可重构集群路由器并行路由分发模型

陈文龙[1]，徐明伟[2]，徐恪[2]

(1. 首都师范大学 信息工程学院，北京 100048；2. 清华大学 计算机科学与技术系，北京 100084)

摘 要：传统一对多点的路由分发方式存在周期长、负载不均衡等问题。可重构集群路由器的体系结构中板卡数大量增加，上述问题更为突出，需要对路由分发算法改进。对现有路由分发方法及可重构路由体系进行分析，设计了树型并行路由分发模型。模型将可重构路由器所有板卡构造成一棵不平衡的分发树，路由从树根向叶子并行层层传递。研究了该模型板卡路由分发速度及负载均衡状况，并设计了模型实现算法及实施步骤。基于 NS2 的实验结果验证了 TPRD 模型的性能优势。

关键词：可重构路由器；路由；并行分发

中图分类号：TP393 **文献标识码**：A **文章编号**：1000-436X(2012)06-0118-07

Parallel route distributing model in reconfigurable cluster router

CHEN Wen-long[1], XU Ming-wei[2], XU Ke[2]

(1. Information Engineering College, Capital Normal University, Beijing 100048, China;

2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Long distribution delay and load unbalance of router cards were the key problems of traditional route distribution techniques. Especially, in reconfigurable cluster routers, the number of linecards significantly increased, which made the problem highlighted. To address this issue, the TPRD(tree-based parallel route distribution) model was proposed after systematically analyzing existing route distribution methods and the cluster router architecture. In TPRD, all cards of cluster router were constructed to an unbalanced distribution tree and routes were transferred form the root node to leaf nodes. The algorithms and deployment approaches for implementation of TPRD were presented. The NS2 simulation results demonstrated TPRD achieved an expected performance.

Key words: reconfigurable router; route; parallel distributing

1 引言

作为互联网最核心的网络设备，路由器经历了单处理器集中式总线结构、多处理器分布式共享总线结构、多处理器分布式交换结构 3 个发展阶段，可重构、可扩展体系结构是路由器发展的下一个阶段[1]。所谓可重构路由器，就是由若干个子路由器级连，重构而成的一个统一的路由器系统。它在功能、性能、接口规模等方面，具有极大的可重构性。它的主要优点包括：保护前期运营投资、提高系统性能、增强设备功能模块的可靠性、简化网络拓扑等。当今主要的路由器厂商已开发出自己的可重构路由器产品，包括：

收稿日期：2011-09-27；**修回日期**：2012-03-15

基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（2009CB320502）；国家高技术研究发展计划（“863”计划）基金资助项目（2009AA01A334）；“高可靠嵌入式系统技术”北京市工程研究中心基金资助项目

Foundation Items: The National Basic Research Program of China (973 Program)(2009CB320502); The National High Technology Research and Development Program of China (863 Program)(2009AA01A334); “High Reliable Embedded System Technology” Engineering Research Center of Beijing

Cisco 公司的 CRS-1[2]、JUNIPER 公司的路由矩阵(routing matrix)[3]、AVICI 公司 TSR[4]等。

分布式路由器体系中，一般通过一对多点的单播分发模式实现系统的路由分发。该模式存在分发周期长、负载不均衡等问题。路由器中大量路由分发，常与设备启动、接口 UP/DOWN、路由协议启动等事件同时发生。此时，一般都伴有高负荷的路由协议计算、路由决策以及路由模块或其他辅助模块的大量协议报文收发。路由分发和上述事件同时抢占处理器资源及板间通信带宽资源，会导致路由器系统性能瞬时降低。当路由器线卡数量较少及路由容量较小时，问题显得并不明显。然而，面对新的网络环境：路由器结构可重构、互联网路由表急剧增加[5,6]等，传统的路由表分发模式面临极大的挑战。

首先，随着路由器体系结构向可重构方向发展，若干子路由器级连。一个路由器系统包含的板卡大量增加，路由表分发时间将明显延长，发送者的负载会急剧增大。其次，随着互联网规模的不断扩大，以及 Multi-Homing 的实施，核心路由器的路由表容量不断增大。与此同时，互联网上对延迟和丢失敏感的新型应用（如 VoIP，流媒体，网络游戏，视频会议）正在大量部署，路由分发慢会影响用户的体验。而且，路由分发慢还会导致大量分组被错误转发，造成线卡处理和链路带宽等网络资源的浪费。这要求路由器在面对更多路由的情况下，更快地将核心路由表同步到各板卡。

从分布式体系结构开始，路由器设计者就开始研究系统的路由同步问题。路由表同步是指分布式路由器中，如何使所有板卡尽快与系统当前的核心路由表(KRT，kernel routing table)保持一致，从而保证报文转发和路由查询的正确性。关于路由器的路由同步方式，相关的研究和实现可以参考国防科技大学一种集群路由器转发表同步框架及关键算法[7]和 JUNIPER 公司的路由矩阵(routing matrix)[3]等。业界普遍采用主动广播更新，全冗余备份存储的路由同步机制。一般由主控卡集中计算出整个可重构路由器的核心路由表，进而通过主动广播更新方式同步所有板卡的路由表，每块板卡都存储核心路由表的完全镜像。

本文的贡献是，面向可重构路由器体系及互联网新环境带来的路由同步问题，设计了树型并行路由分发模型，给出了具体实施方法。理论分析和基于 NS2 的仿真实验，验证了模型能大大减少路由分发时间，并使负载更为均衡。

本文第 2 节给出了可重构路由器典型路由结构模型及传统同步模式下的分发性能，设计了树型路由并行分发（TPRD, tree-based parallel route distribution）模型，推导出 k 叉树分发中每一发送周期达到路由同步的板卡数；第 3 节介绍了 TPRD 模型相关算法及实施步骤，并从同步时间及负载均衡出发分析了不同板卡数对应的最佳 k 值；第 4 节给出了性能分析及基于 NS2 的实验；第 5 节是结束语。

2　并行路由分发

2.1　可重构路由器体系结构模型

可重构路由器是由若干个子路由器级连而成，级连后会形成 2 个层次的板间通信功能模块[8]。一方面是数据层快速交换模块，用于线卡间外部转发分组的快速交换，完成整个路由系统的报文转发；另一方面是控制层通信模块，用于本地收发的协议报文及控制消息的板间传递。每个子路由器等同于现在分布式路由器，其内部通信结构是一个高速以太网交换模块实现控制层通信。可重构体系中，所有子路由器的交换模块将级连在一起，控制层通信结构是一个多级高速以太网交换模块。逻辑上，可认为可重构路由器中任意 2 块板卡之间通过一个以太网交换模块直接进行控制层通信。

图 1 给出了本文可重构路由器结构模型。把首先计算出核心路由表的主控卡称为超级路由主控（SRM, super route mater），由 SRM 发起对整个可重构系统的路由同步工作。本文有如下定义：

m 表示可重构路由器中的子路由器数；

P_i 表示第 i 个子路由器的板卡数；

C_{one} 表示发送一条路由消息的代价，主要包括处理器及带宽消耗等；

q 表示SRM某个时间段内产生的核心路由的数量；

T_{one} 为可重构路由器任意两板间收发一条路由项消息所需时间，也叫传递周期。

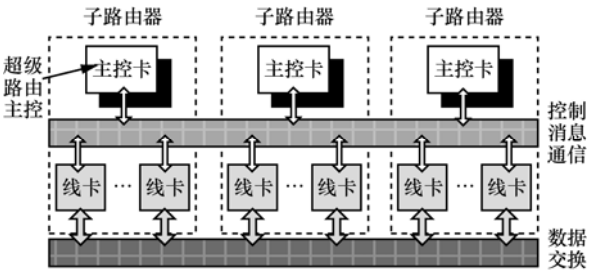


图 1　本文的可重构路由器结构模型

无论采取何种方案，路由同步完成后每块板卡都会接收所有核心路由，而板卡接收一条路由消息的代价总是一样的，所以忽略分析接收代价，只分析发送代价。另外，本文设定可重构路由器路由表同步模型的优劣评价包括 2 个方面：1）从 SRM 将路由分发到系统中所有板卡所需要的时间应尽量短；2）在路由同步的过程中，系统中各板卡为此工作所付出的代价尽可能均衡。

2.2 分析

传统分布式路由同步模式是简单的一对多串行路由分发(SRD，serial route distributing)策略，SRM 依次向每个接收者单播发送路由消息。除去 SRM，共有 $\sum_{i=1}^{m} P_i - 1$ 个路由消息接收者。所以，发送路由消息所需代价 C_{total} 以及系统路由分发时间 T_{total} 分别为 $\left(\left(q\sum_{i=1}^{m} P_i\right)-q\right)C_{\text{one}}$ 和 $\left(\left(q\sum_{i=1}^{m} P_i\right)-q\right)T_{\text{one}}$，它们都随路由消息数和板卡总数呈线性增长。可重构路由器中板卡数大量增加，SRM 发送路由消息需要的处理器负载和带宽都非常大，瞬间极大地影响板卡的处理性能。同时，可重构系统中的其他板卡只负责路由消息接收工作，发送代价为 0，出现负载极度不均的情况，造成资源浪费。另一方面，系统中路由同步时间也较长，容易造成网络设备路由表和转发表不一致及大量分组丢失。

由于路由同步是一对多的分发过程，因此先要权衡组播方式和单播方式的优劣。本文认为单播方式更为合理，主要基于以下考虑：1）路由分发必需是一个可靠传递过程，要有重传和确认机制；2）可重构路由器中，支持多种协议族分组寻径方式；板卡可配置成只支持某种或某几种协议族寻径，所以不是每块板卡都需要所有协议族的路由（本文把所有协议族的寻径策略都统称为路由）；组播将导致组播风暴，给板卡带来大量的无效消息，严重影响内部通信性能；3）组播分发机制，对于发送源 SRM 来说，需要保证所有板卡路由消息接收的可靠性，带来了太大的负载；4）路由消息是有状态的，不能乱序；SRM 要考虑接收者的状态以及链路状况，对于组播来说算法太复杂；5）若专门为路由消息构建组播可靠传输机制，会使得现有的路由器内部通信机制更加复杂，不易实现。

2.3 TPRD 模型

为了提高可重构路由器路由同步时间以及均衡路由消息分发过程中各板卡的负载，本文提出了一种新型的树型路由并行分发模型。SRM 计算出新的核心路由表后，构造路由更新消息，依次向 k 块板卡发送。每个获取到路由消息的板卡再向 k 个接收者发送，直到所有板卡都获取到该路由消息，完成该路由消息在整个路由系统的同步。每个板卡遵循先传播再处理的原则来节省系统同步时间，即板卡在转发路由消息之后再进行本地核心路由表更新处理。可以发现，整个路由分发体系如同一棵 k 叉树。初始状态只有树根节点，即计算出新的核心路由表的 SRM；每个节点只负责对自己子节点的路由消息发送，直到叶子节点。当 k=1 时，所有节点连成一条线，称其为单路蔓延。

由于分发过程中节点对其子节点的路由传递是依次进行的，当给第 2 个子节点传递消息时，第 1 个子节点也开始给它的子节点传递消息。因此，树的形成过程并不是平衡、均匀地向下蔓延。每一个周期，所有未满 k 个子节点的节点，同时向自己的某个子节点传送路由消息。最终形成一棵倾向一侧的不均衡的 k 叉树，树的深度就是路由消息最长的转发跳数，也就是整个系统路由分发的周期数。

图 2 以 2 叉树为例，描绘了路由分发过程中前 4 个周期路由分发情况，形成的是一棵不均衡 2 叉树。任一状态，准备进行消息传递的都是那些尚未满 2 个子节点的节点，即图 2 中实心节点。不失一般性，规定节点总是先将路由消息传递给左子节点，再传给右子节点。所以，对于任一节点，其左子节点为根的子树深度总比右子节点为根的子树深度大 1。可以计算出每经过一个传递周期 T_{one}，能够达到路由同步的节点数分别为 1、2、4、7、12、20、33 等。

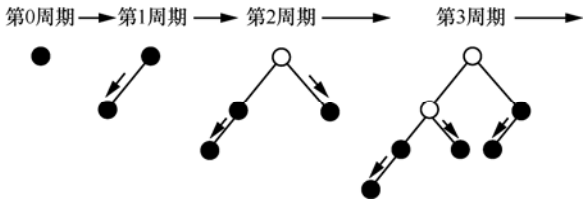


图 2 基于 2 叉树的并行分发示例

定义 1 k 路分发中，路由消息经过 i 个传递周期后，可重构路由器达到同步的线卡数为 $F_k(i)$。

定义 2 k 路分发中，路由消息经过 i 个传递周期后，分发树中有 j 个子节点的节点个数为 $V_k(i, j)$，$0 \leqslant j \leqslant k$。

根据 TPRD 模型路由分发规则，任一传递周期，分发树中只有那些子节点数未满 k 的节点会进行路由分发。所以，n 个周期到 n+1 个周期增加的节点数，等于 n 个周期完成时子节点个数未满 k 的节点数。满足式(1)：

$$F_k(n+1)=F_k(n)+V_k(n,0)+V_k(n,1)+\cdots+V_k(n,(k-1)) \tag{1}$$

而且，第 n 个分发周期后，分发树中子节点数为 0 的节点个数，就是 n−1 个周期到第 n 个周期增加的节点数：$F_k(n)$−$F_k(n$−1)。同理，第 n 个周期后，分发树中子节点数为 1 的节点个数，就是 n−2 个周期到 n−1 个周期增加的节点数 $F_k(n$−1)−$F_k(n$−2)。所以，第 n 个分发周期后，分发树中各类节点满足以下等式：

$$V_k(n,0)=F_k(n)-F_k(n-1)$$
$$V_k(n,1)=F_k(n-1)-F_k(n-2)$$
$$\cdots$$
$$V_k(n,i)=F_k(n-i)-F_k(n-(i+1))$$
$$\cdots$$
$$V_k(n,k-1)=F_k(n-k+1)-F_k(n-k)$$

用上述等式替代式(1)中右边 $V_k(i,j)$项，得到：

$$F_k(n+1)=2F_k(n)-F_k(n-k) \tag{2}$$

所以，$F_k(n)$的计算公式为

$$F_k(n)=\begin{cases}0, & n<0\\ 1, & n=0\\ 2F_k(n-1)-F_k(n-k-1), & n>0\end{cases} \tag{3}$$

所以，TPRD 模型中可重构路由器路由消息的同步周期为 nT_{one}，并满足式(4)，符号定义同前面描述。

$$n=\min\left\{j\in N\mid F_k(j)\geqslant\sum_{i=1}^{m}P_i\right\} \tag{4}$$

大规模路由消息分发事件发生，一般都伴随着大量 BGP update 报文的收发，而 update 报文从线卡上交到 SRM 需要占用大量板间带宽及板卡 CPU 资源。另一方面，TPRD 模型中叶子节点没有发送路由消息的工作，路由分发负载最轻，或者子节点数较少的节点（可称其为近叶子节点）相对会有略少的路由消息传递工作。所以，将涉及 BGP 协议报文收发的线卡设定为分发树的叶子节点或近叶子节点，就可提高系统面对大量 BGP 路由学习情况下的路由同步性能。令 LC_{BGP} 为路由器中 BGP 会话连接收发报文所经线卡，此类线卡数量为 θ。由于 TPRD 模型会将所有线卡对应成序列为从 1 到 $\sum_{i=1}^{m}P_i$ 的节点，并且算法使得序列号大的节点总是后加入到分发树中，成为叶子节点或近叶子节点的可能性极大。所以，可令 θ 个 LC_{BGP} 的分发树节点序号分别为 $\sum_{i=1}^{m}P_i-\theta+1$、$\sum_{i=1}^{m}P_i-\theta+2$、…、$\sum_{i=1}^{m}P_i$，从而达到 LC_{BGP} 路由消息传递负载尽量小的目的。

3　算法实现

TPRD 模型实施的关键在于将可重构路由器所有板卡视为树节点，构造成符合模型思想的分发树，即确定板卡间的父子关系（消息分发关系）。算法主要思想是：在设备刚启动尚未进行核心路由表计算之前，由 SRM 根据当前板卡数计算出路由分发周期数；接着为每块板卡计算其路由消息的下一个转发板卡，即子节点；最后 SRM 将父子关系信息通告各板卡。SRM 核心路由表变化时，将更新消息发给它的下一跳目的板卡；每块板卡收到路由消息后，按预定转发关系将消息依次转发给它的下一跳目的板卡。也就是说，事先已计算每块板卡的子节点号信息，路由分发时只需根据父子关系转发路由消息。为方便描述，符号定义如表 1 所示。

表 1　　**相关符号定义**

符号	描述
D_i	以 $Node_i$ 为根的子树的深度
$S_{(i)(j)}$	$Node_i$ 第 j 个子节点的节点号
c	当前处理的板卡号
t	可重构路由器总共板卡数
n	最大路由传递周期，即整个分发树的树深
k	每节点最多子节点数
Inqueue(N_i)	$Node_i$ 入队列
Equeue(N_i)	$Node_i$ 出队列
Empty_queue()	队列为空

算法假设路由器中所有板卡号都是从 1 开始依次编号，SRM 的序号为 1 且为树根节点。系统各板卡获得路由转发关系包括以下 4 步：

步骤 1　SRM 根据 t 计算路由同步到所有板卡需要的周期 n，n 满足式(4)；

步骤 2　对所有节点的子节点号初始化为无效值，并将树根节点放入处理队列，详见图 3(a)；

步骤 3　从树根节点开始，依次计算节点间的父子关系，详见图 3(b)；

步骤 4　SRM 将每块板卡的子节点信息：$S_{(i)(j)}$，即路由消息的下一跳转发板卡号，通告各板卡。

```
算法 1　TPRD_Init
1 for (i=1; i<=t; i++)
2     for (j=1; j<=k; j++)
3         S(i)(j)= −1;
4 D1 = n;
5 c = 1;
6 Inqueue(N1);
7 return;
```

(a)算法 1

```
算法 2　TPRD_Calculation
1 while (! Empty_queue())
2     get the first node:Ni from the queue;
3     for (m=1; m<= k; m++)
4         if (m > Di)  break;
5         if (c > t)  break;
6         c ++;
7         S(i)(m)= c;
8         Dc = Di − m ;
9         if (Dc > 0)  Inqueue(Nc);
//只有那些子树深度大于 0 的节点才需进队列
10    Equeue(Ni);
11 return;
```

(b)算法 2

图 3　TPRD 模型相关算法

上述步骤后，系统每块板卡都获知了自己的子板卡号（分发树中的子节点），收到路由消息后根据子板卡号进行下一跳转发即可。各板卡路由消息接收处理过程如图 4 所示，为了提高速度，板卡总是先转发路由消息再进行本地存储。子板卡号等于 −1 为无效值，无需处理。

表 2 给出了一个分配算法示例，给出了当系统中有 7 块板卡时，针对 2 路分发通过队列实现板卡父子关系计算的过程。初始化阶段：计算出 7 块板卡需要的分发周期为 3，即分发树的深度为 3，1 号板板卡入队列。队列中存储的节点都是非叶子节点。每次针对队列头部节点计算它的子节点，左子节点的子树深度为父节点减 1，右子节点的子树深度为父节点减 2。节点代表的子树深度大于 0 则需有入队列操作。此后，头部节点出队列。由 SRM 执行上述算法。

上述 TPRD 算法是针对路由系统正常工作时的描述，如果考虑板卡故障及热插拔，则涉及分发树的重计算。只要路由系统感知板卡增删，TPRD 算法就要根据新的板卡数重新进行上述 4 个步骤的工作，并按照新的分发树进行路由同步。限于篇幅，本文不对该问题深入探讨。另外，在实际路由器中（例如 CRS1 路由器和 TSR 路由器），线卡间的物

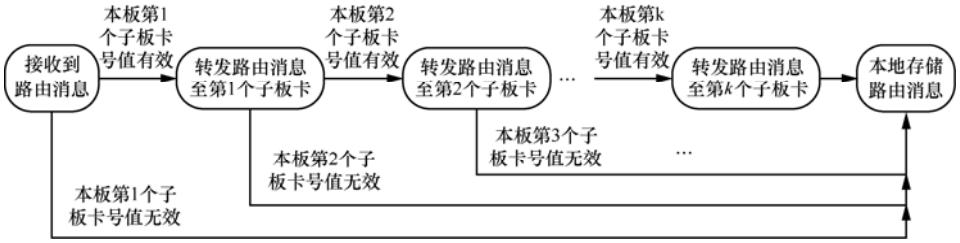


图 4　板卡路由消息接收转发处理流程

表 2　　**TPRD 算法实施示例**

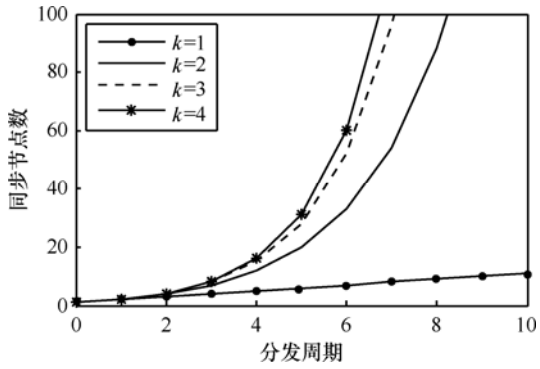
当前处理的最大节点号	队列操作	子节点号设置及队列操作说明
1	3 1	初始化，1 号板卡入队列，其子树深度为 3
3	3　2　1 1　2　3	1 号板卡第 1 子节点：2 号板卡；2 号板卡子树深度(3-1)大于 0，入队列 1 号板卡第 2 子节点：3 号板卡；3 号板卡子树深度(3-2)大于 0，入队列
5	2　1　1 2　3　4	2 号板卡第 1 子节点：4 号板卡；4 号板卡子树深度(2-1)大于 0，入队列 2 号板卡第 2 子节点：5 号板卡；5 号板卡子树深度(2-2)大于 0，不入队列
6	1　1 3　4	3 号板卡第 1 子节点：6 号板卡；6 号板卡子树深度(1-1)大于 0，不入队列 3 号板卡第 2 子节点：无
7	1 4	4 号板卡第 1 子节点：7 号板卡；7 号板卡子树深度(1-1)大于 0，不入队列 4 号板卡第 2 子节点：无

注：$\begin{matrix} j \\ i \end{matrix}$ 表示以 i 号板卡为根子树深度为 j。

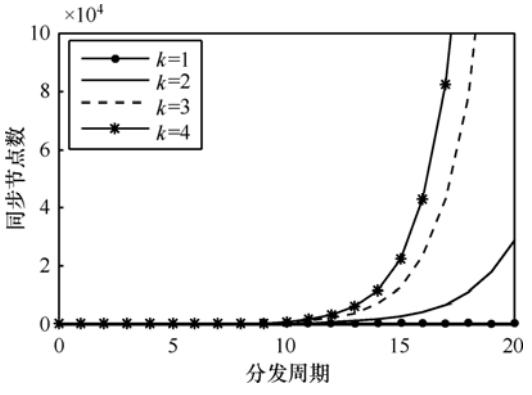
理拓扑将会影响其对等关系属性，而线卡间的交换结构是未来路由器体系结构的另一个研究难点。所以，本文采用了一个较理想的以太网总线交换模型开展研究工作。

4　实验及性能分析

本文从路由同步时间和负载均衡程度来评价路由分发算法。图 5 给出了 TPRD 模型中 k 取值为 1、2、3、4 时，一条路由消息同步的板卡数随发送周期的变化，图 5(a)和图 5(b)分别是针对不同板卡数量的分析。可以看出，板卡数量在 100 以下时，2 路分发较单路蔓延同步速度提高的较快，3 路分发较 2 路分发有少量提高。随着板卡数量增加，3 路分发的优势逐渐体现。当板卡数量达到 10 000 以上时，3 路分发较 2 路分发有明显提高，4 路分发较 3 路分发略有提高。对于负载均衡，在 k 路分发中，每个节点的路由消息传递负载就是与其子节点数相关。负载最重的就是拥有 k 个子节点的节点，负载为 kC_{one}，而负载最轻的就是叶子节点，负载为 0。所以 k 的值越小，整个可重构系统各板卡对于路由消息传递的负载就最均衡。



(a)少量板卡同步分析



(b) 大量板卡同步分析

图 5　分发周期与同步板卡关系

基于上述分析，由于目前可重构路由器的系统规模不会到达万块板卡，加上对板卡负载均衡的考虑，本文建议通过 2 路 TPRD 分发方式实现可重构系统中各板卡的路由同步。当然，随着路由器可重构体系结构的发展，设计者可以平衡同步时间及负载均衡再来选定 k 的值。

模拟实验基于 NS2 进行，测试 TPRD（2 路分发）、TPRD（3 路分发）和 SRD 3 种分发模式下，达到路由同步状态的板卡数与所需时间的关系。每条路由消息为 50byte，图 6(a)和图 6(b)分别是针对 1 条路由分发和 100 条路由分发的测试结果。显然，TPRD 相对传统 SRD 分发模式同步时间有了很大的节省，并且随着板卡数量增多优势更明显。另一方面，TPRD 模式下 2 路分发和 3 路分发的路由同步速度却相差无几。由于 2 路分发相对 3 路分发各板卡负载更为均衡，因此本文认为 TPRD 模型 2 路分发最为合适。这与图 5 的理论分析完全一致。接着，分析了板间有大量 BGP 协议报文传递时路由同步模式的性能。实验中 SRM 与 5 个板卡间有 185kbyte/s 的 BGP 协议报文交互。如前文分析，TPRD 总会设置这些有背景流量的板卡为分发树的叶子节点。实验结果如图 6(c)和图 6(d)所示，与没有 BGP 流量的实验结果基本一致。

5　结束语

面对互联网路由表容量的急剧增长，一个高效的路由分发模型是可重构路由器体系结构设计面临的一个挑战。本文首先对当前普遍实施的一对多点的主动广播更新的路由同步模式进行了性能分析，它存在同步周期长、负载不均衡等问题。于是，基于可重构路由器典型路由体系架构，设计了 TPRD 树型并行路由分发模型来分摊系统路由分发负载、提高路由同步速度。推导出 TPRD 模型的 k 路分发中，任一周期能达到路由同步的板卡数，并分析了 k 路分发时各板卡的负载均衡状况。TPRD 模型考虑了路由分发时伴随大量 BGP update 报文收发的情况，设置进行 BGP 协议报文收发的线卡为分发树的叶子节点，从而保证这些线卡尽少量的路由消息分发负载。论文给出了 TPRD 模型的实现算法和具体实施方法，并就当前可重构路由器的体系结构而言，建议采用 2 路分发方式。基于 NS2 的实验验证了 TPRD 模型的快速路由分发特性。

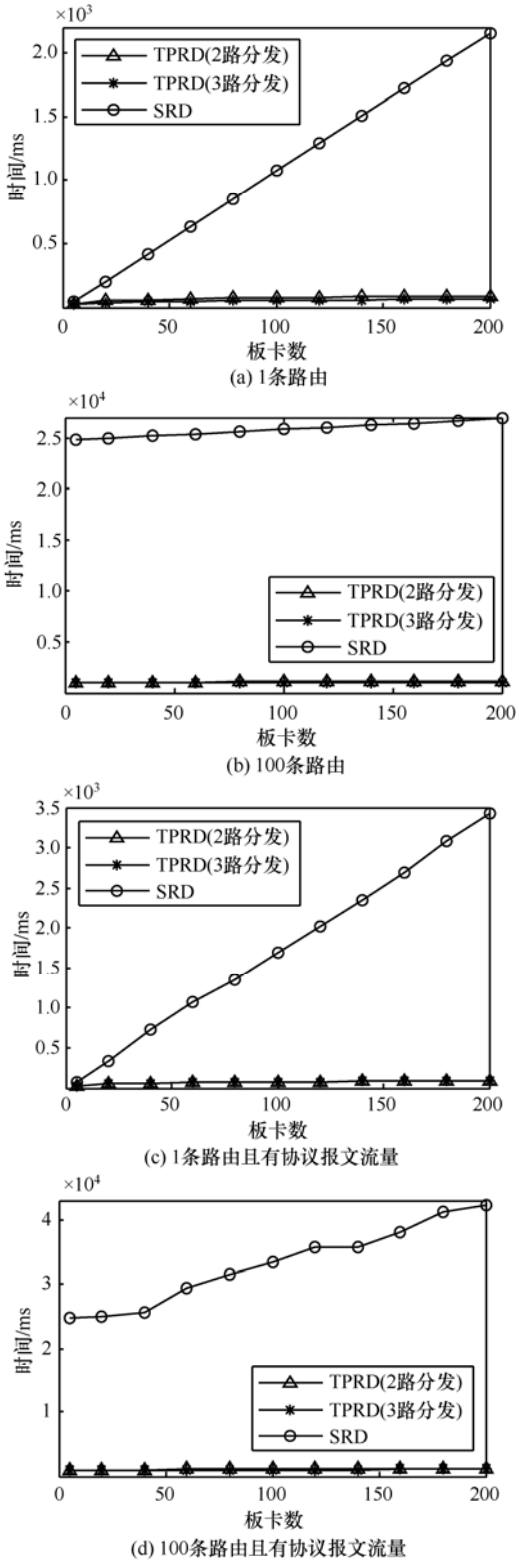


图 6　不同分发模式基于 NS2 的实验结果

可重构路由器的发展尚处于初期发展阶段，还有很多重要问题需要研究，如：分布式路由协议、高速交换网络互连结构、FIB 表高效存储等。本文下一步工作将围绕这些问题展开研究，并将进行可重构路由器原型系统的实现。

参考文献：

[1] 徐恪, 吴建平, 徐明伟. 高等计算机网络:体系结构、协议机制、算法设计与路由器技术[M]. 北京:机械工业出版社, 2009.
XU K, WU J P, XU M W. Advanced Computer Networks: Architecture, Protocol Mechanism, Algorithm Design and Router Technology[M]. Beijing: Mechanism Industry Press, 2009.

[2] Cisco CRS-1 carrier routing system[EB/OL]. http://www.cisco.com, 2011.

[3] TX matrix platform[EB/OL]. http://www.juniper.net, 2010.

[4] The avici TSR: cornerstone of the multi-terabit core[EB/OL]. http://www.avici.com, 2008.

[5] BGP routing table data[EB/OL]. http://bgp.potaroo.net, 2011.

[6] MEYER D, ZHANG L, FALL K. Report from the IAB Workshop on Routing and Addressing[S]. RFC 4984. 2007.

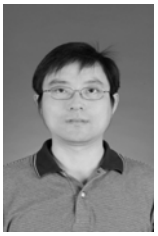
[7] 张晓哲, 卢锡城, 朱培栋. 一种可扩展路由器转发表同步框架及关键算法[J]. 软件学报, 2006, 17(3):445-453.
ZHANG X Z, LU X C, ZHU P D. A synchronization framework and critical algorithm maintaining single image of IP forwarding tables among cluster router's nodes[J]. Journal of Software, 2006, 17(3): 445-453.

[8] 吴鲲. 可扩展路由器软件体系结构研究[D]. 北京: 清华大学,2006.
WU K. Research on Software Architecture for Extensible Router [D]. Beijing: Tsinghua University, 2006.

作者简介：

陈文龙（1976-），男，江西吉安人，博士，首都师范大学讲师，主要研究方向为网络体系结构和网络协议。

徐明伟（1971-），男，辽宁朝阳人，博士，清华大学教授、博士生导师，主要研究方向为网络体系结构、高速路由器体系结构和协议测试。



徐恪（1974-），男，江苏洪泽人，博士，清华大学教授、博士生导师，主要研究方向为下一代互联网、交换和路由结构、P2P 网络和物联网。