(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112671667 B
(45)授权公告日 2022.06.28

(21)申请号 202011405147.0
(22)申请日 2020.12.03
(65)同一申请的已公布的文献号
申请公布号 CN 112671667 A
(43)申请公布日 2021.04.16
(73)专利权人 中国科学院计算技术研究所
地址 100080 北京市海淀区中关村科学院南路6号
(72)发明人 杨晔 姜海洋 谢高岗
(74)专利代理机构 北京律诚同业知识产权代理有限公司 11006
专利代理师 祁建国
(51)Int.Cl.
*H04L 47/215*(2022.01)
*H04L 47/24*(2022.01)
*H04L 12/46*(2006.01)

(56)对比文件
CN 111158905 A,2020.05.15
CN 111158905 A,2020.05.15
US 9158591 B2,2015.10.13
CN 111612373 A,2020.09.01
王健.基于软件定义网络架构的数据中心网络若干关键问题研究.《中国博士学位论文全文数据库》.2016,

审查员 姜颖

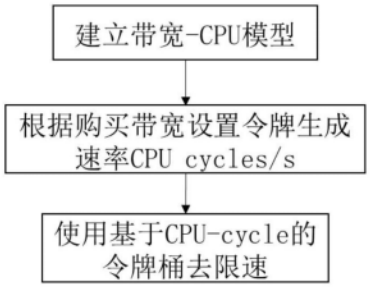
权利要求书1页 说明书5页 附图1页

(54)发明名称

一种虚拟转发设备的转发限速方法及系统

(57)摘要

本发明提出一种虚拟转发设备的转发限速方法及系统,包括通过测量得到服务提供商网络平均数据包大小和流数目作为流量特征,该服务提供商网络包含虚拟转发设备和多个租户,使用流量生成工具生成符合该特征的流量,通过在该虚拟转发设备中转发该流量的数据包并测量达到各个带宽所用于转发的CPU资源,构建CPU资源和带宽的对应关系;根据各租户的基础带宽需求以及该对应关系,为各租户分配基础CPU资源,并得到该服务提供商网络在满足该基础CPU资源后剩余的CPU资源,以为每一位租户按照权重再分配空闲CPU资源,每位租户用于转发的最终CPU资源为其分配的基础CPU资源和空闲CPU资源之和;该虚拟转发设备使用该最终CPU资源对其对应的租户完成数据包的转发。




CN 112671667 B

1.一种虚拟转发设备的转发限速方法,其特征在于,包括:

步骤1、通过测量得到服务提供商网络平均数据包大小和流数目作为流量特征,该服务提供商网络包含虚拟转发设备和多个租户,使用流量生成工具生成符合该流量特征的流量,通过在该虚拟转发设备中转发该流量的数据包并测量达到各个带宽所用于转发的CPU资源,构建CPU资源和带宽的对应关系;

步骤2、根据各租户的基础带宽需求以及该对应关系,为各租户分配基础CPU资源,并得到该服务提供商网络在满足该基础CPU资源后剩余的CPU资源,以为每一位租户按照权重再分配空闲CPU资源,每位租户用于转发的最终CPU资源为其分配的基础CPU资源和空闲CPU资源之和;

步骤3、该虚拟转发设备使用该最终CPU资源对其对应的租户完成数据包的转发。

2.如权利要求1所述的虚拟转发设备的转发限速方法,其特征在于,该步骤3包括:该虚拟转发设备利用该最终CPU资源对租户转发数据包进行限速,超过该最终CPU资源时的数据流量将等待或被丢弃。

3.如权利要求1所述的虚拟转发设备的转发限速方法,其特征在于,该基础CPU资源、该空闲CPU资源和该最终CPU资源的单位均为CPU cycles/s。

4.如权利要求2所述的虚拟转发设备的转发限速方法,其特征在于,该步骤3包括:使用令牌桶或公平队列的方式对租户转发数据包进行限速。

5.如权利要求2所述的虚拟转发设备的转发限速方法,其特征在于,该虚拟转发设备为虚拟路由器或虚拟交换机。

6.一种虚拟转发设备的转发限速系统,其特征在于,包括:

模块1,用于测量得到服务提供商网络平均数据包大小和流数目作为流量特征,该服务提供商网络包含虚拟转发设备和多个租户,使用流量生成工具生成符合该流量特征的流量,通过在该虚拟转发设备中转发该流量的数据包并测量达到各个带宽所用于转发的CPU资源,构建CPU资源和带宽的对应关系;

模块2,用于根据各租户的基础带宽需求以及该对应关系,为各租户分配基础CPU资源,并得到该服务提供商网络在满足该基础CPU资源后剩余的CPU资源,以为每一位租户按照权重再分配空闲CPU资源,每位租户用于转发的最终CPU资源为其分配的基础CPU资源和空闲CPU资源之和,该虚拟转发设备使用该最终CPU资源对其对应的租户完成数据包的转发。

7.如权利要求6所述的虚拟转发设备的转发限速系统,其特征在于,该虚拟转发设备利用该最终CPU资源对租户转发数据包进行限速,超过该最终CPU资源时的数据流量将等待或被丢弃。

8.如权利要求6所述的虚拟转发设备的转发限速系统,其特征在于,该基础CPU资源、该空闲CPU资源和该最终CPU资源的单位均为CPU cycles/s。

9.如权利要求7所述的虚拟转发设备的转发限速系统,其特征在于,该模块2包括:使用令牌桶或公平队列的方式对租户转发数据包进行限速。

10.如权利要求7所述的虚拟转发设备的转发限速系统,其特征在于,该虚拟转发设备为虚拟路由器或虚拟交换机。

# 一种虚拟转发设备的转发限速方法及系统

## 技术领域

[0001] 本发明涉及限速方法，提出了一种软件转发下隔离性增强的限速方法和系统。

## 背景技术

[0002] 随着网络设备向通用化和高性价比方向演进，软件转发逐渐取代了传统的硬件交换机、路由器，在很多场景下如云计算，网络功能虚拟化(NFV)中被大量使用。随着网络流量的剧增、流量类型的丰富，在这些虚拟转发设备中对不同流量在性能上的隔离性提出了更高的要求。

[0003] 软件转发设备与硬件转发设备的功能类似，为网络中数据包根据其五元组信息转发到对应的端口或链路上，典型的实现有虚拟交换机(vSwitch)、虚拟路由器(vRouter)等。软件转发设备的运行模式通常为启动若干轮询(PMD)线程来转发流量，这些PMD线程被绑定到通用服务器上若干专用的CPU核来提供能转发所需的物理计算资源。因此，软件转发设备所能提供的转发性能，即带宽上限，是由专用于转发的CPU和来提供的。

[0004] 但是这些CPU所能提供的转发能力是不固定的，在不同特征流量情况下，所能提供的转发速率/带宽也不同。比如说，一个CPU核能够将包大小为1518字节的数据流转发至10Gbps，但是其转发64字节包大小的数据流却只能转发到2Gbps带宽。这就给租户的带宽保障带来了挑战。

[0005] 目前已有的限速方法都是基于端口速率的，即直接按照带宽指标来限速，高于指标的流量不放行，例如广泛使用的基于每秒比特数(BPS)或每秒数据包数(PPS)的令牌桶策略，在软件转发设备中根据令牌来转发流量，但是这种方法很容易因为流量转发任务对于CPU资源的竞争而失效。很常见的现象就是，当一个流突然改变数据包大小(比如从1518字节更改为64字节)，在软件转发设备中，虽然该流的带宽/转发速率依然保持不变且没有超过限制，但是该流的CPU资源消耗会增加几倍，这就抢占了原本属于别的流的转发CPU资源，使得别的流可用的CPU资源下降，进而降低带宽。

[0006] 这种现象会造成一些流会因为别的流更改了流量特征而被干扰，带宽不稳定，网络性能的隔离性差。而已有的限速方法在实现上都没有考虑流量对CPU资源的竞争和软件转发设备不稳定的处理能力。因此一种资源隔离性好的软件转发限速方法亟待提出。

## 发明内容

[0007] 针对现有技术的不足，本发明提出一种虚拟转发设备的转发限速方法，其中包括：

[0008] 步骤1、通过测量得到服务提供商网络平均数据包大小和流数目作为流量特征，该服务提供商网络包含虚拟转发设备和多个租户，使用流量生成工具生成符合该特征的流量，通过在该虚拟转发设备中转发该流量的数据包并测量达到各个带宽所用于转发的CPU资源，构建CPU资源和带宽的对应关系；

[0009] 步骤2、根据各租户的基础带宽需求以及该对应关系，为各租户分配基础CPU资源，并得到该服务提供商网络在满足该基础CPU资源后剩余的CPU资源，以为每一位租户按照权

重再分配空闲CPU资源，每位租户用于转发的最终CPU资源为其分配的基础CPU资源和空闲CPU资源之和；

[0010] 步骤3、该虚拟转发设备使用该最终CPU资源对其对应的租户完成数据包的转发。

[0011] 所述的虚拟转发设备的转发限速方法，其中该步骤3包括：该虚拟转发设备利用该最终CPU资源对租户转发数据包进行限速，超过该最终CPU资源时的数据流量将等待或被丢弃。

[0012] 所述的虚拟转发设备的转发限速方法，其中该基础CPU资源、该空闲CPU资源和该最终CPU资源的单位均为CPUcycles/s。

[0013] 所述的虚拟转发设备的转发限速方法，其中该步骤3包括：使用令牌桶或公平队列的方式对租户转发数据包进行限速。

[0014] 所述的虚拟转发设备的转发限速方法，其中该虚拟转发设备为虚拟路由器或虚拟交换机。

[0015] 本发明还提出了一种虚拟转发设备的转发限速系统，其中包括：

[0016] 模块1，用于测量得到服务提供商网络平均数据包大小和流数目作为流量特征，该服务提供商网络包含虚拟转发设备和多个租户，使用流量生成工具生成符合该特征的流量，通过在该虚拟转发设备中转发该流量的数据包并测量达到各个带宽所用于转发的CPU资源，构建CPU资源和带宽的对应关系；

[0017] 模块2，用于根据各租户的基础带宽需求以及该对应关系，为各租户分配基础CPU资源，并得到该服务提供商网络在满足该基础CPU资源后剩余的CPU资源，以为每一位租户按照权重再分配空闲CPU资源，每位租户用于转发的最终CPU资源为其分配的基础CPU资源和空闲CPU资源之和，该虚拟转发设备使用该最终CPU资源对其对应的租户完成数据包的转发。

[0018] 所述的虚拟转发设备的转发限速系统，其中该虚拟转发设备利用该最终CPU资源对租户转发数据包进行限速，超过该最终CPU资源时的数据流量将等待或被丢弃。

[0019] 所述的虚拟转发设备的转发限速系统，其中该基础CPU资源、该空闲CPU资源和该最终CPU资源的单位均为CPUcycles/s。

[0020] 所述的虚拟转发设备的转发限速系统，其中该步骤3包括：使用令牌桶或公平队列的方式对租户转发数据包进行限速。

[0021] 所述的虚拟转发设备的转发限速系统，其中该虚拟转发设备为虚拟路由器或虚拟交换机。

[0022] 由以上方案可知，本发明的优点在于：

[0023] (1)隔离性。该方法从根本上消除了不同的流或租户对软件转发设备中CPU资源的竞争，通过按照带宽-CPU资源的对应关系来合理分配CPU资源，每个流或租户能够享有自己所需的CPU资源来做流量转发而不受其他用户的影响，实现了带宽隔离性。

[0024] (2)鲁棒性。鲁棒性是在隔离性的前提下实现的，我们提出的基于CPU-cycle的限速方法在保障了基本的带宽以后，为了充分利用软件转发设备中的CPU资源，将CPU上尚且空闲的资源按权重分给了流或租户，能够一定程度上抵御流量突发造成的丢包。

[0025] (3)平台无关。我们提出是一种方法，其具有平台无关的特性，包括从建立带宽-CPU的对应关系、为不同的流或租户分配CPU资源，都是可以在任何软件实现的转发设备上

实现的，例如vRouter，vSwitch等。

[0026] 综上，本发明关注的是软件转发下的限速方法的带宽隔离性问题，提出了一种基于CPU-cycle的限速方法。限速的指标是硬件CPU资源而不是传统的BPS/PPS等带宽指标，通过构建软转发设备中的带宽-CPU对应关系，为流或租户分配它们所需的转发CPU资源，用限制CPU资源的方式来实现对这些流或租户的限速和隔离性。在CPU资源层面上的分配和隔离保证了这些流或租户之间不受干扰，同时充分分配空闲的CPU资源又能抵御突发流量的丢包影响，实现了隔离且鲁棒地限速。

## 附图说明

[0027] 图1为本发明的方法流程图；

[0028] 图2为本发明实施例示意图；

[0029] 图3为基于CPUcycle的令牌桶机制的每次批处理转发流程图。

## 具体实施方式

[0030] 本设计为了解决软件转发设备中现有的限速方法隔离性差的问题，为了对CPU竞争加以限制和合理分配转发CPU资源，我们提出了一种基于CPU-cycle的限速方法，用转发CPU资源的分配来实现限速的功能。

[0031] 该方法与传统的限速方法最大的区别在于限速的指标。传统的方法直接按照带宽的指标(即BPS或PPS)进行限速，高于带宽指标的流量则丢弃或等待，因而CPU资源的竞争则会严重影响限速效果，损害网络性能的隔离性。而我们则从最基本转发引擎的硬件资源出发，由于软件转发设备的转发速率是由专用的CPU资源提供的，我们提出的基于CPU-cycle的限速方法则不再按照带宽指标来判断流量允不允许转发，而是通过限制租户的转发任务在转发设备中每秒消耗的CPUcycles(CPU周期，表示CPU资源的单位，如一个2.0GHz的CPU核其每秒有2.0G个CPUcycles)来为其限制转发速率。租户可为与虚拟交换机相连接的虚拟机，也可以是与虚拟路由器相连接的各种虚拟设备。

[0032] 为了实现通过限制每个流/租户消耗的CPUcycles/s来限制其转发速率，首先需要建立转发设备内CPUcycles/s与带宽的对应关系。在不同的软件转发设备中，影响该带宽-CPU对应关系的条件大致包含两大类：

[0033] 1)流量特性。数据包大小、并发流的数目都会影响这个对应关系。比如说同样带宽下，64字节数据包消耗的CPU资源更多，而并发流数目多也会导致转发设备中查表的步骤较慢，消耗更多CPU资源。因此为了消除该类条件对带宽-CPU对应关系造成的影响，服务提供商可以在售卖带宽时约束为在特定的数据包大小和特定流数目情况下的带宽，比如在转发设备上采集一段时间(比如一小时)的流量，以获取流经转发设备的流量中平均的数据包大小和流数目。并以此来确定带宽-CPU资源的对应关系。

[0034] 2)平台特性。不同的CPU平台的硬件性能不同，能够转发的带宽也不同，因此平台也是很容易影响带宽-CPU对应关系；软件转发设备中的路由表、流表等规则数目和复杂度也会影响对应关系的构建。因此这些平台特性需要在对应平台上测量得出。

[0035] 在考虑了上述两大特性后，可以构建出基于测量的带宽-CPU对应关系，在该关系指导下，每个流或租户可以得到在实际场景中达到其购买带宽所需的基本CPU资源$C_{basic}$(单

位是CPUcycles/s)，这可以保证基础的带宽。在基础带宽得到保障后，为了充分利用CPU资源，还需要根据当前CPU核上剩余的资源情况，为每条流按照权重再分配空闲的CPU资源$C_{idle}$以充分利用空闲资源实现所有用户最大带宽，这里的$C_{idle}$包含两部分的空闲CPU资源：一类是核上未分配的CPU资源，另一类是已分配但是未使用的资源。最终，每个流或租户，分得的用于转发的CPU资源为$C_{basic}+C_{idle}$，这样保证了在带宽保障和限速层面上既拥有隔离性又具有鲁棒性，能够一定程度上抵御流量突发的影响。

[0036] 在将CPU资源分配完毕后，下一步就是只利用这些分配的CPU资源为对应的流或租户完成流量的转发。可以使用令牌桶、公平队列等方式，利用这些分配好的有限CPU资源来限速，对一个流或租户来说，超过这些单位被分配到的CPU资源处理能力之外的流量将等待或被丢弃。这样就实现了流或租户在硬件CPU资源层面的隔离性，进而实现了带宽的良好隔离。

[0037] 为让本发明的上述特征和效果能阐述的更明确易懂，下文特举实施例，并配合说明书附图作详细说明如下。

[0038] 下面为本发明提出的限速方法的一种实施例。

[0039] 该实施例选用的openvSwitch作为一个典型的软件转发平台，open vSwitch是业界的虚拟交换机标杆，被广泛使用。我们在其上使用基于CPU-cycle的令牌桶实现了该限速方法，令牌桶中的令牌数是剩余CPUcycles，令牌的生成速率单位是CPUcycles/s。

[0040] 限速方法的实施如图1所示，分为如下步骤：

[0041] 1.在服务器平台上(包含多个虚拟机和一个openvSwitch)测量平均流量的带宽-CPU对应关系。测量该对应关系需要先测得一段时间内通过open vSwitch中的流量的平均数据包大小和流数目等特征，然后使用流量生成工具生成符合平均数据包大小和流数目特征的流量，在openvSwitch中转发该流量的数据包并测量达到各个带宽规格所需要用于转发的CPU资源，即$C_{basic}$。

[0042] 2.为每个租户根据其购买带宽分配令牌生成速率(即CPUcycles/s)。通过上述第一步中建立的对应关系，根据购买的带宽找到对应的CPUcycles/s以及空闲CPU资源情况，设置对应令牌桶的令牌生成速率。如图2所示，4个VM在openvSwitch中使用一个2.2GHz的CPU核来做转发，假设它们按照带宽-CPU对应关系分到的$C_{basic}$分别为0.2Gcycles/s、0.2Gcycles/s、0.4G cycles/s、0.4Gcycles/s，还剩余1Gcycles/s的空闲CPU资源，那么在保证隔离性基础上按权重分配，因此每个VM实际设置的令牌生成速率分别为0.36Gcycles/s、0.36Gcycles/s、0.72Gcycles/s、0.72Gcycles/s。

[0043] 3.使用基于CPU-cycle的令牌桶在转发过程中进行限速。令牌桶的机制运行流程如图3所示，每次当有流量需要转发时，先去检查对应的租户的令牌桶内是否还有CPUcycles。如果剩余的CPUcycles大于0，则为其转发一批数据包，并在转发过程结束后，将消耗的CPUcycles在令牌桶中减去；如果剩余的CPUcycles小于0，则跳过本次批转发，为下一个流或租户转发流量。

[0044] 在上述过程中，我们展示了如何用一种基于CPU-cycle的令牌桶机制来实现该限速方法，其中令牌桶内的令牌生成速率和转发过程中令牌桶的作用机制是实现该限速方法的重点。但是该限速方法不仅可以用令牌桶实现，一些其他的机制如公平队列，也能够分配并隔离CPU资源，并实现这种限速方法。

[0045] 通过该方法对软件转发设备中专用于转发的CPU资源的隔离和分配实现了限速的隔离性，使得流和租户不再会互相影响，可以灵活地应用在各种软件转发场景下，为用户提供稳定的带宽保障服务。

[0046] 以下为与上述方法实施例对应的系统实施例，本实施方式可与上述实施方式互相配合实施。上述实施方式中提到的相关技术细节在本实施方式中依然有效，为了减少重复，这里不再赘述。相应地，本实施方式中提到的相关技术细节也可应用在上述实施方式中。

[0047] 本发明还提出了一种虚拟转发设备的转发限速系统，其中包括：

[0048] 模块1，用于测量得到服务提供商网络平均数据包大小和流数目作为流量特征，该服务提供商网络包含虚拟转发设备和多个租户，使用流量生成工具生成符合该特征的流量，通过在该虚拟转发设备中转发该流量的数据包并测量达到各个带宽所用于转发的CPU资源，构建CPU资源和带宽的对应关系；

[0049] 模块2，用于根据各租户的基础带宽需求以及该对应关系，为各租户分配基础CPU资源，并得到该服务提供商网络在满足该基础CPU资源后剩余的CPU资源，以为每一位租户按照权重再分配空闲CPU资源，每位租户用于转发的最终CPU资源为其分配的基础CPU资源和空闲CPU资源之和，该虚拟转发设备使用该最终CPU资源对其对应的租户完成数据包的转发。

[0050] 所述的虚拟转发设备的转发限速系统，其中该虚拟转发设备利用该最终CPU资源对租户转发数据包进行限速，超过该最终CPU资源时的数据流量将等待或被丢弃。

[0051] 所述的虚拟转发设备的转发限速系统，其中该基础CPU资源、该空闲CPU资源和该最终CPU资源的单位均为CPUcycles/s。

[0052] 所述的虚拟转发设备的转发限速系统，其中该步骤3包括：使用令牌桶或公平队列的方式对租户转发数据包进行限速。

[0053] 所述的虚拟转发设备的转发限速系统，其中该虚拟转发设备为虚拟路由器或虚拟交换机。

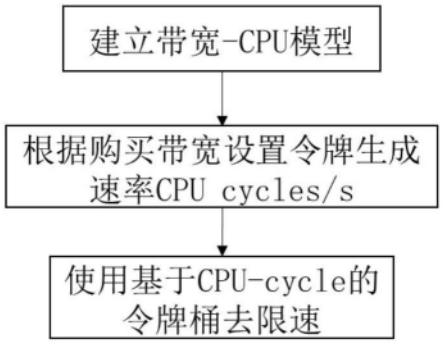



图1

| | VM1 | VM2 | VM3 | VM4 |
|---|---|---|---|---|
| 权重 | 1 | 1 | 2 | 2 |
| 购买带宽 | 1G bps | 1G bps | 2G bps | 2G bps |
| $C_{basic}$ (cycles/s) | 0.2 G | 0.2 G | 0.4 G | 0.4 G |
| $C_{basic} + C_{idle}$ (cycles/s) | 0.2 + 0.16 G | 0.2 + 0.16 G | 0.4 + 0.32 G | 0.4 + 0.32 G |

图2

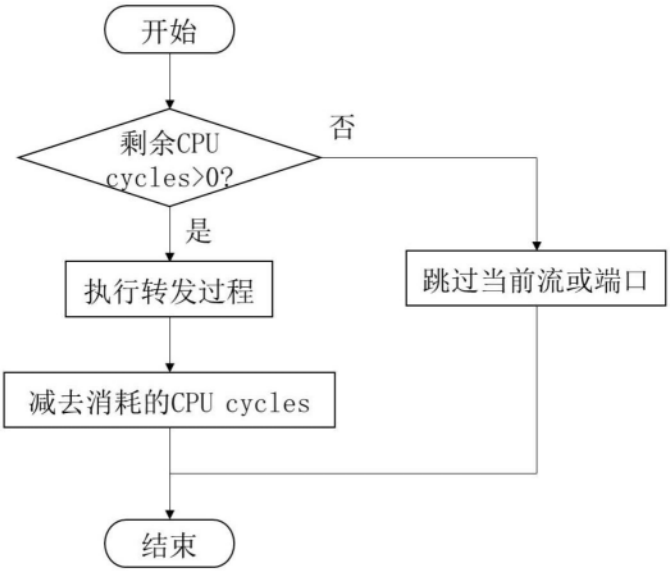



图3