基于多令牌桶的组播拥塞控制

高永辉，蒋 林
（西安邮电学院 电子工程学院，陕西 西安 710061）

摘 要：IP组播是一种有效的数据传输方式，在过去几年中，组播传输机制已经成为一个活跃的研究领域。但由于其自身特性决定了在组播中实现可靠性和拥塞控制非常困难，组播的拥塞控制问题一直没能得到很好的解决，这成为了其发展的瓶颈，不断增加的UDP数据流恶化了TCP控制拥塞的能力，而且是引起高丢包率的原因之一。文中提出了一种可以对局域网内IP分配固定带宽的扩展令牌桶算法，扩展了令牌桶个数，一个令牌桶控制一个IP，消除了共享带宽的缺点，并通过设计电路，建模仿真结果表明可以对路由器交换节点的组播达到准入控制，防止造成网络拥塞，最大限度保证网络Qos，并且设计的电路占用硬件资源少，能够应用于高速电路当中。

关键词：令牌桶；组播；拥塞控制；IP网络；超大规模集成电路
中图分类号：TP302；TP393 **文献标识码**：A **文章编号**：1673-629X(2012)02-0149-04

Multicast Congestion Control Based on Multi Token Barrels

GAO Yong-hui, JIANG Lin
(School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710061, China)

Abstract: IP multicast is an effective data transmission, over the past few years, the multicast transmission mechanism has become an active area of research. However, it is very difficult to realise reliability and to achieve in multicast congestion because of its characteristics, multicast congestion control issues have not been able to get a good solution, it has become a bottleneck in its development. The increasing deterioration of the UDP data stream aggravates the ability of TCP congestion control, and it is one of the reasons to cause high packet loss rate. It presents a way to assign a fixed IP on the LAN bandwidth expansion token bucket algorithm, extends the number of token bucket, a token bucket control of an IP, eliminating the shortcomings of shared bandwidth, and through the design of circuits, modeling and simulation results show it can control the access of multicast and ensure the Qos of network and the design of the circuit occupies less hardware resources, which can be used in high speed circuits.

Key words: token barrel; multicast; congestion control; IP network; VLSI

0 引 言

随着视屏点播、远程教育、电话会议、网络交互式游戏等实时性业务的兴起，组播(Multicast)[1,2]技术得到广泛应用，IP组播是利用一种协议将IP数据包从一个源传送到多个目的地，将信息的拷贝发送到一组地址，到达所有想要接收它的接收者处。这些实时业务一般采用UDP[3,4]协议进行传输。然而由于UDP协议和IP组播都不提供拥塞控制机制(Multicast Congestion Control Mechanisms)[5,6]，导致这些业务和TCP业务共存时出现了带宽占用不公平[7]。在路由器数据包分组转发中，组播会长时间占据端口，导致网络拥塞，严重时瘫痪网络。因此，必须对IP组播进行准入控制，预防组播风暴发生。令牌桶[8]控制机制以其简单、高效性而在网络系统中大量应用[3]。文中根据令牌桶流量控制机制，在路由器入口中，专门针对多播设计一种多令牌桶准入控制机制，通过软件和FPGA功能仿真验证，证实此方法不仅占用资源少，容易硬件实现，而且能有效防止多播"风暴"的发生。

收稿日期：2011-07-20；**修回日期**：2011-10-25
基金项目：陕西省"13115"科技创新工程重大科技专项(2009ZDKG-43)；陕西省教育科研计划项目(2010JK840)
作者简介：高永辉(1985-)，男，硕士研究生，研究方向为集成电路设计；蒋 林，教授，博士，硕士生导师，研究方向为专用集成电路设计。

1 令牌桶算法概述

1.1 令牌桶结构

令牌桶是网络设备的内部存储池，而令牌则是以给定速率填充令牌桶的虚拟信息包。每个到达的数据包分组块都会从令牌桶中取走一定数量的令牌。根据预先设定的匹配规则对报文进行分类，不符合匹配规则的数据包不需要经过令牌桶的处理，直接发送；符合匹配规则的数据包，则需要令牌桶进行处理。当桶中有足够的令牌则报文可以被继续发送下去，同时令牌

桶中的令牌量按数据包的长度作相应的减少；当令牌桶中的令牌不足时，该数据包不能被发送，该数据包将被丢弃（或标记），只有等到桶中生成了新的令牌，后面到来的数据包才可以接着发送。这就可以限制数据包的流量只能是小于等于令牌生成的速度，达到限制流量，预防网络拥塞的目的。令牌桶结构示意图如图 1 所示。

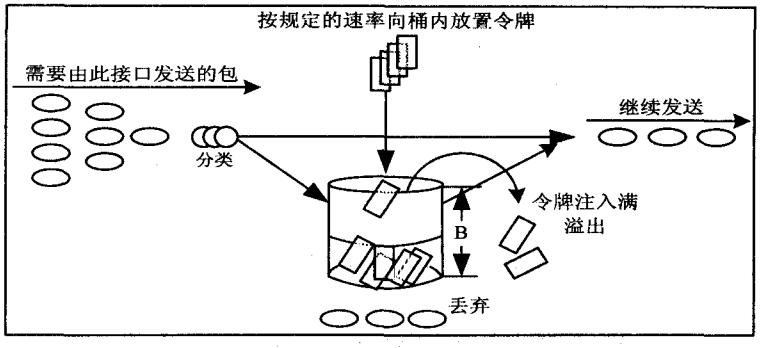

图 1 令牌桶算法结构示意图

1.2 令牌添加

令牌的添加有多种实现方式：

（1）周期性添加，一种是令牌按照恒定的速率（CIR：Committed Information Rate）[8,9]，即每隔 1/CIR 时间添加一次令牌，添加的时间间隔就是令牌桶的容量与添加速率的比值[10]。还有一种是采用两个添加周期 T_1、T_2，且 $T_1 > T_2$，对该种报文在路由器缓存区设置门限 K，当发现缓存区统计值 $q > K$ 时，添加周期切换至 T_2，否则保持按较大的周期 T_2 添加，此方案以占用更多的带宽资源为代价来降低报文丢失率的，该方法需要实时统计每种预先设定的报文在缓存区的个数，同时门限与时钟周期不规则切换很大程度上影响复用器性能，更重要占用带宽资源。

（2）一次性添加，只有当令牌桶中没有令牌才添加，且每次添加的令牌数等于 $t \times$CIR（t 为当前添加时刻与上一次添加时刻的时间差），一次性添加完毕，并不是周期性添加[11]。此方法的令牌桶中当前令牌数很大程度上与 t 值有关，对突发流量处理稍显不足，会短时间内造成大量丢包。

1.3 数据包处理及令牌消耗流程

方式一：根据 IP 数据包分组转发特性，在路由器入口，若包头描述信息包含该数据包总长度或者没有告诉总长度只告诉包头数据块长度，则用包长或包头数据块长度与当前令牌桶中令牌作比较，桶中令牌大于或等于要消耗的令牌数，则允许该数据包所有分组块进入路由器缓存 fifo，否则，丢弃当前该数据包的所有分组块。这种处理方式在突发流量或是一个包的分组块很多或是输入数据流速率高时会造成大量丢包，不宜采用。

方式二：当 IP 数据包到来时观察映射到的令牌桶中令牌数，若非负，则准入该数据包所有分组块，该数据包所有分组块都依到来先后消耗桶中令牌，允许令牌数减为负；若检测发现令牌数为负数，则丢弃当前数据包的所有分组块，直到下一次更新将令牌桶加为正才允许数据包进入缓存 fifo，此令牌借贷方法能有效防止突发流量造成大量丢包。

2 基于多令牌桶组播控制电路设计与实现

2.1 接口信号定义

IP 分组块帧结构包括以下几个：

frame_sport ：（6bit）输入数据块的源端口号。

frame_dport：（140bit）输入数据块的目的端口号。

fradata_state ：输入数据块的状态指示信号，01：包头；10：包尾；00：包中间；11：小于等于 128 字节的短包。

fradata_length：输入数据块的长度指示信号，0 代表 1 字节，1 代表 2 字节，全 1 则代表 128 字节，依次类推。

frame_type ：3'b000：正常数据包；3'b001：dlf_broadcast ；3'b010：unknown IPMC ；3'b011：known IPMC；3'b100：unknown L2MC；3'b101：known L2MC ；3'b110：broadcast 其他：按照 0 处理。

act_ind：（3bit）0 代表单播 ；1：二层组播；2 三层组播；3：二层+三层组播；4：1+1 保护组播。

forward_fradata：（1024bit）输入到转发模块的数据信号。

pulse_in：（6bit）采用独热码编写，将时钟每 6 拍划分为一个时钟周期，每 6 拍发送一个 IP 分组。

包交换网络节点具体实现“多播风暴”的方法是：在交换机的每个入口配置若干个令牌桶（此处设定为 4 个），智能分类器根据一定的优先级和分类原则将到达交换节点的多播数据包分成信号 forward_frame_type 标识的 7 种数据包类型，其中“风暴”控制只处理正常数据包以外的数据包。交换机每个源端口都有两个配置寄存器：reg1 和 reg2。reg1 用来指示到达此输入端口的 6 种“风暴”类型的数据包映射到 4 个令牌桶中的哪一个，例如数据块信息为，且非正常包，要进行“风暴”控制，源端口 0 中对 unknown IPMC 配置显示映射到令牌桶 2 中，查看令牌桶 2 的令牌数，若为负数，则丢弃当前数据包，否则允许其进入交换机缓存 fifo，同时源端口 0 的令牌桶 2 消耗相应的令牌数，而其他令牌桶令牌数保持不变。配寄存器 reg2 用来存储每

个源端口的四个令牌桶的配置,4 个令牌桶的大小均可配置。

每个源端口的令牌桶配置如果做成寄存器,但考虑到交换机输入端口较多,众多的配置寄存器势必会给以后配置参数修改和代码维护带来极大不便,一般大于 1K 的配置寄存器都应该做成存储器,交换机所有端口的配置都在两个存储 RAM 中,即映射对应关系的 mapping_ram,和桶大小参数配置 config_ram。这样,一个时钟节拍就只能读取一个端口的四个令牌桶配置参数,当进行令牌注入时,可以采用端口依次串行更新的方法。即每 6 拍更新一个端口的四个令牌桶。

2.2　电路总体实现结构图

电路总体实现结构图见图 2。

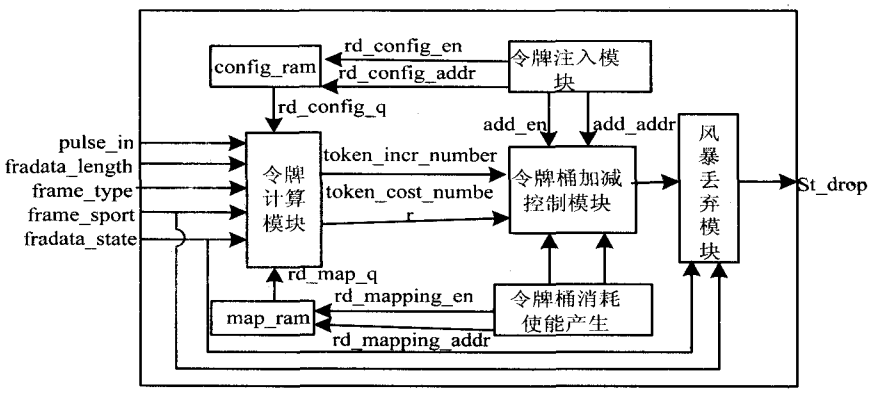

图 2　播"风暴"控制总体电路框图

网络中总线上来的 IP 数据包分成若干个片,一个完整的数据包由包头,若干块包中,一块包尾构成,也可以由一块包头和一块包尾构成,当数据包只有一块时就是短包。总线上到来的数据包可能源端口号是乱序的,例如,端口 0 发送来一个包头,接着端口 1 又发送来一个包头,接着端口 0 的包中才来,总线上的数据包的 IP 分组是乱序的,但是到达每一个源端口的数据包却是严格按照包头—包中—包尾的顺序到达。风暴控制只在每个源端口包头或者短包时判断数据包是否丢弃,因此必须把每个源端口数据包在包头或者短包块的丢弃状态寄存起来,包头判断丢弃,这个包的后续到来的包中、包尾均判断丢弃(因为此时数据包已经不完整了),直到此端口下一次又有新的包头或者短包到来时,将新判断的结果用来更新端口丢弃状态寄存器的值。

2.3　核心电路设计

(1)令牌注入(更新)模块:令牌注入采用上述 1.2 节介绍的令牌添加方法的第一种周期性添加。令牌添加间隔时间(填充速率)计算公式为:

$$T(us)=\frac{T_0\times n_0+T_1\times n_1}{1024\cdot(n_0+n_1)\cdot f}$$

T_0、T_1 表示两个添加周期,且 $T_1 > T_0$。n_0 表示以 T_0 为周期添加令牌连续进行的次数,n_1 表示以 T_1 为周期添加令牌连续进行的次。f 单位(Mhz)为实际中电路时钟频率。同时参数 T_0,T_1,n_0,n_1 的设定与 f 密切相关,f 越大意味着电路工作速度越快,单位时间内数据包吞吐量越大。为防止大量丢包,f 越大,T_0,T_1,n_0,n_1 相对较小,更新速率加快。图 3 是令牌注入模块实现结构图,其中 N 为交换机输入端口个数,每次更新时刻到来时,电路串行对交换机 N 个源端口的令牌桶注入令牌,直到 N 个端口的令牌桶更新完毕,才将加使能 Inc_en 置低。

(2)令牌消耗使能和令牌桶加减控制模块实现框图。

图 4 是令牌桶加减控制电路实现框图,框图的加、减控制严格按照六拍一个时钟周期的时序设计要求。在时钟周期的第四拍(p3)观察读出的令牌桶中令牌数的最高位,这一位是令牌桶正负标志位,将其寄存到该端口的丢弃状态寄存器中,在下一拍再从此端口中读出寄存的丢弃结果作为输出信号 sc_drop。亦即丢弃信号固定在最后一拍得出。

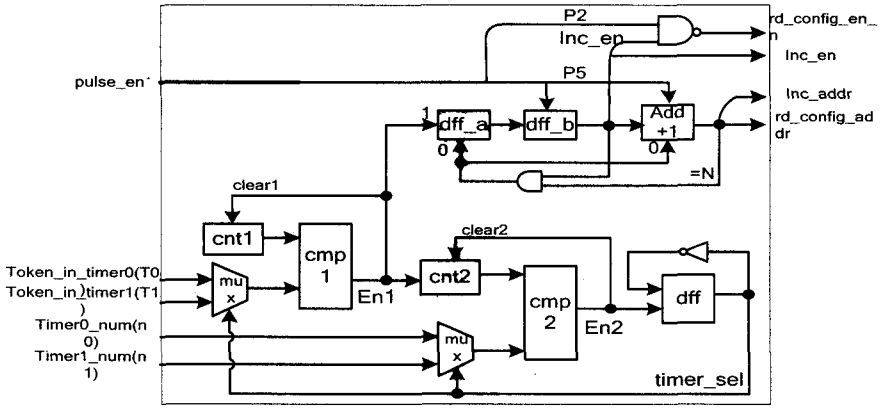

图 3　令牌注入模块实现结构图

在图 4 会发现对数据块字节长度 fradate_length 乘 16(满足 2^n 型,左移 4 位,硬件实现速度很快[12]),这里表示一个令牌相当于 1/16 个字节。具体可以通过公式,以及总线带宽计算出。

图 5 的 adder 模块包含加、减两种操作,当令牌桶加溢出时要用令牌桶最大值回写到令牌数 ram 中,这里说明一点:令牌桶数值存在加溢出和减溢出(均为

高有效),一位无法区分是哪种形式的溢出,这里,将每个令牌桶计数值扩大两位,最高位为减溢出标志位,次高位为加溢出标志位,解决了加、减溢出的区分。其详细电路图如下:

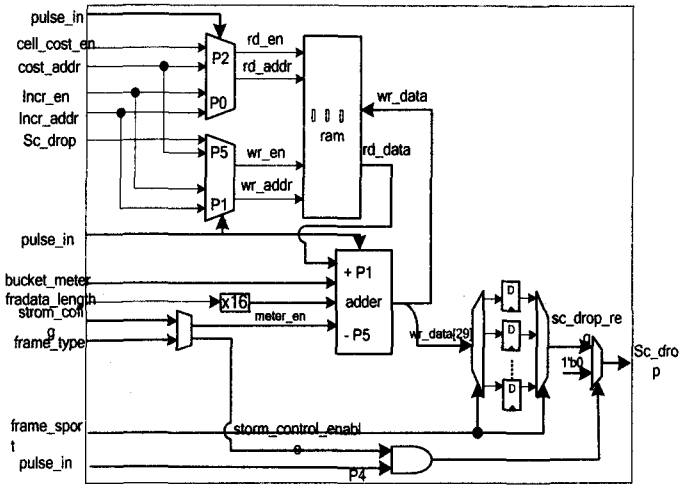

图 4 令牌桶加减控制电路结构图

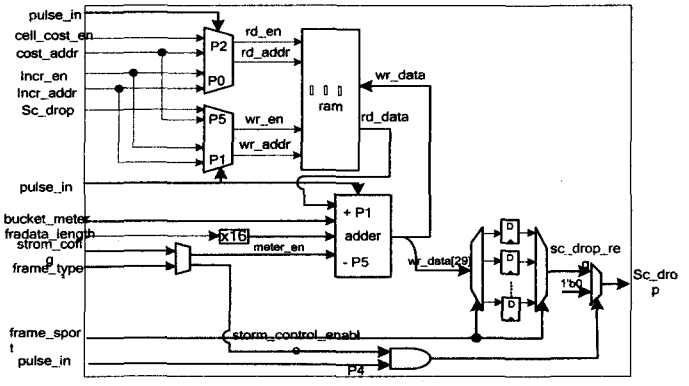

图 5 adder 模块功能图

3 硬件实现仿真结果

对于上述设计电路,采用 verilog 语言描述,并在 synopsys VCS 上进行功能仿真和时序仿真(见图 6),时钟频率为300Mhz,满足时序要求,带宽64kbps时,电路每隔 7.8125us 更新一次,令牌桶减为负数后有可能一次更新仍不能将令牌桶加成正数,需要进行多次更新才可以再次将领令牌加为正数。同时,在组播突发流量时丢弃一定量的数据包来防止组播拥塞,瘫痪整个网络。

波形中当前多播类型映射到源端口 3 的令牌桶 2 中,而令牌桶 2 的令牌高位的符号标志位为 1,表示令牌桶 2 中令牌为负,故令牌桶 2 丢弃当前数据块。Quartus II 综合结果表明电路可以稳定工作在 300Mhz,且占较少资源。

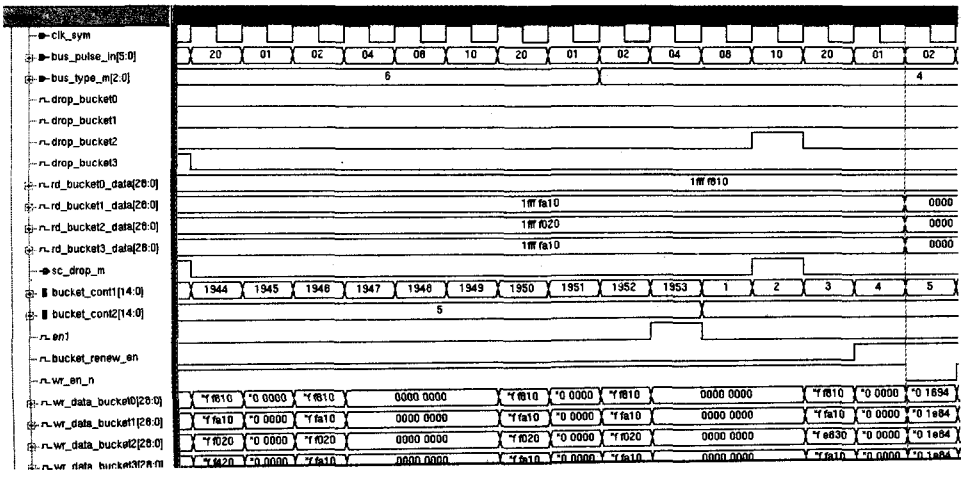

图 6 令牌桶控制多播仿真波形图

4 结束语

文中基于令牌桶算法,在包交换芯片的每个入口采用多个令牌桶对多播进行准入控制,并设计电路,通过软硬件仿真,结果表明该令牌桶算法性能很高,能有效防止"组播"风暴的发生,而且算法本身简单,占用硬件资源少,适合高速集成。文中对"多播"风暴的处理方法可推广到 IP 网络包交换的芯片的设计。

参考文献:

[1] Wang Shuda, Pan Qinghe, Niu Jilai. Congestion control on wireless network technology in MiroSot[J]. Journal of Harbin Institute of Technology, 2005, 37(7):959-961.

[2] 韩礼国,才书训. 流媒体 Qos 端到端自适应控制策略综述[J]. 计算机技术与发展,2006,16(11):246-249.

[3] 何宝宏. IP 网络的服务质量讲座:第 4 讲 IP 网络流量与拥塞控制技术[J]. 中国数据通信,2003,5(5):96-99.

[4] 伊文斌,周贤娟,鄢化彪. uIP TCP/IP 协议分析及其在嵌入式系统中的应用[J]. 计算机技术与发展,2007,17(9):240-244.

[5] Lv Yunfei, Wang Xinggang. Streaming Media Congestion Control Based on Loss Differentiation Algorithm in Wireless Environment[J]. Computer Engineering, 2005, 31(13): 19-21.

[6] Xia Wei, Lin Yaping, Li Chao. Study of congestion control mechanism for wireless networks using expert control[J]. Journal of China Institute of Communication, 2004, 25(1):164-173.

(下转第 216 页)

```
rows += "<td>" + datas.Rows[i][j].ToString()
+ "</td>";
}
htmTable += (rows + "</tr>");
rows = "";
}
}
return getHtmTableHeader() + htmTable+"</table
>";//表头加表数据
}
```

③htmTableToExcel():将报表导出为 Excel 文件;

④getReportYear():获取报表的年月;

⑤isExist():根据 id 等参数,判断当前报表是否已入库;

⑥getReportTitle():根据 id 等参数,获取报表标题;

⑦getReportName():根据 id 等参数,获取当前报表对应的数据库表名。

(6)原油产量月报。

上报公司原油产量月报如表 1 所示:

表 1 上报公司原油产量月报

原油产量月报			
级别:上报公司			日期:2011 年 5 月
指标名称	计量单位	本月	累计
原油产量	吨	26500	126000
1. 新井产量	吨	1632	3303
其中:新井压裂增产量	吨	1632	3303
2. 旧井产量	吨	24868	122694
其中:旧井压裂增产量	吨	1056	1834
3. 试油产量	吨	0	0

5 结束语

基于 Web 的统计报表设计与实现的技术方法在定义各种报表类的基础上,实现了生成复杂的统计汇总数据并按特定格式输出的功能,具有以下几方面的显著特性:

①报表预处理类的设计,将报表复杂数据的处理与生成完全独立出来,减少了耦合度,提高了开发效率;

②报表类的设计,为报表业务提供通用的接口,具有很强的适应性和可维护性;

③实现了报表数据与报表格式的严格分离,充分体现了层次结构的理念;

④报表参数和报表字典的设置,增强了报表的灵活性和通用性,提高了系统性能和开发效率。

在《采油厂生产统计管理信息系统》项目中,采用本技术方法,方便、快捷、准确地实现了采油厂原油生产、钻井生产、主要生产能力、技术经济指标等四大类、三个级别、五十多张统计汇总报表的生成、保存、浏览及导出,实际应用效果良好。

参考文献:

[1] 张亚平,贺庄占. B/S 架构下动态报表的一种实现方式[J]. 计算机技术与发展,2007,17(4):93-95.

[2] Ullman C, Goode C. Beginning ASP . NET Using C#[M]. America:Wrox Press Ltd,2001:383-446.

[3] Joshi B, Dickinson P. Professional ADO. Net Programming [M]. America:Wrox Press Inc. ,2002.

[4] 胡立辉. 基于 B/S 的超级汇总报表处理系统的设计[J]. 计算机工程与设计,2005,26(7):1900-1902.

[5] 周晓光,张文波,苏志远. B/S 结构下动态汇总报表的设计与实现[J]. 北京工商大学学报,2005,23(4):36-39.

[6] 范金花,梁正和. 报表系统中 ETL 通用框架的设计与研究[J]. 计算机技术与发展,2009,19(6):202-205.

[7] 张海波,董槐林. 一种基于 POI 的 Web 表格生成[J]. 计算机技术与发展,2008,18(2):21-23.

[8] 高 鹏. 基于动态 DW+Formula 1 技术的集成式通用报表模型研究[J]. 计算机应用与软件,2009,26(11):137-140.

[9] 刘朝玮,李初福,何小荣,等. 石化企业生产计划图形建模优化系统的动态格式报表设计和实现[J]. 计算机与应用化学,2006,23(2):133-136.

[10] 熊 伟,郭继坤,张仁平,等. 实现"中国式"复杂表头的动态报表[J]. 后勤工程学院学报,2007,23(2):67-71.

[11] 马瑞敏,王成良. WEB 动态报表实现中的参数化过滤技术[J]. 计算机系统应用,2009(2):176-179.

[12] Butler J, Caudill T. ASP. NET Database Programming Weekend Crash Course[M]. America:Hungry Minds Inc. ,2002.

(上接第 152 页)

[7] 谢希仁. 计算机网络 [M]. 北京: 电子工业出版社, 2003: 248-274.

[8] Heinanen J, Guerin R. IETF RFC 2697:A single rate three color marker[R]. Philadelphia,PA,USA:University of Pennsylvania,1999.

[9] Clark D,Fang W. Explicit allocation of best effort packet delivery service[J]. IEEE/ACM Trans on Netw,1998,6(4): 362-373.

[10] 李晓利,郭宇春. Qos 中令牌桶算法实现方式比较[J]. 中兴通信技术,2007,13(3):56-59.

[11] 杨晓强,高 晔. NoC 系统设计的研究[J]. 微电子学与计算机,2008,25(7):176-179.

[12] 杜慧敏,李宥谋,赵全良. 基于 Verilog 的 FPGA 设计基础[M]. 西安:西安电子科技大学出版社,2005:140-145.