(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117407564 A
(43)申请公布日 2024.01.16

(21)申请号 202311388119.6

(22)申请日 2023.10.25

(71)申请人 江苏省未来网络创新研究院
地址 211111 江苏省南京市江宁区秣周东路7号

(72)发明人 张广兴 姜海洋 吴颖 金宇翔 田利荣 梁帅

(74)专利代理机构 南京理工信达知识产权代理有限公司 32542
专利代理师 彭甲临

(51)Int.Cl.
G06F 16/901(2019.01)
G06F 16/953(2019.01)
G06F 16/951(2019.01)
H04L 41/12(2022.01)

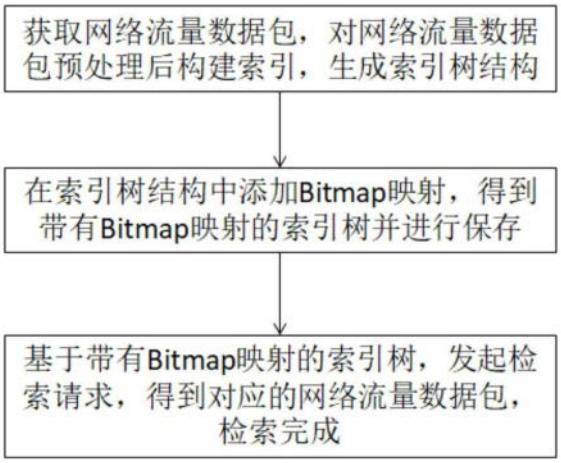
权利要求书2页 说明书6页 附图2页

(54)发明名称

基于HashTrie和Bitmap的网络流量索引优化策略的方法及系统

(57)摘要

本发明涉及一种基于HashTrie和Bitmap的网络流量索引优化策略的方法及系统,该方法包括:获取网络流量数据包,对网络流量数据包预处理后构建索引,生成索引树结构;在索引树结构中添加Bitmap映射,并进行保存;发起检索请求,得到对应的网络流量数据包,检索完成。本发明与现有技术相比,其显著优点是:在HashTrie算法基础上提出了bitmap-HashTrie算法,即在Hash Trie索引上添加bitmap位图映射,该算法融合了bitmap和HashTrie两种算法优势,满足大流量场景下快速创建索引,并且基于bitmap索引导出减少了索引空间占用,同时对于检索通过先匹配bitmap位图进行初步筛选,提升了检索效率。



CN 117407564 A

1.一种基于HashTrie和Bitmap的网络流量索引优化策略的方法,其特征在于:所述方法包括:

获取网络流量数据包,对网络流量数据包预处理后构建索引,生成索引树结构;

在索引树结构中添加Bitmap映射,得到带有Bitmap映射的索引树并进行保存;

基于带有Bitmap映射的索引树,发起检索请求,得到对应的网络流量数据包,检索完成。

2.根据权利要求1所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法,其特征在于:所述预处理包括对网络流量数据包进行清洗、解码、解析处理,经预处理后获得网络流量数据包的五元组信息的属性值。

3.根据权利要求2所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法,其特征在于:所述索引树结构为HashTrie树。

4.根据权利要求1所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法,其特征在于:在所述索引树的每个树节点前面添加Bitmap映射,所述Bitmap映射用于指示每个树节点对应所述五元组信息的任意属性值。

5.根据权利要求4所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法,其特征在于:在所述所引述的每个树节点前面添加Bitmap映射后,采用中序遍历每个树节点,导出属性值不为0的树节点并按序依次写入索引文件,其中:所述索引文件中的偏移量按照导出的树节点的属性值进行排序。

6.根据权利要求5所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法,其特征在于:所述发起检索请求,包括:

输入检索数据,获取检索数据的属性值;

将检索数据的属性值与带有Bitmap映射的索引树进行匹配;

若匹配的,得到检索数据对应的索引位置,提取所述索引位置的网络流量数据包;反之未匹配的,直接返回检索失败。

7.一种基于HashTrie和Bitmap的网络流量索引优化策略的系统,其特征在于:所述系统包括:

预处理模块,用于接收网络流量数据包并进行预处理,提取出网络流量数据包的五元组信息;

索引构建模块,用于对预处理后的网络流量数据包进行索引构建,生成索引树结构;

索引树的bitmap映射模块,用于在索引树的每个树节点前面添加Bitmap映射,

Bitmap映射用于指示任意树节点是否存在于任意属性值上;

索引导出模块:对带有Bitmap映射的索引树进行保存,并在只导出属性值不为0的树节点;

检索模块:根据带有Bitmap映射的索引树,定位到目标网络流量数据包并返回检索结果。

8.一种计算机系统,其特征在于,包括:存储器和处理器;

所述存储器,用于存储程序;

所述处理器,用于执行所述程序,实现如权利要求1至6中任一项所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法的各个步骤。

9.一种可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时,实现如权利要求1至6中任一项所述的基于HashTrie和Bitmap的网络流量索引优化策略的方法的各个步骤。

基于HashTrie和Bitmap的网络流量索引优化策略的方法及系统

技术领域

[0001] 本发明涉及网络流量处理技术领域,特别是一种基于HashTrie和Bitmap的网络流量索引优化策略的方法及系统。

背景技术

[0002] 随着互联网的快速发展,网络流量日益增加,通常小型网络或企业规模的网络都已达到40Gbps,甚至100Gb ps,导致网络数据传输量巨大,进而使得网络流量监测和管理变得越来越重要。为满足网络取证和监管等需求,网络存储系统既要保证流量实时全流量存储并实时创建索引,又要保证在这种大规模网络环境下索引空间占用要足够小的前提下高效地检索数据,并且灵活适用于各种网络环境。但当前的网络流量索引系统技术仍存在以下问题:

[0003] (1)基于包的K叉树算法,虽然它是近年比较流行的网络流量索引方法,但其主要缺陷就是其索引的空间开销较大,尤其是对于诸如IPv6地址等字段长度较长的属性,或者共享前缀特征比较差的流量,其索引结构的空间规模会急剧膨胀。

[0004] (2)中国专利CN113139100B公开了一种基于包的Hash Trie算法,根据网络流量特点在K叉树基础上进行了剪枝,利用数据包被索引属性的某个前缀,而不是完整的属性值,将它和其它类型的数据包(那些索引属性值不同的数据包)区分开来。Hash Trie的数据结构在线提取属性最短不相交前缀建立索引,缩短其参与索引建立的字段长度,进而提升建索引的时空效率,但是它没有考虑实际网络中在有限集合中属性值并非占用全部地址空间,即有些属性值可能为空的情况,在将索引导入文件时,将其为空的属性值也导入磁盘文件,便会造成不必要的时间和空间的浪费。

[0005] 总之,当前的网络存储技术存在索引占用空间大、查询响应慢等问题,难以满足现代网络流量监测和管理的需求。通过以上分析,现有的流量索引方法均存在不同程度的缺陷,不能够真正运用到高速网络流量场景,为此迫切需要设计实现一种在大流量场景索引创建速率够块、索引空间占用足够小情况下能够对历史流量数据进行快速查询,并能够灵活应对不同网络环境下的挑战和变化的方法和系统。

发明内容

[0006] 本发明的目的在于提供一种基于HashTrie和Bitmap的网络流量索引优化策略的方法及系统,在Hash Trie算法基础上提出了bitmap-HashTrie算法,即在Hash Trie索引上添加bitmap位图映射,满足大流量场景下快速创建索引,并且基于bitmap索引导出达到减少索引空间占用的目的。

[0007] 实现本发明目的的技术解决方案为:

[0008] 一种基于HashTrie和Bitmap的网络流量索引优化策略的方法,该方法包括:

[0009] 获取网络流量数据包,对网络流量数据包预处理后构建索引,生成索引树结构;

[0010] 在索引树结构中添加Bitmap映射,得到带有Bitmap映射的索引树并进行保存;

[0011] 基于带有Bitmap映射的索引树,发起检索请求,得到对应的网络流量数据包,检索完成。

[0012] 进一步的,预处理包括对网络流量数据包进行清洗、解码、解析处理,经预处理后获得网络流量数据包的五元组的属性值。

[0013] 进一步的,索引树结构为HashTrie树。

[0014] 进一步的,在索引树的每个树节点前面添加Bitmap映射,Bitmap映射用于指示每个树节点对应五元组信息的任意属性值。

[0015] 进一步的,在所引述的每个树节点前面添加Bitmap映射后,采用中序遍历每个树节点,导出属性值不为0的树节点并按序依次写入索引文件,其中:索引文件中的偏移量按照导出的树节点的属性值进行排序。

[0016] 进一步的,发起检索请求,包括:

[0017] 输入检索数据,获取检索数据的属性值;

[0018] 将检索数据的属性值与带有Bitmap映射的索引树进行匹配;

[0019] 若匹配的,得到检索数据对应的索引位置,提取索引位置的网络流量数据包;反之未匹配的,直接返回检索失败。

[0020] 一种基于HashTrie和Bitmap的网络流量索引优化策略的系统,该系统包括:

[0021] 预处理模块,用于接收网络流量数据包并进行预处理,提取出网络流量数据包的源IP地址、目的IP地址、协议类型信息;

[0022] 索引构建模块,用于对预处理后的网络流量数据包进行索引构建,生成索引树结构;

[0023] 索引树的bitmap映射模块,用于在索引树的每个树节点前面添加Bitmap映射,Bitmap映射用于指示任意树节点是否存在于任意属性值上;

[0024] 索引导出模块:对带有Bitmap映射的索引树进行保存,并在只导出属性值不为0的树节点;

[0025] 检索模块:根据带有Bitmap映射的索引树,定位到目标网络流量数据包并返回检索结果。

[0026] 一种计算机系统,包括存储器和处理器;其中:

[0027] 存储器,用于存储程序;

[0028] 处理器,用于执行程序,实现如基于HashTrie和Bitmap的网络流量索引优化策略的方法的各个步骤。

[0029] 一种可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时,实现如基于HashTrie和Bitmap的网络流量索引优化策略的方法的各个步骤。

[0030] 本发明与现有技术相比,其显著优点是:

[0031] (1)提高空间效率:通过在HashTrie的基础上添加Bitmap映射,减少了索引导出时空间占用;

[0032] (2)实现快速检索:添加Bitmap的位图方式使得检索变得非常快速。通过Bitmap映射可以快速确定是否存在符合条件的节点,对于bitmap匹配不中直接返回检索结果,匹配中的再继续匹配HashTrie等索引结构查询,提高检索效率;

[0033] (3)具备灵活性和可扩展性:Bitmap和属性HashTrie的结合使得索引更加灵活和可扩展。当需要增加新的属性信息或者处理更大规模的数据包时,只需要增加更多的位即可,而属性HashTrie可以很好地适应数据包的变化和增长。这使得索引可以灵活地适应不同网络环境的需求。

[0034] (4)达到可重复性:采用本发明的索引方法和系统可以重复使用,为其他类似系统提供了一种参考和借鉴,使得其更具有可操作性和可复制性。

附图说明

[0035] 图1是本发明的基于HashTrie和Bitmap的网络流量索引优化策略的方法的流程图。

[0036] 图2是本发明中实施例的索引结构从内存到磁盘的转换示意图。

[0037] 图3为本发明中实施例的IP地址检索示意图。

[0038] 图4为本发明的基于HashTrie和Bitmap的网络流量索引优化策略的系统的结构图。

具体实施方式

[0039] 以下结合附图,详细说明本发明的实施方式。

[0040] 如图1所示,一种基于HashTrie和Bitmap的网络流量索引优化策略的方法,该方法包括:

[0041] 获取网络流量数据包,对网络流量数据包预处理后构建索引,生成索引树结构;

[0042] 在索引树结构中添加Bitmap映射,得到带有Bitmap映射的索引树并进行保存;

[0043] 基于带有Bitmap映射的索引树,发起检索请求,得到对应的网络流量数据包,检索完成。

[0044] 具体的,预处理包括对网络流量数据包进行清洗、解码、解析处理,经预处理后获得网络流量数据包的五元组信息的属性值。

[0045] 具体的,索引树结构为HashTrie树。

[0046] 具体的,在索引树的每个树节点前面添加Bitmap映射,Bitmap映射用于指示每个树节点对应五元组信息的任意属性值。

[0047] 具体的,在所引述的每个树节点前面添加Bitmap映射后,采用中序遍历每个树节点,导出属性值不为0的树节点并按序依次写入索引文件,其中:索引文件中的偏移量按照导出的树节点的属性值进行排序。

[0048] 具体的,发起检索请求,包括:

[0049] 输入检索数据,获取检索数据的属性值;

[0050] 将检索数据的属性值与带有Bitmap映射的索引树进行匹配;

[0051] 若匹配的,得到检索数据对应的索引位置,提取索引位置的网络流量数据包;反之未匹配的,直接返回检索失败。

[0052] 一种基于HashTrie和Bitmap的网络流量索引优化策略的系统,该系统包括:

[0053] 预处理模块,用于接收网络流量数据包并进行预处理,提取出网络流量数据包的五元组信息;

[0054] 索引构建模块,用于对预处理后的网络流量数据包进行索引构建,生成索引树结构;

[0055] 索引树的bitmap映射模块,用于在索引树的每个树节点前面添加Bitmap映射,Bitmap映射用于指示任意树节点是否存在于任意属性值上;

[0056] 索引导出模块:对带有Bitmap映射的索引树进行保存,并在只导出属性值不为0的树节点;

[0057] 检索模块:根据带有Bitmap映射的索引树,定位到目标网络流量数据包并返回检索结果。

[0058] 一种计算机系统,包括存储器和处理器;其中:

[0059] 存储器,用于存储程序;

[0060] 处理器,用于执行程序,实现如基于HashTrie和Bitmap的网络流量索引优化策略的方法的各个步骤。

[0061] 一种可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时,实现如基于HashTrie和Bitmap的网络流量索引优化策略的方法的各个步骤。

[0062] 上述模块之间相互配合,共同实现高效的网络流量索引功能。由此可见,本发明的基于HashTrie和Bitmap的网络流量索引优化策略的方法和系统能够广泛应用于各种类型的网络环境中并有效地解决现有技术中存在的缺陷和不足之处通过减少索引导出时空间占用和提升检索效率,本发明的实施例可以为现代网络流量监测和管理提供更加高效和可靠的技术支持。

[0063] 下面结合本发明的实际应用场景,详细说明基于HashTrie和Bitmap的网络流量索引优化策略的方法的操作过程。

[0064] 实施例一:

[0065] (1)数据包接收和预处理:在该环境中,通过交换机或路由器等网络设备接收网络流量数据包,并对网络流量数据包进行预处理;需要进行清洗、解码、解析操作,以便提取出源IP地址、目的IP地址、协议类型的属性信息。

[0066] 需要注意的是,除了本发明举例的五元组信息以外,网络流量数据包中其他形式的信息也同样适用本发明基于HashTrie和Bitmap的网络流量索引优化策略的方法。

[0067] (2)数据包索引构建:将经过预处理的数据包按照一定的规则和算法进行索引构建,得到相应的索引树结构。这里所指的索引树结构优选HashTrie树,也可以选用其他适用的数据结构。

[0068] (3)索引树的Bitmap映射:在索引树的每个树节点前面添加Bitmap映射,Bitmap用于指示某个节点是否存在于某个属性值上。具体来说,Bitmap可以用于指示源IP地址、目的IP地址、协议类型的属性是否存在某个特定的值。例如,对于源IP地址,判断是否有节点存在对应的源IP地址值;对于目的IP地址,判断是否有节点存在对应的目的IP地址值;对于协议类型,判断是否有节点存在对应的协议类型值。

[0069] (4)索引导出:将带有Bitmap映射的索引树存储到磁盘文件中,以便后续的检索和查询操作。在索引导出时,对Bitmap为0的树节点不导出,减少时间和空间的占用。

[0070] 如图2所示,以ip地址为例,为地址10.2.1.1,10.2.10.1,172.171.8.1,172.170.2.1,171.168.1.2,创建的索引结构;bit在内存中创建的基于Bitmap的HashTrie

树,同时并将内存中的树结构按照bitmap映射导入磁盘文件文件。在内存中在每个树节点的前面添加bitmap映射,位图的每一位代表了属性索引值是否存在,置1代表该属性值存在,0代表该属性值不存在,如在树根节点中bitmap的第10个bit置1,则代表对应的树节点中属性值10存在。在索引导出时采用中序遍历树节点,再对每个树节点按从左到右顺序依次写入索引文件,这样做的目的是让索引文件中的偏移量按照属性值从小到大的顺序排列,方便进行范围查找;范围查找时只需找到最小属性值的偏移量和最大属性值的偏移量,而无需一一读取范围内每个属性值的偏移,减少磁盘的随机IO,进而提升检索效率。同时按照内存中每个树节点前面记录的bitmap的位图,只将bitmap位图置1的树节点中属性值的偏移写入文件,减少了空地址占用空间。

[0071] (5)检索和查询:根据相应的索引树和Bitmap映射,快速定位到目标数据包并返回检索结果。具体来说,检索和查询的过程包括以下步骤:

[0072] 首先匹配Bitmap映射;对应没匹配中的直接返回检索失败;只有匹配中的才继续检索并返回结果,这样可以大大提高检索效率。

[0073] 如图3所示,以检索ip地址10.2.1.1为例,在第五步时得到一个长度加偏移的值,其中低位10000为数据在索引文件中的偏移量,0001代表ip地址10.2.1.1共计一个偏移位置,如果为0002即代表有两个此地址的偏移;接下来通过10000这个偏移地址获取到最终符合条件的网络流量数据包,检索完毕。

[0074] 实施例二:

[0075] 本发明适用于运营商网络环境中。在该环境中,通过类似的方式接收网络流量数据包,并对数据包进行预处理。然后采用HashTrie算法构建索引树结构,为每个节点添加Bitmap标记,并实现快速检索和查询功能。

[0076] 与实施例1不同的是,运营商网络环境下的网络流量数据包通常具有更高的复杂性和多样性。因此,在构建索引树时,需要考虑多样的属性信息,例如源IP地址和目的IP地址的子网掩码、协议类型的端口号等。同时,在Bitmap标记时,对更多的属性进行标记,从而能够更精确地定位到符合条件的数据包。

[0077] 实施例三:

[0078] 本发明适用于云平台环境中。在该环境中,通过类似的方式接收网络流量数据包,并对数据包进行预处理。然后采用HashTrie算法构建索引树结构,为每个节点添加Bitmap标记,并实现快速检索和查询功能。

[0079] 与实施例一和实施例二不同的是,云平台环境下的网络流量数据包通常具有更高的动态性和变化性。因此,设计灵活、动态的索引树结构,在Bitmap标记时,考虑更多的属性信息,以便能够更加准确地定位到符合条件的数据包。

[0080] 实施例二和实施例三说明本发明的高效网络流量索引方法和系统具有广泛的适用性,并能够灵活应对不同网络环境下的挑战和变化。

[0081] 需要说明的是:上述本申请实施例先后顺序仅仅为了描述,不代表实施例的优劣。且上述对本申请特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可

能是有利的。

[0082] 本申请中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于装置、设备和存储介质实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0083] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0084] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

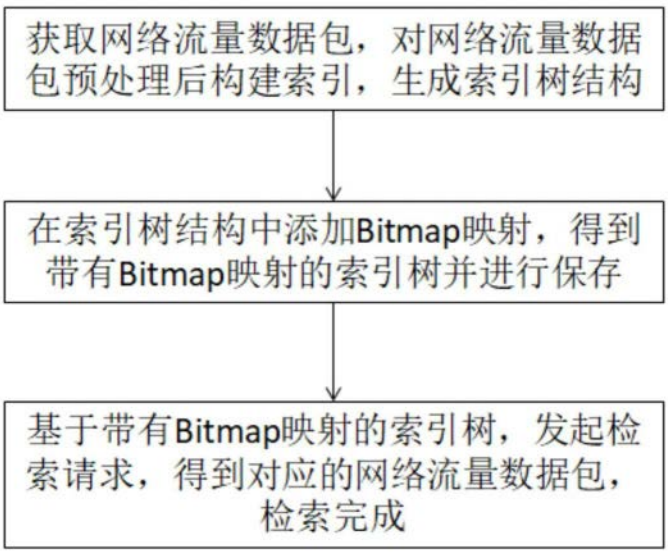

获取网络流量数据包，对网络流量数据包预处理后构建索引，生成索引树结构
在索引树结构中添加Bitmap映射，得到带有Bitmap映射的索引树并进行保存
基于带有Bitmap映射的索引树，发起检索请求，得到对应的网络流量数据包，检索完成

图1

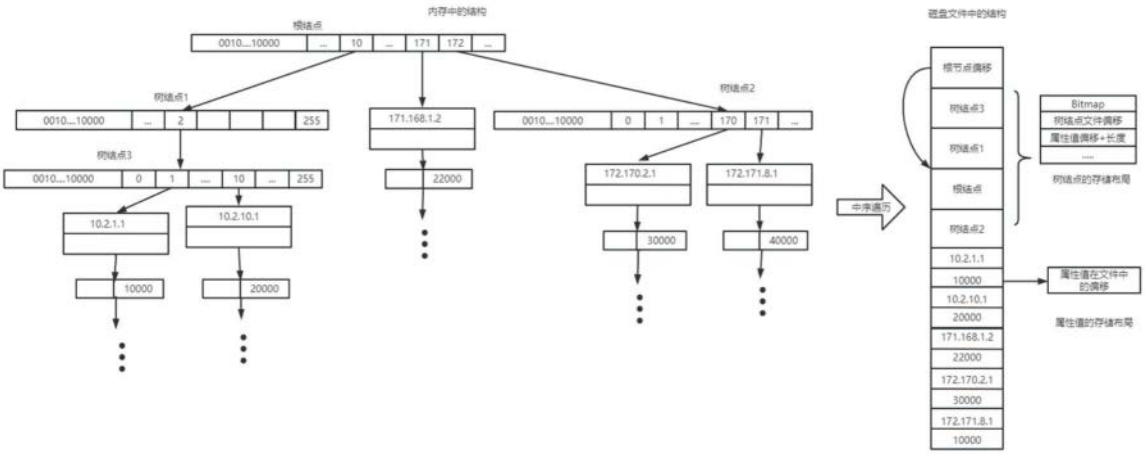

内存中的结构
磁盘文件中的结构
中序遍历
Bitmap
树结点文件偏移
属性值偏移+长度
树结点的存储布局
属性值在文件中的偏移
属性值的存储布局

图2

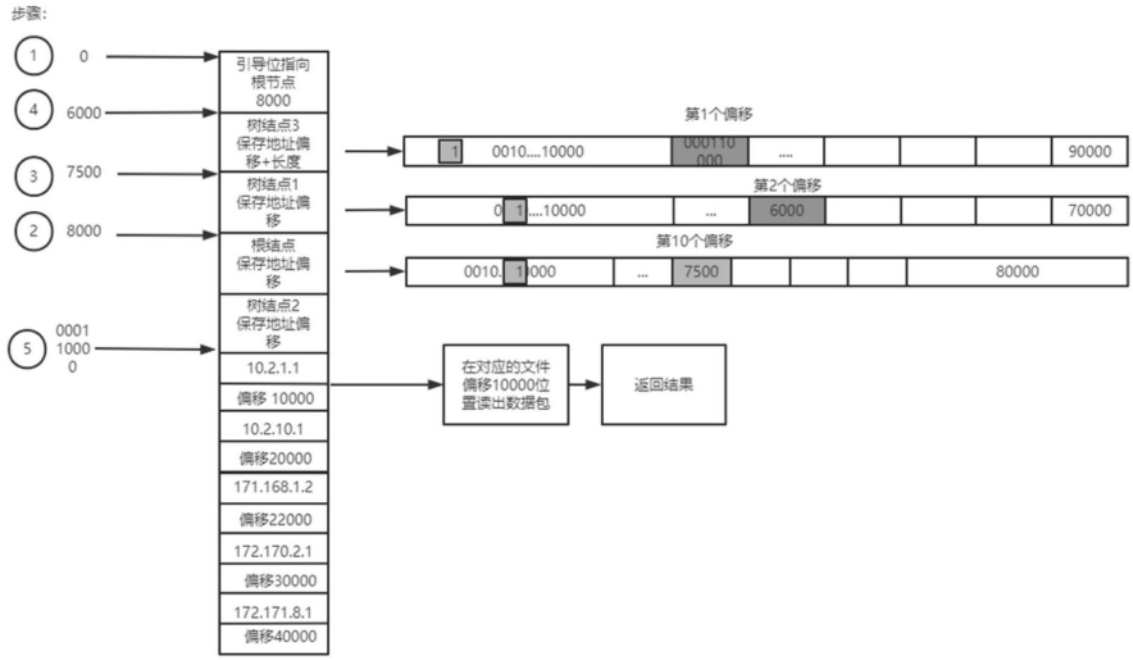

步骤:
1
0
4
6000
3
7500
2
8000
5
0001
1000
0
引导位指向根节点8000
树结点3保存地址偏移+长度
树结点1保存地址偏移
根结点保存地址偏移
树结点2保存地址偏移
10.2.1.1
偏移 10000
10.2.10.1
偏移20000
171.168.1.2
偏移22000
172.170.2.1
偏移30000
172.171.8.1
偏移40000
第1个偏移
1
0010....10000
000110
000
....
90000
第2个偏移
0
1
....10000
...
6000
70000
第10个偏移
0010.
1
000
...
7500
80000
在对应的文件偏移10000位置读出数据包
返回结果

图3

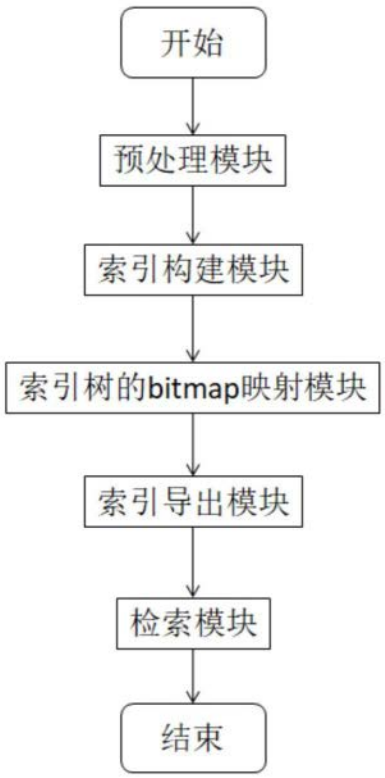

开始
预处理模块
索引构建模块
索引树的bitmap映射模块
索引导出模块
检索模块
结束

图4