流分类技术的研究

张 李，涂晓东，何 诚

(电子科技大学 宽带光纤传输与通信网技术教育部重点实验室 成都 610054)

【摘要】介绍了流分类算法的概念以及对流分类算法的要求；把目前存在的流分类算法分成三类：多维查找转换为一维查找算法、相关区域查找算法、独立区域查找算法，并对各类算法的性能进行了讨论；通过引入并行流分类算法说明了流分类算法的研究重点是减小存储空间和提高更新速度。

关 键 词 流分类；多维查找转换为一维查找算法；相关区域查找算法；独立区域查找算法；并行流分类

中图分类号 TP393 **文献标识码** A

Packet Classification Algorithm of Switch

Zhang Li，Tu Xiaodong，He Cheng

(Key Laboratory of Brodband Optical Fiber Transmission and Communication Networks UEST of China, Ministry of Education Chengdu 610054)

Abstract The conception of packet classification and request for it is introduced. The paper classifies the existed packet classification algorithms into three kinds: conversion into single-field search、dependent field search、independent field search, then compare their performance. The presentation of parallel packet classification indicates how to reduce the storage space and update time is the research importance.

Key words packet classification; conversion into single-field search; dependent field search; independent field search; parallel packet classification

随着英特网的发展，流分类技术得到广泛地应用，如防火墙、基于策略的路由、流量控制、流量整形、计费等等。与流分类有关的策略和规则的集合称为流分类器或简称分类器。分类器中的每一条规则根据数据包头的某些域来定义数据包所属的流。常用的几个域有：目的IP地址、源IP地址、源端口、目的端口、协议类型等。

1 流分类算法的定义和要求

给出流分类的正式定义：分类器有N条规则$\{R_j,1\leqslant j\leqslant N\}$，规则$R_j$由3部分组成：1) 正则表达式$R_j[i]$，$1\leqslant i\leqslant k$；2) 优先级$pri(R_j)$；3) 执行的操作$action(R_j)$。对接收到的每个数据包，可以看作是$k$维空间中的一个点$(P_1,P_2,\cdots,P_k)$。$k$维流分类问题就是要在所有规则中找到与点$(P_1,P_2,\cdots,P_k)$相匹配，并且优先级最高的规则$R_m$，即$pri(R_m)>pri(R_j)$，$\forall j\neq m$，$1\leqslant m\leqslant N$，$1\leqslant j\leqslant N$，且$\forall(1\leqslant i\leqslant k)$，$P_i$与$P_j[i]$相匹配，则称$R_m$是最佳匹配规则。

对流分类算法的要求可以概括为以下5点：1) 处理速度快。目前物理链路速度有了很大的提高，对网络处理器提出了更高的要求，而流分类算法的复杂性使得它成为了限制路由器和交换机处理速度的瓶颈。2) 所需存储容量低。在存储容量要求低的情况下，可采用速度快但是造价昂贵的存储器技术。3) 更新时间短。由于网络上各种数据流量巨大，流分类算法在更新数据结构时必须快，才能适应分类器的频繁变化。4) 规则的定义灵活。好的流分类算法应该支持各种不同形式的规则，包括数据包头各个区域的精确匹配，前缀匹配和范围匹配。5) 规则具有可扩展性。分类器必须能够对分类所依据的域的数目、域的长度以及规则数目进行扩展，应付以后网络的需要，特别是IPv4～IPv6的过渡。

收稿日期：2004－07－01

作者简介：张 李(1980－)，男，硕士生，主要从事宽带IP网络中流分类技术方面的研究；涂晓东(1970－)，男，博士，副教授，主要从事宽带IP网络、光网络、存储网络方面的研究；何 诚(1980－)，男，硕士生，主要从事宽带IP网络中流分类技术方面的研究.

2 流分类算法的分类及比较

流分类算法可以根据不同的原则进行分类，本文根据对多个区域查找之间的关系把现有的流分类算法分成3类。1) N表示规则数目，2) d表示维数，3) w表示每一维的宽度。

2.1 多维查找转换为一维查找的算法

该类算法的特点就是把流分类查找的各个区域连接起来组成一个查找关键字，从而把多维查找的问题转化为一维查找问题。在基于哈希表进行查找的流分类器以及使用TCAM的流分类器中经常使用到这种方法。缺点是得到的查找关键字很大，而且由于多维合成了一维，无法利用规则中各维内部的共性进行优化设计。此类算法的代表是多元空间查找算法(Tuple-Space-Search)[1]，基本思想是把流分类查找问题分解为多个精确匹配问题。首先把d维的规则映射到一个具有d个分量的空间中，空间的第i个分量表明规则第i维的前缀长度，这样各维前缀长度都相同的规则就对应同一个空间，把它们存储在同一个哈希表中。查找时对所有哈希表进行精确匹配查找。这种算法的空间复杂度为$O(N)$，时间复杂度由需要访问的哈希表数目决定，使得查找的时间复杂度不能确定。在最糟的情况下，空间数目可能达到$O(w^d)$，使得查找时间不能接受。更新速度很快，只需要一次哈希访问的时间。实际上许多知名硬件厂商在设计自己的流分类实现方案时都采用了哈希技术。

2.2 相关区域查找算法

这种算法的特点是前一个区域的查找结果会影响到随后要查找的区域的路径，主要优点是可以使用相对简单的生成树结构，缺点是要想达到较快的查找速度就需要对树的结构进行复制或者在数据中加入链路信息，这样就会增加存储器容量，也会使更新更慢。另外大量对存储器的访问都是互相依赖的，导致了不可预测的延迟。此类算法的代表是分层查找树(Hierarchical Tries)，集合归并查找树(Set-Pruning tries)[2]，查找树网格(Grid-of-tries)和智能分层查找树(Hierarchical Intelligent Cuttings，HiCuts)[3]。分层查找树从d维中任取一维生成第一级二叉树，再从剩下的d–1维中任取一维作为第二维，对该二叉树中每一个与规则表中第一维匹配的结点，按照它的第二维建立第二级二叉树，重复上述过程，直到完成每一维的处理。分层查找树的空间复杂度为$O(Ndw)$，时间复杂度为$O(w^d)$。它简单，容易实现，但是查找较慢，并且更新也不快。集合归并查找树是对分层查找树的改进，通过把前缀长度小的结点对应的子树复制到所有前缀长度大于它的结点的子树上，如果某个结点出现规则重复，则取优先级高的规则。这个过程是在所有子树上递推进行的。查找时只需要依次找到所有树上的最长匹配，就能找到相应的规则。时间复杂度为$O(dw)$，空间复杂度为$O(N^d dw)$，可见是以增大存储空间来减少查找时间，扩展性也较差。查找树网格是在分层查找树的基础上，给某些结点增加一个转移指针b(0或1)，它指向另一个子树的一个结点。存在从子树T_y的Y结点到子树T_x的x结点的转移指针的条件是：1) T_x和T_y是同一层上的不同子树，并且指向它们的根结点的指针是同一棵树T上的两个不同的结点(r和s)的下一棵指针。2) 从T_y的根结点到Y再串连上转移指针b的比特串等于从T_x的根结点到x的比特串。3) Y没有等于转移指针b的子结点。4) s是T中离r最近的满足上述条件的父结点。查找树网格避免了由于复制规则导致存储空间扩大和分层查找树的回溯问题。在处理两维流分类问题时，时间复杂度为$O(w)$，空间复杂度为$O(Nw)$，应此它是一种很好的处理两维流分类问题的算法。在处理多维问题时，它也可以用来优化分层查找树的最后两层子树。HiCuts的实现需要建立一种决策树的数据结构，每个叶结点都存储着一些规则，查找时经过决策树找到一个叶结点，再对这个叶结点中的规则进行线性查找找到匹配的规则。决策树的根节点包含了整个d维空间，具体结构可以通过参数来决定。参数binth规定了每个叶结点包含的规则的最大数，当一个结点的规则数大于binth时，就把d维中的某一维平均划分成$NP(C)$份，形成$NP(C)$个子结点。如果结点包含的规则数小于binth，那么该结点就是一个叶结点。HiCuts的时间复杂度为$O(d)$，空间复杂度为$O(N^d)$。可以根据规则的特征调整参数来优化数据结构，降低所需的存储空间，提高查找速度，规则的更新容易实现。缺点是预处理的时间较长，适用于规则较少的情况。

2.3 独立区域查找算法

这类算法首先通过一维查找算法对每个区域单独进行查找，产生中间查找结果，再根据中间结果决定最后多维的查找结果。这样就可以利用各个区域的特点使得查找更加有效,另外存储器的访问是独立的，因此可以并发进行。它的性能很大程度上决定于中间查找结果的编码。代表算法有crossproducting，位向量交集(Bitmap-Intersection)和重复流分类(Recursive Flow Classification，RFC)[4]。Crossproducting先找出各维上不同情况

的组合(每种组合都对应于一条规则)，把它们存储在一个Crossproduct表中，查找是在每个维上单独进行的，最后根据各个维的查找结果对Crossproduct表进行查找，找到相应的规则。时间复杂度为$O(dw)$，空间复杂度为$O(N^d)$。它的缺点在于存储空间要求很大，成级数变化。另外它的更新不方便，每次增加规则都需要重新计算Crossproduct表。位向量交集把所有规则的同一维映射到同一条数轴上，那么每条规则在该数轴上都是一个范围或一个点，给它们都定义一个中间向量，它的位宽为N，对应于N条规则，规则按照优先级排列，根据每条规则在该范围的上的匹配情况给它取值，规则匹配时，向量的对应位取1，否则取0。查找时分别先找到每一维上匹配的中间向量，再把它们进行与运算，找到向量中优先级最高的1位，对应的规则就是匹配的规则。时间复杂度为$O(dw+N/\text{memwidth})$，空间复杂度为$O(dN^2)$。这种算法试图通过增加存储器访问次数来减少所需存储空间，但是效果不明显，而且它并没有解决更新困难的问题。RFC算法在对数据包进行处理时，可以看作是将数据包头部的S比特映射到T比特的类符号的过程。其中$T=\text{lb}N$，$T<<S$，是由N条流分类规则决定的。时间复杂度为$O(d)$，空间复杂度为$O(N^d)$。它具有流分类速度快，能直接支持范围和前缀匹配等优点，局限性在于当规则条数、维数和每维宽度增加时，所需的存储空间太大。同时，由于不同的流分类器具有不同的特征，对于特定的流分类器若该方法所基于的特征不具备或不明显，每一维的长度的压缩量会很小，会严重影响流分类的性能。此外，RFC算法规则的更新非常困难，最糟的情况下需要重建整个数据结构。

3 流分类算法的改进和发展方向

目前对网络处理设备的要求较高，具体到流分类问题上来说就是前面提到的几条要求。上面提到的各种算法均有比较明显的缺点，限制了它们的应用。在文献[5]中介绍了一种并行流分类算法(Parallel Packet Classification，P^{2}C)，通过这种算法，可以看出流分类算法发展的大概方向是要减少存储空间以及提高更新速度。P^{2}C包括3个处理阶段：

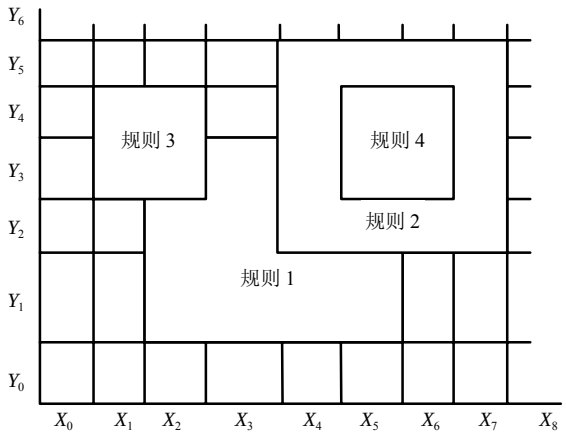


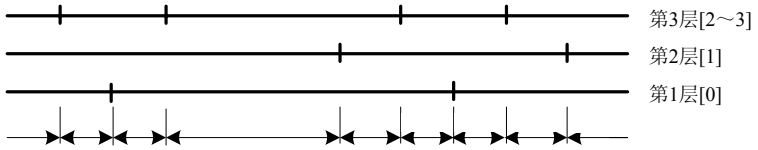
图1 二维流分类所用的规则

1) 预处理阶段。P^{2}C是基于独立区域查找的思想，各个区域的查找是同时进行的。针对Bitmap-Intersection和RFC的缺点，P^{2}C采用了一种新的编码方式来处理中间查找结果，克服了上述两种算法的缺点，减小了所需的存储空间和数据结构之间的相关性，提高了更新速度。下面简要说明，以二维为例，规则如图1所示。

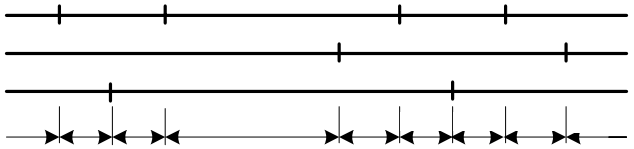
首先对X维进行编码，有三种方式，如图2所示。这三种编码方式具有不同的特点。第一种方法使得数据之间的关联性最小，从而获得最高的更新速率。而第二种和第三种方法可以使得中间查找结果最小，从而需要更小的存储空间。P^{2}C最大的特点就在于这三种编码可以同时对一组规则使用，对每一条规则都可以独立选择。这就使得P^{2}C能够适应广泛的应用和各种环境。其中第一种和第二种编码被认为是标准的方式，可以用来平衡更新速度和存储器的效率。例如对于那些既存在静态规则又有动态规则的防火墙，可以用第二种方法对静态规则进行编码，而用第一种方法对动态规则进行编码。而第三种方法只在极少的情况下使用，通常是在受到很大限制时为了便于执行时采用，例如需要支持很大数量的规则而存储容量上限即将达到时。

2) 中间查找结果产生阶段。在查找时P^{2}C克服了Bitmap-Intersection和RFC的缺点，在后两种方法中与一条规则相关的信息是分布在多个查找结果中的，给更新带来负担。而在P^{2}C中通过一种叫on-the-fly的结构体产生中间查找结果，这种方法不但可以提高存储器效率，而且可以解决更新问题。

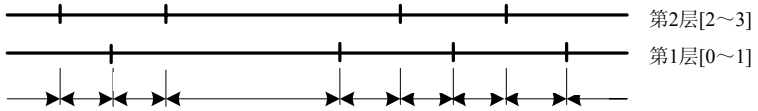
3) 查找匹配规则阶段。P^{2}C规则的查找是在TCAM中进行的，当各个区域的中间查找结果产生后串接起来，对TCAM进行查找，就能找到匹配的规则。在本例中对Y维进行与X维相似的处理后，将中间查找结果串连起来查找TCAM，获得匹配规则。



(a)　P²C编码方式Ⅰ：独立编码



(b)　P²C编码方式Ⅱ：相关编码



(c)　P²C编码方式Ⅲ：减少层数

图2　P²C的编码方式

4　结 束 语

流分类问题是解决QoS、VPN等问题的基础，获得了越来越多的关注，已成为目前宽带IP网络的一个热点问题。只有很好的解决了这个问题，才能使IP网络获得更大的发展空间和应用空间，为广大用户提供更好的服务。

本文研究工作得到中兴通信科研基金资助，在此表示感谢。

参　考　文　献

[1] Srinivasan V, Suri S，Varghese G. Packet classification using tuple space search[J]. Proceedings of ACM Sigcomm Communication Review Archive, 1999, 29(4):135-146

[2] Srinivasan V, Suri S, Varghese G , *et al*. Fast and scalable layer four switching[J]. Proceedings of ACM Sigcomm Communication Review Archive, 1998, 28(4):191-202

[3] Gupta P，McKeown N. Classifying packets with hierarchical intelligent cuttings[J]. IEEE Micro, 2000, 20(1):34-41

[4] Gupta P, McKeown N. Packet classification on multiple fields[J]. ACM SIGCOMM Computer Communication Review Archive, 1999, 29(4): 147-160

[5] van Lunteren J, Engbersen J. Fast and scalable packet classification[J]. IEEE J Selected Areas in Communications, 2003, 21(4):560 -571

编　辑　孙晓丹