



(12) 发明专利申请

(10) 申请公布号 CN 103986667 A

(43) 申请公布日 2014. 08. 13

(21) 申请号 201410190571. 6

(22) 申请日 2014. 05. 07

(71) 申请人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为
总部办公楼

(72) 发明人 贺鹏 丁万夫 林程勇 谢高岗
关洪涛 黄昆

(74) 专利代理机构 北京龙双利达知识产权代理
有限公司 11329

代理人 王君 肖鹂

(51) Int. Cl.

H04L 12/861 (2013. 01)

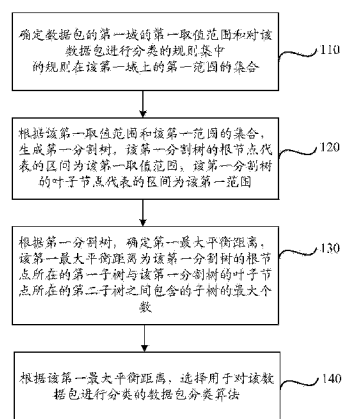
权利要求书4页 说明书27页 附图7页

(54) 发明名称

选择数据包分类算法的方法和装置

(57) 摘要

本发明公开了一种选择数据包分类算法的方法和装置。该方法包括：确定数据包的第一域的第一取值范围和对数据包进行分类的规则集中的规则在该第一域上的第一范围的集合；根据第一取值范围和第一范围的集合，生成第一分割树；根据第一分割树，确定第一最大平衡距离；根据第一最大平衡距离，选择用于对数据包进行分类的数据包分类算法。本发明实施例根据规则集的范围分布情况在数据包分类算法之间进行选择，仅需要创建分割树，避免了对每个数据包分类算法进行分别创建决策树的操作。能够根据该分割树，快速的选择数据包分类算法，提高选择数据包分类算法的方法的效率。



1. 一种选择数据包分类算法的方法,其特征在于,包括:

确定数据包的第一域的第一取值范围和对所述数据包进行分类的规则集中的规则在所述第一域上的第一范围的集合;

根据所述第一取值范围和所述第一范围的集合,生成第一分割树,所述第一分割树的根节点代表的区间为所述第一取值范围,所述第一分割树的叶子节点代表的区间为所述第一范围;

根据所述第一分割树,确定第一最大平衡距离,所述第一最大平衡距离为所述第一分割树的根节点所在的第一子树与所述第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数;

根据所述第一最大平衡距离,选择用于对所述数据包进行分类的数据包分类算法。

2. 根据权利要求1所述的方法,其特征在于,还包括:

确定所述数据包的第二域的第二取值范围和对所述数据包进行分类的所述规则集中的规则在所述第二域上的第二范围的集合;

根据所述第二取值范围和所述第二范围的集合,生成第二分割树,所述第二分割树的根节点代表的区间为所述第二取值范围,所述第二分割树的叶子节点代表的区间为所述第二范围;

根据所述第二分割树,确定第二最大平衡距离,所述第二最大平衡距离为所述第二分割树的根节点所在的第三子树与所述第二分割树的叶子节点所在的第四子树之间包含的所述子树的最大个数;

其中,所述根据所述第一最大平衡距离,选择用于对所述数据包进行分类的数据包分类算法,包括:

确定所述第一最大平衡距离和所述第二最大平衡距离的较大值;

根据所述较大值,选择用于对所述数据包进行分类的数据包分类算法。

3. 根据权利要求2所述的方法,其特征在于,所述第一域为源IP地址域,所述第二域为目的IP地址域,或者所述第一域为目的IP地址域,所述第二域为源IP地址域。

4. 根据权利要求1-3中任一项所述的方法,其特征在于,所述根据所述第一取值范围和所述第一范围的集合,生成第一分割树,包括:

根据所述第一取值范围,生成所述第一分割树的根节点;

根据所述第一取值范围,确定所述第一分割树的根节点的第一分割点;

从所述第一范围的集合中选择第一窄范围的集合,所述第一窄范围的长度与所述第一取值范围的长度的比值小于第一数值;

根据所述第一取值范围、所述第一分割点和所述第一窄范围的集合,生成所述第一分割树的子节点,所述子节点代表的区间与所述第一窄范围相交;

根据所述第一分割树的根节点和所述第一分割树的子节点,生成所述第一分割树,所述第一分割树的叶子节点代表的区间为所述第一窄范围。

5. 根据权利要求4所述的方法,其特征在于,所述根据所述第一取值范围,确定所述第一分割树的根节点的第一分割点,包括:

根据公式 $I_m = \frac{I_{\max} - I_{\min} + 1}{2} + I_{\min} - 1$ 确定所述第一分割点;

其中, I_m 是所述第一分割点, I_{\max} 是所述第一取值范围的最大值, I_{\min} 是所述第一取值范围的最小值。

6. 根据权利要求 4 或 5 所述的方法, 其特征在于, 所述根据所述第一最大平衡距离, 选择用于对所述数据包进行分类的数据包分类算法, 包括:

根据所述第一窄范围的集合, 确定所述第一窄范围的集合对应的第一规则数;

根据所述第一规则数, 确定第一判断值;

根据所述第一最大平衡距离和所述第一判断值, 选择用于对所述数据包进行分类的数据包分类算法。

7. 根据权利要求 6 所述的方法, 其特征在于, 所述根据所述第一规则数, 确定第一判断值, 包括:

根据公式 $X = \frac{\log_2 \text{numRules}}{2}$ 确定所述第一判断值;

其中, X 是所述第一判断值, numRules 是所述第一规则数; 其中,

所述根据所述第一最大平衡距离和所述第一判断值, 选择用于对所述数据包进行分类的数据包分类算法, 包括:

在所述第一最大平衡距离大于所述第一判断值时, 选择 HyperSplit 算法;

在所述第一最大平衡距离小于或者等于所述第一判断值时, 选择 HyperCuts 算法。

8. 根据权利要求 1-7 中任一项所述的方法, 其特征在于, 所述子树为准平衡子树, 所述根据所述第一分割树, 确定第一最大平衡距离, 包括:

根据所述第一分割树, 确定所述第一分割树包含的所述准平衡子树, 所述准平衡子树的 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数;

对所述第一分割树进行深度优先遍历, 确定所述第一分割树的根节点所在的第一准平衡子树和所述第一分割树的叶子节点所在的第二准平衡子树;

确定所述第一准平衡子树和所述第二准平衡子树之间包含的所述准平衡子树的最大个数为所述第一最大平衡距离。

9. 根据权利要求 4-7 中任一项所述的方法, 其特征在于,

所述第一域为源 IP 地址域或目的 IP 地址域, 所述第一数值取 0.05;

所述第一域为端口域或协议域, 所述第一数值取 0.5。

10. 根据权利要求 8 所述的方法, 其特征在于, 所述第二数值取 1.5-1.8 之间任一数值。

11. 一种选择数据包分类算法的装置, 其特征在于, 包括:

第一确定模块, 用于确定数据包的第一域的第一取值范围和对所述数据包进行分类的规则集中的规则在所述第一域上的第一范围的集合;

生成模块, 用于根据所述第一取值范围和所述第一范围的集合, 生成第一分割树, 所述第一分割树的根节点代表的区间为所述第一取值范围, 所述第一分割树的叶子节点代表的区间为所述第一范围;

第二确定模块, 用于根据所述第一分割树, 确定第一最大平衡距离, 所述第一最大平衡距离为所述第一分割树的根节点所在的第一子树与所述第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数;

选择模块,用于根据所述第一最大平衡距离,选择用于对所述数据包进行分类的数据包分类算法。

12. 根据权利要求 11 所述的装置,其特征在于,

所述第一确定模块,还用于确定所述数据包的所述第二域的第二取值范围和对所述数据包进行分类的所述规则集中的规则在所述第二域上的第二范围的集合;

所述生成模块,还用于根据所述第二取值范围和所述第二范围的集合,生成第二分割树,所述第二分割树的根节点代表的区间为所述第二取值范围,所述第二分割树的叶子节点代表的区间为所述第二范围;

所述第二确定模块,还用于根据所述第二分割树,确定第二最大平衡距离,所述第二最大平衡距离为所述第二分割树的根节点所在的第三子树与所述第二分割树的叶子节点所在的第四子树之间的所述子树的最大个数;

其中,所述选择模块,

具体用于确定所述第一最大平衡距离和所述第二最大平衡距离的较大值;

具体用于根据所述较大值,选择用于对所述数据包进行分类的数据包分类算法。

13. 根据权利要求 12 所述的装置,其特征在于,所述第一域为源 IP 地址域,所述第二域为目的 IP 地址域,或者所述第一域为目的 IP 地址域,所述第二域为源 IP 地址域。

14. 根据权利要求 11-13 中任一项所述的装置,其特征在于,所述生成模块,

具体用于根据所述第一取值范围,生成所述第一分割树的根节点;

具体用于根据所述第一取值范围,确定所述第一分割树的根节点的第一分割点;

具体用于从所述第一范围的集合中选择第一窄范围的集合,所述第一窄范围的长度与所述第一取值范围的长度的比值小于第一数值;

具体用于根据所述第一取值范围、所述第一分割点和所述第一窄范围的集合,生成所述第一分割树的子节点,所述子节点代表的区间与所述第一窄范围相交;

具体用于根据所述第一分割树的根节点和所述第一分割树的子节点,生成所述第一分割树,所述第一分割树的叶子节点代表的区间为所述第一窄范围。

15. 根据权利要求 14 所述的装置,其特征在于,所述生成模块,具体用于根据公式

$$I_m = \frac{I_{\max} - I_{\min} + 1}{2} + I_{\min} - 1 \text{ 确定所述第一分割点};$$

其中, I_m 是所述第一分割点, I_{\max} 是所述第一取值范围的最大值, I_{\min} 是所述第一取值范围的最小值。

16. 根据权利要求 14 或 15 所述的装置,其特征在于,所述选择模块,

用于根据所述第一窄范围的集合,确定所述第一窄范围的集合对应的第一规则数;

用于根据所述第一规则数,确定第一判断值;

用于根据所述第一最大平衡距离和所述第一判断值,选择用于对所述数据包进行分类的数据包分类算法。

17. 根据权利要求 16 所述的装置,其特征在于,所述选择模块,具体用于根据公式

$$X = \frac{\log_2 \text{numRules}}{2} \text{ 确定所述第一判断值};$$

其中,X 是所述第一判断值,numRules 是所述第一规则数;其中,所述选择模块,

具体用于在所述第一最大平衡距离大于所述第一判断值时,选择 HyperSplit 算法;

具体用于在所述第一最大平衡距离小于或者等于所述第一判断值时,选择 HyperCuts 算法。

18. 根据权利要求 11-17 中任一项所述的装置,其特征在于,所述子树为准平衡子树,所述第二确定模块,

用于根据所述第一分割树,确定所述第一分割树包含的所述准平衡子树,所述准平衡子树的 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数;

用于对所述第一分割树进行深度优先遍历,确定所述第一分割树的根节点所在的第一准平衡子树和所述第一分割树的叶子节点所在的第二准平衡子树;

用于确定所述第一准平衡子树和所述第二准平衡子树之间包含的所述准平衡子树的最大个数为所述第一最大平衡距离。

19. 根据权利要求 14-17 中任一项所述的装置,其特征在于,

所述第一域为源 IP 地址域或目的 IP 地址域,所述第一数值取 0.05;

所述第一域为端口域或协议域,所述第一数值取 0.5。

20. 根据权利要求 18 所述的装置,其特征在于,所述第二数值取 1.5-1.8 之间任一数值。

选择数据包分类算法的方法和装置

技术领域

[0001] 本发明涉及通信技术领域,并且更具体地,涉及选择数据包分类算法的方法和装置。

背景技术

[0002] 数据包分类技术被广泛应用于各类网络设备中,是构建灵活可编程网络的关键技术。工业界一般采用三态内容寻址存储器(Ternary Content Addressable Memory,简称“TCAM”)来实现数据包分类,其主要缺点有功耗大,容量小,芯片尺寸大,价格昂贵。与TCAM对应的另一种数据包分类方法是算法解决方案。算法解决方案采用高性能算法,快速检索数据包分类规则,达到高性能数据包分类的目的。算法解决方案功耗小,价格低廉,因此,数据包分类算法具有很大的应用场景。

[0003] 不同的数据包分类算法具有不同的性能,因此,正确地选择数据包分类算法能够取得较好的性能。在大量已发明的数据包分类算法中,基于决策树的数据包分类算法是性能最优的一类算法。基于决策树的数据包分类算法的代表算法是HyperSplit算法和HyperCuts算法。但是,HyperCuts和HyperSplit算法存在创建决策树时间过长的的问题。

[0004] 然而,现有技术中需要分别对HyperCuts算法和HyperSplit算法进行创建决策树,并通过比较两种数据包分类算法的建树结果来选择数据包分类算法对数据包进行分类。由于HyperCuts算法和HyperSplit算法创建决策树的时间过长,因此,该选择数据包分类算法的方法耗时过长,效率低。

发明内容

[0005] 本发明实施例提供一种选择数据包分类算法的方法和装置,能够提高选择数据包分类算法的方法的效率。

[0006] 第一方面,提供了一种选择数据包分类算法的方法,该方法包括:确定数据包的第一域的第一取值范围和对该数据包进行分类的规则集中的规则在该第一域上的第一范围的集合;根据该第一取值范围和该第一范围的集合,生成第一分割树,该第一分割树的根节点代表的区间为该第一取值范围,该第一分割树的叶子节点代表的区间为该第一范围;根据该第一分割树,确定第一最大平衡距离,该第一最大平衡距离为该第一分割树的根节点所在的第一子树与该第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数;根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法。

[0007] 结合第一方面,在第一方面的第一种可能的实现方式中,该方法还包括:确定该数据包的第二域的第二取值范围和对该数据包进行分类的该规则集中的规则在该第二域上的第二范围的集合;根据该第二取值范围和该第二范围的集合,生成第二分割树,该第二分割树的根节点代表的区间为该第二取值范围,该第二分割树的叶子节点代表的区间为该第二范围;根据该第二分割树,确定第二最大平衡距离,该第二最大平衡距离为该第二分割树的根节点所在的第三子树与该第二分割树的叶子节点所在的第四子树之间包含的该子树

的最大个数 ;其中,该根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法,包括 :确定该第一最大平衡距离和该第二最大平衡距离的较大值 ;根据该较大值,选择用于对该数据包进行分类的数据包分类算法。

[0008] 结合第一方面的第一种可能的实现方式,在第一方面的第二种可能的实现方式中,该第一域为源 IP 地址域,该第二域为目的 IP 地址域,或者该第一域为目的 IP 地址域,该第二域为源 IP 地址域。

[0009] 结合第一方面或第一方面的第一种至第二种可能的实现方式中的任一种可能的实现方式,在第一方面的第三种可能的实现方式中,该根据该第一取值范围和该第一范围的集合,生成第一分割树,包括 :根据该第一取值范围,生成该第一分割树的根节点 ;根据该第一取值范围,确定该第一分割树的根节点的第一分割点 ;从该第一范围的集合中选择第一窄范围的集合,该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值 ;根据该第一取值范围、该第一分割点和该第一窄范围的集合,生成该第一分割树的子节点,该子节点代表的区间与该第一窄范围相交 ;根据该第一分割树的根节点和该第一分割树的子节点,生成该第一分割树,该第一分割树的叶子节点代表的区间为该第一窄范围。

[0010] 结合第一方面的第三种可能的实现方式,在第一方面的第四种可能的实现方式中,该根据该第一取值范围,确定该第一分割树的根节点的第一分割点,包括 :根据公式 $I_m = (I_{\max} - I_{\min} + 1) / 2 + I_{\min} - 1$ 确定该第一分割点 ;其中, I_m 是该第一分割点, I_{\max} 是该第一取值范围的最大值, I_{\min} 是该第一取值范围的最小值。

[0011] 结合第一方面的第三种可能的实现方式或第四种可能的实现方式,在第一方面的第五种可能的实现方式中,该根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法,包括 :根据该第一窄范围的集合,确定该第一窄范围的集合对应的第一规则数 ;根据该第一规则数,确定第一判断值 ;根据该第一最大平衡距离和该第一判断值,选择用于对该数据包进行分类的数据包分类算法。

[0012] 结合第一方面的第五种可能的实现方式,在第一方面的第六种可能的实现方式中,该根据该第一规则数,确定第一判断值,包括 :根据公式 $X = (\log_2 \text{numRules}) / 2$ 确定该第一判断值 ;其中, X 是该第一判断值, numRules 是该第一规则数 ;其中,该根据该第一最大平衡距离和该第一判断值,选择用于对该数据包进行分类的数据包分类算法,包括 :在该第一最大平衡距离大于该第一判断值时,选择 HyperSplit 算法 ;在该第一最大平衡距离小于或者等于该第一判断值时,选择 HyperCuts 算法。

[0013] 结合第一方面或第一方面的第一种至第六种可能的实现方式中的任一种可能的实现方式,在第一方面的第七种可能的实现方式中,该子树为准平衡子树,该根据该第一分割树,确定第一最大平衡距离,包括 :根据该第一分割树,确定该第一分割树包含的该准平衡子树,该准平衡子树的 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数 ;对该第一分割树进行深度优先遍历,确定该第一分割树的根节点所在的第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树 ;确定该第一准平衡子树和该第二准平衡子树之间包含的该准平衡子树的最大个数为该第一最大平衡距离。

[0014] 结合第一方面的第三种至第六种可能的实现方式中的任一种可能的实现方式,在第一方面的第八种可能的实现方式中,该第一域为源 IP 地址域或目的 IP 地址域,该第一数

值取 0.05 ;该第一域为端口域或协议域,该第一数值取 0.5。

[0015] 结合第一方面的第七种可能的实现方式,在第一方面的第九种可能的实现方式中,该第二数值取 1.5-1.8 之间任一数值。

[0016] 第二方面,提供了一种选择数据包分类算法的装置,该装置包括:第一确定模块,用于确定数据包的第一域的第一取值范围和对该数据包进行分类的规则集中的规则在该第一域上的第一范围的集合;生成模块,用于根据该第一取值范围和该第一范围的集合,生成第一分割树,该第一分割树的根节点代表的区间为该第一取值范围,该第一分割树的叶子节点代表的区间为该第一范围;第二确定模块,用于根据该第一分割树,确定第一最大平衡距离,该第一最大平衡距离为该第一分割树的根节点所在的第一子树与该第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数;选择模块,用于根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法。

[0017] 结合第二方面,在第二方面的第一种可能的实现方式中,该第一确定模块,还用于确定该数据包的第二域的第二取值范围和对该数据包进行分类的该规则集中的规则在该第二域上的第二范围的集合;该生成模块,还用于根据该第二取值范围和该第二范围的集合,生成第二分割树,该第二分割树的根节点代表的区间为该第二取值范围,该第二分割树的叶子节点代表的区间为该第二范围;该第二确定模块,还用于根据该第二分割树,确定第二最大平衡距离,该第二最大平衡距离为该第二分割树的根节点所在的第三子树与该第二分割树的叶子节点所在的第四子树之间的该子树的最大个数;其中,该选择模块,具体用于确定该第一最大平衡距离和该第二最大平衡距离的较大值;具体用于根据该较大值,选择用于对该数据包进行分类的数据包分类算法。

[0018] 结合第二方面的第一种可能的实现方式,在第二方面的第二种可能的实现方式中,该第一域为源 IP 地址域,该第二域为目的 IP 地址域,或者该第一域为目的 IP 地址域,该第二域为源 IP 地址域。

[0019] 结合第二方面或第二方面的第一种至第二种可能的实现方式中的任一种可能的实现方式,在第二方面的第三种可能的实现方式中,该生成模块,具体用于根据该第一取值范围,生成该第一分割树的根节点;具体用于根据该第一取值范围,确定该第一分割树的根节点的第一分割点;具体用于从该第一范围的集合中选择第一窄范围的集合,该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值;具体用于根据该第一取值范围、该第一分割点和该第一窄范围的集合,生成该第一分割树的子节点,该子节点代表的区间与该第一窄范围相交;具体用于根据该第一分割树的根节点和该第一分割树的子节点,生成该第一分割树,该第一分割树的叶子节点代表的区间为该第一窄范围。

[0020] 结合第二方面的第三种可能的实现方式,在第二方面的第四种可能的实现方式中,该生成模块,具体用于根据公式 $I_m = (I_{\max} - I_{\min} + 1) / 2 + I_{\min} - 1$ 确定该第一分割点;其中, I_m 是该第一分割点, I_{\max} 是该第一取值范围的最大值, I_{\min} 是该第一取值范围的最小值。

[0021] 结合第二方面的第三种或第四种可能的实现方式,在第二方面的第五种可能的实现方式中,该选择模块,用于根据该第一窄范围的集合,确定该第一窄范围的集合对应的第一规则数;用于根据该第一规则数,确定第一判断值;用于根据该第一最大平衡距离和该第一判断值,选择用于对该数据包进行分类的数据包分类算法。

[0022] 结合第二方面的第五种可能的实现方式,在第二方面的第六种可能的实现方式

中,该选择模块,具体用于根据公式 $X = (\log_2 \text{numRules})/2$ 确定该第一判断值;其中, X 是该第一判断值, numRules 是该第一规则数;其中,该选择模块,具体用于在该第一最大平衡距离大于该第一判断值时,选择 HyperSplit 算法;具体用于在该第一最大平衡距离小于或者等于该第一判断值时,选择 HyperCuts 算法。

[0023] 结合第二方面或第二方面的第一种至第六种可能的实现方式中的任一种可能的实现方式,在第二方面的第七种可能的实现方式中,该子树为准平衡子树,该第二确定模块,用于根据该第一分割树,确定该第一分割树包含的该准平衡子树,该准平衡子树的 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数;用于对该第一分割树进行深度优先遍历,确定该第一分割树的根节点所在的第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树;用于确定该第一准平衡子树和该第二准平衡子树之间包含的该准平衡子树的最大个数为该第一最大平衡距离。

[0024] 结合第二方面的第三种至第六种可能的实现方式中的任一种可能的实现方式,在第二方面的第八种可能的实现方式中,该第一域为源 IP 地址域或目的 IP 地址域,该第一数值取 0.05;该第一域为端口域或协议域,该第一数值取 0.5。

[0025] 结合第二方面的第七种可能的实现方式,在第二方面的第九种可能的实现方式中,该第二数值取 1.5-1.8 之间任一数值。

[0026] 本发明实施例根据规则集的范围分布情况在数据包分类算法之间进行选择,仅需要创建分割树,避免了对每个数据包分类算法进行分别创建决策树的操作。能够根据该分割树,快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

附图说明

[0027] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0028] 图 1 是根据本发明一个实施例的选择数据包分类算法的方法的示意性流程图。

[0029] 图 2 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。

[0030] 图 3 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。

[0031] 图 4 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。

[0032] 图 5 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。

[0033] 图 6 是根据本发明另一个实施例的选择数据包分类算法的方法的过程的示意性流程图。

[0034] 图 7 是根据本发明一个实施例的选择数据包分类算法的装置的框图。

[0035] 图 8 是根据本发明另一个实施例的装置的示意框图。

具体实施方式

[0036] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施

例,都属于本发明保护的范围。

[0037] 应理解,本发明实施例的技术方案可以应用于各种通信系统,例如:全球移动通讯(Global System of Mobile communication, GSM)系统、码分多址(Code Division Multiple Access, CDMA)系统、宽带码分多址(Wideband Code Division Multiple Access, WCDMA)系统、通用分组无线业务(General Packet Radio Service, GPRS)、长期演进(Long Term Evolution, LTE)系统、LTE 频分双工(Frequency Division Duplex, FDD)系统、LTE 时分双工(Time Division Duplex, TDD)、通用移动通信系统(Universal Mobile Telecommunication System, UMTS)或全球互联微波接入(Worldwide Interoperability for Microwave Access, WiMAX)通信系统等。

[0038] 图1是根据本发明一个实施例的选择数据包分类算法的方法的示意性流程图。图1的方法可以由选择数据包分类算法的装置300执行。

[0039] 110,确定数据包的第一域的第一取值范围和对数据包进行分类的规则集中的规则在该第一域上的第一范围的集合。

[0040] 数据包的第一域可以为该数据包包头的多个域中的一个域,该第一域的第一取值范围可以为该第一域可以取得的最大范围。该第一取值范围为该第一域的固有属性值,可以根据第一域的类型确定第一域的取值范围。例如,端口域的取值范围为 $[0, 65535]$,IP地址域的取值范围为 $[0, 2^{(32)}-1]$ 。

[0041] 数据包包头中的域的个数为对该数据包进行分类的规则维度。例如,一个数据包包头的域包括源IP地址,目的IP地址,源端口号,目的端口号和协议类型。则对该数据包进行分类的数据包分类规则为:

[0042] 153.0.0.0/8 224.0.0.0/8 0:65535 80:80 TCP → DROP (规则1)

[0043] 其中,153.0.0.0/8表示该规则1在源IP地址域上的范围,224.0.0.0/8表示该规则1在目的IP地址域上的范围,0:65535表示该规则1在源端口域上的范围,80:80表示该规则1在目的端口域上的范围。该规则1为5维规则,表示若该数据包包头的域满足源IP地址符合153.0.0.0/8,目的IP地址符合224.0.0.0/8,源端口处于0:65535,目的端口号为80,协议号是TCP协议时,则执行丢弃操作。

[0044] 该规则集可以包括多个规则,该规则集中的每个规则在数据包的多个域上对应多个范围。每个规则均可以包括在第一域上的第一范围。该第一范围的集合可以表示为规则集的多个规则在第一域上的多个第一范围的集合。例如,规则集中的N个规则在第一域上的第一范围的集合包含的第一范围的个数可以为N个,也可以少于N个,本发明实施例并不限于此。

[0045] 可选地,作为另一实施例,规则集中的规则在该第一域上的第一范围的集合的获取方式可以为通过扫描规则集中的每个规则的范围得出。例如,可以通过查询代码,查询第一范围,并可以通过存储代码,将查询到的第一范围进行存储,以便于在创建第一分割树时使用。

[0046] 120,根据该第一取值范围和该第一范围的集合,生成第一分割树,该第一分割树的根节点代表的区间为该第一取值范围,该第一分割树的叶子节点代表的区间为该第一范围。

[0047] 具体地,根据该第一取值范围和该第一范围的集合,生成的第一分割树可以用来

衡量规则集中的第一范围分布的均匀性。例如,第一范围的分布越均匀,则该第一分割树越接近于满二叉树。

[0048] 根据该第一取值范围,可以确定该第一分割树的根节点代表的区间,根据该第一取值范围和该第一范围的集合,可以确定该第一分割树的子节点代表的区间。

[0049] 应理解,该第一分割树包含一个根节点,该第一分割树的根节点代表一个区间,该根节点代表的一个区间可以为数据包的第一域的取值范围。例如,数据包的域有四个比特,则该域的取值范围为 $[0, 15]$ 。根据该域的取值范围可以确定根节点代表的区间为 $[0, 15]$ 。

[0050] 还应理解,该第一分割树包含至少一个叶子节点,该第一分割树的每个叶子节点可以代表一个区间,该第一分割树的至少一个叶子节点可以代表与至少一个叶子节点对应的至少一个区间。该第一分割树的至少一个叶子节点的至少一个区间可以为该规则集中的规则在第一域的第一范围。

[0051] 130,根据第一分割树,确定第一最大平衡距离,该第一最大平衡距离为该第一分割树的根节点所在的第一子树与该第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数。

[0052] 可选地,作为另一实施例,该子树可以为满足一定条件的二叉树。平衡距离可以表示分割树的根节点所在的子树到分割树的叶子节点所在的子树中间包含的子树的个数。最大平衡距离可以表示该分割树所有的平衡距离中最大的一个。根节点所在的子树、叶子节点所在的子树和包含的子树可以相同也可以不同,也可以为同时满足一定条件的二叉树,本发明实施例并不限于此。

[0053] 该第一分割树的根节点所在的第一子树、该第一分割树的叶子节点所在的第二子树和两者中间包含的子树可以相同也可以不相同,也可以为同时满足一定条件的二叉树,本发明实施例并不限于此。

[0054] 该第一分割树的每个叶子节点都对应一个平衡距离,则该平衡距离中的最大值可以为第一最大平衡距离。

[0055] 140,根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法。

[0056] 可选地,作为另一实施例,基于决策树的数据包分类算法是性能较优的一个数据包分类算法。基于决策树的数据包分类算法可以包括 HyperSplit 算法和 HyperCuts 算法,选择数据包分类算法的方法可以为从该 HyperSplit 算法和 HyperCuts 算法中选择合适的数据包分类算法,对数据包进行分类。

[0057] 可选地,作为另一实施例,根据该第一最大平衡距离,选择数据包分类算法,可以通过分析该第一最大平衡距离来选择数据包分类算法,也可以将该第一最大平衡距离与某一阈值进行比较,根据比较的结果选择数据包分类算法,本发明实施例并不限于此。

[0058] 本发明实施例根据规则集的范围分布情况在数据包分类算法之间进行选择,仅需要创建分割树,避免了对每个数据包分类算法进行分别创建决策树的操作。能够根据该分割树,快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0059] 图 2 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。如图 2 所示,可选地,该方法还包括:

[0060] 150,确定数据包的第二域的第二取值范围和对该数据包进行分类的规则集中的

规则在该第二域上的第二范围的集合；

[0061] 160, 根据该第二取值范围和该第二范围的集合, 生成第二分割树, 该第二分割树的根节点代表的区间为该第二取值范围, 该第二分割树的叶子节点代表的区间为该第二范围；

[0062] 170, 根据该第二分割树, 确定第二最大平衡距离, 该第二最大平衡距离为该第二分割树的根节点所在的第三子树与该第二分割树的叶子节点所在的第四子树之间包含的子树的最大个数；

[0063] 应理解, 步骤 150, 160, 170 的技术方案与步骤 110, 120, 130 的技术方法相同, 此处为避免重复, 在权利要求和说明书中省略确定第二最大平衡的步骤。

[0064] 其中, 140, 根据该第一最大平衡距离, 选择用于对数据包进行分类的数据包分类算法, 包括：

[0065] 140A, 确定该第一最大平衡距离和该第二最大平衡距离的较大值；

[0066] 140B, 根据该较大值, 选择用于对该数据包进行分类的数据包分类算法。

[0067] 可选地, 作为另一实施例。本发明实施例可以针对一个域的范围分布情况进行创建分割树, 确定该分割树的最大平衡距离；也可以针对两个域的范围分布情况分别创建两个分割树。根据该两个分割树, 分别确定最大平衡距离。根据该两个最大平衡距离中的较大值来选择数据包分类算法。可选地, 也可以针对多个域的范围分布情况分别创建对应的多个分割树。根据该多个分割树, 分别确定对应的多个最大平衡距离。根据该多个最大平衡距离中的最大值来选择数据包分类算法, 本发明实施例并不限于此。

[0068] 本发明实施例根据规则集在多个域上的范围的分布情况在数据包分类算法之间进行选择, 创建对应的多个分割树。分别确定对应的多个最大平衡距离, 根据该多个最大平衡距离的最大值选择数据包分类算法。能够提高选择数据包分类算法的方法的准确性。该方法避免了对每个数据包分类算法进行分别创建决策树的操作, 提高选择数据包分类算法的方法的效率和准确性。

[0069] 可选地, 作为另一个实施例。该第一域可以为源 IP 地址域, 该第二域可以为目的 IP 地址域, 或者该第一域为目的 IP 地址域, 该第二域为源 IP 地址域。根据源 IP 地址域和目的 IP 地址域对应的第一范围和第二范围, 分别生成两个对应的分割树。根据该两个分割树, 则分别确定源 IP 地址域对应的最大平衡距离 D_{src} 和目的 IP 地址域对应的最大平衡距离 D_{dst} 。根据该两个最大平衡距离中的较大值来选择数据包分类算法对数据包进行分类。

[0070] 本发明实施例提供的选择数据包分类算法的方法, 根据源 IP 地址域对应的范围和目的 IP 地址域对应的范围分别创建分割树, 确定该两个分割树对应的两个最大平衡距离中的较大值。比较较大值和从源 IP 地址域对应的范围和目的 IP 地址域对应的范围中选择的满足条件的窄范围对应的规则的个数, 从 HyperSplit 算法和 HyperCuts 算法中选择数据包分类算法。该方法避免了对每个数据包分类算法进行分别创建决策树的操作, 能够快速的选择数据包分类算法, 提高选择数据包分类算法的方法的效率。

[0071] 图 3 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。如图 3 所示, 根据第一取值范围和第一范围的集合, 生成第一分割树, 包括：

[0072] 121, 根据该第一取值范围, 生成该第一分割树的根节点。

[0073] 具体地, 该第一分割树的每个节点代表一个区间, 该第一取值范围可以作为该第

一分割树的根节点所代表的区间。例如,数据包的域有四个比特,则该域的取值范围为 [0, 15]。根据该域的取值范围可以确定根节点代表的区间为 [0, 15]。

[0074] 122,根据该第一取值范围,确定该第一分割树的根节点的第一分割点。

[0075] 可选地,该第一分割点可以根据经验从该第一取值范围中包含的数值中选择一个数值作为该第一分割点,也可以取该第一取值范围的中间值作为该第一分割点,也可以通过计算确定该第一分割点,本发明实施例并不限于此。

[0076] 可选地,作为另一实施例。选择该第一分割点的计算方法可以为:

$$[0077] \quad I_m = \frac{I_{\max} - I_{\min} + 1}{2} + I_{\min} - 1$$

[0078] 其中, I_m 是该第一分割点, I_{\max} 是该第一取值范围的最大值, I_{\min} 是该第一取值范围的最小值。例如,该第一取值范围是 [0, 15], 则可以表示为 $[I_{\min}, I_{\max}]$ 。其中, $I_{\min} = 0$, $I_{\max} = 15$, 根据上述公式计算,可以得到 $I_m = 7$ 。

[0079] 123,从第一范围的集合中选择第一窄范围的集合,该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值。

[0080] 可以基于第一范围的集合生成第一分割树,该第一范围可以为长度较短的第一窄范围。该第一窄范围可以满足该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值。也可以为基于其他方法定义的自定义的范围,本发明实施例并不限于此。

[0081] 可选地,作为另一实施例。第一域为源 IP 地址域或目的 IP 地址域,该第一数值可以取 0.05。第一域为端口域或协议域,该第一数值可以取 0.5。例如,若规则集中规则的第一范围表示为 (F_L, F_H) , 则该第一范围满足第一窄范围的条件可以为:1) 如果是源 IP 地址域或目的 IP 地址域,则 $(F_H - F_L + 1) / \text{len}(I) < 0.05$ 。2) 如果是端口域或者协议域,则 $(F_H - F_L + 1) / \text{len}(I) < 0.5$ 。其中, I 表示数据包的第一域的第一取值范围, $\text{len}(I)$ 表示该第一域的第一取值范围的长度。

[0082] 可选地,作为另一实施例。该第一数值的取值也可以取其他数值,例如 0.045、0.046、0.047、0.048、0.049、0.051、0.052 等。满足基于该其他数值确定的长度较短的窄范围可以生成分割树,并根据该分割树可以准确地选择数据包分类算法。对于该第一数值的取值,本发明实施例并不限于此。

[0083] 124,根据该第一取值范围、该第一分割点和该第一窄范围的集合,生成该第一分割树的子节点,该子节点代表的第一范围与该第一窄范围相交。

[0084] 应理解,该第一分割树的子节点包括除根节点的其他节点,该其他节点包括该第一分割树的叶子节点和该第一分割树的中间节点。

[0085] 根据该第一取值范围和该第一分割点,可以生成该第一分割树的根节点的左节点和右节点。具体地,该第一分割点可以把该第一分割树的根节点代表的区间分为该第一分割树的根节点的左右节点分别代表的两个区间。例如,该第一取值范围为 [0, 15], 根据上述公式可以确定分割点为 7。根据该分割点 7, 确定该根节点的左右节点代表的区间分别为 [0, 7] 和 [8, 15]。

[0086] 可选地,作为另一实施例。该第一分割树除去根节点和该根节点的左右节点的其他子节点的生成过程可以为与该根节点的左右节点的生成过程相同,即通过确定分割点,将父节点的区间分为该父节点的左右节点的两个区间。

[0087] 生成的该第一分割树的子节点可以满足该子节点代表的区间与该第一窄范围相交,并删除不满足条件的子节点。可选地,相交的定义可以为:若范围 1 为 (F_{1L}, F_{1H}) , 范围 2 为 (F_{2L}, F_{2H}) , 若不满足 $F_{1H} < F_{2L}$ 或 $F_{2H} < F_{1L}$, 则称该范围 1 和范围 2 相交。

[0088] 以上生成该第一分割树的子节点的迭代过程直到该第一分割树的子节点代表的区间为该第一窄范围时,停止迭代过程。将所代表的区间为第一窄范围的子节点可以作为该第一分割树的叶子节点。例如,该第一取值范围为 $[0, 15]$, 可以确定分割点为 7。根据该分割点 7, 确定该根节点的左右节点代表的区间分别为 $[0, 7]$ 和 $[8, 15]$ 。分别确定左右节点作为父节点的分割点 3 和 11, 则该左节点的左右节点所代表的两个区间可以为 $[0, 3]$ 和 $[3, 7]$, 该右节点的左右节点所代表的两个区间可以为 $[8, 11]$ 和 $[12, 15]$, 依次类推可依次确定分割树中每个子节点所代表的区间依次为 $[0, 7]$ 、 $[8, 15]$ 、 $[0, 3]$ 、 $[3, 7]$ 、 $[8, 11]$ 、 $[12, 15]$ 、 $[0, 1]$ 、 $[2, 3]$ 、 $[8, 9]$ 和 $[10, 11]$ 。若第一窄范围包括 $[2, 3]$ 、 $[8, 9]$ 和 $[12, 15]$, 在保证该第一分割树的叶子节点所代表的区间为该第一窄范围且该子节点代表的区间与该第一窄范围相交, 可以将不满足相交条件的节点删除。因此, 依次确定的子节点代表的区间删除 $[3, 7]$ 、 $[0, 1]$ 和 $[10, 11]$ 。因此可以确定该第一分割树的叶子节点代表的区间为 $[2, 3]$ 、 $[8, 9]$ 和 $[12, 15]$ 。

[0089] 125, 根据该第一分割树的根节点和该第一分割树的子节点, 生成该第一分割树, 该第一分割树的叶子节点代表的区间为该第一窄范围。

[0090] 本发明实施例根据规则集的范围分布情况在数据包分类算法之间进行选择, 仅根据规则集的范围创建分割树, 通过该分割树的叶子节点代表的区间量化范围分布的均匀性。该方法避免了对每个数据包分类算法进行分别创建决策树的操作, 能够根据该分割树的量化结果, 快速的选择数据包分类算法, 提高选择数据包分类算法的方法的效率。

[0091] 可选地, 作为另一实施例, 该生成第一分割树的方法可以由代码实现。具体实现方式可以如下列代码所示:

[0092]

```

Algorithm 1 Create_SplitTree
 $\phi$  为根节点
 $Q = \{\phi\}$ ; 设置队列
while !Q.empty() do
     $N \leftarrow Q.pop\_front()$ ;
     $I_m \leftarrow \lfloor \frac{I_{\max} - I_{\min} + 1}{2} \rfloor + I_{\min} - 1$ ; 生成左右孩子节点  $N_L$  和  $N_R$ 
    for  $r \in N.S$  do
        if  $r$  与  $[I_{\min}, I_m]$  相交 then
             $N_L.S \leftarrow N_L.S \cup \{r\}$ ;
             $N_L.L(r) += N.L(r)$ ;
        end if
        if  $r$  与  $[I_m + 1, I_{\max}]$  相交 then
             $N_R.S \leftarrow N_R.S \cup \{r\}$ ;
             $N_R.L(r) += N.L(r)$ ;
        end if
    end for
    if  $N_R.S == N_L.S$  then
        删除节点  $N_L$  和  $N_R$ 
    end if
    if  $N_L.numRule == 0$  (or  $N_R.numRule == 0$ ) then
        删除  $N_L$  (or  $N_R$ )
    end if
    将剩余的节点放入到队列  $Q$  结尾;
end while

```

[0093] 可选地, 该子节点可以包括以下至少一种信息:

[0094] 1) 该节点所代表的区间, $N.[I_{\min}, I_{\max}]$; 可选地, 对于规则中的某个范围, 分割树的根节点所代表的范围等于该第一取值范围。

[0095] 2) 所有与 $N.[I_{\min}, I_{\max}]$ 相交的窄范围集合 $N.S$;

[0096] 3) 第一窄范围的规则数 $N.L(r)$, 包含窄范围 $r \in N.S$ 的规则数且

$$[0097] \quad \sum_{r \in N.S} N.L(r) = N.numRules$$

[0098] 本发明实施例提供的选择数据包分类算法的方法, 根据规则集的范围创建分割树, 比较该分割树的最大平衡距离和根据窄范围对应的规则数确定的判断值, 从 HyperSplit 算法和 HyperCuts 算法中选择数据包分类算法。该方法避免了对每个数据包分类算法进行分别创建决策树的操作, 能够快速的选择数据包分类算法, 提高选择数据包分类算法的方法的效率。

[0099] 图 4 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。如图 4 所示, 根据第一最大平衡距离, 选择用于对该数据包进行分类的数据包分类算法, 包括:

[0100] 141, 根据该第一窄范围的集合, 确定该第一窄范围的集合对应的第一规则数。

[0101] 应理解, 一条规则在一个域上的窄范围的个数为一个。但是窄范围可以相同, 因此, 规则集的 N 个规则对应最多 N 个窄范围, 而一个窄范围可以对应多个规则, N 为正整数。例如, 规则 1: $[1, 3][2, 4]$, 规则 2: $[1, 3][2, 2]$, 则两个规则在第一域上的窄范围有一个, 即

[1, 3]。该窄范围 [1, 3] 对应两个规则。

[0102] 可选地,作为一个实施例。该第一窄范围对应的第一规则数的获取方式可以为通过扫描该第一窄范围得出。例如,可以通过查询代码,查询该第一窄范围,并可以通过存储代码,将查询到的第一规则数进行存储。

[0103] 142,根据该第一规则数,确定第一判断值。

[0104] 可选地,该第一判断值的确定方法可以根据规则集的规则数计算得到,该计算公式可以经过多次实验总结得出;另外,也可以通过基于预先分析确定某个预定值作为判断值,本发明并不限于此。

[0105] 143,根据该第一最大平衡距离和该第一判断值,选择用于对该数据包进行分类的数据包分类算法。

[0106] 可选地,可以通过比较该第一最大平衡距离和该第一判断值,来选择数据包分类算法,也可以通过其他计算方式来选择数据包分类算法。例如,可以通过数学运算获得阈值,通过判断该阈值的大小来选择数据包分类算法,本发明并不限于此。

[0107] 本发明实施例根据规则集的范围分布情况在数据包分类算法之间进行选择。仅根据规则集的范围创建分割树,根据该分割树确定体现范围分布的均匀性的最大平衡距离。根据最大平衡距离和判断值的比较结果,选择数据包分类算法。避免了对每个数据包分类算法进行分别创建决策树的操作,快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0108] 可选地,作为另一实施例。该第一判断值的确定方法可以根据第一规则数计算得到,计算公式可以为:

$$[0109] \quad X = \frac{\log_2 \text{numRules}}{2}$$

[0110] 其中, X 是该判断值, numRules 是该第一规则数。

[0111] 可选地,作为另一实施例。基于决策树的数据包分类算法可以包括 HyperSplit 算法和 HyperCuts 算法,选择方法则为从该两者中选择合适的包分类算法,对数据包进行分类。在比较该最大平衡距离和判断值时,当该第一最大平衡距离大于该第一判断值时,选择 HyperSplit 算法;在该第一最大平衡距离小于或者等于该第一判断值时,选择 HyperCuts 算法。

[0112] 根据本发明实施例提供的选择数据包分类算法的方法,根据规则集的范围分布情况,创建分割树。比较该分割树的最大平衡距离和根据规则集的规则数确定的判断值,从 HyperSplit 算法和 HyperCuts 算法中选择包分类算法。避免了对每个包分类算法进行分别创建决策树的操作,能够快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0113] 图 5 是根据本发明另一个实施例的选择数据包分类算法的方法的示意性流程图。如图 5 所示,该子树为准平衡子树,该根据该第一分割树,确定第一最大平衡距离,包括:

[0114] 131,根据该第一分割树,确定该第一分割树包含的该准平衡子树,该准平衡子树的 k+1 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数。

[0115] 该第一分割树可以包含多个准平衡子树,其中每个准平衡子树可以相同也可以不

同。该准平衡子树的定义方法可以如下：

[0116]

$$\frac{(k+1) \text{ 层节点数}}{k \text{ 层节点数}} \geq \text{Bratio}$$

[0117] 其中, Bratio 为该第二数值。根据二叉树的特征, k+1 层的节点数与 k 层的节点数的比值的最大值为 2。当 Bratio = 2 时, k+1 层的节点数与 k 层的节点数的比值等于 2, 则该准平衡子树是满二叉树。当 Bratio < 2 时, k+1 层的节点数与 k 层的节点数的比值大于等于 Bratio, 则表示从满二叉树中剔除一些节点所形成的二叉树。

[0118] 可选地, Bratio 可以取 1.5 至 1.8 之间任意数值。例如, 本实施例的 Bratio 可以取 1.52、1.55、1.65、1.75、1.78 等等。Bratio 取值在 1.5 至 1.8 之间为最佳实施例。Bratio 也可以取其他数值使得准平衡子树满足一定条件, 本发明实施例并不限于此。例如, Bratio 可以取 1.4、1.45、1.83、1.85 等接近 1.5 或 1.8 的数值。

[0119] 132, 对该第一分割树进行深度优先遍历, 确定该第一分割树的根节点所在的第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树。

[0120] 该第一分割树的根节点所在的第一准平衡子树可以与该第一分割树的叶子节点所在的第二准平衡子树相同, 也可以不相同。该第一准平衡子树和该第二准平衡子树可以为满足上述准平衡子树的定义的二叉树。

[0121] 根据深度优先遍历方法, 依次深度遍历该第一分割树的子节点, 可以确定该第一分割树的根节点所在的第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树。

[0122] 133, 确定该第一准平衡子树和该第二准平衡子树之间包含的该准平衡子树的最大个数为该第一最大平衡距离。

[0123] 平衡距离可以指叶子节点所在的某颗准平衡子树距离包含树的根节点的准平衡子树之间的准平衡子树的个数。最大平衡距离可以指, 二叉树中平衡距离最大的一个。

[0124] 应理解, 本发明实施例中的第一分割树的叶子节点的个数可以为多个。第一分割树的叶子节点所在的第二准平衡子树的个数也为多个。该多个第二准平衡子树和该第一准平衡子树可以确定对应的多个第一平衡距离。从该多个第一平衡距离中可以取数值最大的作为该第一最大平衡距离。

[0125] 可选地, 作为另一个实施例。该确定最大平衡距离的方法可以由代码实现。具体实现方式可以如下列代码所示：

[0126]

```

Algorithm 2 Measure(root, binth, step)
    global  $D_{\max} = 0$ ;
    {leaf} = getBalanceTreeLeaves(root);
    if step >  $D_{\max}$  then
         $D_{\max} = \text{step}$ ;
    end if
    for leafNode ∈ {leaf} do
        if ChildrenCount(leafNode) == 1 then
            while ChildrenCount(leafNode) == 1 do
                leafNode ← getChildren(leafNode);
            end while
        end if
        if leafNode.numRule < binth or
           ChildrenCount(leafNode) == 0 then
            continue;
        end if
        Measure(leafNode, binth, step+1);
    end for

```

[0127] 其中, *root* 为分割树的根节点, *binth* 为决策树算法的叶子节点中规则数目的上限, getBalanceTreeLeaves(*N*)、ChildrenCount(*Node*) 和 getChildren(*N*) 函数的实现过程可以通过其他代码来实现, 所实现的功能分别为 getBalanceTreeLeaves(*N*) 表示判断并筛选该子树是否满足准平衡子树条件且获取 *N* 为根节点子树的准平衡子树的叶子节点的子节点的集合。ChildrenCount(*Node*) 为获取 *Node* 节点的孩子节点个数。getChildren(*N*) 为获取 *N* 节点的孩子节点。

[0128] 根据本发明实施例提供的选择数据包分类算法的方法, 根据规则集的范围创建分割树。比较该分割树的最大平衡距离和根据窄范围对应的规则数确定的第一判断值, 从 HyperSplit 算法和 HyperCuts 算法中选择数据包分类算法。该方法避免了对每个包分类算法进行分别创建决策树的操作, 能够快速的选择数据包分类算法, 提高选择数据包分类算法的方法的效率。

[0129] 图 6 是根据本发明另一个实施例的选择数据包分类算法的方法的过程的示意性流程图。如图 6 所示, 该选择数据包分类算法的方法的过程如下:

[0130] 在 S201 中, 确定规则集的规则中包括源 IP 地址域对应的范围和目的 IP 地址域对应的范围。

[0131] 该规则集的规则中包括源 IP 地址域对应的范围和目的 IP 地址域对应的范围。扫描该规则集的规则, 确定规则集的规则中包括源 IP 地址域对应的范围和目的 IP 地址域对应的范围。

[0132] 该规则集可以包括多个规则, 该规则集中的每个规则在数据包的多个域上对应多个范围。每个规则均可以包括在源 IP 地址域上的范围。

[0133] 可选地, 作为另一实施例, 规则集的规则在该源 IP 地址域对应的范围和目的 IP 地址域对应的范围的获取方式可以为通过扫描规则集中的每个规则的范围得出。例如, 可以通过查询代码, 查询范围, 并可以通过存储代码, 将查询到的范围进行存储, 以便于在创建

分割树时使用。

[0134] 在 S202 中,对源 IP 地址域对应的范围和目的 IP 地址域对应的范围分别创建分割树。

[0135] 从源 IP 地址域对应的范围中,选择满足范围的长度与该源 IP 地址域的取值范围的长度的比值小于 0.05 的源 IP 窄范围。从目的 IP 地址域对应的范围中,选择满足范围的长度与该目的 IP 地址域的取值范围的长度的比值小于 0.05 的目的 IP 窄范围。

[0136] 根据源 IP 地址域的取值范围,生成分割树的根节点,该根节点代表的区间为该源 IP 地址域的取值范围。例如,数据包的域有四个比特,则该域的取值范围为 [0, 15]。根据该域的取值范围可以确定根节点代表的区间为 [0, 15]。

[0137] 将该根节点的区间取中间值,将该中间值作为分割点,生成该根节点的左右节点代表的两个区间。即生成该分割树的根节点的左右节点。选择该第一分割点的计算方法可以为:

$$[0138] \quad I_m = \frac{I_{\max} - I_{\min} + 1}{2} + I_{\min} - 1$$

[0139] 其中, I_m 是该第一分割点, I_{\max} 是该第一取值范围的最大值, I_{\min} 是该第一取值范围的最小值。例如,该第一取值范围是 [0, 15], 则可以表示为 $[I_{\min}, I_{\max}]$ 。其中, $I_{\min} = 0$, $I_{\max} = 15$, 根据上述公式计算,可以得到 $I_m = 7$ 。因此,该根节点的左节点代表的区间为 [0, 7], 该根节点的右节点代表的区间为 [8, 15]。

[0140] 该分割树除去根节点和该根节点的左右节点的其他子节点的生成过程可以为与该根节点的左右节点的生成过程相同,即通过确定分割点,将父节点的区间分为该父节点的左右节点的两个区间。例如。根节点的左节点代表的区间为 [0, 7], 根据公式计算得到分割点 3。因此,可以得到该左节点的左右两个子节点代表的区间分别为 [0, 3] 和 [4, 7]。

[0141] 以下执行迭代过程,直到该分割树的某个子节点代表的区间为上述选择出的源 IP 窄范围,删除不包含该源 IP 窄范围的子节点。所代表的区间为该源 IP 窄范围的子节点可以作为该分割树的叶子节点。

[0142] 相同的,可以根据以上方法创建根据目的 IP 地址域对应的范围创建的另一个分割树。

[0143] 在 S203 中,根据源 IP 地址域对应的范围创建的分割树,确定该分割树的最大平衡距离 D_{src} ;根据目的 IP 地址域对应的范围创建的分割树,确定该分割树的最大平衡距离 D_{dst} 。

[0144] 确定根据源 IP 地址域对应的范围创建的分割树中包含的准平衡子树,该准平衡子树满足 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值,该第二数值的取值范围为 1.5 至 1.8 之间。该第一分割树可以包含多个准平衡子树,该准平衡子树的定义方法可以如下:

[0145]

$$\frac{(k+1) \text{ 层节点数}}{k \text{ 层节点数}} \geq Bratio$$

[0146] 其中, $Bratio$ 为该第二数值。当 $Bratio = 2$ 时, $k+1$ 层的节点数与 k 层的节点数的比值等于 2, 则该准平衡子树是满二叉树。当 $Bratio < 2$ 时, $k+1$ 层的节点数与 k 层的节点

数的比值大于等于 Bratio,则表示从满二叉树中剔除一些节点所形成的二叉树。

[0147] 确定根节点所在的准平衡子树与多个叶子节点所在的多个准平衡子树之间包含的准平衡子树的多个个数为多个平衡距离,从多个平衡距离中选择最大的作为该根据源 IP 地址域对应的范围创建的分割树的最大平衡距离 D_{src} 。

[0148] 应理解,平衡距离可以指叶子节点所在的某颗准平衡子树距离包含树的根节点的准平衡子树之间的准平衡子树的个数。最大平衡距离可以指,二叉树中平衡距离最大的一个。

[0149] 还应理解,本发明实施例中的分割树的叶子节点的个数可以为多个。该分割树的叶子节点所在的准平衡子树的个数也为多个。该叶子节点所在的多个准平衡子树和该根节点所在的准平衡子树可以确定对应的多个平衡距离。从该多个平衡距离中可以取数值最大的作为该最大平衡距离。

[0150] 相同的,可以根据以上方法确定根据目的 IP 地址域对应的范围创建的分割树的最大平衡距离 D_{dst} 。

[0151] 在 S204 中,取 D_{src} 和 D_{dst} 中的较大值 D_m 。

[0152] 在 S205 中,当 $D_m > \frac{\log_2 numRules}{2}$ 时,选择 HyperSplit 算法。其中,numRules 为该源 IP 窄范围和目的 IP 窄范围对应的规则的个数。

[0153] 在 S206 中,当 $D_m \leq \frac{\log_2 numRules}{2}$ 时,选择 HyperCuts 算法。

[0154] 基于决策树的数据包分类算法可以包括 HyperSplit 算法和 HyperCuts 算法,选择方法则为从该两者中选择合适的包分类算法,对数据包进行分类。

[0155] 根据本发明实施例提供的选择数据包分类算法的方法,根据规则集中的源 IP 地址域对应的范围和目的 IP 地址域对应的范围分别创建分割树,确定该两个分割树对应的两个最大平衡距离中的较大值。比较较大值和源 IP 地址域对应的范围和目的 IP 地址域对应的范围中满足窄范围条件的范围对应的规则的个数,从 HyperSplit 算法和 HyperCuts 算法中选择包分类算法。该方法避免了对每个包分类算法进行分别创建决策树的操作,能够快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0156] 上文中结合图 1 至图 6,详细描述了根据本发明实施例的选择数据包分类算法的方法,下面将结合图 7,详细描述根据本发明实施例的选择数据包分类算法的装置。

[0157] 图 7 示出了根据本发明实施例的处理数据包的装置的示意性框图。图 7 的装置可以实现图 1-6 中描述的选择数据包分类算法的方法的技术方案。如图 7 所示,该选择数据包分类算法的装置 300 包括:

[0158] 第一确定模块 310,用于确定数据包的第一域的第一取值范围和对该数据包进行分类的规则集中的规则在该第一域上的第一范围的集合。

[0159] 数据包的第一域可以为该数据包包头的多个域中的一个域,该第一域的第一取值范围可以为该第一域可以取得的最大范围。该第一取值范围为该第一域的固有属性值,可以根据第一域的类型确定第一域的取值范围。例如,端口域的取值范围为 $[0, 65535]$,IP 地址域的取值范围为 $[0, 2^{(32)}-1]$ 。

[0160] 数据包包头中的域的个数为对该数据包进行分类的规则维度。例如,一个数据

包包头的域包括源 IP 地址,目的 IP 地址,源端口号,目的端口号和协议类型。则对该数据包进行分类的数据包分类规则为:

[0161] 153.0.0.0/8 224.0.0.0/8 0:65535 80:80 TCP->DROP (规则 1)

[0162] 其中,153.0.0.0/8 表示该规则 1 在源 IP 地址域上的范围,224.0.0.0/8 表示该规则 1 在目的 IP 地址域上的范围,0:65535 表示该规则 1 在源端口域上的范围,80:80 表示规则 1 在目的端口域上的范围。该规则 1 为 5 维规则,表示若该数据包包头的域满足源 IP 地址符合 153.0.0.0/8,目的 IP 地址符合 224.0.0.0/8,源端口处于 0:65535,目的端口号为 80,协议号是 TCP 协议时,则执行丢弃操作。

[0163] 该规则集可以包括多个规则,该规则集中的每个规则在数据包的多个域上对应多个范围。每个规则均可以包括在第一域上的第一范围。该第一范围的集合可以表示为规则集的多个规则在第一域上的多个第一范围的集合。例如,规则集中的 N 个规则在第一域上的第一范围的集合包含的第一范围的个数可以为 N 个,也可以少于 N 个,本发明实施例并不限于此。

[0164] 可选地,作为另一实施例,规则集中的规则在该第一域上的第一范围的集合的获取方式可以为通过扫描规则集中的每个规则的范围得出。例如,可以通过查询代码,查询第一范围,并可以通过存储代码,将查询到的第一范围进行存储,以便于在创建第一分割树时使用。

[0165] 生成模块 320,用于根据该第一取值范围和该第一范围的集合,生成第一分割树,该第一分割树的根节点代表的区间为该第一取值范围,该第一分割树的叶子节点代表的区间为该第一范围。

[0166] 具体地,根据该第一取值范围和该第一范围的集合,生成的第一分割树可以用来衡量规则集中的第一范围分布的均匀性。例如,第一范围的分布越均匀,则该第一分割树越接近于满二叉树。

[0167] 根据该第一取值范围,可以确定该第一分割树的根节点代表的区间,根据该第一取值范围和该第一范围的集合,可以确定该第一分割树的子节点代表的区间。

[0168] 应理解,该第一分割树包含一个根节点,该第一分割树的根节点表示一个区间,该根节点代表的一个区间可以为数据包的第一域的取值范围。例如,数据包的域有四个比特,则该域的取值范围为 [0, 15]。根据该域的取值范围可以确定根节点代表的区间为 [0, 15]。

[0169] 还应理解,该第一分割树包含至少一个叶子节点,该第一分割树的每个叶子节点可以代表一个区间,该第一分割树的至少一个叶子节点可以代表与至少一个叶子节点对应的至少一个区间。该第一分割树的至少一个叶子节点的至少一个区间可以为该规则集中的规则在第一域的第一范围。

[0170] 第二确定模块 330,用于根据该第一分割树,确定第一最大平衡距离,该第一最大平衡距离为该第一分割树的根节点所在的第一子树与该第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数。

[0171] 可选地,作为另一实施例,该子树可以为满足一定条件的二叉树。平衡距离可以表示分割树的根节点所在的子树到分割树的叶子节点所在的子树中间包含的子树的个数。最大平衡距离可以表示该分割树所有的平衡距离中最大的一个。根节点所在的子树、叶子节点所在的子树和包含的子树可以相同也可以不同,也可以为同时满足一定条件的二叉树,

本发明实施例并不限于此。

[0172] 该第一分割树的根节点所在的第一子树、该第一分割树的叶子节点所在的第二子树和两者中间包含的子树可以相同也可以不相同,也可以为同时满足一定条件的二叉树,本发明实施例并不限于此。

[0173] 该第一分割树的每个叶子节点都对应一个平衡距离,则该平衡距离中的最大值可以为第一最大平衡距离。

[0174] 选择模块 340,用于根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法。

[0175] 可选地,作为另一实施例,基于决策树的数据包分类算法是性能较优的一个数据包分类算法。基于决策树的数据包分类算法可以包括 HyperSplit 算法和 HyperCuts 算法,选择数据包分类算法的装置可以为从该 HyperSplit 算法和 HyperCuts 算法中选择合适的数据包分类算法,对数据包进行分类。

[0176] 可选地,作为另一实施例,根据该第一最大平衡距离,选择数据包分类算法,可以通过分析该第一最大平衡距离来选择数据包分类算法,也可以将该第一最大平衡距离与某一阈值进行比较,根据比较的结果选择数据包分类算法,本发明实施例并不限于此。

[0177] 本发明实施例提供的选择数据包分类算法的装置,根据规则集的范围分布情况在数据包分类算法之间进行选择,仅需要创建分割树,避免了对每个数据包分类算法进行分别创建决策树的操作。能够根据该分割树,快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0178] 可选地,作为另一实施例。该选择数据包分类算法的装置 300 中,

[0179] 第一确定模块 310,还用于确定该数据包的第二域的第二取值范围和对该数据包进行分类的该规则集中的规则在该第二域上的第二范围的集合;

[0180] 生成模块 320,还用于根据该第二取值范围和该第二范围的集合,生成第二分割树,该第二分割树的根节点代表的区间为该第二取值范围,该第二分割树的叶子节点代表的区间为该第二范围;

[0181] 第二确定模块 330,还用于根据该第二分割树,确定第二最大平衡距离,该第二最大平衡距离为该第二分割树的根节点所在的第三子树与该第二分割树的叶子节点所在的第四子树之间的该子树的最大个数;

[0182] 应理解,310,320,330 在创建第一分割树与创建第二分割树的过程相同,确定第一最大平衡距离与确定第二最大平衡距离的过程也相同。此处为避免重复,在权利要求和说明书中省略确定装置 320,320,330 确定第二最大平衡的步骤。

[0183] 其中,选择模块 340,用于根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法,包括:

[0184] 340A,具体用于确定该第一最大平衡距离和该第二最大平衡距离的较大值;

[0185] 340B,具体用于根据该较大值,选择用于对该数据包进行分类的数据包分类算法。

[0186] 可选地,作为另一实施例。本发明实施例的装置 300 可以针对一个域的范围分布情况进行创建分割树,确定该分割树的最大平衡距离;也可以针对两个域的范围分布情况分别创建两个分割树。根据该两个分割树,分别确定最大平衡距离。根据两者的较大值来选择数据包分类算法。可选地,也可以针对多个域的范围分布情况分别创建对应的多个分

割树。根据该多个分割树,分别确定对应的多个最大平衡距离。根据该多个最大平衡距离中的最大值来选择数据包分类算法,本发明实施例并不限于此。

[0187] 本发明实施例提供的选择数据包分类算法的装置,根据规则集在多个域上的范围的分布情况在数据包分类算法之间进行选择,创建对应的多个分割树。分别确定对应的多个最大平衡距离,根据该多个最大平衡距离的最大值选择数据包分类算法。能够提高选择数据包分类算法的方法的准确性。该装置避免了对每个数据包分类算法进行分别创建决策树的操作,提高选择数据包分类算法的方法的效率和准确性。

[0188] 可选地,作为另一个实施例。该第一域可以为源 IP 地址域,该第二域可以为目的 IP 地址域,或者该第一域为目的 IP 地址域,该第二域为源 IP 地址域。根据源 IP 地址域和目的 IP 地址域对应的第一范围和第二范围,分别生成两个对应的分割树。根据该两个分割树,则分别确定源 IP 地址域对应的最大平衡距离 D_{src} 和目的 IP 地址域对应的最大平衡距离 D_{dst} 。根据该两个最大平衡距离中的较大值来选择数据包分类算法对数据包进行分类。

[0189] 本发明实施例提供的选择数据包分类算法的装置,根据源 IP 地址域对应的范围和目的 IP 地址域对应的范围分别创建分割树,确定该两个分割树对应的两个最大平衡距离中的较大值。比较较大值和从源 IP 地址域对应的范围和目的 IP 地址域对应的范围中选择的满足条件的窄范围对应的规则的个数,从 HyperSplit 算法和 HyperCuts 算法中选择数据包分类算法。该装置避免了对每个数据包分类算法进行分别创建决策树的操作,能够快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0190] 可选地,作为另一实施例。该生成模块 320 生成第一分割树的过程如下:

[0191] 该生成模块 320,具体用于根据该第一取值范围,生成该第一分割树的根节点。

[0192] 具体地,该第一分割树的每个节点代表一个区间,该第一取值范围可以作为该第一分割树的根节点所代表的区间。例如,数据包的域有四个比特,则该域的取值范围为 $[0, 15]$ 。根据该域的取值范围可以确定根节点代表的区间为 $[0, 15]$ 。

[0193] 该生成模块 320,具体用于根据该第一取值范围,确定该第一分割树的根节点的第一分割点。

[0194] 可选地,该第一分割点可以根据经验从该第一取值范围中包含的数值中选择一个数值作为该第一分割点,也可以取该第一取值范围的中间值作为该第一分割点,也可以通过计算确定该第一分割点,本发明实施例并不限于此。

[0195] 可选地,作为另一实施例。选择该第一分割点的计算方法可以为:

$$[0196] \quad I_m = \frac{I_{\max} - I_{\min} + 1}{2} + I_{\min} - 1$$

[0197] 其中, I_m 是该第一分割点, I_{\max} 是该第一取值范围的最大值, I_{\min} 是该第一取值范围的最小值。例如,该第一取值范围是 $[0, 15]$,则可以表示为 $[I_{\min}, I_{\max}]$ 。其中, $I_{\min} = 0$, $I_{\max} = 15$,根据上述公式计算,可以得到 $I_m = 7$ 。

[0198] 该生成模块 320,具体用于从该第一范围的集合中选择第一窄范围的集合,该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值。

[0199] 可以基于第一范围的集合生成第一分割树,该第一范围可以为长度较短的第一窄范围。该第一窄范围可以满足该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值。也可以为基于其他方法定义的自定义的范围,本发明实施例并不限于此。

[0200] 可选地,作为另一实施例。第一域为源 IP 地址域或目的 IP 地址域,该第一数值可以取 0.05。第一域为端口域或协议域,该第一数值可以取 0.5。例如,若规则集中规则的第一范围表示为 (F_L, F_H) , 则该第一范围满足第一窄范围的条件可以为:1) 如果是源 IP 地址域或目的 IP 地址域,则 $(F_H - F_L + 1) / \text{len}(I) < 0.05$ 。2) 如果是端口域或者协议域,则 $(F_H - F_L + 1) / \text{len}(I) < 0.5$ 。其中, I 表示数据包的第一域的第一取值范围, $\text{len}(I)$ 表示该第一域的第一取值范围的长度。

[0201] 可选地,作为另一实施例。该第一数值的取值也可以取其他数值,例如 0.045、0.046、0.047、0.048、0.049、0.051、0.052 等。满足基于该其他数值确定的长度较短的窄范围可以生成分割树,并根据该分割树可以准确地选择数据包分类算法。对于该第一数值的取值,本发明实施例并不限于此。

[0202] 该生成模块 320,具体用于根据该第一取值范围、该第一分割点和该第一窄范围的集合,生成该第一分割树的子节点,该子节点代表的第一范围与该第一窄范围相交。

[0203] 应理解,该第一分割树的子节点包括除根节点的其他节点,该其他节点包括该第一分割树的叶子节点和该第一分割树的中间节点。

[0204] 根据该第一取值范围和该第一分割点,可以生成该第一分割树的根节点的左节点和右节点。具体地,该第一分割点可以把该第一分割树的根节点代表的区间分为该第一分割树的根节点的左右节点分别代表的两个区间。例如,该第一取值范围为 $[0, 15]$, 根据上述公式可以确定分割点为 7。根据该分割点 7,确定该根节点的左右节点代表的区间分别为 $[0, 7]$ 和 $[8, 15]$ 。

[0205] 可选地,作为另一实施例。该第一分割树除去根节点和该根节点的左右节点的其他子节点的生成过程可以为与该根节点的左右节点的生成过程相同,即通过确定分割点,将父节点的区间分为该父节点的左右节点的两个区间。

[0206] 生成的该第一分割树的子节点可以满足该子节点代表的区间与该第一窄范围相交,并删除不满足条件的子节点。可选地,相交的定义可以为:若范围 1 为 (F_{1L}, F_{1H}) , 范围 2 为 (F_{2L}, F_{2H}) , 若不满足 $F_{1H} < F_{2L}$ 或 $F_{2H} < F_{1L}$, 则称这范围 1 和范围 2 相交。

[0207] 以上生成该第一分割树的子节点的迭代过程直到该第一分割树的子节点代表的区间为该第一窄范围时,停止迭代过程。将所代表的区间为第一窄范围的子节点可以作为该第一分割树的叶子节点。例如,该第一取值范围为 $[0, 15]$, 可以确定分割点为 7。根据该分割点 7,确定该根节点的左右节点代表的区间分别为 $[0, 7]$ 和 $[8, 15]$ 。分别确定左右节点作为父节点的分割点 3 和 11,则该左节点的左右节点所代表的两个区间可以为 $[0, 3]$ 和 $[3, 7]$,该右节点的左右节点所代表的两个区间可以为 $[8, 11]$ 和 $[12, 15]$,依次类推可依次确定分割树中每个子节点所代表的区间依次为 $[0, 7]$ 、 $[8, 15]$ 、 $[0, 3]$ 、 $[3, 7]$ 、 $[8, 11]$ 、 $[12, 15]$ 、 $[0, 1]$ 、 $[2, 3]$ 、 $[8, 9]$ 和 $[10, 11]$ 。若第一窄范围包括 $[2, 3]$ 、 $[8, 9]$ 和 $[12, 15]$,在保证该第一分割树的叶子节点所代表的区间为该第一窄范围且该子节点代表的区间与该第一窄范围相交,可以将不满足相交条件的节点删除。因此,依次确定的子节点代表的区间删除 $[3, 7]$ 、 $[0, 1]$ 和 $[10, 11]$ 。因此可以确定该第一分割树的叶子节点代表的区间为 $[2, 3]$ 、 $[8, 9]$ 和 $[12, 15]$ 。

[0208] 该生成模块 320,具体用于根据该第一分割树的根节点和该第一分割树的子节点,生成该第一分割树,该第一分割树的叶子节点代表的区间为该第一窄范围。

[0209] 本发明实施例提供的选择数据包分类算法的装置,根据规则集的范围分布情况在数据包分类算法之间进行选择,仅根据规则集的范围创建分割树,通过该分割树的叶子节点代表的区间量化范围分布的均匀性。该装置避免了对每个数据包分类算法进行分别创建决策树的操作,能够根据该分割树的量化结果,快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0210] 可选地,作为另一实施例,该生成模块的生成第一分割树的方法可以由代码实现。具体实现方式可以如下列代码所示:

[0211]

```

Algorithm 1 Create_SplitTree
 $\phi$ 为根节点
 $Q = \{\phi\}$ ; 设置队列
while !Q.empty() do
 $N \leftarrow Q.pop\_front()$ ;
 $I_m \leftarrow \lfloor \frac{I_{\max} - I_{\min} + 1}{2} \rfloor + I_{\min} - 1$ ; 生成左右孩子节点 $N_L$ 和 $N_R$ 
for  $r \in N.S$  do
if  $r$  与  $[I_{\min}, I_m]$  相交 then
 $N_L.S \leftarrow N_L.S \cup \{r\}$ ;
 $N_L.L(r) += N.L(r)$ ;
end if
if  $r$  与  $[I_m + 1, I_{\max}]$  相交 then
 $N_R.S \leftarrow N_R.S \cup \{r\}$ ;
 $N_R.L(r) += N.L(r)$ ;
end for
if  $N_R.S == N_L.S$  then
删除节点 $N_L$ 和 $N_R$ 
end if
if  $N_L.numRule == 0$  (or  $N_R.numRule == 0$ ) then
删除 $N_L$  (or  $N_R$ )
end if
将剩余的节点放入到队列 $Q$ 结尾;
end while

```

[0212] 可选地,该子节点可以包括以下至少一种信息:

[0213] 1) 该节点所代表的区间, $N.[I_{\min}, I_{\max}]$; 可选地,对于规则中的某个范围,分割树的根节点所代表的范围等于该第一取值范围。

[0214] 2) 所有与 $N.[I_{\min}, I_{\max}]$ 相交的窄范围集合 $N.S$;

[0215] 3) 第一窄范围的规则数 $N.L(r)$, 包含第一窄范围 $r \in N.S$ 的第一规则数且

[0216]
$$\sum_{r \in N.S} N.L(r) = N.numRules$$

[0217] 本发明实施例提供的选择数据包分类算法的装置,根据规则集的范围创建分割树,比较该分割树的最大平衡距离和根据窄范围对应的规则数确定的判断值,从HyperSplit算法和HyperCuts算法中选择包分类算法。该装置避免了对每个包分类算法进行分别创建决策树的操作,能够快速的选择性能较优的包分类算法,提高选择数据包分类

算法的方法的效率。

[0218] 可选地,作为另一实施例。该选择模块 340,

[0219] 用于根据该第一窄范围的集合,确定该第一窄范围的集合对应的第一规则数;

[0220] 应理解,一条规则在一个域上的窄范围的个数为一个。但是窄范围可以相同,因此,规则集的 N 个规则对应最多 N 个窄范围,而一个窄范围可以对应多个规则,N 为正整数。例如,规则 1:[1, 3][2, 4],规则 2:[1, 3][2, 2],则两个规则在第一域上的窄范围有一个,即 [1, 3]。该窄范围 [1, 3] 对应两个规则。

[0221] 可选地,作为一个实施例。该第一窄范围对应的第一规则数的获取方式可以为通过扫描该第一窄范围得出。例如,可以通过查询代码,查询该第一窄范围,并可以通过存储代码,将查询到的第一规则数进行存储。

[0222] 该选择模块 340,用于根据该第一规则数,确定第一判断值。

[0223] 可选地,该第一判断值的确定方法可以根据规则集的规则数计算得到,该计算公式可以经过多次实验总结得出;另外,也可以通过基于预先分析确定某个预定值作为判断值,本发明并不限于此。

[0224] 该选择模块 340,用于根据该第一最大平衡距离和该第一判断值,选择用于对该数据包分类算法进行分类的数据包分类算法。

[0225] 可选地,可以通过比较该第一最大平衡距离和该第一判断值,来选择数据包分类算法,也可以通过其他计算方式来选择数据包分类算法。例如,可以通过数学运算获得阈值,通过判断该阈值的大小来选择数据包分类算法,本发明并不限于此。

[0226] 本发明实施例提供的选择数据包分类算法的装置,根据规则集的范围分布情况在数据包分类算法之间进行选择。仅根据规则集的范围创建分割树,根据该分割树确定体现范围分布的均匀性的最大平衡距离。根据最大平衡距离和判断值的比较结果,选择数据包分类算法。避免了对每个数据包分类算法进行分别创建决策树的操作,快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0227] 可选地,作为另一实施例,该第一判断值的确定方法可以根据第一规则数计算得到,计算公式可以为:

$$[0228] \quad X = \frac{\log_2 \text{numRules}}{2}$$

[0229] 其中, X 是该判断值, numRules 是该第一规则数。

[0230] 可选地,作为另一实施例。基于决策树的数据包分类算法可以包括 HyperSplit 算法和 HyperCuts 算法,选择方法则为从该两者中选择合适的包分类算法,对数据包进行分类。在比较该最大平衡距离和判断值时,当该第一最大平衡距离大于该第一判断值时,选择 HyperSplit 算法;在该第一最大平衡距离小于或者等于该第一判断值时,选择 HyperCuts 算法。

[0231] 根据本发明实施例提供的选择数据包分类算法的装置,根据规则集的范围分布情况,创建分割树。比较该分割树的最大平衡距离和根据规则集的规则数确定的判断值,从 HyperSplit 算法和 HyperCuts 算法中选择包分类算法。避免了对每个包分类算法进行分别创建决策树的操作,能够快速的选择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0232] 可选地,作为另一实施例。该第二确定模块 330,

[0233] 用于根据该第一分割树,确定该第一分割树包含的该准平衡子树,该准平衡子树的 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数。

[0234] 该第一分割树可以包含多个准平衡子树,其中每个准平衡子树可以相同也可以不同。该准平衡子树的定义方法可以如下:

[0235]

$$\frac{(k+1) \text{ 层节点数}}{k \text{ 层节点数}} \geq \text{Bratio}$$

[0236] 其中,Bratio 为该第二数值。根据二叉树的特征, $k+1$ 层的节点数与 k 层的节点数的比值的最大值为 2。当 $\text{Bratio} = 2$ 时, $k+1$ 层的节点数与 k 层的节点数的比值等于 2,则该准平衡子树是满二叉树。当 $\text{Bratio} < 2$ 时, $k+1$ 层的节点数与 k 层的节点数的比值大于等于 Bratio,则表示从满二叉树中剔除一些节点所形成的二叉树。

[0237] 可选地,Bratio 可以取 1.5 至 1.8 之间任意数值。例如,本实施例的 Bratio 可以取 1.53、1.55、1.6、1.65、1.75 等等。Bratio 取值在 1.5 至 1.8 之间为最佳实施例。Bratio 也可以取其他数值使得准平衡子树满足一定条件,本发明实施例并不限于此。例如,Bratio 可以取 1.45、1.48、1.83、1.85 等接近 1.5 或 1.8 的数值。

[0238] 该第二确定模块 330,用于对该第一分割树进行深度优先遍历,确定该第一分割树的根节点所在的第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树。

[0239] 该第一分割树的根节点所在的第一准平衡子树可以与该第一分割树的叶子节点所在的第二准平衡子树相同,也可以不相同。该第一准平衡子树和该第二准平衡子树可以为满足上述准平衡子树的定义的二叉树。

[0240] 根据深度优先遍历方法,依次深度遍历该第一分割树的子节点,可以确定该第一分割树的根节点所在的第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树。

[0241] 该第二确定模块 330,用于确定该第一准平衡子树和该第二准平衡子树之间包含的该准平衡子树的最大个数为该第一最大平衡距离。

[0242] 平衡距离可以指叶子节点所在的某颗准平衡子树距离包含树的根节点的准平衡子树之间的准平衡子树的个数。最大平衡距离可以指,二叉树中平衡距离最大的一个。

[0243] 应理解,本发明实施例中的第一分割树的叶子节点的个数可以为多个。第一分割树的叶子节点所在的第二准平衡子树的个数也为多个。该多个第二准平衡子树和该第一准平衡子树可以确定对应的多个第一平衡距离。从该多个第一平衡距离中可以取数值最大的作为该第一最大平衡距离。

[0244] 可选地,作为另一个实施例。该确定最大平衡距离的方法可以由代码实现。具体实现方式可以如下列代码所示:

[0245]

```

Algorithm 2 Measure(root, binth, step)
    global  $D_{\max} = 0$ ;
    {leaf} = getBalanceTreeLeaves(root);
    if step >  $D_{\max}$  then
         $D_{\max} = step$ ;
    end if
    for leafNode  $\in$  {leaf} do
        if ChildrenCount(leafNode) == 1 then
            while ChildrenCount(leafNode) == 1 do
                leafNode  $\leftarrow$  getChildren(leafNode);
            end while
        end if
        if leafNode.numRule < binth or
            ChildrenCount(leafNode) == 0 then
            continue;
        end if
        Measure(leafNode, binth, step + 1);
    end for

```

[0246] 其中, root 为分割树的根节点, binth 为决策树算法的叶子节点中规则数目的上限, getBalanceTreeLeaves(N)、ChildrenCount(Node) 和 getChildren(N) 函数的实现过程可以通过其他代码来实现, 所实现的功能分别为 getBalanceTreeLeaves(N) 表示判断并筛选该子树是否满足准平衡子树条件且获取 N 为根节点子树的准平衡子树的叶子节点的子节点的集合。ChildrenCount(Node) 为获取 Node 节点的孩子节点个数。getChildren(N) 为获取 N 节点的孩子节点。

[0247] 根据本发明实施例提供的选择数据包分类算法的装置, 根据规则集的范围创建分割树。比较该分割树的最大平衡距离和根据窄范围对应的规则数确定的第一判断值, 从 HyperSplit 算法和 HyperCuts 算法中选择数据包分类算法。避免了对每个包分类算法进行分别创建决策树的操作, 能够快速的选择数据包分类算法, 提高选择数据包分类算法的方法的效率。

[0248] 另外需要说明的是, 图 7 中的各个部件可以由硬件实现, 或者在硬件基础上通过软件实现。

[0249] 图 8 是本发明另一实施例的选择数据包分类算法的选择设备 80 的框图。图 8 的选择设备 80 包括处理器 81 和存储器 82。处理器 81 和存储器 82 通过总线系统 83 相连。

[0250] 处理器 81 控制选择设备 80 的操作。存储器 82 可以包括只读存储器和随机存取存储器, 并向处理器 81 提供指令和数据。存储器 82 的一部分还可以包括非易失性随机存取存储器 (NVRAM)。选择设备 80 的各个组件通过总线系统 83 耦合在一起, 其中总线系统 83 除包括数据总线之外, 还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见, 在图中将各种总线都标为总线系统 83。

[0251] 处理器 81 可能是一种集成电路芯片, 具有信号的处理能力。上述的处理器 81 可以是通用处理器、数字信号处理器 (DSP)、专用集成电路 (ASIC)、现成可编程门阵列 (FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处

理器也可以是任何常规的处理器等。处理器 81 读取存储器 82 中的信息,结合其硬件控制处理设备 80 的各个部件。

[0252] 图 1- 图 6 的方法可以在图 8 的选择设备 80 中实现,或者图 7 的选择数据包分类算法的装置可以由图 8 的选择设备 80 实现。为避免重复,不再详细描述。

[0253] 具体地,处理器 81,用于通过总线 83 调用存储在存储器 82 中的代码,用于确定数据包的第一域的第一取值范围和对该数据包进行分类的规则集中的规则在该第一域上的第一范围的集合;用于根据该第一取值范围和该第一范围的集合,生成第一分割树,该第一分割树的根节点代表的区间为该第一取值范围,该第一分割树的叶子节点代表的区间为该第一范围;用于根据该第一分割树,确定第一最大平衡距离,该第一最大平衡距离为该第一分割树的根节点所在的第一子树与该第一分割树的叶子节点所在的第二子树之间包含的子树的最大个数;用于根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法。

[0254] 根据本发明实施例提供的选择数据包分类算法的选择设备,根据规则集的范围分布情况,创建分割树,比较该分割树的最大平衡距离和根据规则集的规则数确定的判断值,从 HyperSplit 算法和 HyperCuts 算法中选择包分类算法,避免了对每个包分类算法进行分别创建决策树的操作,能够快速的选择性能较优的包分类算法,提高选择数据包分类算法的方法的效率。

[0255] 上述本发明实施例揭示的方法可以应用于处理器 81 中,或者由处理器 81 实现。处理器 81 可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器 81 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 81 可以是通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存取存储器 (Random Access Memory, RAM)、闪存、只读存储器 (Read-Only Memory, ROM)、可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 920,处理器 81 读取存储器 920 中的信息,结合其硬件完成上述方法的步骤。

[0256] 可选地,作为另一实施例,处理器 81 还用于确定该数据包的第二域的第二取值范围和对该数据包进行分类的该规则集中的规则在该第二域上的第二范围的集合;还用于根据该第二取值范围和该第二范围的集合,生成第二分割树,该第二分割树的根节点代表的区间为该第二取值范围,该第二分割树的叶子节点代表的区间为该第二范围;还用于根据该第二分割树,确定第二最大平衡距离,该第二最大平衡距离为该第二分割树的根节点所在的第三子树与该第二分割树的叶子节点所在的第四子树之间的该子树的最大个数;其中,处理器 81 用于根据该第一最大平衡距离,选择用于对该数据包进行分类的数据包分类算法,包括:具体用于确定该第一最大平衡距离和该第二最大平衡距离的较大值;具体用于根据该较大值,选择用于对该数据包进行分类的数据包分类算法。

[0257] 可选地,作为另一实施例。该第一域为源 IP 地址域,该第二域为目的 IP 地址域,或者该第一域为目的 IP 地址域,该第二域为源 IP 地址域。

[0258] 可选地,作为另一实施例,处理器 81,具体用于根据该第一取值范围,生成该第一分割树的根节点;具体用于根据该第一取值范围,确定该第一分割树的根节点的第一分割点;具体用于从该第一范围的集合中选择第一窄范围的集合,该第一窄范围的长度与该第一取值范围的长度的比值小于第一数值;具体用于根据该第一取值范围、该第一分割点和该第一窄范围的集合,生成该第一分割树的子节点,该子节点代表的区间与该第一窄范围相交;具体用于根据该第一分割树的根节点和该第一分割树的子节点,生成该第一分割树,该第一分割树的叶子节点代表的区间为该第一窄范围。

[0259] 可选地,作为另一实施例。处理器 81,具体用于根据该第一取值范围,确定该第一分割树的根节点的第一分割点,包括:

[0260] 根据公式 $I_m = \frac{I_{\max} - I_{\min} + 1}{2} + I_{\min} - 1$ 确定该第一分割点;

[0261] 其中, I_m 是该第一分割点, I_{\max} 是该第一取值范围的最大值, I_{\min} 是该第一取值范围的最小值。

[0262] 可选地,作为另一实施例。处理器 81,用于根据该第一窄范围的集合,确定该第一窄范围的集合对应的第一规则数;用于根据该第一规则数,确定第一判断值;用于根据该第一最大平衡距离和该第一判断值,选择用于对该数据包进行分类的数据包分类算法。

[0263] 可选地,作为另一实施例。处理器 81,用于根据该第一规则数,确定第一判断值,包括:

[0264] 根据公式 $X = \frac{\log_2 \text{numRules}}{2}$ 确定该第一判断值;

[0265] 其中, X 是该第一判断值, numRules 是该第一规则数;其中,处理器 81,

[0266] 具体用于在该第一最大平衡距离大于该第一判断值时,选择 HyperSplit 算法;具体用于在该第一最大平衡距离小于或者等于该第一判断值时,选择 HyperCuts 算法。

[0267] 可选地,作为另一实施例。处理器 81,用于根据该第一分割树,确定该第一分割树包含的该准平衡子树,该准平衡子树的 $k+1$ 层的节点数与 k 层的节点数的比值大于或者等于第二数值, k 为大于或者等于 1 的正整数;用于对该第一分割树进行深度优先遍历,确定该第一分割树的根节点所在第一准平衡子树和该第一分割树的叶子节点所在的第二准平衡子树;用于确定该第一准平衡子树和该第二准平衡子树之间包含的该准平衡子树的最大个数为该第一最大平衡距离。

[0268] 可选地,作为另一实施例。该第一域为源 IP 地址域或目的 IP 地址域,该第一数值取 0.05;该第一域为端口域或协议域,该第一数值取 0.5。

[0269] 可选地,作为另一实施例。该第二数值取 1.5-1.8 之间任一数值。

[0270] 根据本发明实施例提供的选择数据包分类算法的选择设备,根据规则集中的源 IP 地址域对应的范围和目的 IP 地址域对应的范围分别创建分割树,确定该分割树对应的两个最大平衡距离中的较大值。比较较大值和源 IP 地址域对应的范围和目的 IP 地址域对应的范围中满足窄范围条件的范围对应的规则的个数,从 HyperSplit 算法和 HyperCuts 算法中选择包分类算法。避免了对每个包分类算法进行分别创建决策树的操作,能够快速的选择

择数据包分类算法,提高选择数据包分类算法的方法的效率。

[0271] 应理解,在本发明的各种实施例中,上述各过程的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不对本发明实施例的实施过程构成任何限定。

[0272] 另外,本文中术语“系统”和“网络”在本文中常被可互换使用。本文中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0273] 应理解,在本发明实施例中,“与A相应的B”表示B与A相关联,根据A可以确定B。但还应理解,根据A确定B并不意味着仅仅根据A确定B,还可以根据A和/或其它信息确定B。

[0274] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0275] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0276] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、装置或单元的间接耦合或通信连接,也可以是电的,机械的或其它的形式连接。

[0277] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本发明实施例方案的目的。

[0278] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以是两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0279] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到本发明可以用硬件实现,或固件实现,或它们的组合方式来实现。当使用软件实现时,可以将上述功能存储在计算机可读介质中或作为计算机可读介质上的一个或多个指令或代码进行传输。计算机可读介质包括计算机存储介质和通信介质,其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是计算机能够存取的任何可用介质。以此为例但不限于:计算机可读介质可以包括RAM、ROM、EEPROM、CD-ROM或其他光盘存储、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期

望的程序代码并能够由计算机存取的任何其他介质。此外。任何连接可以适当的成为计算机可读介质。例如，如果软件是使用同轴电缆、光纤光缆、双绞线、数字用户线 (DSL) 或者诸如红外线、无线电和微波之类的无线技术从网站、服务器或者其他远程源传输的，那么同轴电缆、光纤光缆、双绞线、DSL 或者诸如红外线、无线和微波之类的无线技术包括在所属介质的定影中。如本发明所使用的，盘 (Disk) 和碟 (disc) 包括压缩光碟 (CD)、激光碟、光碟、数字通用光碟 (DVD)、软盘和蓝光光碟，其中盘通常磁性的复制数据，而碟则用激光来光学的复制数据。上面的组合也应当包括在计算机可读介质的保护范围之内。

[0280] 总之，以上所述仅为本发明技术方案的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

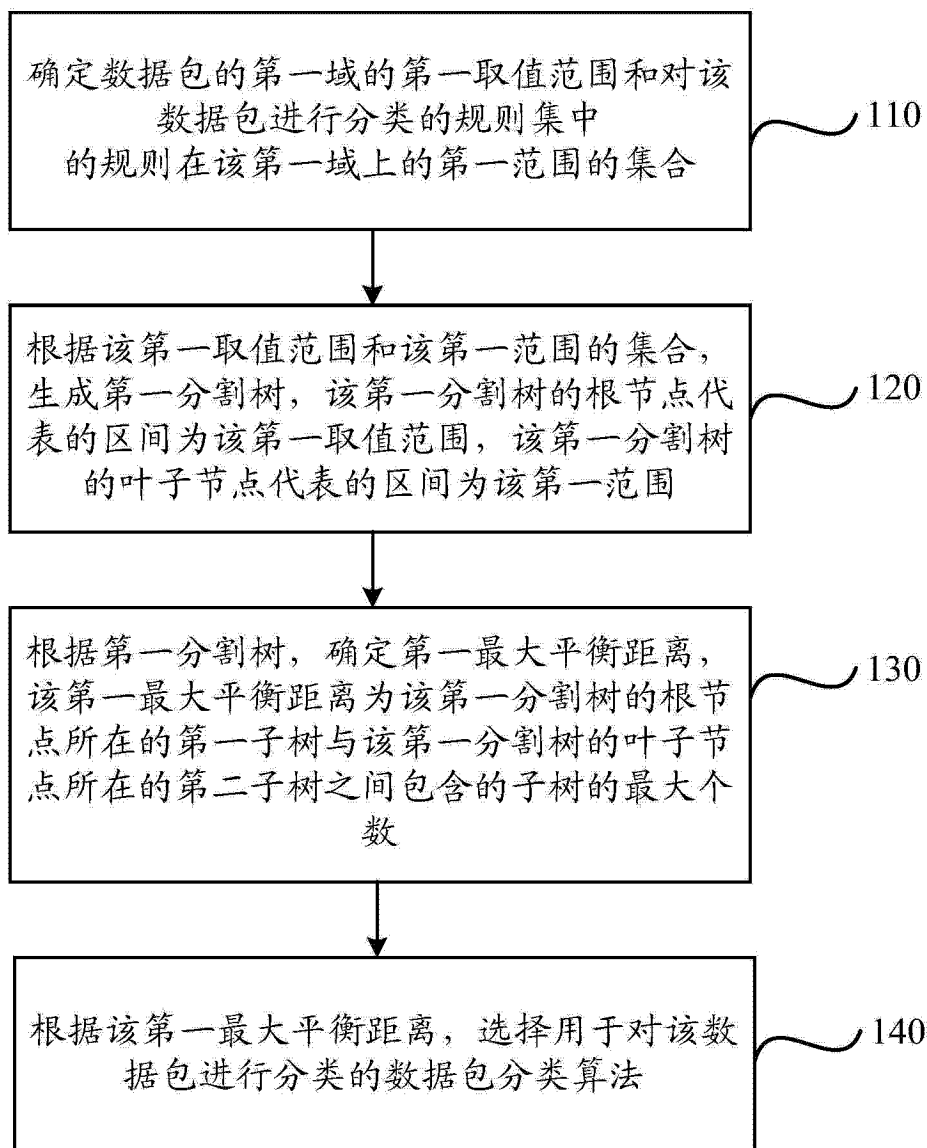


图 1

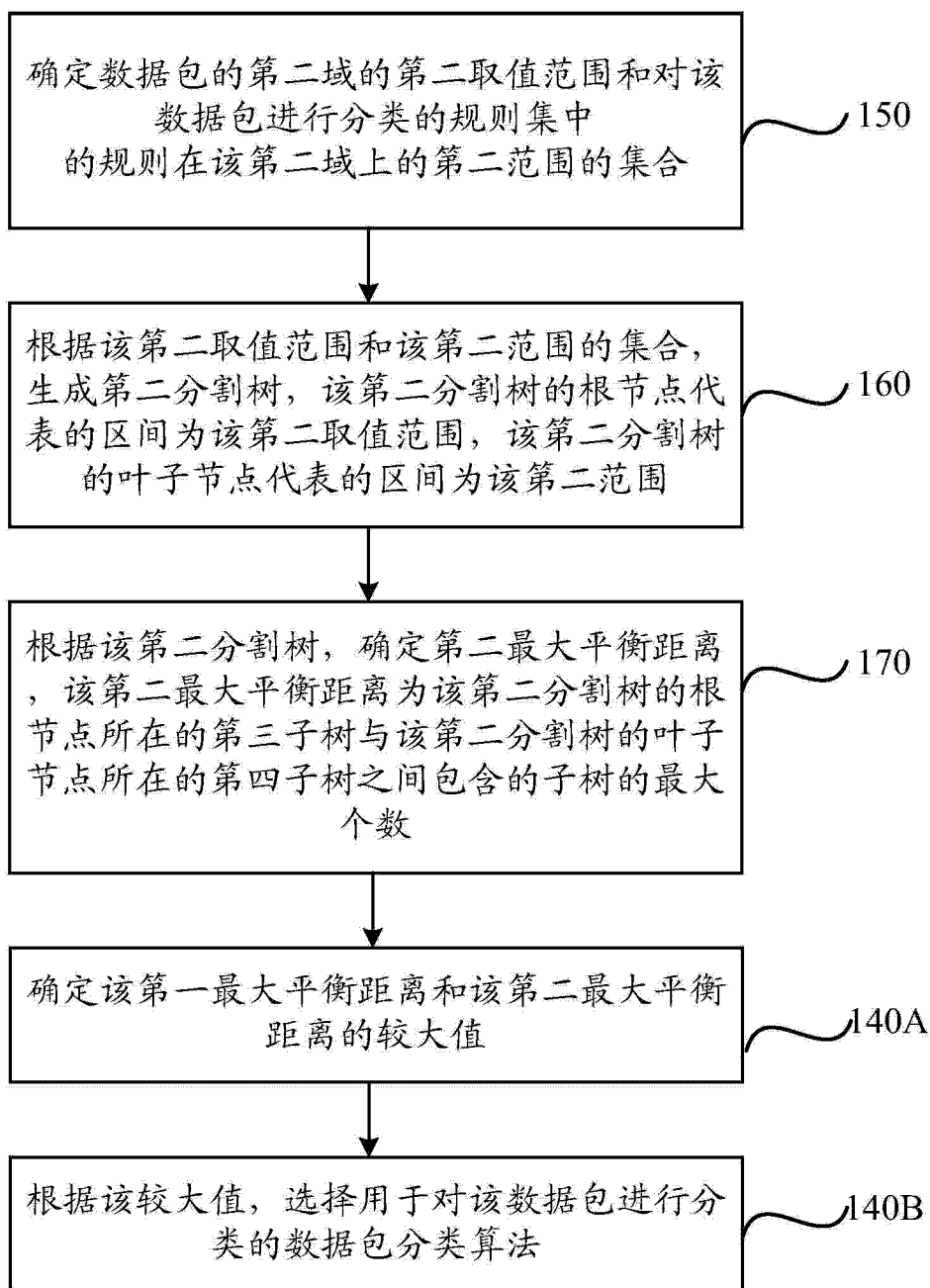


图 2

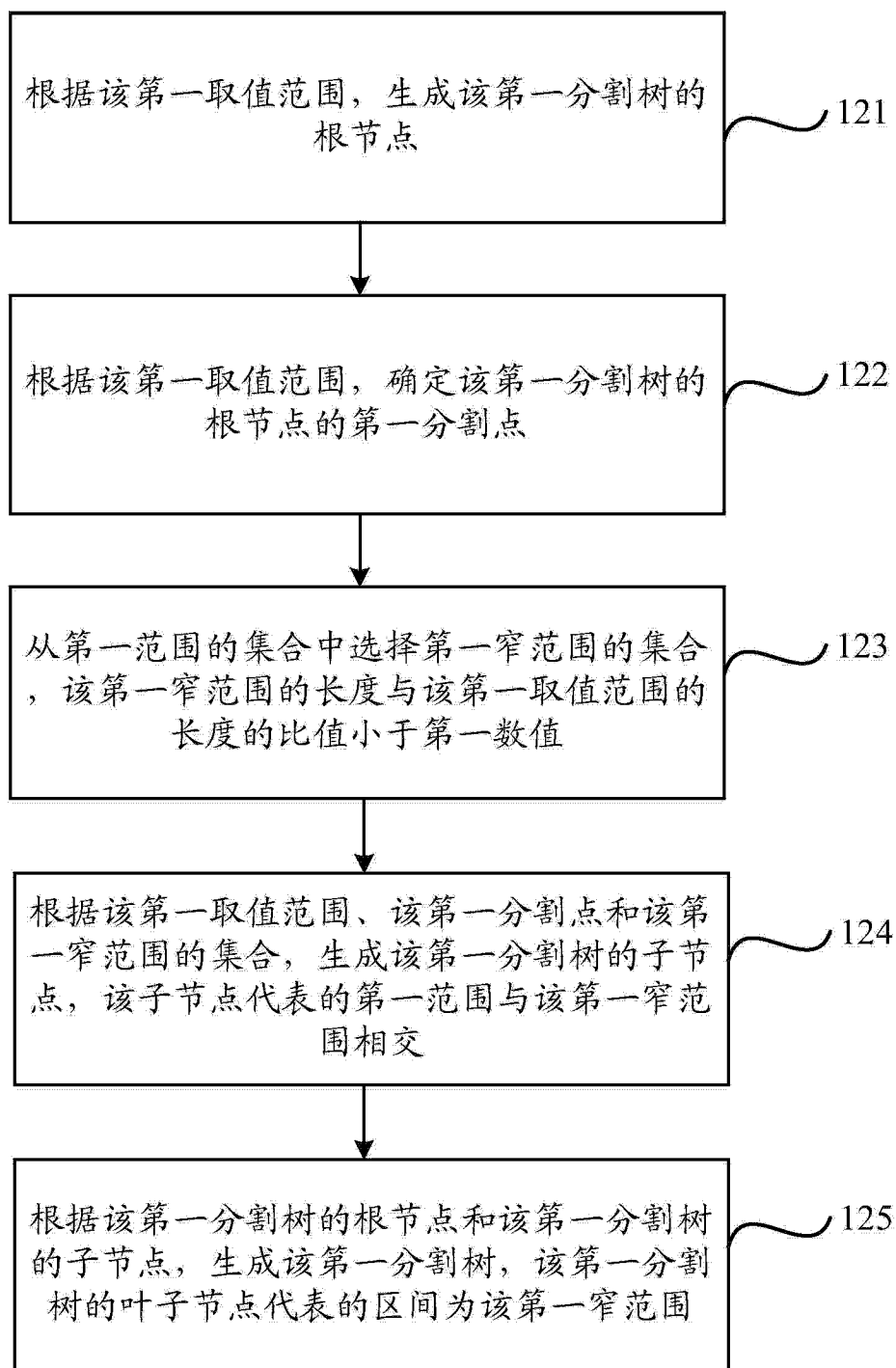


图 3

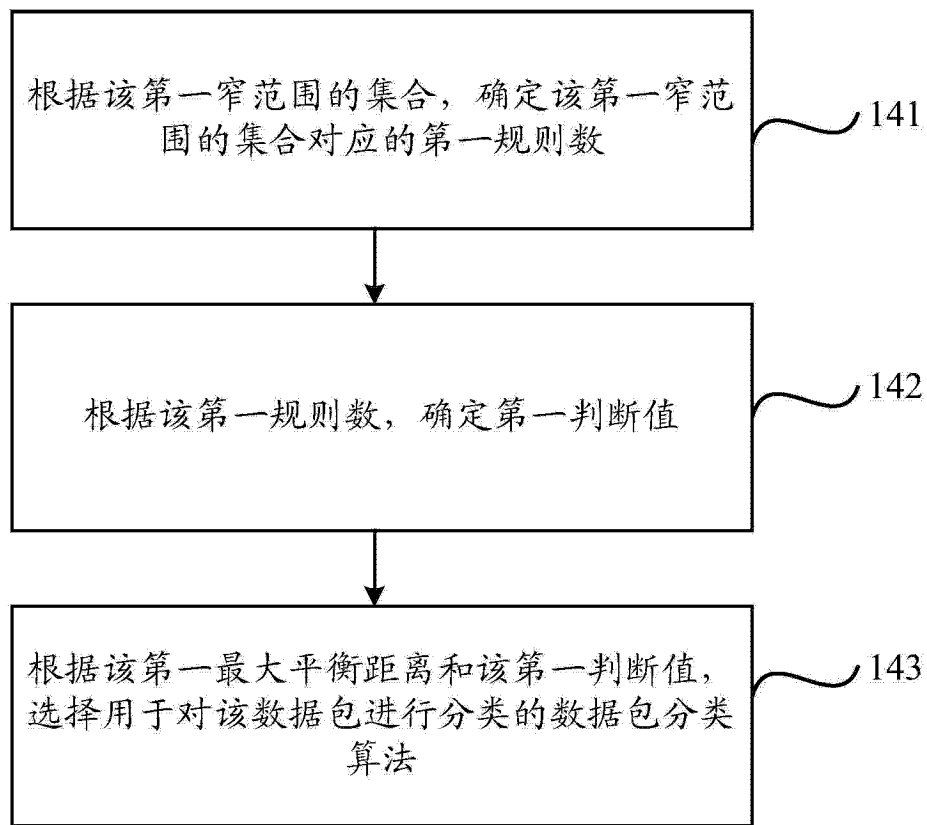


图 4

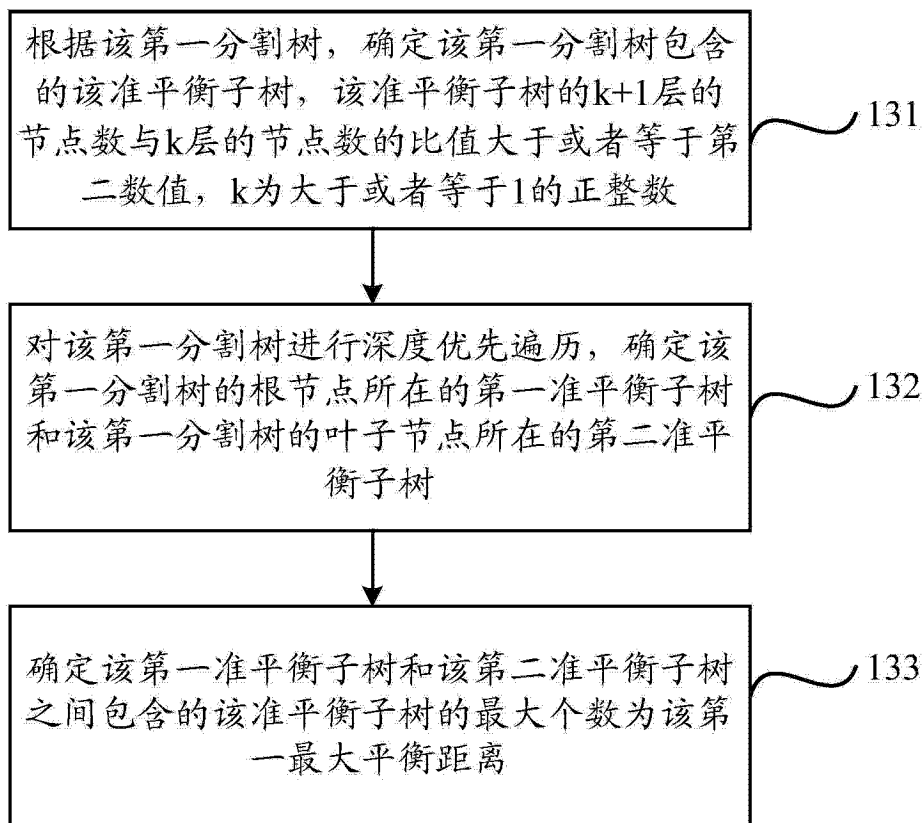


图 5

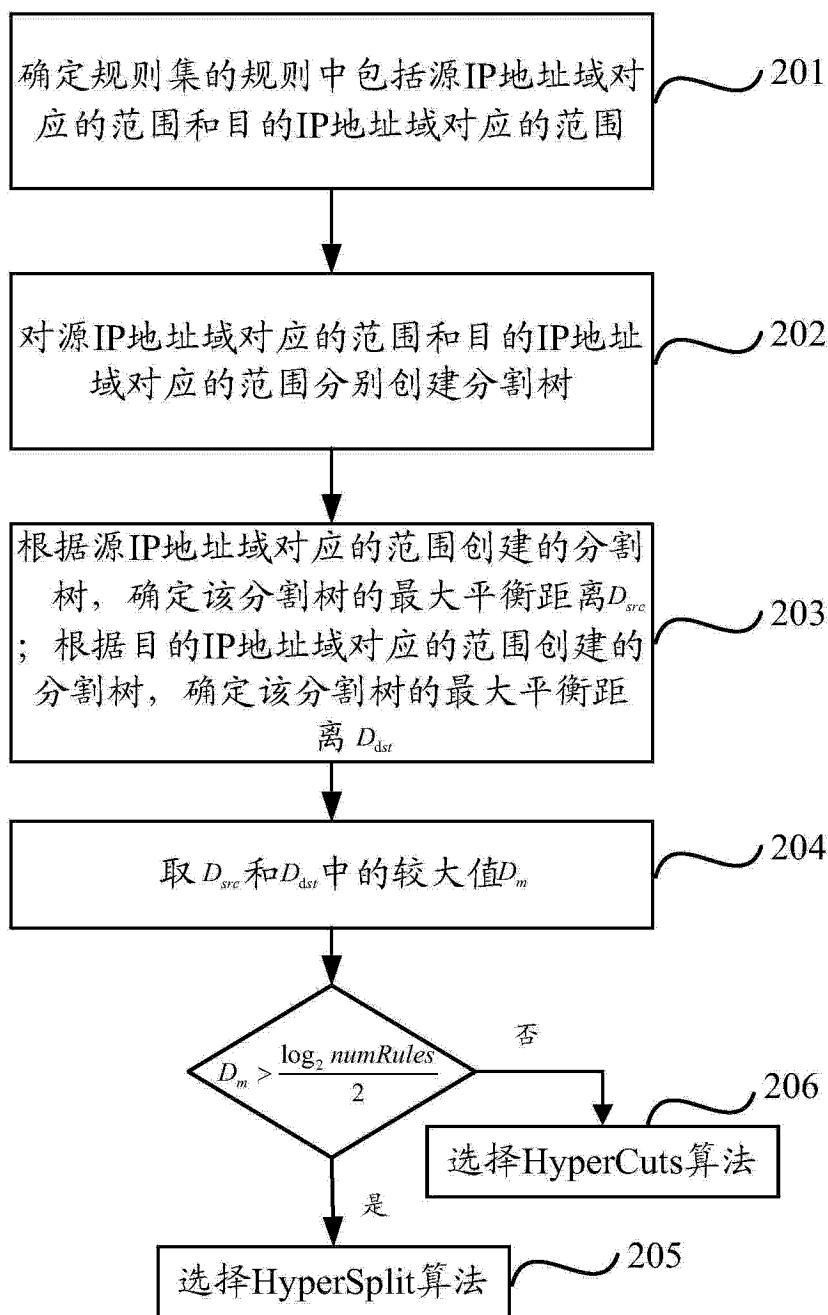


图 6

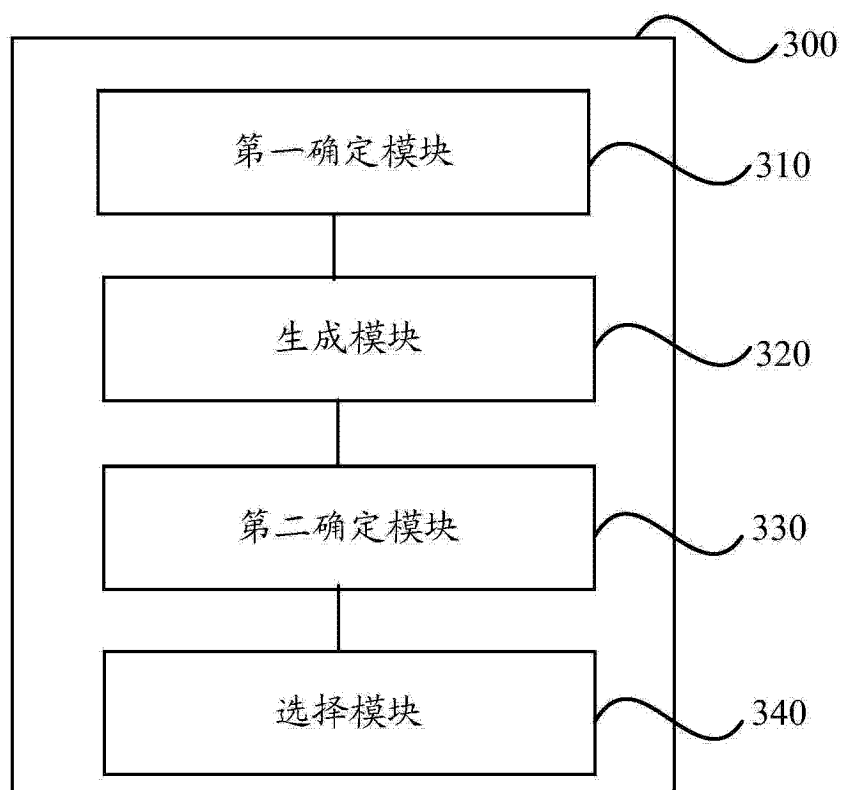


图 7

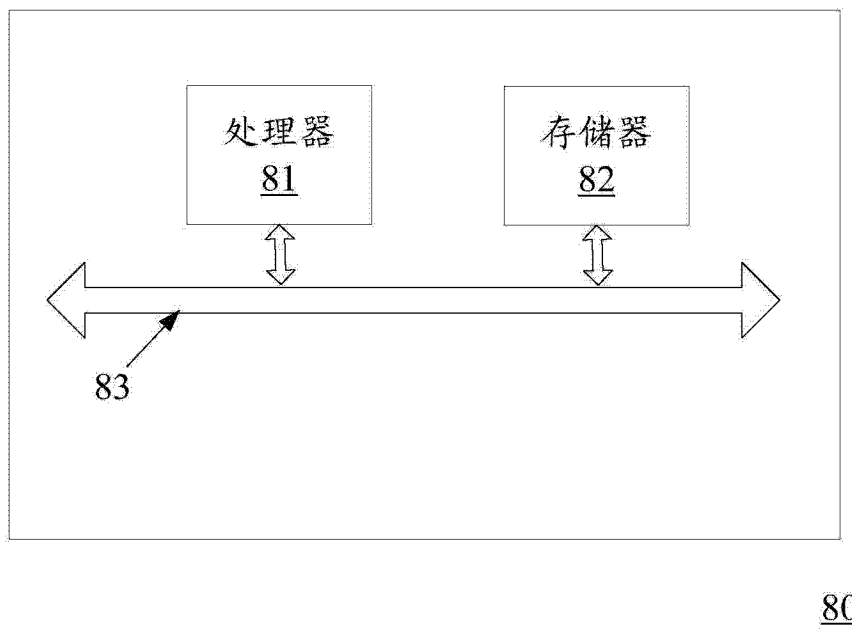


图 8