

基于端用户可控的 IP 网络路由体系结构和算法*

樊秀梅⁺, 陈常嘉

(北方交通大学 电子信息工程学院,北京 100044)

IP Routing Architecture and Algorithm Based on End_Users Control

FAN Xiu-Mei⁺, CHEN Chang-Jia

(School of Electronics and Information Engineering, Northern Jiaotong University, Beijing 100044, China)

+ Corresponding author: Phn: 86-10-62783596, Fax: 86-10-62771138, E-mail: xmfan@csnet1.cs.tsinghua.edu.cn

<http://www.njtu.edu.cn>

Received 2002-05-15; Accepted 2002-08-14

Fan XM, Chen CJ. IP routing architecture and algorithm based on end_users control. *Journal of Software*, 2003,14(5):1023~1028.

<http://www.jos.org.cn/1000-9825/14/1023.htm>

Abstract: IP routing architecture and algorithms of Internet is a sort of key technology. But in the practical application, the actual routing architecture and algorithms have still some problems. The purpose of this paper is to provide a sort of IP routing architecture and algorithms based on end_users control to resolve this problem. In this architecture, routing decision will be done mostly by end_users, but the task of router would be predigested to two simple functions, notify network information and assort with users decision. This paper will achieve above object through users self-organizing behaving. This architecture can adapt the extended requirement of network size and application need, and form a sort of distributed, scalable routing architecture and effective routing algorithm.

Key words: IP routing architecture; self-organizing; end_users control

摘 要: IP 网络的路由体系结构及算法是网络有效运行的关键技术.现行的路由体系结构及算法在实际应用中存在着一些问题.针对该问题,提出一种端用户可控的 IP 网络路由体系结构和具体的路由算法.在提出的端用户可控的路由体系中,利用用户级别的自组织路由方法来简化路由器的负荷,增强端用户的智能性.模拟仿真实验表明,该路由体系的使用,将使路由器的任务简化为通报网络信息和协调用户决断这两个较为简单的功能,且路由选择的决断考虑到端用户的实际需求.该体系结构可以更好地适应网络规模和应用需求的不断扩大,形成一个分布式、扩展性较好的路由体系和有效的路由算法.

关键词: IP 网络路由体系;自组织;端用户可控

中图法分类号: TP393 文献标识码: A

Internet 作为连接世界各大网络的一个松散性互连网络,其实现世界范围连通性的关键技术之一就是路由技术.在目前的 Internet 中,路由器上有多种不同的路由协议,这些协议使用不同的路由算法.为了使网络协调运

第一作者简介: 樊秀梅(1967-),女,山西山阴人,博士,副教授,主要研究领域为计算机网络,IP 网络 QoS 体系结构,系统性能评价.

行,路由器之间需要交换大量的链路和网络状态等信息,并计算、更新路由表,其可扩展性较差。Internet 的巨大增长速度使得现有的路由体系出现了可扩展性、鲁棒性等方面的不足,为了更好地适应新的网络应用需求,需要重新构架一个新的、扩展性较好的路由体系和有效的路由算法。

目前,Internet 所使用的两种主要路由协议是距离矢量路由协议和链路状态路由协议,虽然它们都属于动态路由协议^[5],但这两种路由协议都不能完全适应复杂程度快速增长的 Internet。从算法本质来说,传统的路由协议都把获得网络的确切描述作为其路由计算的必要条件。但随着网络规模的不断扩大,为了满足这个条件就需要付出越来越大的代价。为此,本文提出了一种基于用户的 Internet 路由解决方案。

1 问题的引出与算法特点描述

对于现有的 Internet 路由协议,对于 RIP^[2],OSPF^[3],BGP^[4]等来讲,都是通过路由器之间交换网络信息来建立全网路由的,我们一般称这样的协议为基于路由器的网络路由协议。当网络上只有少数几个路由器,特别是这些路由器是由单个管理实体所维护时,这种路由协议的优点特别明显。当网络规模扩大,而且参与网络建设的实体越来越多时,这种协议方式就显得过于笨重,不能满足需求。

总的来说,Internet 的路由选择已经从本质上脱离了传统电信网络的集中控制,带有了自组织的味道,但这种自组织仅仅停留在路由器级别。随着网络规模的扩大,网络组成的多样化和网络速率的提高,简化网络结点的构成方式和管理方式将成为减小网络成本、提高网络利用率和增强网络可靠性的一个关键因素。本文试图说明,通过用户级别的自组织路由是达到上述目标的一个有效手段。

在本文提出的端用户可控的路由体系中,路由选择的决断将主要由终端用户作出,路由器的任务将简化为通报网络信息和协调用户决断这两个较为简单的功能。这样一来,大量复杂的、较为慢速的计算工作将被推到网络边缘的各个终端去完成,充分利用未来网络中用户终端的剩余处理能力来构成强大的网络计算实体,从而大大减少了网络本身实施同样计算规模所必须花费的成本;另一方面,用户所能感受的,是确实的端到端网络服务性能,因此由用户作出路由选择决断将能够更有效地把路由选择和网络服务质量结合起来,最大限度地满足网络用户需求。

目前见到的大多数相关研究,都是将算法应用与已知的路由问题有机地结合起来,去解决具体的路由选择问题,很少有人将其发展成一个规模的、实用的路由协议。文献[1]首先突破了上面的视点框架,试图将蚂蚁算法的思想应用于基于终端的大规模网络路由上,并取得了成功。该算法的核心思想是,每个在线用户反向通报自己的存在,网络节点按分岔口的数目衰减用户的通报值,并将收到最大通报值的入口作为该节点通向该用户的最佳出口。该方法实际上可以看成是由终端发起的一种改型的 RIP 协议。该算法的不足在于,除了要求网络在双向链路的对称性以外,其路由的选择实际上仍是由路由器作出的。按照网络的服务质量动态选择路由时,仍然需要网络间的大量信息通报和复杂的路由器本地计算。

基于上面的各种考虑,本文给出了一种基于单路径的、基本上遵从最小跳准则的终端路由协议,并通过仿真证实了该协议的正确性和有效性。之所以选择单路径和最小跳准则,是为了与传统的 RIP 或 OSPF 协议进行比较。我们的仿真工作表明,本文提出的终端路由协议,在性能上远好于传统的基于路由器的网络路由协议,尤其是在网络的分组丢失率、分组平均时延和动态适应等方面。具体详见第 3 节。

2 端用户可控的分布式、可扩展路由体系与算法描述

2.1 端用户可控的分布式、可扩展路由体系的基本思想

端用户可控的分布式、可扩展路由体系是受蚂蚁算法的启发,并改造蚂蚁算法而提出的。本路由体系提出了网络蚂蚁的概念,一类网络蚂蚁称为路由探测蚂蚁,简单地说就是一个探测分组,它能够记忆它所经过的路径,并且进一步有可能在每一处节点采集一些有用的路径信息样品。每个终端在发送数据时,同时向同一目的地发送一定比例的路由探测蚂蚁。路由探测蚂蚁随机地选择到某一子网的出行端口,并在传输过程中记录所经过的节点信息,诸如节点 IP 地址、链路时延、链路带宽利用率等,将其标记在分组头上以利于为目的终端的路由

选择提供决策信息.当从一个网络终端 s 发出的沿多个不同路径爬行的路由探测蚂蚁到达目的终端 d 时,目的地比较这些探测蚂蚁头上所携带的信息,根据自己想要的性能参数矢量作出从 s 到 d 的路由决断后,通过另一类网络蚂蚁——路由染色蚂蚁按所选定的路由返回,沿途通知经过的所有路由器,即留下信息素.由于每个子网的终端会收到大量从不同源终端沿不同路径到达的路由探测蚂蚁,因此原则上说,一个路由器可能在所有的方向上都会收到到达一个给定子网的路由染色蚂蚁,并对该路由器通过该输出口到达该子网的路由表的对应元素处留下信息素.为了协调终端的这种决断分散性,在每一个路由器都对它所能到达的子网(终端)建立了一个对应于该子网的信息素向量.路由染色蚂蚁更新每一个经过节点的路由表(有 k 个出口的路由器为每个可到达子网保持一个有 k 个元素的向量,作为该路由器的路由表.路由表是对应每个子网不同端口的性能矢量参数值,这里沿用蚂蚁算法中的名称,称为信息素),假定在开始状态每个端口路径上的信息素是一个相同值.这样后来的分组就可以根据信息素的大小(或加上其他什么参数),依照概率选择传输路径.按照此过程,经过一定的时间,网络可以做到最优的路由选择.

一个信息素向量的元素与路由器的输出口一一对应,每个向量元素的值是该输出口到达该子网的信息素.从一个给定方向收到的路由染色蚂蚁,只能改变其所对应的向量元素的信息素.在一般的信息转发过程中,路由器则从对应于分组到达子网的信息素向量中选择信息素最大的输出口发出该信息分组.

上述过程在原则上并不要求网络对称,它只要求路由染色蚂蚁能够到达所选定路径的每一个路由器,而不一定必须按照路由探测蚂蚁同样的对称路由返回,但当网络对称时,实现较为简单.

2.2 用户可控的分布式、可扩展路由算法的具体实现

本文提出一个用户可控的分布式、可扩展路由算法,简称为 TDF.该算法使用 3 类分组,其具体的分组行为如下描述.

2.2.1 路由探测蚂蚁的行为

路由探测蚂蚁在网络传输时进行两种行为:

① 记录信息行为:当路由探测蚂蚁在网络中传输时,将经过的每一个节点的有用信息记录在分组头上,以利于终端用户据此对网络到用户自身路由的选择和调整.

② 选择路径行为:为了发现新的路由,路由探测蚂蚁的路径选择有一定的随机性,即随机选择路由表里对应每个子网的不同端口.

2.2.2 路由染色蚂蚁的行为

包括 3 部分:产生路由染色蚂蚁、路由染色蚂蚁返回源端和信息素调整行为.

① 产生路由染色蚂蚁:当路由探测蚂蚁到达接收端时,接收端缓存并比较某一固定时间段内,来自于同一源的路由探测蚂蚁所记录的信息,选出接收端认为适合自己要求的、性能良好的路由探测蚂蚁,以此产生路由染色蚂蚁.

② 路由染色蚂蚁返回:根据路由染色蚂蚁头上的历史记录,路由染色蚂蚁按照其来时经过的路由器反向返回源端.

③ 信息素调整行为:当一个反向返回的路由染色蚂蚁经过某一节点时,该节点应对输入链路的相应端口的已有信息素进行增强更新.而信息素是会随着时间的推移而挥发的,所以,也要对对应于同一子网的其他端口的信息素进行减少更新.如果值小于 0,则表示为 0;如果值大于或等于 0,则表示为实际值.

信息素根据如下公式被更新:

$$\tau_{ij}(t+n) = \rho \cdot \tau_{ij}(t) + \Delta \tau_{ij}^k, \tag{1}$$

$$\tau_{ij}(t+n) = (1-\rho) \cdot \tau_{ij}(t). \tag{2}$$

这里,

$$\Delta \tau_{ij} = \sum_{k=1}^m \Delta \tau_{ij}^k, \Delta \tau_{ij}^k$$

是在时间段 $[t, t+n]$ 之间,第 k 个蚂蚁放在边 (i, j) 上的信息素量. ρ 是坚持率(有效率),这样 $(1-\rho)$ 就是挥发率.

2.2.3 数据分组行为

数据分组的传输是根据路由表里对应每个子网不同端口的信息素来选择路径的.在本文的分组交换网络中,数据分组选择信息素最大的端口输出数据.当数据分组到达终端时,就被终端用户接收,不再产生任何行为.

3 仿真结果与分析

为了验证上面想法的可行性和实用性,我们做了大量的仿真.仿真结果表明,本文提出的路由体系和算法(TDF)在用户关心的多种网络性能参数上都优于现有的 OSPF 路由协议.本文的仿真是在相同网络流量和相同网络环境情况下对 TDF 和 OSPF 所作的比较.对同样实验条件的多次重复实验证明了算法的可重复性.

3.1 仿真平台简介

我们使用 C++编写了一个仿真平台 Nsimulator.这里使用了文献[1]中流量场景的概念.流量场景是指能够反映网络在一段时间内由用户加载的数据流量的模板.该平台可以在特定的流量场景下仿真同一网络在不同的路由算法下的工作情况,并对诸如分组产生数、分组丢失数、网络流量及分组到达时间等参数进行统计、分析.在流量场景的设计中,我们考虑了网络中可能遇到的各种问题,如业务的突发性,链路的中断、节点的失效等.

为方便研究,在仿真中将网络模型进行了一定的抽象简化.简化的模型主要考虑了以下几点:

- 每一个节点都与其他几个节点相连.
- 每一个节点都可以发出数据包,并可接受其他节点向它发来的数据分组.
- 每一个节点还要完成数据分组转发的过程.
- 规定了分组的最大跳数.

这个模型可看作是 Internet 网络的抽象模型.在实际应用中,情况可能要更加复杂一些,但是,这种抽象在进行路由算法和拥塞控制算法分析的时候应该是可以概括通信网的主要特征的.在下面的论述中,将主要以这样的模型来进行网络路由算法的讨论而不涉及过多的具体技术实现细节问题.

在不特别注明的情况下,本文中的路由器和网络节点的概念是统一的.

3.2 仿真中所用的参数与有关的计算公式

除非特殊说明,本文中的仿真时间为 40s,链路缓存为 50 个分组,数据分组发送的最大间隔时间为 1ms,对应于每个源的节点缓存为 10 个探测分组,分组传输的最大跳数设定为 16.本文中所说的网络流量值是指仿真期间网络中总的传输分组数.

在本文的仿真中,考虑了由于距离引起的传输时延,其具体的时延计算采用如下公式:

$$\text{TotalDelay}=\text{ServiceDelay}+\text{TransDelay};$$

$$\text{ServiceDelay}=\text{PacketSize}/\text{Bandwidth};$$

$$\text{TransDelay}=\text{PathLength}/200(\text{注:假定以光速传输为标准}).$$

本文中所有的仿真统计数据都是在仿真开始 8s 后才开始统计的.为了更好地模拟网络的突发行为和动态行为,在 9s 时,在原有的数据发送基础上又加入一个流量场景,其分组发送的时间间隔为原有间隔的一半.

3.3 仿真中所用的网络拓扑图与拓扑结构的具体数据

为了增强仿真的实际性,我们选用了具有实际背景的模式,即类似于英国公共 SDH 干线网的拓扑结构.这个拓扑结构在许多文章中都引用过,具有一定的代表性.该拓扑结构有 30 个节点和 110 条链路.在仿真中,由于没有收集到 SDH 干线网的具体网络环境数据,因此根据其实际的节点距离、节点度数等信息规定了不同节点和链路的不同属性(缓存、带宽、长度等).

对 TDF 和 OSPF 所施加的流量场景是根据实际情况生成的.在生成新的流量场景过程中有两个重要参数:分组发送的平均间隔时间 T_m 和平均包长度 L_m .这两个参数控制了网络流量场景的特征.本文选用的流量场景中,节点以随机时间间隔向随机选择的其他节点发送随机长度的数据包.时间间隔服从平均分布,其统计平均为 T_m ;包长服从平均分布,其统计平均为 L_m .在本文的仿真中,我们选取 T_m 的值为 1 000ms,选取 L_m 的值为 800Byte.

TDF 中发送的路由探测蚂蚁和生成的路由染色蚂蚁的包长都采用统一大小,为 100Byte.在 TDF 中,路由探测蚂蚁的发送间隔选定为 10 000ms.

为了比较本文提出的 TDF 路由算法和现在 Internet 网络中普遍采用的 OSPF 算法在分组丢失率、分组传输的平均时延以及动态适应性等方面的性能,我们在不同网络流量环境下做了大量的仿真实验,实验结果如下.

3.4 仿真实验结果

3.4.1 网络输入速率变化时的仿真结果

图 1 和图 2 是当网络中的流量输入速率变动时的仿真统计结果.这里分组丢失率 p 的计算公式为

$$p = \frac{\sum_{i \in T} D_i}{\sum_{i \in T} G_i}, \tag{3}$$

$$D_i = G_i - A_i, \tag{4}$$

其中, D_i 表示在时刻 $t \in [0, T]$ 时的分组丢失数, G_i 表示在时刻 $t \in [0, T]$ 时的分组产生数, A_i 表示在时刻 $t \in [0, T]$ 时的分组到达数.

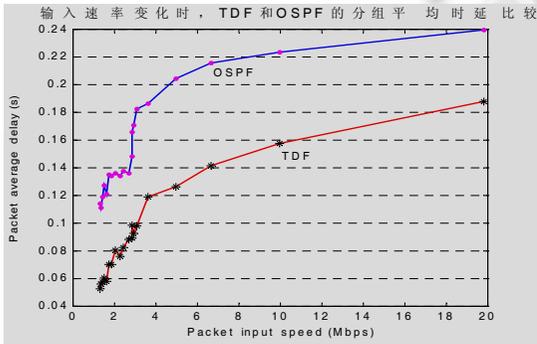


Fig.1 Packet average delay compare

图 1 输入速率变动时的分组平均时延比较

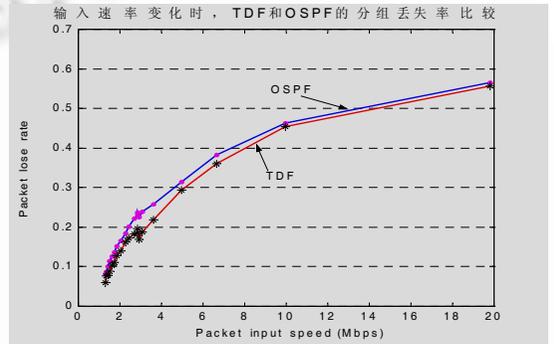


Fig.2 Packet lose rate compare

图 2 输入速率变动时的分组丢失率比较

仿真结果表明,TDF 在时延性能参数、分组丢失率性能上是优于 OSPF 的.

3.4.2 TDF 形成的最短路与 OSPF 的比较

在图 3 中由 TDF 形成的优化路径里,约有 90%是网络的最短路径,且随着网络流量的增大呈上升趋势,而只有约 10%的路径是根据网络的动态变化而发生变动的.这既保持了网络的稳定性,又提高了网络的有效性.

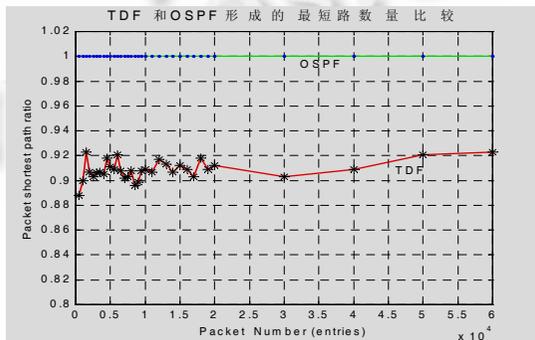


Fig.3 The compare of the shortest path amount

图 3 TDF 形成的最短路与 OSPF 的比较

3.4.3 TDF 和 OSPF 在网络链路动态变化时的性能比较

网络结构是动态变化的,一个好的路由体系应能很好地考虑到这种动态性,使其具有实用性.本文的路由体系和算法是具有这种动态适用性的.在仿真时间段内(60s),分组的平均发送速率约为 10.7Mbps,节点以随机的间

隔向随机选择的其他节点发送随机长度的数据包.时间间隔服从平均分布,其统计平均为 1 000ms,包长服从平均分布,其统计平均为 800Byte.在 9.0s 始,让链路 74,75,76,77,80,81,82,83 断掉,这时节点 21 只剩下一个相连的节点 17,下面的仿真结果就是反映如此情况下的分组丢失率、分组平均时延算法对动态变化的自适应性.

图 4 显示了链路断掉前后分组的丢失情况的对比.从图中可以看出,同一算法在链路断掉后,其分组的丢失都有所增加,这是可以想象和理解的.从图 4 也可以看到,链路断掉后,TDF 可以较快地调整状态,使丢失率下降.其动态调整性能较好,能较好地适应网络状态的变化.

图 5 显示了链路断掉前后分组平均时延情况比较.从图中可以看出,同一算法在链路断掉后,其分组的平均时延都有所增加,这是因为动态变化后有一个重新路由的时间和过程.从图 5 也可以看到,链路断掉后,TDF 可以较快地调整状态,使分组平均时延下降.其动态调整性能较好,能较好地适应网络状态的变化.

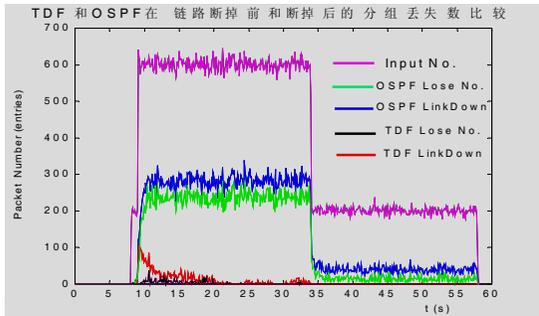


Fig.4 Packet lose rate compare under dynamic variety

图 4 动态变化时分组丢失率的比较

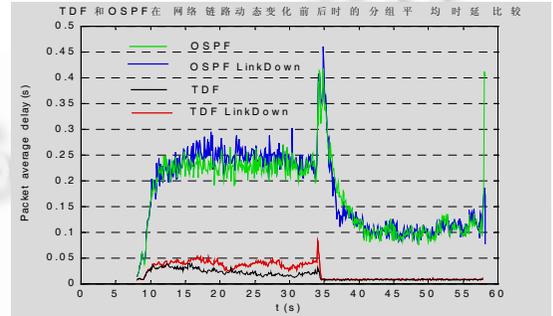


Fig.5 Packet average delay compare under dynamic variety

图 5 动态变化时分组平均时延的比较

4 小 结

为了适应 Internet 的快速增长和壮大,本文提出了一种新的路由体系,称为端用户可控的分布式、可扩展路由体系.同时也提出了具体的、扩展性较好的分布式路由算法(TDF).理论分析和仿真表明,该算法在路由性能上优于 Internet 现用的 OSPF,且具有非常好的可扩展性和动态适应性.与传统路由算法不同的是,TDF 路由不是建立在对全网拓扑完全了解的基础之上,路由形成是基于用户的,通过大量独立运行的主体互相协作而自组织地涌现出来.这种涌现行为从算法分析上属于分布式运算,但是与距离矢量路由和链路状态路由的分布式运算相比,该算法具有更细的时间粒度和更快的反应速度.同时,该算法对路由器硬件的配置要求也很低.从本质上改变了现有的路由风格.同样实验条件下的多次重复实验证明了该算法的可重复性.

本文提出的路由算法比传统路由算法大大向前跨越了一步.它良好的动态性、可扩展性和鲁棒性,使运营商节约了资本投入,使用户有了更广泛的选择自由,充分体现了以人为本、意思自治的商业理念,更加适应新一代 Internet 的需求.

References:

- [1] Zhang L, Ren Y, Shan XM. Selg-Organized criticality in multi-agent cooperative routing. In: Bosman G, ed. Proceedings of the ICNF. Gainesville, 2001. <http://www.doce-conferences.ufl.edu/icnf/>.
- [2] Hedrick C. Routing information protocol, 1988. <http://www.faqs.org/rfcs/rfc1058.html>.
- [3] Moy J. OSPF Version 2, 1998. <http://www.faqs.org/rfcs/rfc2328.html>.
- [4] Rekhter Y, Li T. A border gateway protocol 4 (BGP-4), 1995. <http://www.ietf.org/rfc/rfc1771.txt>.
- [5] Su X. Dynamic multi-path routing: asymptotic approximation and simulations. In: SIGMETRICS 2001/Performance 2001. 2001. <http://www.cs.umd.edu/conference/sigmetrics2001/>.