

文章编号:1007-5321(2021)02-0026-07

DOI:10.13190/j.jbupt.2020-128

# 基于网络演算理论的 Damper 调度机制

任首首, 刘冰洋, 王 闯, 孟 锐, 刘 轩

(北京华为数字技术有限公司, 北京 100195)

**摘要:** 提出了一种新的报文调度机制, 以实现网络端到端确定性低时延和低抖动. 该机制利用网络演算理论来提供逐跳的时延上界指导, 并通过逐跳增加 Damper 的方式对报文进行主动等待, 实现对报文的逐跳聚合整形, 消除了传统网际互联协议(IP)网络中由于波峰叠加造成的时延恶化. 该机制不要求网络设备之间时间同步, 不需要核心节点维护逐流状态, 可支持大规模传输距离, 在大规模 IP 网络中扩展性好.

**关 键 词:** 确定性低时延; 确定性低抖动; 网络演算; Damper

**中图分类号:** TN911.22

**文献标志码:** A

## A Damper Scheduling Mechanism Based on Network Calculus Theory

REN Shou-shou, LIU Bing-yang, WANG Chuang, MENG Rui, LIU Xuan

(Beijing Huawei Digital Technologies Company Limited, Beijing 100195, China)

**Abstract:** A new packet scheduling mechanism, which can provide bounded latency and jitter for Internet protocol (IP) network, is proposed. With the guidance of network calculus, the Damper scheduling mechanism performs per-hop shaping by actively holding packets in the damper module per-hop and thereby eliminating the delay deterioration. The Damper mechanism requires neither time synchronization among network devices, nor per-flow status maintenance on core devices. It supports large-scale transmission distances and thus can be extended on large-scale IP networks.

**Key words:** deterministic latency; deterministic jitter; network calculus; Damper

近年来,增强现实/虚拟现实(AR/VR, augmented reality/virtual reality)、远程医疗、智能制造、自动驾驶等新兴应用悄然而至. 相比传统的 Web 浏览、语音等业务,这类新应用对网络提出了更为严苛的要求. 不仅要求网络为业务应用提供确定性的带宽,还要保证端到端确定性低时延和确定性低抖动<sup>[1]</sup>. 例如,在远程医疗场景下,为保证医生感受不到明显地延迟和抖动,要求端到端时延不超过 20 ms,抖动低于 200  $\mu\text{s}$ <sup>[1-3]</sup>. 又如,在第 3 代合作伙伴计划(3GPP, 3rd generation partnership project)第 5 代移动通信系统(5G, the fifth generation of mobile communications system)需求文稿<sup>[4]</sup>中公开的很多工业控制场景中,控制器需要远程控制机械臂完成很

多精细的操作,要求控制器和机械臂之间的时延小于 1 ms,抖动小于 1  $\mu\text{s}$ <sup>[4-5]</sup>. 传统的国际互联协议(IP, Internet protocol)网络采用“尽力而为”的服务模式,不同流之间会发生报文碰撞和挤压,导致报文的端到端时延和抖动不可控,因此难以为上述服务提供端到端的确定性服务.

笔者基于 Damper 的概念<sup>[6]</sup>,给出了贴近实际网络设备的 Damper 转发架构和完善的 Damper 理论模型,以实现网络端到端确定性低时延和确定性低抖动. 笔者进一步量化分析了 Damper 模型的端到端时延性能和抖动性能,并给出了相关仿真验证结果;最后分析了 Damper 机制在具体产品实现上可能面临的挑战.

收稿日期: 2020-08-20

作者简介: 任首首(1988—),男,主任工程师, E-mail: renshoushou@huawei.com.

# 1 确定性的概念和业界进展

## 1.1 确定性的概念

长久以来,业界关于 IP 服务质量 (QoS, quality of service) 的研究大多聚焦在端到端的带宽或者流的平均速率的确定性上,很多技术都可以保证某条流端到端带宽的确定性。笔者在确定性带宽的基础上主要关注网络如何为业务流提供端到端的确定性低时延和确定性低抖动的保障。带宽确定性是指网络为某条流提供的端到端服务速率不低于所要求的带宽。时延确定性是指对于某条流内的任意一个包,其在网络中经历的端到端时延均不超过某个值,即网络为该流保证确定的时延上界。某条流的抖动则是指该流的报文所可能经历的时延上界和时延下界的差值。

带宽或者流平均速率是一个宏观时间尺度上的(通常在分钟级甚至小时级)具有统计意义的概念,不涉及每个包的具体行为。相比之下,某条流在网络中的时延不确定性和抖动是微观时间尺度(通常在微秒甚至亚微秒级别)上的报文竞争所导致的。例如,在某台网络设备内部,来自不同入端口的报文会在相同的出端口上发生碰撞或者干扰,从而产生竞争和排队。进一步,多条流之间的突发碰撞会导致单流的波形变化,多流波峰叠加形成排队时延和抖动;而流量突发的相互影响、恶性循环、开放增长会导致时延和抖动随跳数呈线性增长。因此,尽管传统的 QoS 保障机制可以保证确定性带宽,但是仍然无法保障时延和抖动的确定性,需要一种新的调度机制来为网络提供此能力。

## 1.2 业界研究进展

20 世纪 90 年代,互联网工程任务组 (IETF, the Internet engineering task force) 提出用综合服务模型 (InteServ, integrated service)<sup>[7]</sup> 的架构来保障 IP 的 QoS。该架构通过控制面信令为每条流沿路预留相应的带宽资源;在数据面上逐跳做逐流调度,提供带宽和时延保障,不同的数据面调度方法有着不同的时延性能。一类比较常见的 InteServ 调度方法是基于一般处理器的共享模型 (GPS, general processor sharing) 衍生而来的调度方法,包括 Packetized GPS, WF2Q, Virtual Clock, SFQ 等<sup>[8]</sup>,这类算法需要设备维护较大的逐流状态开销;另一类常见的调度方法是基于轮询 (RR, round robin) 的调度方法,如 Weighted RR, Deficit RR, Modified Deficit RR, Urgan-

cy RR 等<sup>[8]</sup>。这类算法的端到端时延上界可以利用网络演算理论<sup>[9]</sup> 计算得到,但是通常情况下时延上界都比较大,并且随着流数量的增加,时延恶化很严重。概括来讲,InteServ 调度方法可以提供端到端确定性的时延保障,但是都需要网络设备维护逐流的状态,甚至维护逐流队列,因此其可扩展性很差,难以在大规模 IP 网络中使用。在 InteServ 之后,IETF 又提出用区分服务 (DiffServ, differentiated services)<sup>[10]</sup> 的转发架构来保障 IP 网络的 QoS。该架构的基本思想是在网络边缘设备上将进入网络的流分成不同的类型,并为数据包打上不同的类型标记,如不同优先级等,网络节点通过检查包头来确定对包进行何种处理。由于高优先级流量之间的竞争依然存在,所以该架构仍然无法保障端到端的时延和抖动的确定性。

近年来,电气与电子工程师协会 (IEEE, institute of electrical and electronics engineers) 和 IETF 都分别成立了网络确定性相关的标准化工作组。IEEE 时延敏感网络 (TSN, time-sensitive networking) 工作组提出了一系列用于以太网的标准,包括 802.1Qbv, 802.1Qch, 802.1Qcr<sup>[11]</sup> 等。但是这些标准技术要么要求所有网络节点时间同步,要么仅支持短距离链路,要么需要节点维护逐流状态,难以应用于大规模 IP 网络中。IETF Detnet 工作组则还没有明确的数据面转发方案<sup>[12]</sup>。工业界的一些其他保证零抖动的网络如 Powerlink, EtherCAT 等,采用了特殊的以太网标准,也无法在现有 IP 网络中使用;另外,基于网络演算理论进行确定性分析也渐渐成为业界热点。Li 等<sup>[13]</sup> 基于 Time triggered network 进行了相关分析,但是该网络的可扩展性很差。Zhang 等<sup>[14]</sup> 基于网络演算理论分析了 TSN 相关机制的性能,但是未提出新的调度方法,仅仅是对 TSN 机制的理论性能分析,并未解决前述提到的 TSN 自身的问题。Grigorjew 等<sup>[15-16]</sup> 假设用某种方法可以保证数据流每逐跳的时延上界,以此获得确定的端到端 (E2E, end-to-end) 时延上界,但是由于流之间的相互挤压,使流量突发随着跳数的增加而不断增大,且未给出具体方法说明如何解决该问题并保证在非首跳节点上的时延上界,可行性值得商榷。

## 1.3 主要贡献

基于前人提出的 Damper 概念,笔者对架构和流程做了进一步设计,以保证 E2E 的确定性时延和抖动。笔者的主要贡献包括以下 3 方面:① 结合具

体网络设备,给出了实现 Damper 模型的多种设计架构,弥补了业界在此方向上的空白;② 结合一种最贴近实际网络产品的可实现架构,给出了具体的报文处理流程和理论方法,还给出了如何能够实现异步调度的方法,极大地增加了其在大规模网络中的可扩展性;③ 从理论角度与传统 DiffServe 算法的性能进行了对比,给出了普适、精确的公式化结论并给出了初步的仿真验证结果。

## 2 网络演算简介

网络演算(NC, network calculus)<sup>[9]</sup>是一种基于最小加代数(min-plus algebra)的确定性排队理论。在给定流量输入模型和网络设备服务能力的条件下,基于网络演算理论可以计算出报文在网络的端到端排队时延上界。

1) 到达曲线。把某条流活跃过程中的任意时间点定义为0时刻,选取任意时间长度 $t$ ,如果在 $(0, t]$ 时间段内,该流到达的数据量不超过某个函数 $\alpha(t)$ ,则称 $\alpha(t)$ 为该流的到达曲线。到达曲线描述的是流的到达行为,一条流的到达曲线描述了该流的实际到达流量的上包络,任意时间段内达到的流量均不会超过该包络。在分组网络中,对于一个持续速率为 $r$ 的流,其任一瞬间可能突发一个大小不超过 $b$ 的数据量,那么其流量到达曲线通常可以描述为 $\alpha(t) = rt + b$ 。该模型称为固定比特率(CBR, constant bit rate)到达模型。CBR模型是分组网络中最简单的,也是网络演算理论应用时最容易计算的到达曲线模型<sup>[9]</sup>。

2) 服务曲线。当设备某个出端口队列不为空时,以任意时间为0时刻,选取任意 $t$ , $(0, t]$ 内设备处理数据量的下限不低于函数 $\beta(t)$ ,称 $\beta(t)$ 为该出口的服务曲线。服务曲线描述的是设备的服务能力,是设备在某个出口上服务能力的下包络。在队列不为空时,任意时间段内设备能够服务的数据不低于服务曲线。对于分组网络的转发设备(路由器和交换机等),通常可以将其某个物理接口的整体服务能力描述为 $\beta(t) = \max\{C(t - T), 0\}$ ,其中 $C$ 代表该接口的总带宽, $T$ 代表该接口服务第一比特的延迟,对于某个接口整体而言, $T$ 通常为0。对于通过该接口的某条具体的流 $i$ 而言,在该端口上得到的服务曲线 $\beta_i(t) = \max\{r_i(t - T_i), 0\}$ ,其中 $r_i \leq C$ ,代表该流得到的统计服务速率; $T_i$ 则代表该流的第一比特得到服务的延迟。这在多队列分组网络中很容

易理解,当多个队列都不为空时,1个分组到达后不一定能够立即被处理,调度器可能需要先调度其他队列,经过 $T_i$ 时间才调度该队列,此后保障该队列的调度总量不低于 $r_i t$ 。这类服务模型被称为RL(rate latency)模型。目前几乎所有的调度器都可以被建模为RL模型,该模型也是网络演算理论应用时常用到的服务曲线模型。

3) 时延上界。对于某条流,如果给出了其到达曲线 $\alpha(t) = rt + b$ 和某个节点为该流提供的服务曲线 $\beta_i(t) = \max\{r_i(t - T_i), 0\}$ ,那么该流在该节点经历的时延存在某个上界值,该上界值为 $\alpha(t)$ 和 $\beta_i(t)$ 在时间轴上的最大距离 $T_i + \frac{r_i}{b}$ ,即对于该流的任意一个报文,其节点中经历的排队时延不会超过该上界值。笔者后续将会用这个结论作为排队时延的指导。

## 3 Damper 理论模型和转发架构

### 3.1 Damper 的概念

Damper的初始概念源于Verma<sup>[6]</sup>提出的基本构想,即在下游节点上通过主动等待的方式,弥补不同报文在上游节点经历的时间差,以保证所有报文逐跳经历的时间均相同,从而消除了流波峰叠加带来的不确定性。图1所示为Damper的概念模型。对于任意一个上游节点某个出口的队列系统,下游节点上总有一个相应的Damper模块与之对应。假设已知任意报文在上游某出口对应的队列系统中可能经历的时延上界为 $D$ ,那么对于某个报文 $i$ ,如果其实际经历的时延 $q < D$ ,那么下游对应的Damper模块需要主动控制住该报文 $D - q$ 时间,然后才会将该报文释放给本节点的队列系统。通过该模型可以保证任意一个报文从进入上游节点的排队系统到被下游节点的Damper释放,经历的时延刚好等于 $D$ ,为报文保证了确定的时延上界。进一步地,对于任意一条流,Damper可以保证其进入队列系统是什么样,从Damper出来仍然是什么样,不会受到其他流的影响,从而消除逐跳波峰叠加带来的时延不确定

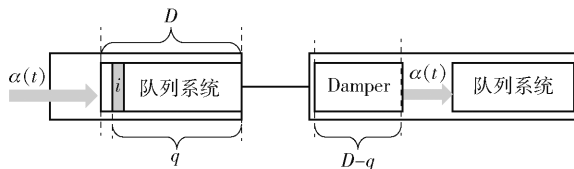


图1 Damper 的概念模型



性. 需要注意的是, 这里的时延上界  $D$  是一个与上游队列系统紧耦合的值, 对于不同的队列系统而言, 时延上界可以不同.

### 3.2 基于 Damper 的转发架构

基于图 1 的概念, 结合实际的路由器设备, 提出了 3 种更为详细的 Damper 转发架构, 如图 2 所示. 简单起见, 图中每台设备只示意性地给出了 4 个接口. 根据图 1 的概念模型, 最直观的架构是图 2(a) 所示的上行板独享 Damper 的架构. 该架构下, 所有的 Damper 模块都在设备的上行板实现, 并且每个不同的入接口都有一个独立的 Damper 模块. 但是, 对于实际的路由器设备来讲, 通常上行板的能力很弱, 比较难实现 Damper 这种复杂的功能. 图 2(b) 将 Damper 模块放到了下行板上. 由于去往同一个出接口的报文可能是从不同的入口进来的, 而不同入口对应了不同的上游节点, 所以, 从图 1 所示概念模型得出的直观架构中, 每个出接口上都需要  $N$  个 Damper 模块, 分别对应设备的  $N$  个接口. 尽管下行板的能力比上行板要强很多, 但是在图 2(b) 所示的设计中, 整台设备需要  $N^2$  个 Damper 模块.

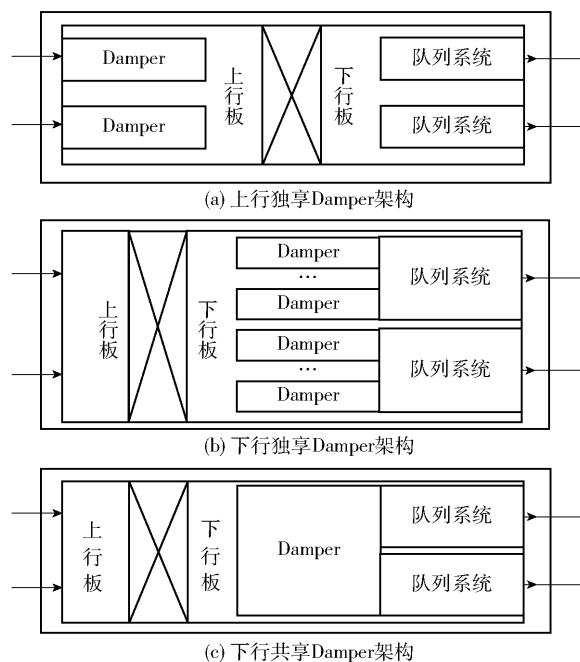


图2 基于 Damper 的 3 种转发架构

从图 1 所示的模型可以看出, Damper 的能力是针对不同报文经历的实际时延等待相应时长的, 只需要知道将报文控制的时间, 可以不关心报文从哪个入口进来. 因此, 如果报文中携带了相关时间信息, 那么可以对下行独享 Damper 架构进行改进. 如

图 2(c) 所示, 设备整体只需要 1 个 Damper 模块, 该模块根据报文携带的时间信息对报文进行主动等待, 无须关心报文属于哪条流, 从哪个上游节点来. 类似地, 共享的 Damper 模块也可以放在上行板, 不再赘述.

### 3.3 转发流程

结合最贴近产品能力的转发架构 [见图 2(c)], 给出了详细的 Damper 模块和报文转发流程, 如图 3 所示. 一个 Damper pair  $h$  由上游节点  $h$  的队列系统、下游节点  $h+1$  的上行板和交换结构 (SW, switch fabric) 以及下游节点  $h+1$  上的 Damper 模块 3 部分组成. 在该模型下, Damper 模块的作用是保证任意一个经过该 pair 的报文, 在该 pair 内部经历的时间均严格等于一个固定值  $D^h$ . 对于 2 个固定的设备来讲, 传输线路通常是固定距离, 其引起的时延为常数, 因此图 3 所示模型中的一个 pair 不包括传输线路部分和传输时延, 有

$$D^h = Q^h + P^{h+1} \quad (1)$$

$$D^h = q^h + p^{h+1} + d^{h+1} \quad (2)$$

其中:  $P^{h+1}$  为报文在设备  $h+1$  的上行板和交换结构中经历的时延的上界, 称为最大处理时延;  $p^{h+1}$  为报文在上行板和交换结构中经历的实际时延. 对于网络设备而言, 通常可以认为  $P$  是一个固定的常数.  $Q^h$  为报文在设备  $h$  相应出口上的排队系统中可能经历的时延上界. 一旦给定了设备  $h$  某个接口的链路带宽以及经过该接口的流量平均速率和最大突发度, 根据第 2 节中的结论可以计算得到  $Q^h$ .  $q^h$  为报文在设备  $h$  的排队系统中经历的实际时延.  $d^{h+1}$  则代表了该报文需要在设备  $h+1$  的 Damper 模块中等待的时间. 将报文从设备  $h+1$  的 Damper 模块释放的时刻称为该报文的 Eligibility time, 记为  $E^{h+1}$ . 对于任意一个报文, 从  $E^{h+1}$  之后到被设备  $h+1$  发送之前经历的这段时间是属于 pair  $h+1$  的.

需要注意的是, 这里假设采用 Damper 机制进行调度的流都是具有确定性需求的流, 通常这类业务流需要向网络承诺其平均速率和最大突发度; 否则任何方法都无法为其保证确定性. 在 Damper 机制中, 确定性流量要比 best effort 流量具备更高的优先级. 图 3 所示的模型中只给出了一个 pair, 实际中, 节点  $h+1$  可以有多个入接口, 分别对应不同上游节点的出口队列系统, 即节点  $h+1$  上的 Damper 可以和多个上游节点处接口队列系统组成多个不同的 pair, 每个 pair 有着不同的时延上界. 另外, 未考

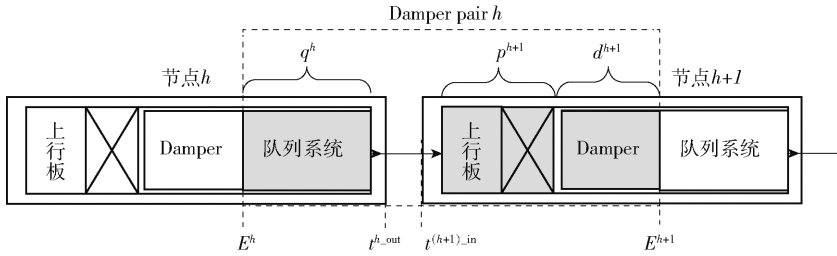


图3 基于下行共享 Damper 架构的模型

虑模型中报文在出入接口上的串行化时间,实际应用中可以通过增加  $D$ ,将这部分时间囊括进来,不会对模型的基本原理产生影响。

在图3所示的模型下,报文的转发流程如下:

- 1) 设备  $h$  收到某报文  $i$ ;
- 2) 设备  $h$  解析报文  $i$ ,并根据报文携带的相关时间信息,计算报文  $i$  在设备  $h$  上的释放时间  $E_i^h$  (在不引起歧义的情况下,后文中会省略下标  $i$ );
- 3) 设备  $h$  将报文  $i$  放入 Damper 模块,等待至  $E^h$  时刻,将该报文释放给本设备的队列系统;
- 4) 设备  $h$  通常是以先进先出 (FIFO, first in first out) 的调度方式发出报文,发出之前需要将报文  $i$  在本跳上 Damper 机制相关的时间信息封装到报文中;
- 5) 设备  $h+1$  收到报文,按照步骤 2) ~ 4) 的方法处理报文。

上述转发流程中最重要的一步就是根据报文  $i$  携带的相关时间信息计算其在本跳释放时间。从 Damper 的概念直观来看,需要报文携带  $E^h$  和  $Q^h$ ,有

$$E^{h+1} = E^h + P^{h+1} + Q^{h+1} + x \quad (3)$$

其中  $x$  为链路传输时延。由于  $E^h$  和  $E^{h+1}$  为报文在不同设备上的时间,所以设备间的时间须同步。但进一步分析可以发现,Damper 模块的目标是补偿报文  $i$  实际经历的时间和最大时延  $D$  之间的差值,因此  $E^{h+1}$  可以通过式(4)来计算。

$$E^{h+1} = t^{(h+1),in} + P^{h+1} + [Q^h - (t^{h,out} - E^h)] \quad (4)$$

其中:  $t^{(h+1),in}$  为设备  $h+1$  收到该报文的时刻,  $t^{h,out}$  为设备  $h$  发出该报文的时刻。可以看出,式(4)中括

号内的3项作差后得到了1个相对时间,即报文在设备  $h$  队列系统中的时延上界和实际经历时间的差值。因此,从式(4)来看,Damper 机制不需要设备之间的时间同步,报文中只需要携带  $E^h$ ,  $t^{h,out}$  和  $Q^h$  信息,设备  $h+1$  即可根据这些信息计算出报文的  $E^{h+1}$ 。

## 4 性能分析

### 4.1 端到端时延和抖动

基于图3所示的模型,可以得到如图4所示的端到端 Damper 模型系统。该网络系统共包括  $H$  跳网络设备,其时延上界(排除线路传输时延)表示为

$$D^{E2E} = \sum_{h=1}^H D^h \quad (5)$$

其中:  $D^1 - D^{H-1}$  分别对应了从第1跳开始的  $H-1$  个 Damper 对。对于  $D^H$ ,可以将其看做只有排队系统而没有下游 Damper 模块的不完整对,因此其上界也可以按照排队系统时延上界的计算方法得到  $D^H = Q^H$ 。从式(5)可以看出,Damper 系统可以为流提供确定性的端到端时延保证。

在 Damper 机制下,网络设备  $1 - H - 1$  不会产生抖动,该系统的端到端抖动仅来源于最后1条,因此端到端的抖动为

$$J^{E2E} = D^H \quad (6)$$

根据不同的接入流量,结合首跳整形技术和网络演算的理论,可以将这个抖动的大小控制在十几微秒,甚至几微秒之内。

基于端到端的 Damper 系统,用图5所示的场景

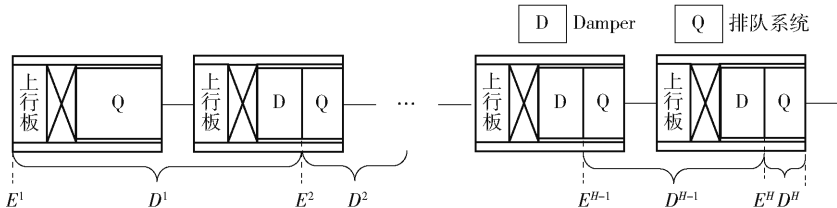


图4 端到端的 Damper 系统

对 Damper 机制进行了仿真验证. 图中所有设备出口带宽均为 10 Gbit/s, 各流均为确定性流, 采用最高优先级. 其中, 流 1 的平均速率为 2.24 Gbit/s, 最大突发为 1.4 KB; 流 2 和流 3 的平均速率均为 6.72 Gbit/s, 最大突发为 4.2 KB; 流 4 和流 5 的平均速率为 3.36 Gbit/s, 最大突发为 2.1 KB. 另外, 还有一些 best effort 流量(图中未画出)经过各跳设备, 采用低优先级. 各跳出接口的排队时延上界( $Q$ 值)按照网络演算理论进行设置, 均为  $5 \mu\text{s}$ .

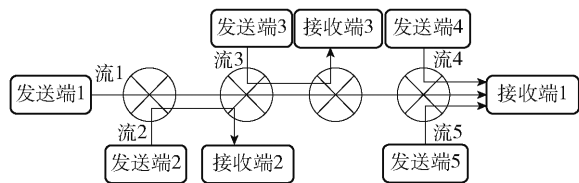


图 5 仿真场景

通过实验观测了流 1 中报文端到端的排队时延. 由于流 2~5 分别在各跳设备和目标流经历相同的出接口, 所以可以模拟出高优先级流之间的干扰效果. 经过观测并统计了流 1 长时间内报文的 E2E 排队时延情况, 如图 6 所示. 可以看出, 即使在有干扰流存在的情况, 基于 Damper 调度机制, 仍然可以保证目标流端到端的排队时延不超过  $\sum Q^h = 20 \mu\text{s}$ , 并且抖动也不超过末跳的最坏时延  $Q^4 = 5 \mu\text{s}$ .

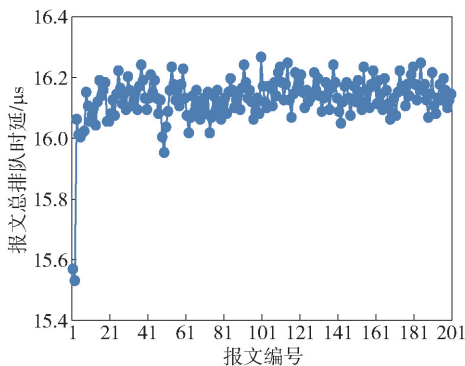


图 6 报文 E2E 排队时延(部分)

## 4.2 与传统 DiffServ 的对比

根据文献[17-18]中的结论, 传统 DiffServ 端到端排队时延的上界为

$$D^{\text{diff}} = \frac{\tau H}{1 - (H-1)v} \quad (7)$$

其中:  $\tau$  为经过某个出口上所有流初始突发度之和的串行化时延(初始突发度总和除以链路带宽)的

最大值,  $v$  为最大链路利用率. 可以看出, 只有  $v < \frac{1}{H-1}$  时, 该方法才存在时延上界, 即 DiffServ 方法的最大利用率不超过  $\frac{1}{H-1}$ .

Damper 机制下, 根据网络演算的理论, 每跳的排队时延上界为  $Q^h = \tau^h$ . 从网络演算理论可知,  $\tau^h$  不超过经过设备  $h$  的所有流出口的突发度之和的串行化时间, 所以有  $\tau^h < \tau$ .

$$D^D \leq \tau H \quad (8)$$

对比式(7)和式(8)可以发现, 在网络接入流量相同的条件下, Damper 的时延要小于 DiffServ 的时延. 该前提下, Damper 对于链路利用率没有要求, 即可达到 100% 的链路利用率. 因此, Damper 可接入的确定性流量带宽是 DiffServ 的  $H-1$  倍.

## 4.3 可扩展性

相比 TSN 已有的确定性调度方法, Damper 机制具有很强的可扩展性, 具体表现包括: ①网络核心节点(除入口节点之外的其他网络设备)不需要维护逐流的状态, 只需要维护逐接口的状态, 即各个接口的最大排队时延; ②各节点的队列系统采用 FIFO 队列即可, 调度复杂度为  $O(1)$ ; ③网络设备之间无须时间同步, 异步调度即可; ④不关心链路传输时延, 可支持长距离链路. 综上所述, Damper 机制可以应用于大规模 IP 网络中, 可扩展性很强.

## 5 Damper 面临的挑战

Damper 理论模型假设设备可以在报文的释放时间时刻准确地将报文释放给本设备的队列系统, 不能过早也不能过晚. 但是, 在实际的路由器或者交换机设备中, 要实现这一点非常困难. 因此, 如何利用现有设备实现或者近似实现 Damper 理论模型, 还需要解决如下问题:

1) Damper 机制要求设备有门控能力, 并且门控的粒度须无限精细, 才能保证在准确的时间打开门控并且释放报文; 但是目前的设备只能做到十几微秒级别的门控粒度, 无法保证精确释放每个报文;

2) Damper 模块需要缓存报文, 而现在设备的通用方法是通过队列的方式来缓存未被调度的报文. 由于设备缓存不同报文的 Eligibility time 可能跨度比较大, 如果门控粒度很细, 就要求队列的个数足够多, 才能够覆盖不同报文的 Eligibility time 的时间跨度; 目前的设备中, 队列资源是有限的, 通常为

几十个;

3) 由于存在着“多打一”的情况,所以在同一时刻,可能会有很多报文需要释放,需要非常高的芯片速率,需要超出现有设备能力范围的芯片加速比。

## 6 结束语

传统 IP 聚合调度下,多流的突发碰撞导致流的波形变化,波峰叠加,形成排队时延和抖动;而流间突发相互影响,恶性循环,开放增长,导致时延和抖动随跳数超线性增长。笔者提出一种基于网络演算的 Damper 异步转发机制,可以通过逐跳整形的方式消除流碰撞引起的时延不确定性,为网络提供端到端的确定性能力。相比传统的 DiffServ 方式, Damper 机制对于确定性流量的接入规格可以提升  $H-1$  倍。

## 参考文献:

- [1] 3GPP. 3GPP Release 15: TR 21. 915—2019 [S/OL]. 2019[2020-08-12]. <https://www.3gpp.org/release-15>.
- [2] Popovski P, Trillingsgaard K F, Simeone O, et al. 5G wireless network slicing for eMBB, URLLC, and mMTC: a communication-theoretic view [J]. IEEE Access, 2018: 55765-55779.
- [3] Fettweis G, Boche H, Wiegand T, et al. The tactile internet-ITU-T technology watch report [R]. Geneva: ITU, 2014.
- [4] 3GPP. 3GPP technical specification group services and system aspects: service requirements for the 5G system: 2017 [S/OL]. [2020-08-10]. [https://www.3gpp.org/ftp/Specs/archive/22\\_series/22.261/](https://www.3gpp.org/ftp/Specs/archive/22_series/22.261/).
- [5] 左旭彤,王莫为,崔勇. 低时延网络: 架构,关键场景与研究展望[J]. 通信学报, 2019, 40(8): 22-35.  
Zuo Xutong, Wang Mowei, Cui Yong. Low-latency networking: architecture, key scenarios and research prospect[J]. Journal on Communications, 2019, 40(8): 22-35.
- [6] Verma D. Guaranteeing delay jitter bounds in packet switching networks [C] // Proceedings of IEEE TriCom'91. Chapel Hill: [s. n.], 1991.
- [7] Braden R, Clark D, Shenker S. RFC1633: integrated services in the Internet architecture: an overview [R/OL]. (1994-09-16) [2020-08-14]. <http://www.doc88.com/p-1364841430117.html>.
- [8] 王重钢,隆克平,龚向阳,等. 分组交换网络中队列

调度算法的研究及其展望[J]. 电子学报, 2001, 29(4): 553-559.

Wang Chonggang, Long Keping, Gong Xiangyang, et al. The study and perspective of queue scheduling algorithms in packet switching networks [J]. Chinese Journal of Electronics, 2001, 29(4): 553-559.

- [9] Le Boudec J Y, Thiran P. Network calculus: a theory of deterministic queueing systems for the Internet[M]. [S. l.]: Springer Science and Business Media, 2001.
- [10] Blake S, Black D, Carlson M, et al. RFC 2475: an architecture for differentiated services [R/OL]. (1998-09-16) [2020-08-10]. <http://www.doc88.com/p-6833728128856.html>.
- [11] Specht J, Samii S. Urgency-based scheduler for time-sensitive switched ethernet networks [C] // 2016 28th Euromicro Conference on Real-Time Systems (ECRTS). Toulouse: IEEE, 2016: 75-85.
- [12] Finn N, Thubert P, Varga B, et al. Deterministic networking architecture, draft-ietf-detnet-architecture-02 [J]. IETF Internet-Draft, 2017: 1-43.
- [13] Li Z, Wan H, Zhao B, et al. Dynamically optimizing end-to-end latency for time-triggered networks [C] // Proceedings of the ACM SIGCOMM 2019 Workshop on Networking for Emerging Applications and Technologies. Beijing: ACM, 2019: 36-42.
- [14] Zhang J, Chen L, Wang T, et al. Analysis of TSN for industrial automation based on network calculus [C] // 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). Zaragoza: IEEE, 2019: 240-247.
- [15] Grigorjew A, Metzger F, Hoßfeld T, et al. Bounded latency with bridge-local stream reservation and strict priority queueing [C] // 2020 11th International Conference on Network of the Future (NoF). [S. l.]: IEEE, 2020: 55-63.
- [16] Grigorjew A, Metzger F, Hoßfeld T, et al. Technical report on bridge-local guaranteed latency with strict priority scheduling [R]. Berlin: [s. n.], 2020: 1-10.
- [17] Charny A, Le Boudec J Y. Delay bounds in a network with aggregate scheduling [C] // International Workshop on Quality of Future Internet Services. Berlin: Springer, 2000: 1-13.
- [18] Zhang Z L, Duan Z, Hou Y T. Fundamental trade-offs in aggregate packet scheduling [C] // Proceedings Ninth International Conference on Network Protocols. [S. l.]: IEEE, 2001: 129-137.