可以得到各表项的访问次数。

(2)保存信息条目的访问次数:由于采样率 Sr 通常比较低,因而在一定时期内,进入到处理器的关联某个条目的数据包通常较少。而且,淘汰策略只要做到能够尽可能地选择最不经常使用的表项即可。因此对应一个表项,可采用一个字节的计数器来表示它的访问次数,或者说存储访问次数等级信息。该计数器在算法中用 Visit Counter 表示。

(3)确定"最近"所代表的时间:Sr 通常设置得较低,在这种情况下,各条目的计数器变化相对较慢。因此可以选择从计数器清零开始,到存在某个条目的计数器值超过它的计数容量的一半时的时间,作为一个处理周期。显然这个周期和网络流量相关,并且这个周期不是常量。

确定了信息条目的访问频度存储机制后,还需标记每个条目的状态位,该状态位用以表示该条目是否在硬件表中。

当存在某个条目的计数器值超过它的计数容量的一半时,系统需要进行相应处理。系统首先需要确定硬表是否已用尽。如果硬表未处于满状态,则不实施淘汰操作;如果硬件表已满,且软表的条目大于硬表,则检查软表中各条目的访问频度等级计数器。如果存在一个不在硬表中的条目 A,其访问频级值大于硬表中某一对应条目 B 的访问频级值,那么将硬表中条目 B 移出硬表,而把条目 A 写入硬表。

需要强调的是,二层转发表和三层路由表的表项通常采取 Hash 算法得到表项位置的索引值,这样条目 B 必须满足两个约束:和条目 A 不在同一页面,在页面内的序号和条目 A 相同。

3.4 本方法的工作过程

归纳起来,本方法主要有 3 部分操作:初始化过程、表项学习或建立过程、表项管理和置换过程。具体实现用流程图表示,如图 2 所示,其中子图(a)描述初始化过程,子图(b)描述表项学习或建立过程,子图(c)描述表项管理和置换过程。

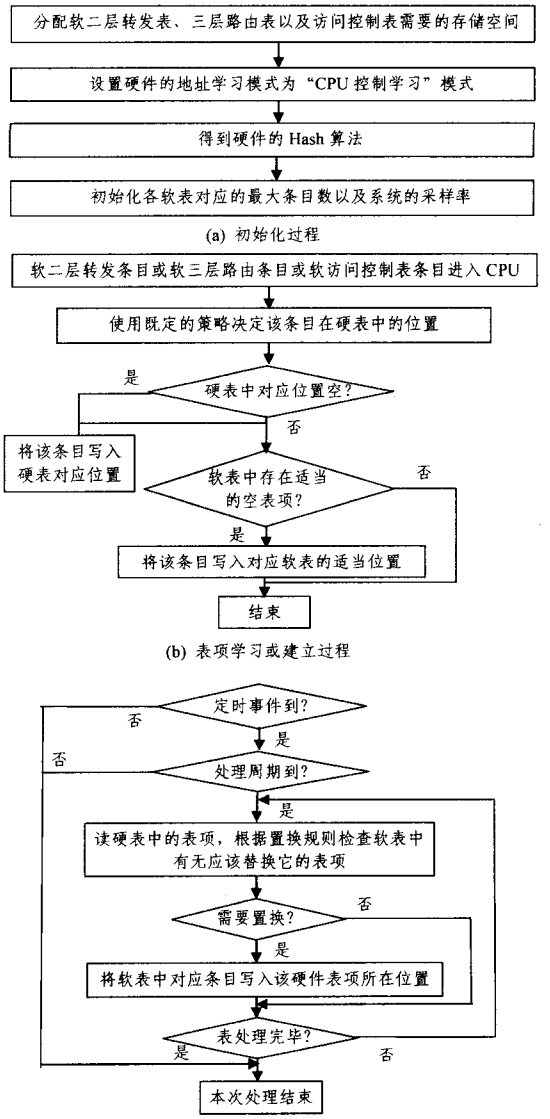

图 2 算法完整过程描述

4 实现和测试

本方法在我们自主开发的 M8416E 电信级以太网设备(以下简称 M8416E 设备)上实现。M8416E 设备采用 10U 标准机箱,宽度×深度×高度=436mm×260mm×440mm。M8416E 设备背板带宽 256G,单一交换控制盘交换容量为 128G,转发性能为 144Mpps。M8416E 设备具有两个交换控制槽位、4 个业务槽位,从下往上槽位号依次为 1,2…6,1,2,3

和 4 四个槽位为业务槽位，5 和 6 两个槽位为主控槽位，两块主控卡同时在位时实行 1:1 保护。该设备最多可支持 96 个百兆端口、72 个千兆接口以及 4 个 10GE 接口。

M8416E 设备设计的硬件资源数量分别为二层转发表容量 32k、三层路由表容量 12k 以及访问控制列表容量 4k。主控卡上使用的交换芯片支持 SFLOW 功能，数据包进入 CPU 的采样率可在$(6.0*10^{-8}, 3.9*10^{-3})$区间内选取。

需要特别说明的是，由于 M8416E 设备具有两块主控卡，采取 1:1 主备保护模式，而备用主控卡的处理器大部分时间基本处于空闲状态，因此本方法的大部分统计、查找、排序、比较等运算在备用主控卡上执行，以节省主用主控卡的 CPU 处理时间。而相应的软二层转发表、三层路由表和访问控制列表在主用主控卡和备用主控卡上各存储完全一样的一套，软表和硬表之间的数据交换则在主用主控卡和备用主控卡上同时进行，以便系统在主用主控卡和备用主控卡之间倒换时仍然能正常工作。在包含本方法的系统设计完成后，需搭建相应的测试环境对本方法进行测试验证。采取的测试网络拓扑如图 3 所示。

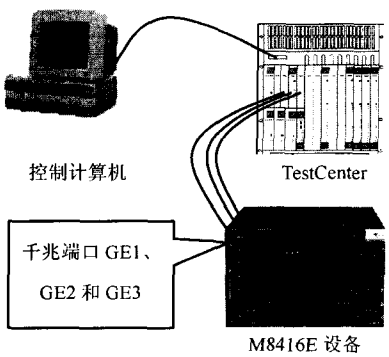

图 3　测试网络拓扑

图 3 中，测试仪表中 1 个百兆快速以太网口（管理端口）与控制计算机相连接。同时测试仪表中有 3 个千兆以太网接口分别与 M8416E 电信级以太网设备的 24GE 业务卡中的 3 个端口 GE1，GE2 和 GE3 相联，其中 GE3 用作测试流输入端口，GE2 用作测试流输出端口，GE1 用作不期望数据流接收端口。当测试虚拟二层转发表功能时，GE1，GE2 和 GE3 三个端口同属于所有已创建的 Vlan，GE1 用于监测查不到目的端口的洪泛数据包的数量。测试虚拟三层路由表功能时，GE1，GE2 和 GE3 三个端口不同属于任何已创建的 Vlan，缺省路由的目的端口设置为 GE1，则 GE1 用于监测查不到路由而走缺省路由转发的数据包的数量。测试虚拟访问控制列表功能时，可不使用 GE1。

由控制计算机利用自编的伪随机数据流发生器生成测试的数据流，然后配置到测试仪表上，由测试仪表发数据流到 M8416E 电信级以太网设备上，观察和统计仪表上收到的数据情况，然后再利用前述的评价方法分别计算出 P_{l2dlf}，P_{l3dlf} 以及 R_{en}。测试虚拟二层地址表时，共进行 6 次测试，这 6 次测试用到的测试流包含的二层转发地址的个数分别为 32k，40k，48k，56k，64k 和 96k，每次测试的数据流包含的二层转发地址是连续的。测试虚拟三层路由表时，也进行 6 次测试，这 6 次测试用到的测试流包含的三层路由的个数分别为 12k，16k，20k，24k，28k 和 32k。测试虚拟访问控制列表时，同样进行 6 次测试，这 6 次测试用到的测试流包含的访问控制列表数目分别为 4k，5k，6k，7k，8k 和 10k。为便于比较分析，每次测试得到两个结果，分别是没有虚拟机制时的结果和使用虚拟方法的结果。将测试结果绘制成统计图，得到图 4，其中子图(a)表示虚拟二层转发表的测试结果，子图(b)表示虚拟三层路由表的测试结果，子图(c)表示虚拟访问控制列表的测试结果。测试结果表明，采用该虚拟硬件方法，系统转发性能显著的提高。

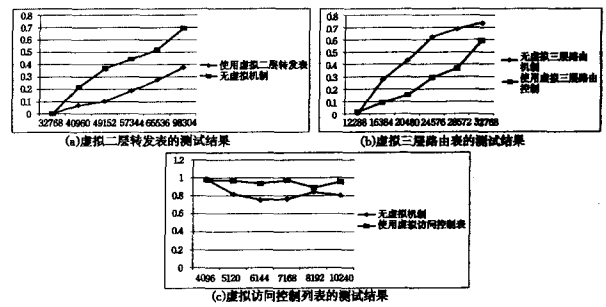

图 4　3 个硬件设备的测试结果

结束语　电信级以太网系统在网上应用时，面临着网络对硬件资源的需求不断增长带来的挑战。虽然通过更换或升级硬件可以解决硬件资源不足的问题，但采用此方案需要巨大的硬件投入。同时，放弃现有设备会带来资源的巨大浪费，但是通过采用虚拟硬件机制缓解硬件资源不足的问题，使现网设备可以在硬件资源要求高的系统设计环境下使用，这在应用层面是有重要意义的。设计相应的虚拟硬件资源算法并在电信级以太网设备上加以实现、评测和确认。从实验结果来看，采用此方法对系统的转发性能有明显的改善。

参 考 文 献

[1] 艾瑞咨询. 未来几年美国社会化网络用户增长稳定[EB/OL]. http://news.iresearch.cn/charts/75302.shtml

[2] 艾瑞咨询. 2001-2010 年中国互联网用户数及增长率[EB/OL]. http://www.iresearch.com.cn/consulting/online_media/Graph.asp?id=9449

[3] Chen Nan. Carrier Ethernet The Next Ethernet Wave[OL]. http://forum2005.cnw.com.cn/2004/down/Carrier_ethernet.pdf

[4] Gittens M, Kim Y, Godwin D. The vital few versus the trivial many: examining the Pareto principle for software[C]// Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC'05). 2005

[5] 庞丽萍. 操作系统原理(第三版)[M]. 华中科技大学出版社

[6] Nanda S, Chiueh T. A survey on virtualization technologies [D]. tr-179. Department of Computer Science, State University of New York, February 2005

[7] 刘谋用，葛霏光. 操作系统研制中虚拟硬件的实现[J]. 浙江大学学报，2001(4)：351-355

[8] 李向华，张小芳，陆起涌. 基于虚拟存储的嵌入式存储系统的设计方法[J]. 半导体技术，2003，28(8)：30-32

[9] Lakshmi N, Andrea C, Remzi H. Dependability Analysis of Virtual Memory Systems[C]// Proceedings of the 2006 International Conference on Dependable Systems and Networks (DSN'06). 2006：355-364

[10] Gao Jun, Steenkiste P. An access control architecture for programmable routers[C]// Open Architectures and Network Programming Proceedings. April 2001：15-24

[11] IEEE Computer Society. Virtual Bridged Local Area Networks [S]. IEEE Std 802.1q. 1998

[12] Wang M, Li B, Li Z. sFlow: towards resource-efficient and agile service federation in service overlay networks[C]// Proceedings Distributed Computing Systems. 2004：628-635