



(12) 发明专利申请

(10) 申请公布号 CN 113419973 A

(43) 申请公布日 2021.09.21

(21) 申请号 202110672103.2

(22) 申请日 2021.06.17

(71) 申请人 杭州迪普科技股份有限公司  
地址 310051 浙江省杭州市滨江区通和路  
68号中财大厦6楼

(72) 发明人 孙军伟 秦德楼 赵旭东

(74) 专利代理机构 北京博思佳知识产权代理有限公司 11415  
代理人 陈蕾

(51) Int.Cl.  
G06F 12/02 (2006.01)

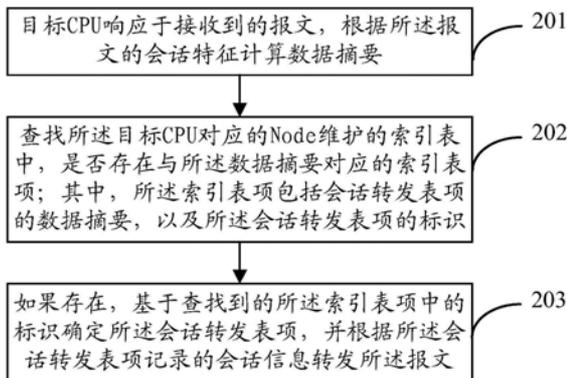
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种报文转发方法及装置

(57) 摘要

本申请提供一种报文转发方法及装置,该方法包括:目标CPU响应于接收到的报文,根据所述报文的会话特征计算数据摘要;查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。基于以上技术方案,通过在网络设备的每个Node上都维护有索引表,使得每个CPU在处理报文时,可以查找该CPU对应Node上的索引表,实现CPU尽可能访问本地内存,减少远端内存访问,提高了网络设备的处理性能,提升了报文转发效率。



1. 一种报文转发方法,应用于网络设备,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表;所述方法包括:

目标CPU响应于接收到的报文,根据所述报文的会话特征计算数据摘要;

查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

2. 根据权利要求1所述的方法,所述网络设备还搭载了具有RSS功能的网卡;

所述目标CPU响应于接收到的报文之前,所述方法还包括:

所述网卡基于RSS功能,计算所述报文的五元组对应的哈希值;

根据所述计算出的哈希值,确定接收所述报文的的目标CPU,并将所述报文发送至所述目标CPU。

3. 根据权利要求1所述的方法,所述方法还包括:

响应于所述网络设备的初始化请求,获取所述Node的总数,并基于所述Node的总数为每个Node分配会话,以及与所述会话的数量对应的内存空间;

为每个Node分配用于存放所述索引表的内存空间。

4. 根据权利要求3所述的方法,所述方法还包括:

将所述会话按链表方式关联起来,并初始化为空闲链表;

将所述索引表初始化为空闲状态。

5. 根据权利要求1所述的方法,所述会话转发表包括空闲链表以及使用链表;所述空闲链表存储了空闲的会话转发表项,所述使用链表存储了已经添加的会话转发表项。

6. 根据权利要求1所述的方法,所述标识用于指示所述会话转发表项的内存地址,相对于所述会话转发表起始内存地址的地址偏移;

所述基于查找到的所述索引表项中的标识确定会话转发表项,包括:

根据所述索引表项中记录的所述地址偏移,确定所述会话转发表项的内存地址。

7. 根据权利要求1所述的方法,所述索引表项中还存储了报文的五元组;

所述基于查找到的所述索引表项中的指针确定会话转发表项之前,所述方法还包括:

比较所述报文的五元组与查找到的所述索引表项中的五元组是否一致,如果一致,基于查找到的所述索引表项中的指针确定会话转发表项。

8. 根据权利要求1-6任一所述的方法,所述方法还包括:

如果不存在与所述数据摘要对应的索引表项,则获取所述空闲链表中的空闲会话,保存所述报文对应的会话信息至所述空闲会话,并将所述空闲会话转移至使用链表;

获取所述空闲索引表,将所述空闲会话的内存地址相对于上述会话转发表起始内存地址的地址偏移作为标识,以及基于所述报文的会话特征计算出的数据摘要,记录在所述空闲索引表中,并将所述空闲索引表更新为使用状态;

根据所述报文对应的会话信息转发所述报文。

9. 根据权利要求5所述的方法,所述方法还包括:

遍历所述Node维护的会话转发表中的使用链表,检测是否有会话老化标记;

如果有,将所述使用链表中的使用会话,转移到空闲会话链,并将所述Node维护的索引表更新为空闲状态。

10.一种报文转发装置,应用于网络设备,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表;所述装置包括:

计算单元,用于在目标CPU响应于接收到的报文时,根据所述报文的会话特征计算数据摘要;

查找单元,用于查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

转发单元,用于在存在所述索引表项时,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

## 一种报文转发方法及装置

### 技术领域

[0001] 本申请涉及通信技术领域,尤其涉及一种报文转发方法及装置。

### 背景技术

[0002] 请参见图1,图1是示出的三种CPU访问内存的结构示意图。

[0003] 如图1所示,单核CPU直接访问唯一的内存,CPU内部各VCPU访问内存的时间几乎相同。而对于多核CPU来说,在UMA结构(Uniform Memory Access,统一内存访问)下,各CPU通过共享一个北桥来访问唯一的内存,虽然各CPU访问内存的时间几乎相同,但会受限于北桥响应时间的瓶颈。

[0004] 而当多核CPU采用NUMA(Non Uniform Memory Access,非统一内存访问)架构之后,每个CPU既可以访问自己的内存,也可以访问其他CPU的内存,CPU对本地内存访问的时间,远大于对其他CPU进行远端内存访问的时间。

### 发明内容

[0005] 有鉴于此,本申请提供一种报文转发方法及装置,通过使CPU访问对应Node中维护的索引表,避免产生过多的远端内存访问,从而提升网络设备的性能。

[0006] 具体地,本申请是通过如下技术方案实现的:

[0007] 第一方面,本申请提出一种报文转发的方法,应用于网络设备,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表;该方法包括:

[0008] 目标CPU响应于接收到的报文,根据所述报文的会话特征计算数据摘要;

[0009] 查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

[0010] 如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0011] 第二方面,本申请还提出一种报文转发的装置,应用于网络设备,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表;该装置包括:

[0012] 计算单元,用于在目标CPU响应于接收到的报文时,根据所述报文的会话特征计算数据摘要;

[0013] 查找单元,用于查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

[0014] 转发单元,用于在存在所述索引表项时,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0015] 本申请的实施例提供的技术方案可以包括以下有益效果：

[0016] 通过在网络设备的每个Node上都维护有索引表,使得每个CPU在处理报文时,可以查找该CPU对应Node上的索引表,避免查找其他Node上的索引表而产生过多的远端内存访问,从而实现CPU尽可能访问本地内存,减少远端内存访问,提高了网络设备的处理性能,提升了报文转发效率。

[0017] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

### 附图说明

[0018] 图1是本申请一示例性实施例示出的三种CPU访问内存的结构示意图；

[0019] 图2是本申请一示例性实施例示出的一种报文转发方法的流程图；

[0020] 图3是本申请一示例性实施例示出的另一种报文转发方法的流程图；

[0021] 图4是本申请一示例性实施例示出的一种的报文转发装置所在电子设备的硬件结构图；

[0022] 图5是本申请一示例性实施例示出的一种报文转发装置的框图。

### 具体实施方式

[0023] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

[0024] 在本申请使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本申请。在本申请和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。还应当理解,本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0025] 应当理解,尽管在本申请可能采用术语第一、第二、第三等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本申请范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。取决于语境,如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0026] 目前,对于单核CPU来说,其内部每个VCPU访问内存的时间几乎相同。此时,网络报文的会话转发表项也都分配在唯一的Node(内存节点)上,不同VCPU对报文进行转发的业务性能不存在差别。

[0027] 对于采用UMA架构的多核CPU来说,各CPU共享唯一的内存,访问内存上存储的会话转发表项的时间也几乎相同,业务性能不存在差别,但会受限于北桥响应时间的瓶颈。

[0028] 而NUMA架构为减轻对总线和内存的带宽需求,以内存访问的不一致性为代价,导致NUMA结构对进程调度算法的要求较高,需要尽量减少跨Node的内存访问次数,以提升系统性能。

[0029] 可见,对于采用NUMA架构的多核CPU来说,若仍延用唯一内存时索引表的分配方

案,则各CPU访问此索引表会产生过多的远端访问内存,影响设备的性能。

[0030] 例如,如图1所示,假设索引表分配在CPU1对应的内存上,则CPU0、CPU2接收到报文之后,需要查找索引表进行报文转发时,需要访问CPU1对应的内存,则会产生过多的远端内存访问,导致报文处理的效率大打折扣,也会由于远端内存访问的时间过久,占用CPU的资源,影响CPU处理其他业务,进而降低设备的整体性能。

[0031] 有鉴于此,本申请提供一种在网络设备的每个Node上都维护了索引表,使得每个CPU在处理报文时,可以查找该CPU对应Node上的索引表,实现本地内存访问的技术方案。

[0032] 在实现时,该网络设备搭载了采用NUMA架构的多个CPU,该网络设备包括至少一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表;

[0033] 目标CPU接收到报文之后,根据该报文的会话特征计算数据摘要;

[0034] 查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

[0035] 如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0036] 例如,所述标识用于指示所述会话转发表项的内存地址,相对于所述会话转发表起始内存地址的地址偏移;目标CPU可以根据所述索引表项中记录的所述地址偏移,确定所述会话转发表项的内存地址,并根据所述会话转发表项记录的会话信息转发所述报文。

[0037] 在以上技术方案中,通过在网络设备的每个Node上都维护有索引表,使得每个CPU在处理报文时,可以查找该CPU对应Node上的索引表,避免查找其他Node上的索引表而产生过多的远端内存访问,从而实现CPU尽可能访问本地内存,减少远端内存访问,提高了网络设备的处理性能,提升了报文转发效率。

[0038] 接下来对本申请实施例进行详细说明。

[0039] 请参见图2,图2是本申请一示例性实施例示出的一种报文转发方法的流程图,如图2所示,所述方法应用于网络设备,包括以下步骤:

[0040] 步骤201:目标CPU响应于接收到的报文,根据所述报文的会话特征计算数据摘要;

[0041] 步骤202:查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

[0042] 步骤203:如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0043] 其中,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表。

[0044] 在本实施例中,目标CPU响应于接收到的报文,根据报文的会话特征计算数据摘要;

[0045] 例如,目标CPU响应于接收到的报文,获取报文的会话特征,并基于预设的哈希算法,计算该报文的会话特征对应的哈希值。

[0046] 其中,目标CPU可以是一个物理CPU,也可以是VCPU,本申请对此并不限定,本领域技术人员可以根据实际需要划分出的Node和CPU进行关联,只需保证每个CPU具有一个对

应的Node即可。

[0047] 通过将内存划分出多个Node,并将每个Node关联一个或多个CPU,使得每个CPU都对应有一个Node,从而提升CPU对其本地内存的访问速度。

[0048] 另外,网络设备可以从多个CPU中确定出处理该报文的目标CPU,并将报文发送至该目标CPU。

[0049] 例如,RSS(Receive Side Scaling)是一种网卡驱动技术,网络设备可以利用RSS功能,实现在多CPU系统下,将网卡接收到的报文在多个CPU之间高效分发。

[0050] 在示出的一种实施方式中,所述网络设备还搭载了具有RSS功能的网卡;

[0051] 所述网卡基于RSS功能,计算所述报文的五元组对应的哈希值;

[0052] 根据所述计算出的哈希值,确定接收所述报文的目标CPU,并将所述报文发送至所述目标CPU。

[0053] 例如,网卡可以对接收到的报文进行解析,获取五元组信息;再通过预设的哈希算法计算五元组信息对应的哈希值;取哈希值的最低有效位用于索引间接寻址表RETA(redirection table);其中,间接寻址表RETA中保存了用于分配数据报文到不同的CPU接收处理的索引值。通过将网络流量分载到多个CPU上,可以降低单个CPU的占用率。

[0054] 在本实施例中,目标CPU查找自身对应的Node维护的索引表中,是否存在与根据接收到的报文的会话特征计算出的数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识。

[0055] 例如,目标CPU基于根据接收到的报文的会话特征计算出的数据摘要,匹配自身对应的Node维护的索引表,查找是否存在与该数据摘要相一致的索引表项;该查找到的索引表项中记录了报文对应的会话转发表项的数据摘要,以及会话转发表项的标识。

[0056] 其中,可以在每个Node分配出两个连续的内存空间,分别用于维护会话转发表以及索引表。会话转发表中的会话转发表项记录了报文转发需要的会话信息,索引表中的索引表项记录了报文对应的会话转发表项的数据摘要,以及会话转发表项的标识。

[0057] 另外,当查找到存在与所述数据摘要对应的索引表项之后,为确保匹配的准确性,可以进一步确定报文的五元组与所述索引表项中记录的五元组是否相同。

[0058] 在示出的一种实施方式中,所述索引表项中还存储了报文的五元组;

[0059] 进一步的,比较所述报文的五元组与查找到的所述索引表中的五元组是否一致,如果一致,基于查找到的所述索引表项中的指针确定会话转发表项。

[0060] 对于每个Node中维护的会话转发表以及索引表,可以由在网络设备在执行初始化操作时进行分配。

[0061] 在示出的一种实施方式中,响应于所述网络设备的初始化请求,获取所述Node的总数,并基于所述Node的总数为每个Node分配会话,以及与所述会话的数量对应的内存空间;为每个Node分配用于存放所述索引表的内存空间。

[0062] 例如,在网络设备开启NUMA功能并进行初始化时,假设检测到的Node数量为n,设备支持的会话总数为m,则可以为每个Node分配m/n个会话,并在Node中分配对应的内存空间作为会话池,用于存储分配的会话;同时,还可以为每个Node分配用于存放索引表的内存空间。

[0063] 进一步的,在示出的一种实施方式中,将所述会话按链表方式关联起来,并初始化

为空闲链表;将所述索引表初始化为空闲状态。

[0064] 例如,在每个Node分配会话完成后,将分配的会话按照链表的方式串联起来,并接在空闲链表头上。通过线性链表的方式,可以实现对会话的灵活管理。

[0065] 具体的,在示出的一种实施方式中,所述会话转发表包括空闲链表以及使用链表;所述空闲链表存储了空闲的会话转发表项,所述使用链表存储了已经添加的会话转发表项。

[0066] 在本实施例中,如果存在与所述数据摘要对应的索引表项,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0067] 例如,当确定存在与数据摘要对应的索引表项之后,获取该索引表项中记录的会话转发表项的标识,根据该标识获取会话转发表项记录的会话信息,并根据该会话转发表项记录的会话信息转发报文。

[0068] 在示出的一种实施方式中,所述标识用于指示所述会话转发表项的内存地址,相对于所述会话转发表起始内存地址的地址偏移;

[0069] 进一步的,根据所述索引表项中记录的所述地址偏移,确定所述会话转发表项的内存地址。

[0070] 例如,会话转发表的起始内存地址,加上地址偏移量,即为上述报文对应的会话转发表项的内存地址。

[0071] 在示出的一种实施方式中,如果不存在与所述数据摘要对应的索引表项,则获取所述空闲链表中的空闲会话,保存所述报文对应的会话信息至所述空闲会话,并将所述空闲会话转移至使用链表;

[0072] 获取所述空闲索引表,将所述空闲会话的内存地址相对于上述会话转发表起始内存地址的地址偏移作为标识,以及基于所述报文的会话特征计算出的数据摘要,记录在所述空闲索引表中,并将所述空闲索引表更新为使用状态;

[0073] 根据所述报文对应的会话信息转发所述报文。

[0074] 在CPU对首次接收到的报文执行上述操作之后,完成了该报文对应的会话的建立,更新了报文对应的会话转发信息;

[0075] 当后续再次接收到该报文时,Node维护的会话转发表中已经记录了该报文对应的会话转发表项,索引表中也记录了该报文对应的会话转发表项的数据摘要,以及会话转发表项的标识,CPU只需执行上述步骤201~203即可直接使用会话信息完成报文转发。

[0076] 在示出的一种实施方式中,遍历所述Node维护的会话转发表中的使用链表,检测是否有会话老化标记;

[0077] 如果有,将所述使用链表中的使用会话,转移到空闲会话链,并将所述Node维护的索引表更新为空闲状态。

[0078] 例如,可以在每个Node关联的CPU上,分配一个监控会话使用链的线程,当会话老化时,线程检测到会话老化标记,将索引表更新为空闲状态,并将会话从使用链上摘下,挂到空闲链上。

[0079] 在以上技术方案中,通过在网络设备的每个Node上都维护有索引表,使得每个CPU在处理报文时,可以查找该CPU对应Node上的索引表,避免查找其他Node上的索引表而产生

过多的远端内存访问,从而实现CPU尽可能访问本地内存,减少远端内存访问,提高了网络设备的处理性能,提升了报文转发效率。

[0080] 请参见图3,图3是本申请一示例性实施例示出的另一种报文转发方法的流程图。

[0081] 如图3所示,在示出的一种实施方式中,报文转发的方法包括以下步骤:

[0082] S301:目标CPU接收报文;

[0083] 其中,所述报文可以由网络设备搭载的具有RSS功能的网卡,通过计算所述报文的五元组对应的哈希值,确定接收所述报文的的目标CPU。

[0084] S302:根据所述报文的会话特征计算数据摘要;

[0085] 具体的,目标CPU可以获取报文的会话特征,并基于预设的哈希算法,计算该报文的会话特征对应的哈希值。

[0086] S303:查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;

[0087] 其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识。

[0088] 值得说明的是,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表。

[0089] 通过将内存划分出多个Node,并将每个Node关联一个或多个CPU,使得每个CPU都对应有一个Node,从而提升CPU对其本地内存的访问速度。

[0090] S304:如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项;

[0091] 其中,所述标识用于指示所述会话转发表项的内存地址,相对于所述会话转发表起始内存地址的地址偏移。

[0092] 具体的,可以根据所述索引表项中记录的所述地址偏移,确定所述会话转发表项的内存地址,根据该内存地址获取所述会话转发表项记录的会话信息。

[0093] S305:根据所述会话转发表项记录的会话信息转发所述报文。

[0094] S306:如果不存在与所述数据摘要对应的索引表项,则获取所述空闲链表中的空闲会话,保存所述报文对应的会话信息至所述空闲会话,并将所述空闲会话转移至使用链表;

[0095] 其中,所述会话转发表包括空闲链表以及使用链表;所述空闲链表存储了空闲的会话转发表项,所述使用链表存储了已经添加的会话转发表项。

[0096] S307:获取所述空闲索引表,将所述空闲会话的内存地址相对于上述会话转发表起始内存地址的地址偏移作为标识,以及基于所述报文的会话特征计算出的数据摘要,记录在所述空闲索引表中,并将所述空闲索引表更新为使用状态;

[0097] 其中,会话转发表的起始内存地址,加上地址偏移,即为该获取到的空闲会话的内存地址。

[0098] S308:根据所述报文对应的会话信息转发所述报文。

[0099] 在以上技术方案中,通过在网络设备的每个Node上都维护有索引表,使得每个CPU在处理报文时,可以查找该CPU对应Node上的索引表,避免查找其他Node上的索引表而产生过多的远端内存访问,从而实现CPU尽可能访问本地内存,减少远端内存访问,提高了网络设备的处理性能,提升了报文转发效率。

[0100] 与上述方法实施例相对应,本申请还提供了装置的实施例。

[0101] 与上述方法实施例相对应,本申请还提供了一种报文转发装置的实施例。本申请的报文转发装置的实施例可以应用在电子设备上。装置实施例可以通过软件实现,也可以通过硬件或者软硬件结合的方式实现。以软件实现为例,作为一个逻辑意义上的装置,是通过其所在电子设备的处理器将非易失性存储器中对应的计算机程序指令读取到内存中运行形成的。从硬件层面而言,如图4所示,为本申请一示例性实施例示出的一种的报文转发装置所在电子设备的硬件结构图,除了图4所示的处理器、内存、网络接口、以及非易失性存储器之外,实施例中装置所在的电子设备通常根据该电子设备的实际功能,还可以包括其他硬件,对此不再赘述。

[0102] 请参见图5,图5是本申请一示例性实施例示出的一种报文转发装置的框图,如图5所示,所述报文转发装置500可以应用在前述图4所示的电子设备中,包括:

[0103] 计算单元501,用于在目标CPU响应于接收到的报文时,根据所述报文的会话特征计算数据摘要;

[0104] 查找单元502,用于查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

[0105] 转发单元503,用于在存在所述索引表项时,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0106] 其中,所述报文转发装置500可以应用于网络设备,所述网络设备搭载了采用NUMA架构的多个CPU,所述网络设备至少包括一个Node;其中,每个Node至少对应一个CPU,每个Node维护了会话转发表以及索引表;

[0107] 在一实施例中,所述网络设备还搭载了具有RSS功能的网卡;

[0108] 所述网卡基于RSS功能,计算所述报文的五元组对应的哈希值;

[0109] 根据所述计算出的哈希值,确定接收所述报文的的目标CPU,并将所述报文发送至所述目标CPU。

[0110] 在一实施例中,所述报文转发装置500还包括:

[0111] 分配单元,用于响应于所述网络设备的初始化请求,获取所述Node的总数,并基于所述Node的总数为每个Node分配会话,以及与所述会话的数量对应的内存空间;用于为每个Node分配用于存放所述索引表的内存空间。

[0112] 在一实施例中,所述报文转发装置500还包括:

[0113] 初始化单元,用于将所述会话按链表方式关联起来,并初始化为空闲链表;用于将所述索引表初始化为空闲状态。

[0114] 在一实施例中,所述会话转发表包括空闲链表以及使用链表;所述空闲链表存储了空闲的会话转发表项,所述使用链表存储了已经添加的会话转发表项。

[0115] 在一实施例中,所述标识用于指示所述会话转发表项的内存地址,相对于所述会话转发表起始内存地址的地址偏移;

[0116] 所述转发单元503包括:

[0117] 地址计算单元,根据所述索引表项中记录的所述地址偏移,确定所述会话转发表项的内存地址。

- [0118] 在一实施例中,所述索引表项中还存储了报文的五元组;
- [0119] 所述报文转发装置500还包括:
- [0120] 比较单元,用于比较所述报文的五元组与查找到的所述索引表中的五元组是否一致,如果一致,基于查找到的所述索引表项中的指针确定会话转发表项。
- [0121] 在一实施例中,所述报文转发装置500还包括:
- [0122] 第一获取单元,用于在如果不存在与所述数据摘要对应的索引表项时,获取所述空闲链表中的空闲会话,保存所述报文对应的会话信息至所述空闲会话,并将所述空闲会话转移至使用链表;
- [0123] 第二获取单元,用于获取所述空闲索引表,将所述空闲会话的内存地址相对于上述会话转发表起始内存地址的地址偏移作为标识,以及基于所述报文的会话特征计算出的数据摘要,记录在所述空闲索引表中,并将所述空闲索引表更新为使用状态;
- [0124] 第二转发单元,用于根据所述报文对应的会话信息转发所述报文。
- [0125] 在一实施例中,所述报文转发装置500还包括:
- [0126] 检测单元,用于遍历所述Node维护的会话转发表中的使用链表,检测是否有会话老化标记;
- [0127] 转移单元,用于在如果有老化标记时,将所述使用链表中的使用会话,转移到空闲会话链,并将所述Node维护的索引表更新为空闲状态。
- [0128] 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,对于客户端设备实施例、装置实施例而言,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。
- [0129] 对于装置实施例而言,由于其基本对应于方法实施例,所以相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理模块,即可以位于一个地方,或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本申请方案的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。
- [0130] 上述实施例阐明的装置、装置、模块或模块,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机,计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任意几种设备的组合。
- [0131] 与上述方法实施例相对应,本说明书还提供了一种电子设备的实施例。该电子设备包括:处理器以及用于存储机器可执行指令的存储器;其中,处理器和存储器通常通过内部总线相互连接。在其他可能的实现方式中,所述设备还可能包括外部接口,以能够与其他设备或者部件进行通信。
- [0132] 在本实施例中,通过读取并执行所述存储器存储的与用户身份验证逻辑对应的机器可执行指令,所述处理器被促使:
- [0133] 目标CPU响应于接收到的报文,根据所述报文的会话特征计算数据摘要;

[0134] 查找所述目标CPU对应的Node维护的索引表中,是否存在与所述数据摘要对应的索引表项;其中,所述索引表项包括会话转发表项的数据摘要,以及所述会话转发表项的标识;

[0135] 如果存在,基于查找到的所述索引表项中的标识确定所述会话转发表项,并根据所述会话转发表项记录的会话信息转发所述报文。

[0136] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本申请的真正范围和精神由下面的权利要求指出。

[0137] 应当理解的是,本申请并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

[0138] 以上所述仅为本申请的较佳实施例而已,并不用以限制本申请,凡在本申请的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本申请保护的范围之内。

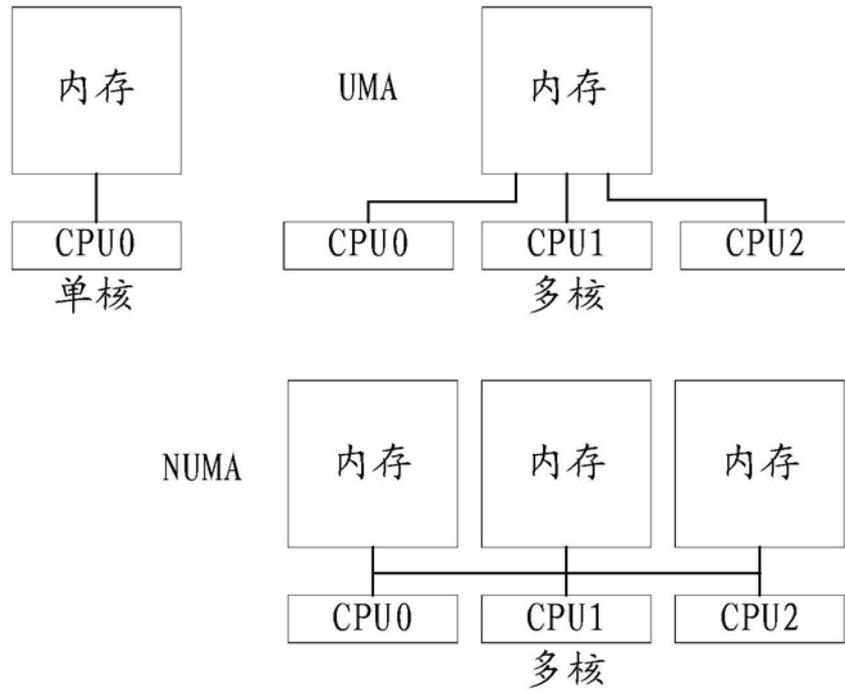


图1

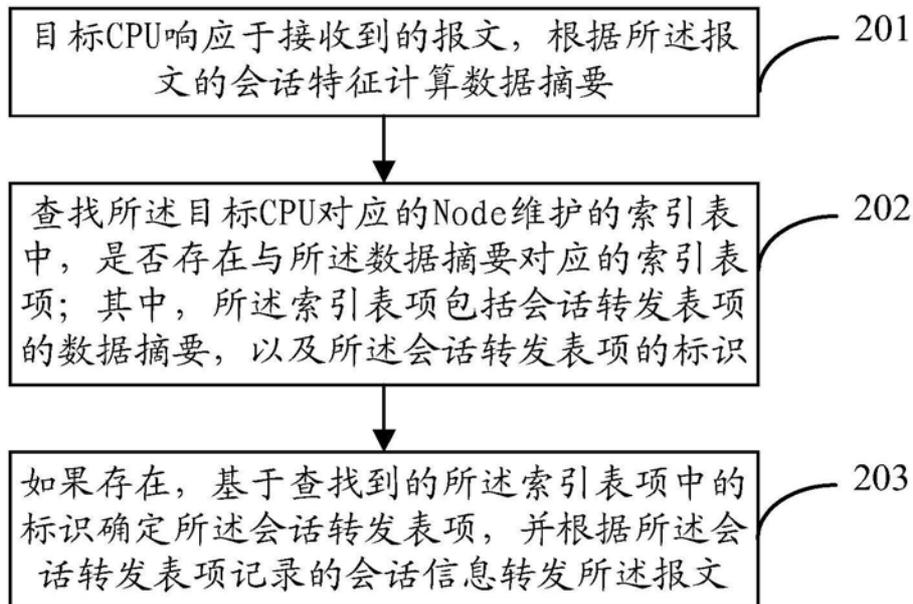


图2

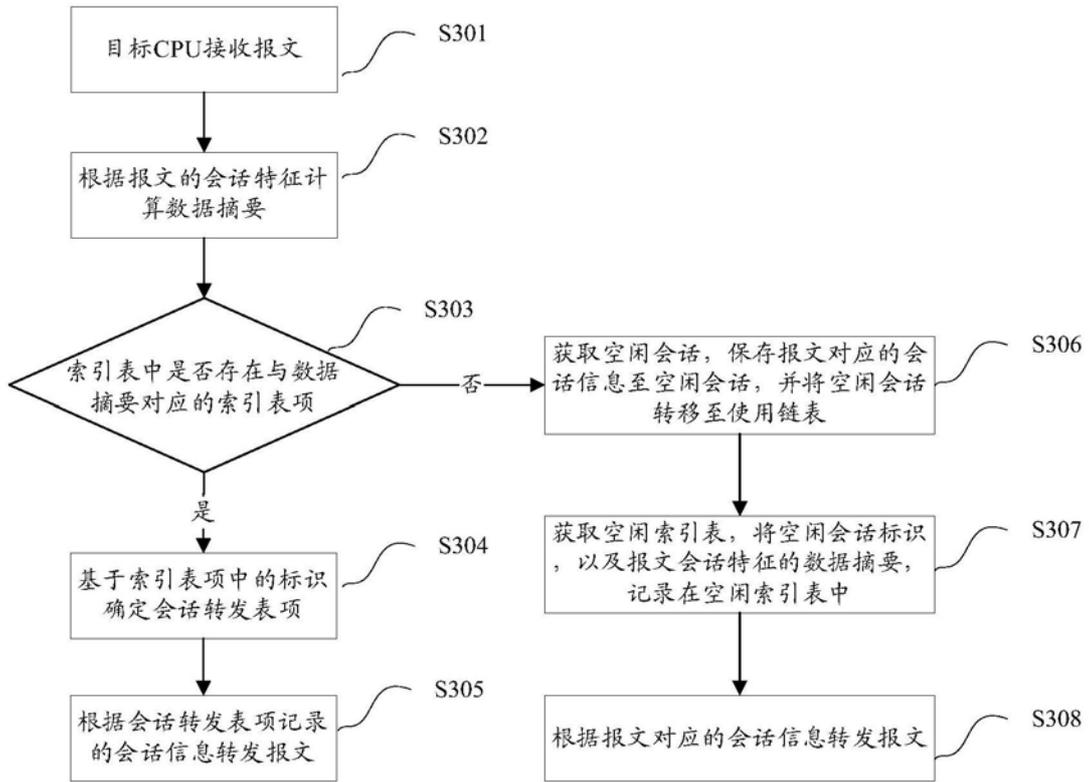


图3

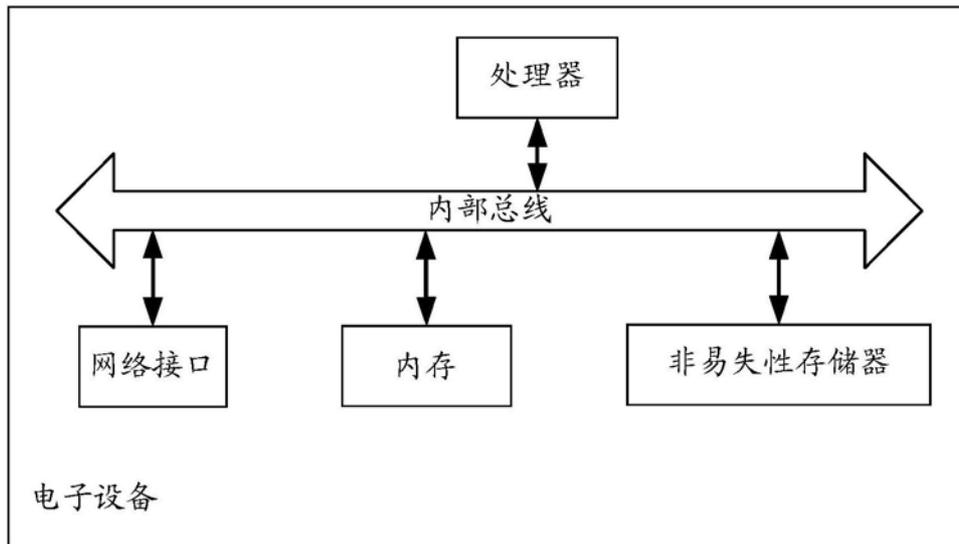


图4

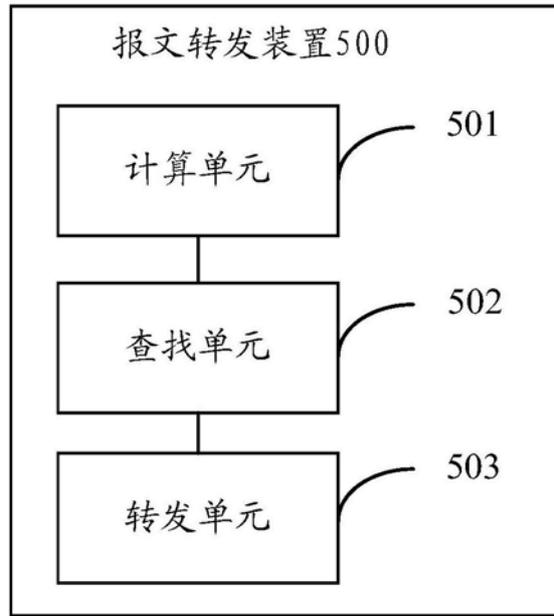


图5