

DOI: 10.7652/xjtuxb201310017

利用域转换的三态内容寻址存储器报文分类算法

田乐, 陈庶樵, 黄慧群, 马腾

(国家数字交换系统工程技术研究中心, 450002, 郑州)

摘要: 针对基于三态内容寻址存储器(TCAM)的报文分类存在范围扩张导致空间利用率较低的问题,提出了一种利用域转换的报文分类算法(DTRM)。首先将规则集独立的范围预编码算法中范围规则编码所需的比特数量由 2^k-1 替换为任意值,从而能够利用 TCAM 中的所有冗余位进行编码,实现新范围域的构建,然后利用范围规则的分布特征,以规则集能够表示为较少的 TCAM 表项为原则,设计域转换函数,将规则集原始范围域转换为新构建的范围域。报文分类时,利用域转换函数将报文头部转换为新构建范围域中的某一数值或范围,并与 TCAM 表项进行并行比较,最终得到分类结果。仿真结果表明,与并行报文分类算法相比,DTRM 算法的范围扩张因子由 1.6 减少至 1.21,TCAM 空间利用率由 63%增加至 82%,同时支持规则的增量更新。

关键词: 报文分类;三态内容寻址存储器;范围扩张;冗余位;域转换

中图分类号: TN914.42 **文献标志码:** A **文章编号:** 0253-987X (2013)10-0097-06

A Packet Classification Approach for Ternary Content-Addressable Memory Using Domain Transformation

TIAN Le, CHEN Shuqiao, HUANG Huiqun, MA Teng

(National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

Abstract: A novel algorithm using domain transformation, called DTRM, is proposed to solve the range expansion problem of ternary content-addressable memory (TCAM), in which the rules containing range fields have to be represented by multiple TCAM entries. Firstly, the number of encoding bits of 2^k-1 needed in the database independent range pre-encoding algorithm is replaced by any value so that the extra bits in TCAM entries can be fully used to encode and to construct a new range domain; Secondly, the relationships among rules are used to design a domain transformation function, and the original domain of each field is transformed to a new domain so that the classifier can be expressed by a small number of TCAM entries. When a packet arrives, the packet header is transformed to the new domain with the corresponding approach, and the transformed result is then compared with all the TCAM entries to get match rules. Simulation results and comparisons with the parallel packet classification algorithm show that the proposed algorithm not only reduces range expansion from 1.6 to 1.21 and increases the storage efficiency from 63% to 82%, but also gains better update performance.

Keywords: packet classification; ternary content-addressable memory; range expansion; extra bits; domain transformation

网络服务的日益多样化和不断增加的网络安全需求,要求网络设备能够提供新的处理机制,如基于策略的路由、流量测量、资源预留、拥塞控制、负载均衡等,这些技术的实现都离不开对网络数据报文快速、准确地分类^[1]。

三态内容寻址存储器(ternary content-addressable memory, TCAM)是一种基于内容查找的存储器^[2-3],具有查询速度快、匹配时间固定等优点,目前广泛应用于报文分类领域。TCAM 不直接支持范围规则匹配,通常采用前缀扩展(prefix expansion, PE)的方式将一条范围规则转换为多条 TCAM 表项,从而完成范围规则匹配^[4]。一条位宽为 w 的范围规则,最坏情况需要转换成 $2w-2$ 条 TCAM 表项。如果规则在多个维度上均存在范围扩张,那么 TCAM 表项数量将呈指数级增长,如五元组规则中源端口和目的端口均为范围字段,最坏情况下一条五元组规则需要转换为 $(2 \times 16 - 2)^2 = 900$ 条 TCAM 表项。文献[5]对一些现实应用的规则集进行了统计,发现由于范围规则匹配问题,一条规则平均需要转换成 6.3 条 TCAM 表项,导致较低的 TCAM 空间利用率。由于 TCAM 的容量较小且价格昂贵,每比特价格大约是 SRAM 的 20 倍和 SDRAM 的上百倍^[6],因此减少规则集范围扩张,提高 TCAM 空间利用率具有非常重要的现实意义。

为解决范围规则匹配问题,文献[7]提出了基于格雷码的范围编码算法 SRGE,范围扩张为 $2w-4$ 。PE 和 SRGE 编码后的 TCAM 表项宽度为 w ,目前理论上已经证明,该类编码算法的范围扩张存在极限值 w ,无进一步降低的空间^[8]。由于 TCAM 表项宽度相对固定,因此存储规则后一般会剩余部分比特未使用。为此,文献[9-11]采用 TCAM 中的冗余位对规则集进行编码。文献[9]提出了规则集独立的范围预编码算法 DIRPE,当 TCAM 表项宽度为 2^w-1 时,每个范围规则只需转换为一条 TCAM 表项,由于 TCAM 中的冗余位数量有限,作者采用分段编码的方式,以增加表项数量为代价降低对冗余位数量的要求。文献[10]提出了一种基于比特向量映射的编码方法 Bit-Map,对范围进行分段并且将分段后的范围值映射为索引向量中的某个比特。在 Bit-Map 算法的基础上,文献[11]提出了并行报文分类算法 PPC,采用层次化编码的方式,每个分层里的范围值互不相交,对每个分层进行单独编码。DIRPE、Bit-Map 和 PPC 算法利用 TCAM 中的冗余位进行编码,范围扩张均小于 PE 和 SRGE 算法。

然而,由于上述算法对编码比特数量有严格的要求,无法利用到 TCAM 表项中的每一个冗余位,因此影响了 TCAM 空间利用率的进一步提升。

本文提出了一种能够利用 TCAM 中所有冗余位的编码算法(domain transformation for range match, DTRM),进一步减少了规则集范围扩张,保证了较高的 TCAM 空间利用率。

1 DTRM 算法介绍

本文所提出的 DTRM 算法主要分为域构建和域转换 2 部分。

(1)域构建:在 DIRPE 编码的基础上,通过取消每个比特串编码所需比特数量必须为 2^k-1 的限制,提出了一种支持任意比特数量的编码方法 Smart-DIRPE,利用 Smart-DIRPE 对 TCAM 中的所有冗余位进行编码,构建新的范围域 $D(F)'$ 。

(2)域转换:根据 Smart-DIRPE 编码特点和规则集中范围规则的分布特征,设计域转换函数,实现规则集原始范围域 $D(F)$ 到新构建范围域 $D(F)'$ 的转换,使得规则集能够转换为较少数量的 TCAM 表项,减少范围扩张。

1.1 范围域构建

文献[9]提出了利用 TCAM 冗余位进行编码的 DIRPE 算法,对于位宽为 w 的范围字段,将 w 个比特划分为 m 个较短的比特串分别编码,每个比特串编码结果合并后得到最终编码。每个比特串采用表 1 所示的编码方法,第 i 个比特串编码所需的比特数量为 $(2^{k_i}-1)$,DIRPE 编码后 TCAM 表项宽度为 $w'=(2^{k_1}-1)+(2^{k_2}-1)+\cdots+(2^{k_m}-1)$,其中 $w=k_1+k_2+\cdots+k_m$ 。

表 1 DIRPE 编码方法

范围	$=i$	$\geq i$	$< i$	$[i, j]$
编码	$0^{2^k-i-1}1^i$	$x^{2^k-i-1}1^i$	$0^{2^k-i}x^{i-1}$	$0^{2^k-1-j}x^{j-i}1^i$

由于 DIRPE 编码对比特数量有严格要求,因此编码后通常会有部分比特位未使用,造成了比特位的闲置。为充分利用 TCAM 中的冗余位,本文通过取消每个比特串编码所需比特数必须为 2^k-1 的限制,提出了一种支持任意比特数量的编码算法 Smart-DIRPE。假设当前可用的比特数量为 d ,将 d 个比特划分为 m 个比特串,第 i 个比特串包含的比特数量为 d_i-1 ,并且 $d=\sum d_i-m$,得到编码向量 $c=(d_1, d_2, \cdots, d_m)$;采用表 1 所示编码方法并用 d_i-1 替换 2^k-1 ,最后将 m 个比特串编码结果合并

得到最终编码结果,记为 $sd(m, \mathbf{c})$, 利用该编码构建新的范围域 $D(F, \mathbf{c})$, $D(F, \mathbf{c}) = [0, \prod_{i=1}^m c(i) - 1]$ 。

$sd(m, \mathbf{c})$ 编码可以表示为一个深度为 $m+1$ 、第 i ($1 \leq i \leq m$) 层节点的度均为 $c(i) = d_i$ 的树形结构, 称为 $sd(m, \mathbf{c})$ 树。定义 $D(\mathbf{v})$ 为节点 $\mathbf{v}$ 表示的范围, $E(\mathbf{v})$ 为 $D(\mathbf{v})$ 的编码结果, 以图 1 所示的 $sd(2, 2, 5)$ 树为例, $sd(m, \mathbf{c})$ 树具有以下性质。

(1) $sd(m, \mathbf{c})$ 树的根节点 v_r 表示的范围为 $D(F, \mathbf{c})$, 编码为 $E(v_r) = \{x\}^d$ 。如 $sd(2, 2, 5)$ 树的根节点表示的范围为 $D(v_r) = [0, 2 \times 5 - 1] = [0, 9]$, 编码为 $E(v_r) = xxxxx$ 。

(2) $sd(m, \mathbf{c})$ 树为有序树, 第 i 层共有 $n = d_0 d_1 \cdots d_{i-1}$ ($d_0 = 1$) 个节点, 从左至右依次标记为 v_{ij} ($1 \leq j \leq n$), 节点表示的范围为 $D(v_{ij}) = [d_{im}(j-1), d_{im}j - 1]$, 其中 d_{im} 满足 $d_{im} = d_i d_{i+1} \cdots d_m$ 。如 $sd(2, 2, 5)$ 树第 2 层有 $n = d_0 d_1 = 2$ 个节点, 从左至右依次为 v_{21} 和 v_{22} , $D(v_{21}) = [0, 4]$, $D(v_{22}) = [5, 9]$ 。

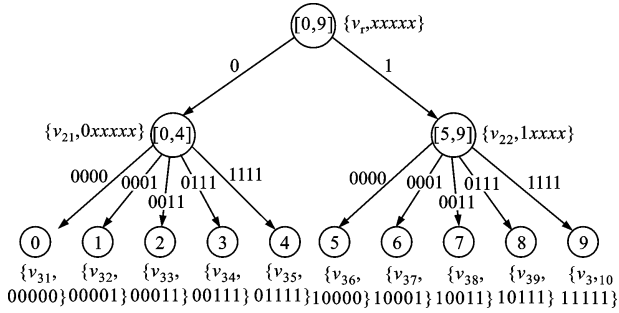


图 1 $sd(2, 2, 5)$ 树示意图

(3) 非叶子节点 v_{ij} 有 d_i 个孩子节点, 记为 $ch(v_{ij}, t)$ ($1 \leq t \leq d_i$), 非叶子节点 v_{ij} 到其孩子节点 $ch(v_{ij}, t)$ 的边 $e_{i,t}: v_{ij} \rightarrow ch(v_{ij}, t)$ 的编码为 $E(e_{i,t}) = \{0\}^{d_i-t} \{1\}^{t-1}$ 。非叶子如节点 v_{22} 有 $d_2 = 5$ 个孩子节点, v_{22} 到孩子节点 $ch(v_{ij}, 4) = v_{39}$ 的边 $e_{2,4}: v_{22} \rightarrow ch(v_{22}, 4)$ 的编码为 $E(e_{2,4}) = 0111$ 。

(4) 根节点 v_r 到节点 v_{ij} 有唯一路径 $path(v_{ij})$, $path(v_{ij})$ 由自上而下的 $(i-1)$ 条边连接而成。叶子节点编码为 $E(path(v_{ij}))$, 非叶子节点 v_{ij} 编码为 $E(path(v_{ij}))\{x\}^{c_{i,m}}, c_{i,m} = d_i + d_{i+1} + \cdots + d_m$ 。如叶子节点 v_{39} 编码为 $E(v_{39}) = 10111$, 非叶子节点 v_{22} 的编码为 $E(v_{22}) = 1xxxx$ 。

1.2 域转换

定义 1 域转换函数 T_c 表示域 $D(F)$ 到域 $D(F, \mathbf{v})'$ 的一种映射关系, 域转换函数 T_c 满足以下条件: $\forall x, y \in D(F), T_c(x), T_c(y) \in D(F, \mathbf{c})'$, 如果 $x < y$, 那么 $\max T_c(x) < \min T_c(y)$, 如果 $y = x + 1$, 那么

$$\max T_c(x) + 1 = \min T_c(y)。$$

定理 1 $\forall a, b, x \in D(F)$, 且 $x \in [a, b]$, 那么 $T_c(x) \subseteq [\min T_c(a), \max T_c(b)]$ 。

定理 2 规则集 $S \subseteq D(F)$, $\forall [a, b] \subseteq D(F)$, 利用转换函数 T_c 将范围 $[a, b]$ 转换为 $[\min T_c(a), \max T_c(b)]$, 得到等价规则集 $S' \subseteq D(F, \mathbf{c})'$ 。对任意报文 P , 规则集 S 与 S' 有相同的规则决策。

本文根据规则集内部范围规则的分布特征, 将规则集原始范围域分割为互不重叠的范围, 并转换为 $sd(m, \mathbf{c})$ 树的节点, 以实现规则集原始范围域 $D(F)$ 到新构建范围域 $D(F, \mathbf{c})'$ 的转换。以图 2 所示的域转换为例, 具体描述如下。

(1) $D(F)$ 转换为 $sd(m, \mathbf{c})$ 树的根节点 v_r , 如 $D(F) = [0, 7]$ 转换为 $sd(2, 2, 5)$ 树的根节点 $D(F, \mathbf{c})' = [0, 9]$ 。

(2) 范围 $[a, b] \subseteq D(F)$ 转换为 $sd(m, \mathbf{v})$ 树的节点 v_{ij} , 节点 v_{ij} 有 d_i 个孩子节点, 如果 $d_i < (b-a+1)$, 选择 d_i-1 个分割点 $a \leq x_1 < x_2 < \cdots < x_{d_i-1} < b$, 将范围 $[a, b]$ 分割为 $[a, x_1], [x_1+1, x_2], \cdots, [x_{d_i-1}+1, b]$ 共 d_i 个子范围, 并分别转换为节点 v_{ij} 的 d_i 个孩子节点 $ch(v_{ij}, t)$, $1 \leq t \leq d_i$ 。如选择分割点 $x = 4$, 将范围 $[0, 7]$ 分割为 $[0, 4]$ 和 $[5, 7]$ 共 2 个子范围, 分别转换为节点 v_r 的孩子节点 v_{21} 和 v_{22} 。

(3) 范围 $[a, b] \subseteq D(F)$ 转换为 $sd(m, \mathbf{v})$ 树的节点 v_{ij} , 节点 v_{ij} 有 d_i 个孩子节点, 如果 $d_i \geq (b-a+1)$, 选择 $(b-a)$ 个分割点 $1 \leq x_1 < x_2 < \cdots < x_{b-a} < d_i$, 将 d_i 个孩子节点依次划分为 $\{ch(v_{ij}, 1), \cdots, ch(v_{ij}, x_1)\}, \{ch(v_{ij}, x_1+1), \cdots, ch(v_{ij}, x_2)\}, \cdots, \{ch(v_{ij}, x_{b-a}+1), \cdots, ch(v_{ij}, d_i)\}$ 共 $(b-a+1)$ 个孩子节点集合, 节点 v_{ij} 与孩子节点集合 $\{ch(v_{ij}, x_i+1), \cdots, ch(v_{ij}, x_{i+1})\}$ 的 $(x_{i+1}-x_i)$ 条边合并, 并且编码为 $\{0\}^{d_i-x_{i+1}} \{x\}^{x_{i+1}-x_i-1} \{1\}^{x_i}$ 。范围 $[a, b]$ 的 $(b-a+1)$ 个数值分别转换为 $(b-a+1)$ 个孩子节点集合。如选择分割点 $x_1 = 2$ 和 $x_2 = 4$, 将节点 v_{22} 的 $d_i = 5$ 个孩子节点依次划分为 $\{v_{36}, v_{37}\}, \{v_{38}, v_{39}\}$ 和

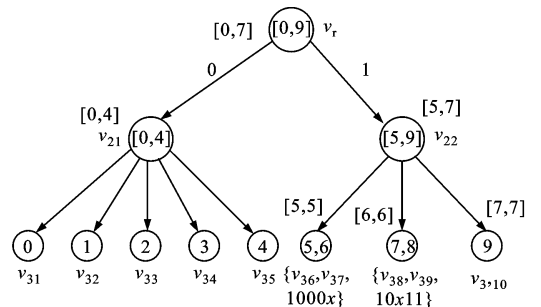


图 2 域 $D(F) = [0, 7]$ 到 $D(f, 2, 5)' = [0, 9]$ 转换示意图

$v_{3,10}$ 共 3 个孩子节点集合, 节点集合 $\{v_{36}, v_{37}\}$ 和 $\{v_{38}, v_{39}\}$ 分别编码为 1000 x 和 10 x 11, 范围 $[5, 7]$ 的 3 个数值 $\{5, 6, 7\}$ 分别转换为这 3 个孩子节点集合。

(4) 重复步骤 2、3, 直到完成对所有 $x \in D(F)$ 的转换。

定义 2 $S \cap [l, r]$ 表示规则集 S 在 $[l, r]$ 上的投影规则集合, $\forall R = [a, b] \in S$, 规则 R 在 $[l, r]$ 上的投影规则为 $R_g \in S \cap [l, r]$, 当 $R = [a, b] \cap [l, r] \neq \emptyset$ 时, $R_g = [\max(l, a), \min(r, b)]$, 否则 $R_g = \emptyset$ 。

定义 3 域转换函数 T_c 将规则集 $S \cap [l, r]$ 转换为等价规则集 $(S \cap [l, r])'$, 如果 $(S \cap [l, r])'$ 有最少的 TCAM 表项, 那么称该域转换函数 T_c 为规则集 $S \cap [l, r]$ 的最优域转换函数, 记为 $PA(l, r, c)$ 。

定理 3 范围 $[l, r] \subseteq D(F)$ 经最优转换函数 $PA(l, r, c_i)$ 转换为 $sd(m, c)$ 树的节点 v_{ij} , 记规则集 $(S \cap [l, r])'$ 的 TCAM 表项数量为代价函数 $\text{cost}(l, r, c_i)$, 其中 $c = (d_1, d_2, \dots, d_m)$, $c_i = (d_1, d_{i+1}, \dots, d_m)$, 代价函数 $\text{cost}(l, r, c_i)$ 满足如下条件。

(1) 当 $(l - r + 1) > d_i$ 时, $\text{cost}(l, r, c_i) = \min\{\text{cost}(l, x_1, c_{i+1}) + \text{cost}(x_1 + 1, x_2, c_{i+1}) + \dots + \text{cost}(x_{d_{i-1}} + 1, r, c_{i+1}) - \sum |S \odot [x_i + 1, x_j]| (j - i - 1)\}$, 其中 $\{i, j\} \in [0, 1, \dots, d_i]$, $j > i + 1$, $l \leq x_1 < \dots < x_{d_{i-1}} < x_{d_i} = r$, $S \odot [x_i + 1, x_j]$ 表示范围为 S_{ij} 的规则集合, $[x_i + 1, x_j] \subseteq S_{ij} \subset [x_{i-1} + 1, x_{j+1}]$ 。

(2) 当 $(r - l + 1) \leq d_i$ 时, $\text{cost}(l, r, c_i) = |s \cap [l, r]|$ 。若比特串向量为空向量 $c_\emptyset$, 则 $\text{cost}(l, r, c_\emptyset) = \infty$ 。

定理 3 证明如下。

(1) 当 $(l - r + 1) > d_i$ 时, 范围 $[l, r]$ 被分割为 $[l, x_1], [x_1 + 1, x_2], \dots, [x_{d_{i-1}} + 1, r]$ 共 d_i 个子范围, 并分别转换为节点 v_{ij} 的 d_i 个孩子节点。由 $sd(m, c)$ 编码特点可知, $PA(l, r, c_i)$ 可以通过最优转换函数 $PA(l, x_1, c_{i+1}), PA(x_1 + 1, x_2, c_{i+1}), \dots, PA(x_{d_{i-1}} + 1, 2^w - 1, c_{i+1})$ 得到最优解, 分别在 $PA(x_i + 1, x_{i+1}, c_{i+1})$ ($0 \leq i \leq d_i$) 的编码结果前加上前缀 $\{0\}^{b_i - i} \{1\}^i$ 即可得到 $PA(l, r, c_i)$ 的编码结果。如果 $[x_i + 1, x_j] \subseteq R \subset [x_{i-1} + 1, x_{j+1}]$, 且 $\forall R \subseteq (S \cap [l, r])$, 那么等价规则 R' 在 $[x_i + 1, x_{i+1}]$ ($i - 1 \leq t \leq j$) 中参与计数, 规则 R' 表示为 $\text{cost}(x_{i-1} + 1, x_i, c_{i+1}) + \text{cost}(x_j + 1, x_{j+1}, c_{i+1}) + (j - i)$ 条表项。根据表 1 所示编码方法, 规则 R' 实际上只需要表示为 $\text{cost}(x_{i-1} + 1, x_i, c_{i+1}) + \text{cost}(x_j + 1, x_{j+1}, c_{i+1}) + 1$ 条表项, 因此规则 R' 的表项数量多记了 $(j - i - 1)$ 。

(2) 当 $(r - l + 1) \leq d_i$ 时, 范围 $[l, r]$ 的 $(r - l + 1)$ 个数值分别转换为节点 v_{ij} 的 $(r - l + 1)$ 个孩子节点集合, 由 $sd(m, c)$ 编码特点可知, $\forall R \subseteq (S \cap [l, r])$, 等价规则 R' 只需要表示为一条 TCAM 表项, 因此 $\text{cost}(l, r, c) = |S \cap [l, r]|$ 。

1.3 多维域转换

多维规则的范围扩张为各维度范围扩张的乘积, 为使规则集 S 有最小的范围扩张, 必须充分考虑维度 $(F_1, F_2, \dots, F_d)$ 的范围扩张对规则集 S 范围扩张的影响。本文在一维域转换方法的基础上, 提出了基于一维权重的多维规则集域转换方法, 通过对规则加权求得近似最优解。

定义 4 $\forall R \in S$, 规则 R 分别在维度 $(F_1, F_2, \dots, F_d)$ 上范围扩张后的 TCAM 表项数量为 $n_i(R)$ ($1 \leq i \leq d$), 记规则 R 在维度 F_p 上的权重为 $W_p(R) = n_1(R) \cdots n_{p-1}(R) n_{p+1}(R) \cdots n_d(R)$ ($1 \leq p \leq d$)。

定义 5 范围 $[l, r] \subseteq D(F_p)$, 如果维度 F_p 上的域转换函数 $T_{p,c}$ 使得规则集 $S \cap [l, r]$ 的等价规则集 $(S \cap [l, r])'$ 在维度 F_p 上的权重之和最小, 那么称该域转换函数为规则集 $S \cap [l, r]$ 在维度 F_p 上的最优域转换函数, 记为 $PA_p(l, r, c)$ 。

定理 4 范围 $[l, r] \subseteq D(F_p)$, 规则集 $S \cap [l, r]$ 经维度 F_p 上的最优转换函数 $PA_p(l, r, c)$ 转换为 $sd(m, c)$ 树的节点 v_{ij} , 记等价规则集 $(S \cap [l, r])'$ 在维度 F_p 上的权重之和为维度 F_p 上的代价函数 $\text{cost}_p(l, r, c_i)$, 其中 $c = (d_1, d_2, \dots, d_m)$, $c_i = (d_i, d_{i+1}, \dots, d_m)$, 代价函数 $\text{cost}_p(l, r, c_i)$ 计算如下。

(1) 当 $(l - r + 1) > d_i$ 时, $\text{cost}_p(l, r, c_i) = \min\{\text{cost}_p(l, x_1, c_{i+1}) + \text{cost}_p(x_1 + 1, x_2, c_{i+1}) + \dots + \text{cost}_p(x_{d_{i-1}} + 1, r, c_{i+1}) - \sum w_p(r_{ij})(j - i - q)\}$, 其中 $\{i, j\} \in [0, 1, \dots, d_i]$, $j > i + 1$, $l \leq x_1 < \dots < x_{d_{i-1}} < x_{d_i} = r$, $r_{ij} \in S \odot [x_i + 1, x_j]$, $S \odot [x_i + 1, x_j]$ 表示规则集 S 中范围为 S_{ij} 的规则集合, $[x_i + 1, x_j] \subseteq S_{ij} \cap [x_{i-1} + 1, x_{j+1}]$ 。

(2) $(r - l + 1) \leq d_i$ 时, $\text{cost}_p(l, r, c_i) = \sum w_p(r), r \in |s \cap [l, r]|$ 。

如果比特串向量为空向量 $c_\emptyset$, 则 $\text{cost}_p(l, r, c_\emptyset) = \infty$ 。

1.4 规则更新

对于利用规则集内部特征进行编码的算法如 Bit-Map 和 PPC, 当规则更新时, 如果加入的规则包含新的范围值, 则需要对规则集重新编码, 但对于 DTRM 算法, 由域转换定义可知, 算法对每一个数

值均进行了编码,因此当有新的规则添加时,只需要对规则进行单独编码,就可支持规则的增量更新。只有当随着规则的不断添入和删除造成编码效率低下时,才需要对规则集重新编码。

2 实验仿真

实验采用主频为 2.8 GHz 的 Intel(R)Pentium (R)4 CPU,内存为 2.0 GB,操作系统为 Windows XP,并采用 VMware Workstation 7.0 搭建环境,测试规则集由 Classbench^[12]生成,包括 3 类(ACL1、FW1、IPC1)共 9 个规则集,规则集容量分别为 10^3 个(1 K)、 5×10^3 个(5 K)和 10×10^3 个(10 K)。

设置 TCAM 表项宽度为 144 bit,对上述 9 个规则集分别采用 PE、SRGE、DIRPE、Bit-Map、PPC 以及 DTRM 算法进行范围编码,统计编码后规则集的 TCAM 表项数量。为尽可能减少范围扩张,根据 DIRPE 的范围扩张与编码向量之间的关系^[7],DIRPE 编码中源、目的端口均采用 {2,2,3,3,3,3} 编码;对编码算法 Bit-Map 和 PPC,根据规则集内部范围值及其互相关系,确定 TCAM 表项中的源、目的端口宽度。

表 2 列出了规则容量分别为 1 K 和 10 K 时的 TCAM 表项数量,可以看出,对于同一规则集,PE 和 SRGE 算法的 TCAM 表项数量远大于规则数量,造成了 TCAM 存储空间的浪费,DIRPE、Bit-Map 以及 PPC 算法的 TCAM 表项数量次之,DTRM 算法的 TCAM 表项数量接近规则数量,具有最小的范围扩张。

规则集	TCAM 表项数量					
	PE	SRGE	DIRPE	Bit-Map	PPC	DTRM
FW1_1K	2 610	2 531	1 470	1 648	1 403	900
ACL1_1K	1 227	1 218	1 071	1 064	1 026	920
IPC1_1K	1 560	1 485	1 303	1 207	1 167	973
FW1_10K	30 727	29 796	23 781	19 643	14 645	10 367
ACL1_10K	12 812	11 715	11 043	11 505	10 978	9 645
IPC1_10K	15 104	12 381	10 573	12 463	11 987	9 215

规则集中范围规则约占规则总数的 10%,比较范围规则的范围扩张,能够更为真实地反映出 DTRM 算法的编码性能。表 3 列出了规则容量大约为 5 000 时范围规则的编码结果。FW1_5K 的范围规则扩张因子从 PE 算法的 24.52 降低为 3.15,两者相差 8 倍之多,即使相比于 PPC 算法的 7.16,

范围扩张因子也减少了近 50%;ACL1_5K 规则集的范围规则扩张因子从 PE 算法的 3.42 降低为 1.26,两者相差约 2.7 倍,相比于 PPC 算法的 2.47,范围扩张因子减少了近 50%。由此可见,由于 DTRM 充分挖掘并利用了规则集内部特征和 TCAM 表项中的冗余位,相比其他范围编码算法,DTRM 具有更小的范围扩张因子。

表 3 6 种算法的范围编码结果					
编码方法	规则集	位宽/bit		表项数量	范围扩张因子
		源端口	目的端口		
PE	ACL1_5K	16	16	1 435	3.42
	FW1_5K	16	16	11 156	24.52
	IPC1_5K	16	16	3 367	9.35
SRGE	ACL1_5K	16	16	1 391	3.31
	FW1_5K	16	16	10 691	23.50
	IPC1_5K	16	16	3 011	8.36
DIRPE	ACL1_5K	16	43	1 082	2.58
	FW1_5K	34	34	5 688	12.50
	IPC1_5K	34	34	2 118	5.88
Bit-Map	ACL1_5K	1	33	1 219	2.92
	FW1_5K	20	33	5 672	12.47
	IPC1_5K	16	25	1 886	5.24
PPC	ACL1_5K	1	21	1 011	2.47
	FW1_5K	12	23	3 258	7.16
	IPC1_5K	11	18	1 324	3.68
DTRM	ACL1_5K	16	56	641	1.26
	FW1_5K	34	38	1 432	3.15
	IPC1_5K	36	36	939	2.60

逐次增加冗余位数量,对规则集 FW1_5K 分别进行 DTRM、Bit-Map 及 PPC 编码,统计不同冗余位数量下的 TCAM 表项数量,计算 TCAM 表项与实际规则数量的比值,得到范围扩张因子。当冗余位数量为 0 时,Bit-Map、PPC 及 DTRM 编码均等价为首级编码,范围扩张因子均为 3.42。图 3 为范围扩张随冗余位数量的变化情况,可以看出,随着冗余位数量的增加,TCAM 表项数量不断减小,在相同冗余位数量下,DTRM 的范围扩张最小。如当冗余位数量增加到 70 时,Bit-Map、PPC 及 DTRM 编码的范围扩张因子分别为 2.2、1.6 及 1.22。

为进一步理解 DTRM 对范围扩张的影响,在冗余位数为 60 时,对规则集 FW1_10K 分别进行 DTRM 和 Bit-Map 编码,统计具有不同范围扩张因子的规则占规则总数的比例 α ,得到规则集范围扩张分布情况,如图 4 所示。从图 4 可以看出,相对于

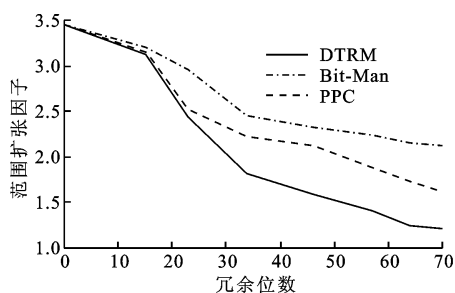


图3 范围扩张随冗余位数的变化情况

Bit-Map 编码, DTRM 的范围扩张分布向左移动, 即范围扩张较大的规则数量相对减少, 而范围扩张较小的规则数量有所增加。由此可以说明, 与 Bit-Map 算法相比, 采用 DTRM 编码后规则集 FW1_10K 有更小的范围扩张, FW1_10K 规则集可以表示为数量更少的 TCAM 表项。

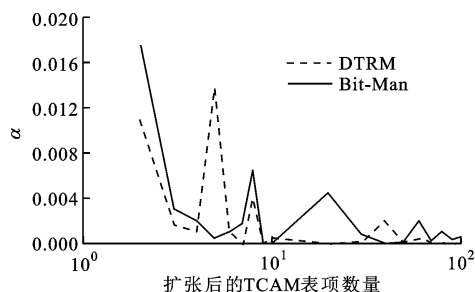


图4 FW1-10K 范围扩张分布情况

3 结 论

本文利用 TCAM 表项中的冗余位和规则集内部特征, 提出了一种利用域转换的范围规则匹配方法 DTRM。首先针对 DIRPE 算法对 TCAM 冗余位数量有严格限制的不足, 提出了一种支持任意比特数量的编码算法 Smart-DIRPE; 根据 Smart-DIRPE 编码和规则间的相互关系, 设计转换函数实现对不同范围值的扩展或者移动, 减少规则集的范围扩张。本算法能够有效减少范围扩张, 提高 TCAM 空间利用率, 且支持规则的增量更新。本文的下一步工作是找到针对多维规则集的最优域转换方法, 或者证明这是一个 N-P 问题^[13]。

参考文献:

- [1] KOMPILLA R R, LEVCHENKO K, SNOEREN A C. Every microsecond counts: tracking fine-grain latencies with a lossy difference aggregator [C]// Proceedings of ACM SIGCOMM. New York, USA: ACM, 2009: 255-266.
- [2] NETLOGIC. Knowledge-based processors datasheet [EB/

OL]. [2013-02-24]. <http://www.broadcom.com/products/Knowledge-Based-Process-ors/Layers-2-4/NL91024.html>.

- [3] IDT. 7 Network search engine datasheet [EB/OL]. [2009-08-20]. <http://www.idt.com/products/Knowledge-Based-Processors/75K72100.html>.
- [4] SRINIVASAN V, VARGHESE G, SURI S. Fast and scalable layer for switching [C]// Proceedings of ACM SIGCOMM. New York, ACM, 1998: 191-202.
- [5] SPITZNAGEL E, TAYLOR D, TURNER J. Packet classification using extended TCAMs [C]// Proceedings of 11th IEEE International Conference on Network Protocols. Piscataway, NJ, USA: IEEE, 2003: 120-131.
- [6] 朱国胜, 余少华. 基于 TCAM 的范围匹配方法: C-TCAM [J]. 通信学报, 2012(1): 31-37.
ZHU Guosheng, YU Shaohua. Range matching method based on TCAM: C-TCAM [J]. Journal on Communications, 2012(1): 31-37.
- [7] BREMLER-BARR A, HENDLER D. Space efficient TCAM-based classification using gray coding [J]. IEEE Transactions on Computers, 2012, 61(1): 18-30.
- [8] ROTTENSTREICH O, KESLASSY I. Worst-case TCAM rule expansion [C]// Proceedings of INFOCOM. Piscataway, NJ, USA: IEEE, 2010: 1-5.
- [9] LAKSHMINARAYANAN K, RANGARAJAN A, VENKATACHARY S. Algorithms for advanced packet classification with ternary CAMs [C]// Proceedings of ACM SIGCOMM. New York, USA: ACM, 2005: 193-204.
- [10] LIU Huan. Efficient mapping of range classifier into Ternary-CAM [C]// Proceedings of 10th Symposium on High Performance Interconnects. Piscataway, NJ, USA: IEEE, 2002: 95-100.
- [11] BREMLER-BARR A, HAY D, HENDLER D. Layered interval codes for TCAM-based classification [C]// Proceedings of INFOCOM. Piscataway, NJ, USA: IEEE, 2009: 1305-1313.
- [12] TAYLOR D E, TURNER J S. ClassBench: a packet classification benchmark [J]. IEEE/ACM Transactions on Networking, 2007, 15(3): 499-511.
- [13] DAVID S J, JEFFREY D. A worst-case performance bounds for simple one-dimensional packing algorithms [J]. SIAM Journal on Computing, 1974, 3(4): 487-494.

(编辑 刘杨)