第 37 卷 第 24 期
Vol.37 No.24

计 算 机 工 程
Computer Engineering

2011 年 12 月
December 2011

• 网络与通信 •

文章编号：1000—3428(2011)24—0086—02 文献标识码：A 中图分类号：TP393

基于网络数据流局部性的连接管理优化方法

熊 兵[1,2]，廖年冬[1]，李 峰[1]，陈晓苏[2]

(1. 长沙理工大学计算机与通信工程学院，长沙 410114；
2. 华中科技大学计算机科学与技术学院，武汉 430074)

摘 要：根据高速网络环境下连接管理的性能需求，提出一种连接管理优化方法，即将 MTF 启发法应用于连接表。网络数据流的局部性特点表现为，属于同一个连接的一组数据包可能在短时间内集中到达。基于此，应用 MTF 启发法优化连接表的查找操作，形成 MTF 连接表。给出优化后的连接管理算法流程。借助实际高速网络数据流，对 MTF 连接表优化方法进行性能评估。实验结果表明，MTF 连接表的查找性能明显优于传统的排序连接表。

关键词：高速网络；连接管理；网络数据流局部性；MTF 启发法

Connection Management Optimization Method Based on Network Dataflow Locality

XIONG Bing[1,2], LIAO Nian-dong[1], LI Feng[1], CHEN Xiao-su[2]

(1. School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China;
2. School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

【Abstract】Targeting the performance requirements of connection management in high-speed networks, an optimization of connection management is proposed, applying Move to Front(MTF) heuristic to improve the lookup operation of connection table. The locality property of network traffic is analyzed, which is exhibited as a group of packets within a connection probably arrive in group in a short time. Based on the property, the MTF heuristic is applied to optimize the lookup operation of connection table, and get the MTF connection table. The implementation of optimized connection management is described. The optimization and the MTF connection table, with physical high-speed network traffic are evaluated. Experimental results indicate that the lookup performance of MTF connection table is superior to that of the traditional sorted connection table.

【Key words】high-speed network; connection management; network dataflow locality; Move to Front(MTF) heuristic

DOI: 10.3969/j.issn.1000-3428.2011.24.028

1 概述

连接管理是数据包处理领域的基础关键功能之一，可用于构建入侵检测[1]、内容审计和网络地址转换等各类网络数据流分析系统。在连接管理中，维护并发连接会话的哈希表称为连接表。

在高速网络环境下，连接表的查找操作需要花费大量的开销：(1)数据包到达非常密集，每个数据包都需要查找连接表，查找操作相当频繁；(2)并发连接数量可达数十万之多，连接表异常庞大，单次查找开销很大。因此，要实现高效的连接管理，就必须优化连接表的查找算法。

目前已有不少工作研究连接表的优化方法。针对网络处理中的哈希表，文献[2]利用扩展的布鲁姆过滤器(Bloom Filter)，提出一种新的哈希表数据结构和查找算法。该哈希表可支持快速查找，但需要大量额外的辅助内存。针对入侵检测系统中的连接表，文献[3]设计了哈希函数，但在解决哈希冲突时，并没有充分利用网络数据流特性。为了提升连接表的超时处理性能，文献[4]改进并实现了一种 Patricia 算法来组织连接会话。文献[5]提出每次将访问的连接移到表项尾部的优化方法。然而，这样会降低连接表的查找性能。

针对高速网络环境下连接管理的性能需求，本文提出一种基于网络数据流局部性的连接管理优化方法，即将 Move to Front(MTF)启发法应用于连接表。

2 连接管理优化方法

2.1 网络数据流局部性

网络数据流局部性[6]是包交换网络中的一种现象，通常表现为：无论从时间还是从空间的角度上看，网络数据流中各数据包中的 IP 地址和端口分布都并不均匀。网络数据流局部性可分为时间局部性和空间局部性 2 个方面。时间局部性是指网络中的相邻数据包很可能高度相关。这就是说，在网络数据流中一个从主机 A 到主机 B 的数据包，往往紧随另一个从主机 A 到主机 B 或主机 B 到主机 A 的数据包。时间局部性体现了网络数据流的短期特性。空间局部性是指网络数据流在主机中的分布往往很不均匀。抽样观察表明：大约 90%的数据流分布在 10%的主机上[7]。空间局部性体现了网络数据流的长期特性。

在计算机网络中，局部性特点是人为设计的协议和传输

基金项目：国家自然科学基金资助项目(6097311)；湖南省教育厅科学研究基金资助项目(11C0036)；长沙理工大学人才引进基金资助项目

作者简介：熊 兵(1981－)，男，讲师、博士，主研方向：网络与信息安全，图像处理，模式识别；廖年冬，讲师、博士；李 峰，教授、博士；陈晓苏，教授

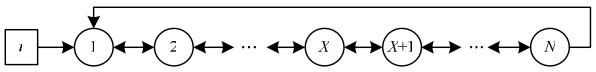
收稿日期：2011-06-01 **E-mail：**xiongbing@csust.edu.cn

数据的应用程序导致产生的自然结果。目前 Internet 上最为普遍的 Web 应用，会产生大量的文件下载活动；日益普及的网络电话、网络电视、视频聊天等流媒体应用，会产生持续的大块数据传输活动；传统的 FTP 服务和电子邮件服务，会产生相对集中的数据上传下载活动。这些活动都会产生特定主机对之间的大量突发数据包。

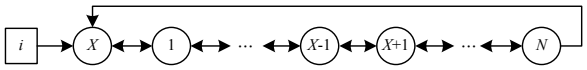
2.2　MTF 启发法

从连接管理的角度看，网络数据流局部性体现为：(1)连接的一个数据包到达后，它的下一个数据包可能很快到达；(2)属于同一个连接的一组数据包很可能在短时间内集中到达。这对于连接表查找而言意味着：(1)如果连接表中的某个连接记录被访问，那么该记录可能很快再次被访问；(2)一个连接记录很可能在短时间内多次连续被访问。

为了减少平均查找开销，可应用 MTF 启发法对连接表查找进行优化。具体操作方法是：在每次查找连接表命中一个连接后，将该连接移动到所在哈希链的最前面，而哈希链中该连接前的所有连接后退一个位置。应用 MTF 启发法的哈希链称为 MTF 哈希链。图 1 给出了 MTF 哈希链的工作原理。图 1(a)表示哈希值为 i 的连接 X 上到达一个数据包前第 i 号哈希链的状态；图 1(b)表示该数据包刚处理完后第 i 号哈希链的状态。此时连接 X 已被移至第 i 号哈希链的最前面。



(a)数据包到达前的第 i 号哈希链



(b)数据包处理后的第 i 号哈希链

图 1　第 i 号 MTF 哈希链的工作原理

通过应用 MTF 启发法，连接表的查找性能有较大幅度的提升。经过一系列访问之后，连接表中的所有活跃连接都会被移到靠近所在哈希链链头的位置。显然，基于 MTF 启发法的查找操作具有 2 个基本特征：(1)对于非活跃连接上的少量数据包，其查找长度将会稍微增加；(2)对于活跃连接上的大量数据包，其查找长度将显著减少。因此，连接表的平均查找长度将会明显减少，整体查找性能得到较大提高。

3　连接管理实现

应用 MTF 启发法的连接表称为 MTF 连接表。利用 MTF 连接表进行连接管理的大致处理步骤为：

(1)收到一个网络数据包后，提取数据包的关键字段信息，进而计算出对应连接的关键字，即连接标识符；

(2)根据连接标识符，查找连接表，取出对应的连接；

(3)针对具体应用需求根据数据包信息更新连接内容；

(4)将连接插入连接表中对应哈希链的链头位置。

优化后的连接管理实现流程用伪代码描述如下所示。该算法的思想是根据到达的数据包 P，查找连接表 Table，并更新对应的连接内容。

```
1. 收到数据包 P 后，提取关键字段，诸如源/目标 IP SIP/DIP 及源/目标端口 SPT/DPT;
2. 用 P.SIP、P.DIP、P.SPT 及 P.DPT 计算 P 的连接标示符 CID;
3. 用 hash 函数计算 P.CID 的 hash 值 H;
4. For each 连接表 Table[H]中的连接 C;
5.    if C.CID 匹配 P.CID，then
6.       释放连接表 Table[H]中的连接 C
7.       break;
8. if 没有发现 C，then
9.    if P 开始一个新的连接，then
10.         分配一个新的连接 C;
11.   else
12.         终止数据包 P and exit;
13. 用 P 更新 C;
14. if C 被终止，then
15.   删除 C;
16. else
17.   在连接表 Table[H]的链头位置插入 C;
```

定义 1(连接端标识符 CEI) 若以 IP 代表网络层首部中的 IP 地址，PT 代表传输层首部中的端口号，则二元组 CEI(IP, PT)称为连接端标识符。

定义 2(CEI 的大小关系) 设连接的 2 个连接端为：CEI_1(IP, PT)和 CEI_2(IP, PT)，$CEI_1<CEI_2$，当且仅当：$CEI_1.IP<CEI_2.IP$ 或者 $CEI_1.IP=CEI_2.IP$ 且 $CEI_1.PT<CEI_2.PT$。若有 $CEI_1<CEI_2$，则称 CEI_1 为较小端，CEI_2 为较大端。

定义 3(连接标识符 CID) 设 CEI_S 和 CEI_B 分别代表一个给定连接的较小端和较大端，则二元组 $CID(CEI_S,CEI_B)$称为连接标识符。

4　实验结果与分析

从上述算法流程可以看出：连接管理性能主要取决于连接表的查找操作。因此，实验对改进的 MTF 连接表和传统的排序连接表进行性能对比。传统的排序连接表采用链式排序法解决哈希冲突，即将每个表项中的所有连接链接起来，并按照连接标识符进行排序。实验从测试数据集中逐个读取网络数据包，统计每个数据包查找连接表时的查找长度，每读取预设数量的数据包时计算平均查找长度。在统计查找长度时，由于 2 种连接表查找失败时的查找长度相同，因此只统计查找成功时的情况。

由于大多数公布的数据流样本都做了 IP 地址匿名化处理，不满足本研究工作的需求，因此实验选用江苏省计算机网络技术重点实验室发布的 NENC-20070101 作为测试数据集。该数据集利用专用采集平台从一条 3 Gb/s 链路上捕获而得，采集时间为 1 min。该数据集中的 TCP 包有 16.5×10^6 个左右，TCP 连接有 363 195 条，其中，92 631 条正常结束。数据集中的并发 TCP 连接数基本上以 3 000 个/s 的速度线性增长，最终数量达 170 000 个左右。

图 2 给出 2 种连接表的平均查找长度(哈希表长度为 2^{16})。从中可以看出，MTF 连接表的平均查找长度比排序连接表小。随着读入数据包的增加和并发连接数的增长，排序连接表的平均查找长度波动幅度较大，整体上逐步增长，基本上一直都远大于 MTF 连接表的平均查找长度。而 MTF 连接表的平均查找长度始终比较稳定，基本上能够保持在 1.5 以内。

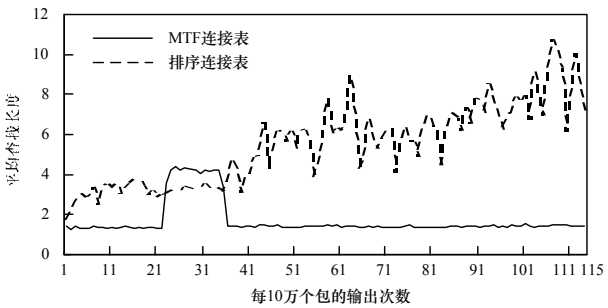


图 2　2 种连接表的平均查找长度

(下转第 90 页)