



在线 SDN 控制器性能监控器设计与研究

黄序富¹, 刘川², 林亦雷³, 金耀辉^{1,4}, 罗萱⁴

- (1. 上海交通大学区域光纤通信网与新型光通信系统国家重点实验室 上海 200240;
2. 中国电力科学研究院 北京 100192; 3. 国网上海市电力公司信息通信公司 上海 200120;
4. 上海交通大学网络信息中心 上海 200240)

摘要:随着网络规模和网络数据流量的变大,软件定义网络控制器性能问题可能成为网络瓶颈,为了在动态负载情况下仍然可以获取控制器准确的性能状态,设计了一个在线的性能监控器框架,并定义了控制器时延性能衡量指标 Lpdex。在一款广泛使用的 FloodLight 控制器中实现了性能监控器的简单原型,经过实验评估,验证了控制器性能监控器的可行性和 Lpdex 可在线准确地反映控制器在不同负载情况下时延性能的变化。

关键词:软件定义网络;性能监控器;性能指标

doi: 10.11959/j.issn.1000-0801.2015090

Design and Study on Online Software Defined Networking Controller Performance Monitor

Huang Xufu¹, Liu Chuan², Lin Yilei³, Jin Yaohui^{1,4}, Luo Xuan⁴

- (1. State Key Lab of Advanced Optical Communication Systems and Networks, Shanghai Jiaotong University, Shanghai 200240, China; 2. China Electric Power Research Institute, Beijing 100192, China;
3. Shanghai State Grid Information & Telecommunication Company, Shanghai 200120, China;
4. Network & Information Center, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: With network scale and network traffic growing up, the performance of SDN controller can probably be bottleneck, in order to still obtain correct performance status on dynamical workload, an online performance monitor frame was designed and Lpdex, a controller latency performance metric, was defined. Then a simple prototype of performance monitor was implemented on FloodLight, a widely used controller, after experimental evaluation, controller performance monitor was validated feasible and Lpdex can indicate controller latency performance accurately online with workload varying.

Key words: software defined networking, performance monitor, performance metric

收稿日期:2014-09-08;修回日期:2015-02-15

基金项目:国家自然科学基金资助项目(No. 61371084),国家电网公司2014年科技项目“软件定义网络(SDN)体系及其电力应用关键技术研究”基金资助项目,上海市科委科技创新计划交大基金资助项目(No. 13511502302)

Foundation Items: The National Natural Science Foundation of China (No. 61371084), 2014 SGCC Technology Project “Research on Key Technologies of SDN System and Power Application”, Shanghai Committee of Science and Technology Innovation Program with SJTU (No. 13511502302)

论文引用格式:黄序富,刘川,林亦雷等. 在线 SDN 控制器性能监控器设计与研究. 电信科学, 2015090

Huang X F, Liu C, Lin Y L, et al. Design and study on online software defined networking controller performance monitor. Telecommunications Science, 2015090



1 引言

软件定义网络 (software defined networking, SDN) 是由美国 Stanford (斯坦福) 大学和加州大学 Berkeley (伯克利) 分校提出的 OpenFlow^[1] 发展起来的。SDN 作为新型网络架构, 通常认为具备 3 个特点^[2]: 第一, 控制平面和转发平面分离; 第二, 控制平面为集中式逻辑控制; 第三, 控制平面向上层提供应用程序接口 (application programming interface, API), 以实现网络业务自动化编排和部署。因为这些特性, SDN 可以极大地提升网络资源的利用率、简化网络管理、降低运营成本和促进网络应用创新, 已在工业界和学术界引起了广泛的研究。目前, SDN 技术的应用场景^[3] 也非常广泛, 包括企业网络、数据中心、光网络、无线网络以及电力通信网络等。经过大量研究和实践, 已对 SDN 技术和应用前景有较高评价。

同样因为 SDN 的网络架构, 使得 SDN 控制器处于核心位置, 控制器需要同时处理数据平面和上层应用的所有请求, 因此控制器性能很可能成为瓶颈。首先, 根据参考文献[3], 在网络规模或者网络数据流量变大时, 数据平面建立流表请求数变大, 控制器性能很可能会成为网络瓶颈; 此外, 根据参考文献[4], 控制器响应时延会给数据流增加额外的时延, 而有很多场景如在数据中心、电力通信网络要求低时延; 再者, 控制器上层的网络应用运行是否正常, 依赖于控制器性能状态是否正常。所以, 控制器性能为非常重要的问题, 这对于优化 SDN 也具有重要作用。目前, 一般认为 SDN 控制器最主要的功能是控制数据平面的转发, 主要用以下 3 个指标衡量控制器性能^[5]: 响应时延, 即控制器响应一条数据建立请求 (packet-in) 所需时间; 吞吐量, 控制器单位时间内响应 packet-in 数目; 系统开销, 即控制器的 CPU 和内存开销。

然而若要更加系统地研究控制器性能问题, 需要应对多个挑战。首先, 控制器性能受多个因素影响, 如运行设备、控制器转发算法、控制器线程数目和数据平面拓扑复杂度; 再者, 如何定义和选择合理的控制器性能指标, 如通常使用平均响应时延, 但仅仅用一个平均值过于简单, 它会抹去响应时延分布的细节情况, 可能出现这种情况, 大多数的响应时延很低, 而只有少部分响应时延特别高, 得出平均值也较高, 认为控制器时延性能差, 而这是不恰当的结论, 因为此时大多数响应时延性能还是很低; 最后, 通常网络流量处于动态变化中, 引起的控制器负载也在一直动态变化, 未知的负载也会对控制器性能造成影响。

基于控制器性能在 SDN 中的重要性, 同时为了在动态负载情况下仍然可以获取控制器准确的性能状态, 本文设计了一个在线软件定义网络控制器性能监控器框架, 并定义了一个可准确衡量控制器性能的指标 Lpdex (latency performance index)。根据设计的框架, 本文在目前工业界和学术界都广泛使用的一款控制器 FloodLight^[6] 中做了控制器性能监控器的简化实现。利用实现的控制器监控器, 首先验证该监控器具有可行性, 然后研究分析新定义的控制器性能指标 Lpdex 的具体含义和参数的选择, 最后研究 Lpdex 是否可以在动态负载情况下准确地反映出控制器的性能状态。经过实际评估分析, 本文设计的控制器性能监控器和定义的性能指标 Lpdex 是可行的, 能够准确地表征控制器性能状态, 为控制器运营管理人员提供了一个直观可靠的控制器性能状态参考指标。最后, 简要讨论性能监控器的其他扩展应用, 比较目前已有的控制器性能研究工作, 分析了本文工作的创新性。

2 方案设计

为实现可广泛使用的在线软件定义网络控制器性能监控器, 设计方案需要解决以下一些难点。

其一, 可扩展性问题。目前单个控制器的性能总是有限的, 业界有很多解决方案是使用多控制器。针对这个问题, 本文设计方案采用的是分布式控制器性能监控模块。第二个难点为, 控制器性能监控器模块对控制器性能影响最小化。因为监控模块会对控制器添加一些额外的负载, 对控制器性能会有部分影响, 所以监控器模块架构需要尽可能地最小化监控模块影响。针对此问题, 本文监控器设计框架为在控制器端仅添加分布式的原始性能信息采集客户端, 而后所有原始性能信息汇聚到一个集中式的信息管理中心。在管理中心内, 对原始性能数据进行存储、计算处理, 提取出本文定义的性能监控指标 Lpdex 以及监控数据, 也可以用于其他的一些应用 (第 4 节)。本文设计的在线 SDN 控制器性能监控器的框架如图 1 所示。

其中, 本文控制器内设计的采集器采集的时延是控制器端的时延, 如图 2 所示, 虽然对整个 packet-in 时延而言, 这仅是其中的一部分, 但本文仍仅测量这部分时延, 主要基于以下几点原因。首先, 性能监控器最关注的是控制器, 监控其内部时延更为重要, 而且根据分析时延具体组成, 控制器与交换机之间的传输时延一般为实际变化较小的部分, 传输时延一般也较小, 而控制器处理时延根据控制

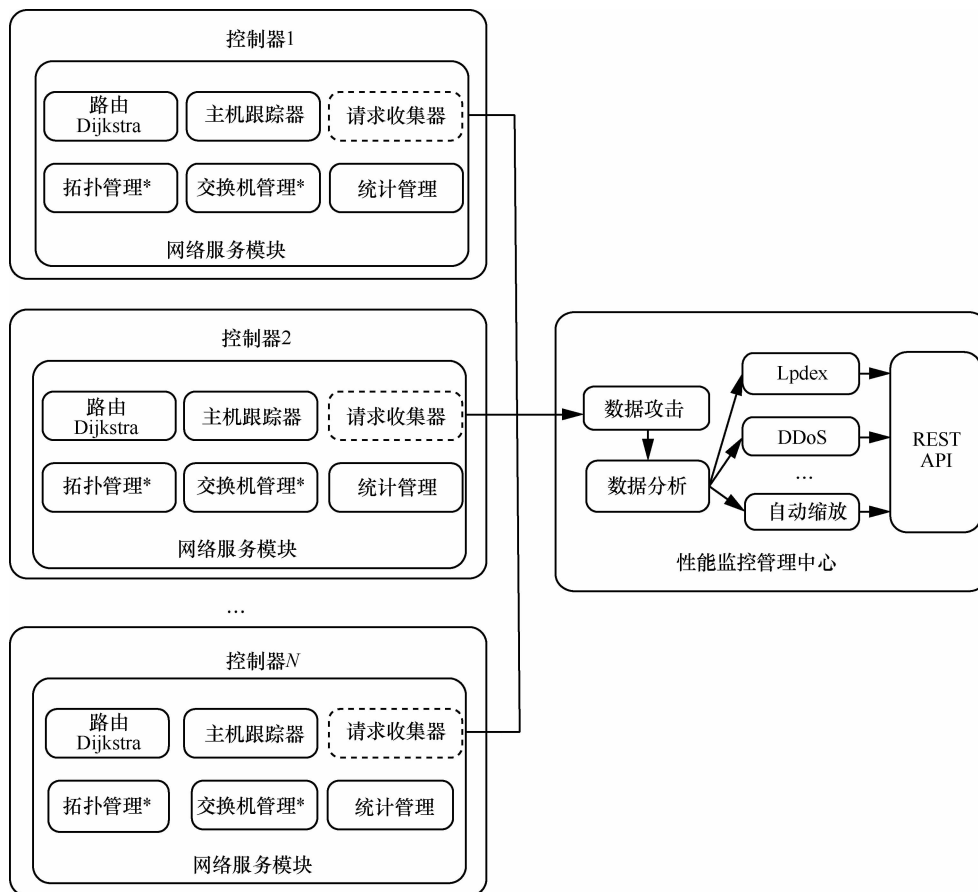


图1 在线SDN控制器性能监控器框架

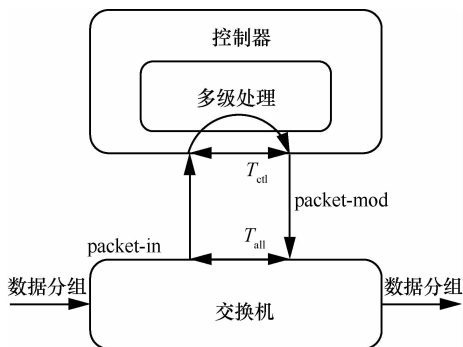


图2 控制器端监控时延和交换机端监控时延

器性能变化会较大,这部分时延也占整个时延的大部分,这部分将在实验部分具体验证。

为了解决平均响应时延可能不能准确表征控制器性能状态的问题,类似于Apdex^[7]定义应用性能指标用于解决平均时延可能不能全面反映应用的时延性能,定义了一个Lpdex指标来衡量控制器时延性能。Lpdex的原理是在一定的时间内,基于统计的规律,把packet-in响应请求的响应时延划分为3类:第1类,若是响应时延小于或等于

自定义的目标响应时延(target latency),则为满意的响应时延,可得1分;第2类,若是响应时延大于目标响应时延,但小于或等于4倍的目标响应时延,则为可容忍的响应时延,可得0.5分;第3类,若是响应时延大于4倍的目标响应时延,则归为不能容忍的响应时延,得0分。Lpdex即在这段时间内得分总和除以这段时间内packet-in请求数目的总和(本文中的一段时间内,选择的是秒级的单位时间粒度,若是粒度过细可以进行数据汇聚,得出更加粗粒度的Lpdex),如式(1)所示:

$$Lpdex = (N_{sat} + 0.5 \times N_{tol}) / N_{all} \quad (1)$$

其中, N_{sat} 为一段时间内满意的响应时延数目, N_{tol} 为一段时间内可容忍的响应时延数目, N_{all} 为该段时间内packet-in请求数目的总和。由式(1)可知,Lpdex的取值范围是0~1,其值越大,表示满意的响应时延数目越多,交换机满意度越高,控制器性能状态越佳,反之则性能状态越差。

3 实现与评估

根据上述控制器性能监控器的设计框架,在目前业内



使用比较广泛的一款开源控制器 FloodLight 中做了一个简化的性能监控器原型实现。因为 FloodLight 控制器是功能模块化的控制器,本文的实现方法即在内部添加了一个监控模块。这个监控模块的主要原理为:在 packet-in 请求进入时添加时间戳,完成处理后再记录时间戳,这样即可知道在控制器内处理一条 packet-in 请求的时延,即完成了控制器时延性能采集的功能;然后监控模块分别统计每秒钟满意响应时延的数目、可容忍响应时延的数目、不可容忍响应时延的数目和请求总数等参数。根据这些数据即可计算出定义的性能参数 Lpdex 和其他的一些性能参数。具体修改的 FloodLight 控制器代码已开源在github^[9]中。

根据 FloodLight 实现的性能监控器,搭建了一个实验平台进行控制器监控器合理性验证、Lpdex 的具体含义与目标响应时延的设定和动态负载情况下的 Lpdex 3 个实验。其中 FloodLight 控制器运行在双核 CPU 和 4 GB 内存的虚拟机内,虚拟机操作系统为 Ubuntu14.04,Java 版本为 1.7。具体实验配置和结果如下文所述。

3.1 控制器监控器合理性验证

关于在交换机端测量 packet-in 的响应时延,目前已有 Cbench^[9]工具可以很方便地测试出。Cbench 模拟的测试环境如图 3 所示,其工作原理为模拟多个 OpenFlow 交换机向控制器发送 packet-in 请求,然后在交换机上统计请求所需时延和单位时间内收到的响应数目。

使用 Cbench 可以方便地测量出交换机侧的响应时延,使用本文的性能监控器可测试出在控制器侧的响应时延。最终两者的响应时延和比例结果如图 4 所示。

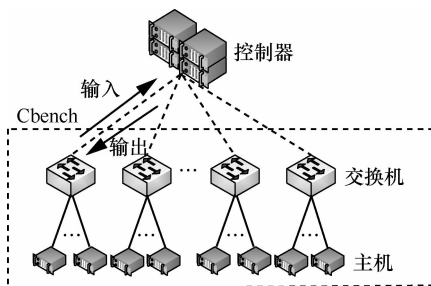


图 3 Cbench 模拟环境

由图 4 可知,两种响应时延都是在启动过程中逐渐降低的,而且在控制器侧测量的响应时延比较接近从交换机侧测量的响应时延,经过多次测量,在控制器端测量的平均响应时延占整个时延的 60%以上(图 4 中的具体数值为 65.17%),由此验证了在控制器端测量响应时延是可行的。

3.2 Lpdex 的具体含义与目标响应时延的设定

在验证了控制器端测量响应时延具有实际意义后,还需要分析自定义的控制器性能参数 Lpdex 的具体含义以及其参数目标响应时延的设定。类似验证控制器性能监控的有效性,使用 Cbench 发送 packet-in 请求,然后在控制器中统计 10 s 内经过控制器的响应时延,观测其 CDF,根据设定的不同目标响应时延,计算 Lpdex。具体的响应时延的 CDF 和不同目标响应时延 Lpdex 值的设定如图 5 所示。

根据响应时延的 CDF 可知,Lpdex 总是根据目标参数把 CDF 划分为 3 部分,各个部分分别被赋予不同的权重即可得出值,所以 Lpdex 实际上是简化的加权 CDF。Lpdex 相比于直接的平均值,仍然部分保留统计分布的规律,因此比仅使用平均值更能反映出总体规律,可更加准确地反映控制器性能的状态。另外,目标参数的设定可根据具体的要

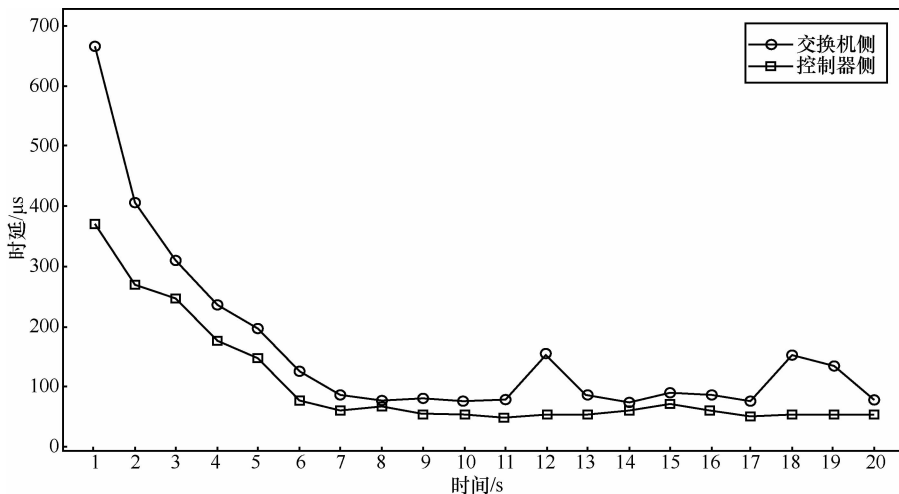


图 4 交换机侧响应时延和控制器侧响应时延

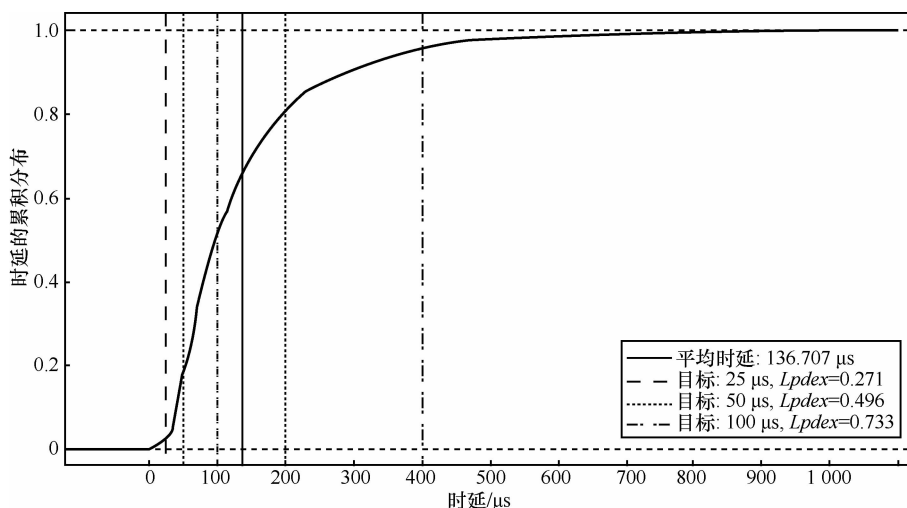


图 5 响应时延的 CDF 图和不同目标参数的 Lpdex

求设定,如有严格的响应时延要求,则可以设定较小的值,反之可设定更大的值。即使设定好初始值后,具体的值也可以根据实际情况在控制器性能监控器中做反馈调节。其中一种参考调节方式为对数据流体验的时延采样调整目标参数,例如采样的数据流响应时延为满意范围,而 Lpdex 却较低,则可以适当加大目标参数,具体的定量关系不在本文的研究范围。至于初始值,可以根据实际需求设定一个大概的值,例如,若在某些场景下,流可容忍的额外时延很低,则定义目标响应时延为较小的值,本文初步设定目标响应时延参数为 $25\ \mu\text{s}$,后文也使用这个目标响应参数。

3.3 动态负载情况下的 Lpdex

在线监控器总是工作于动态负载情况下,必须分析其能否准确反映出不同负载情况下的控制器性能状态。由于 Cbench 不能够定量控制发送 packet-in 请求的速率,本文将控制主机定量地发送大量 ARP 请求,从而在交换机侧触发一定量的 packet-in 请求,实现不同负载的情况。为了模拟 OpenFlow 网络,Mininet^[10]是一个非常不错的选择,本文模拟的简单树型结构数据平面如图 6 所示。

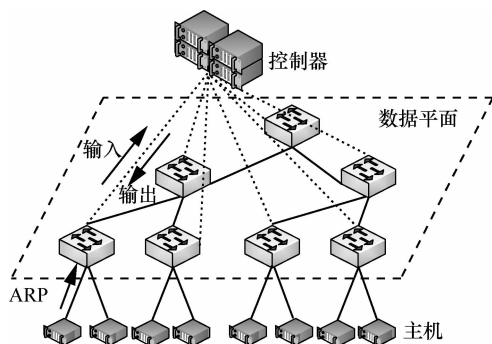


图 6 Mininet 模拟 OpenFlow 网络

在模拟完成 OpenFlow 网络之后,在主机端每秒定量地产生 ARP 请求,根据产生的量的不同对控制器施加不同的负载,图 7 为对控制器动态地施加负载压力时,性能监控器 Lpdex 的时序变化情况。图 8 为控制器在不同 ARP 负载下的 Lpdex 箱线图统计以及处理 packet-in 请求吞吐量的箱线图统计。

从图 7 可以明显看出,在一定负载范围内,Lpdex 虽然会有些波动变化,但都保持在较高的水平,而当负载达到某一个临界值时,控制器先尽力处理 packet-in 请求,若请求负载一直很高,则会出现大量的响应时延变大,Lpdex 将会急剧降低,也反映出控制器时延性能快速下降。若这时除去负载,根据 Lpdex 的请求可以看出其控制器时延性能也逐渐恢复到较高的水平。根据图 8 可知,Lpdex 和吞吐量具有相似的规律,吞吐量验证了 Lpdex 随负载变化的规律。由此,本文的在线控制器性能监控器及 Lpdex 可以在动态负载情况下,准确地表征控制器的时延性能状态变化,Lpdex 为性能监控的可参考指标。

4 相关研究工作

本文设计的在线 SDN 控制器性能监控器除了可使用 Lpdex 准确地反映当前控制器的时延性能外,还可有其他的一些应用。例如,packet-in 限流反馈功能,在监控器中统计发现某台交换机或某台主机发送 packet-in 请求的速率过快,可对该端口或主机进行 packet-in 限流;packet-in 的 DDos 攻击或其他异常检测功能,若统计发现在一段较为平稳正常的状态,突然有某些主机持续不断地发送大量的 packet-in 请求,则可以对这些请求进行分析,监测其是否

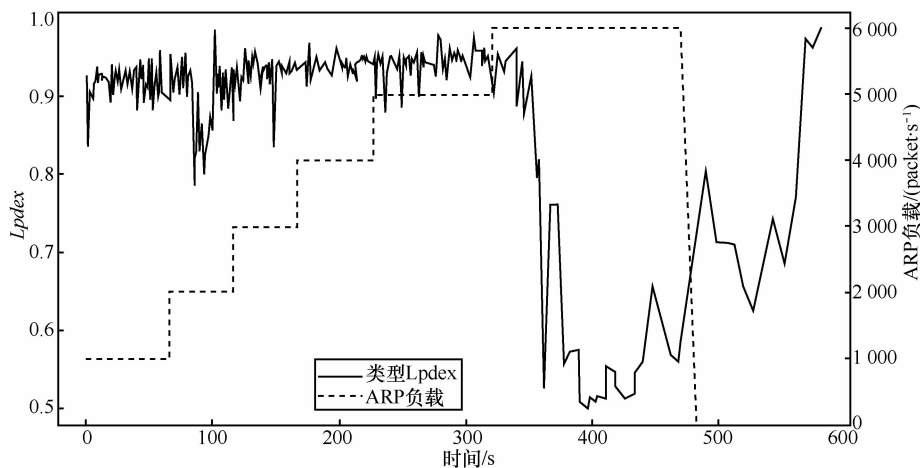


图7 不同 ARP 负载 Lpdex 变化时序

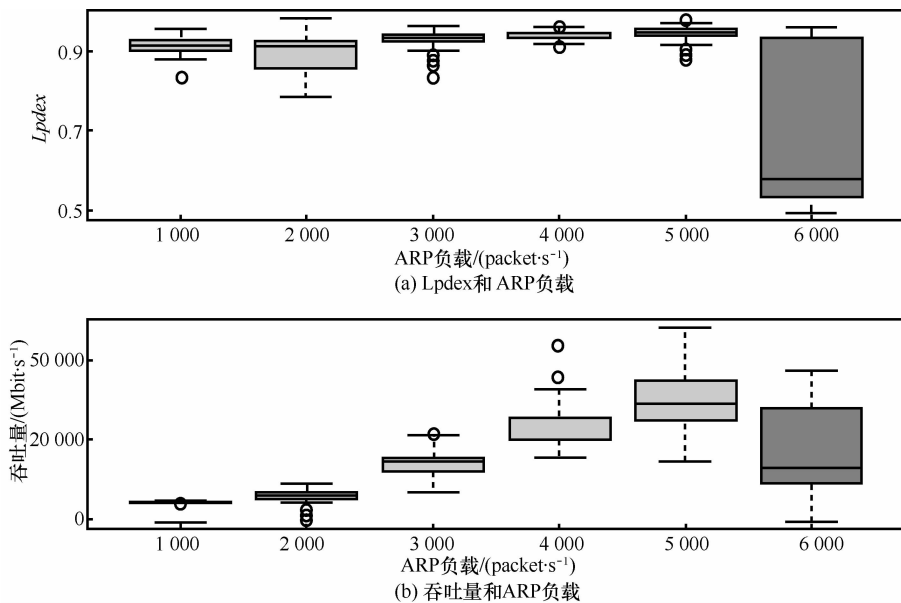


图8 不同 ARP 负载的 Lpdex 和吞吐量

为 DDoS 攻击源或是否有其他异常情况;控制器线程或实例自动扩展功能,在正常的一些请求情况下,工作负载压力较大时,Lpdex 和其他的一些性能指标变差,可以考虑自动扩展控制器的线程或者添加控制器实例,或考虑能否增加主机的 CPU 和内存资源等,这样可使得控制器保持较好的时延性能。

本文设计的在线 SDN 控制器性能监控器和控制器性能监控指标 Lpdex 都是针对通用的 SDN 控制器框架,在 SDN 中具有广泛的应用场景。例如,该设计系统可运用在 SDN 架构的电力通信网络中。电力通信网络作为智能电网的基础支撑平台,对于整个智能电网正常安全运行具有重要作用。运用 SDN 架构的电力通信网络可以解决以传统 SDH 技术为基础的电力通信网络无法满足其“流量可观可

控,资源动态分配”的需求。在运用 SDN 架构的电力通信网络中,SDN 控制器性能状态对于电力网络正常安全运行至关重要,使用本文设计的在线 SDN 控制器性能监控器和控制器监控指标 Lpdex,即可准确地反映控制器在动态负载情况下的性能状态。

基于控制器性能的重要性,目前学术界也有关于控制器性能方面的相关研究工作。目前关于控制器性能的主要研究工作有参考文献[5,10~13]。参考文献[5]中,主要就控制器的吞吐量和响应时延性能做了初步的研究,对 NOX 控制器提出改进,提出其多线程版本,并设计了控制器性能的测量工具 Cbench;参考文献[11]中建立 OpenFlow 排队论模型,然后分析数据分组在系统中的逗留时间和分组丢失概率;参考文献[12]主要研究控制器的设计架构对控制

器性能的影响,从而可以在控制器系统架构设计上提供指导;参考文献[13]主要使用网络微积分从理论上研究了一个分级部署控制器的性能,从数学模型中获得根控制器的缓存大小和上界值;参考文献[14]中,为测量更加细粒度(到每个交换机级别)的控制器响应时延性能,设计了类似于Cbench的测量工具OPCBenchmark并验证了其有效性。本文与这些工作的主要不同点有两点:其一,本文为研究控制器性能,从在线监控的角度出发,可以实时地获取动态负载下的控制器性能状态,而非从理论上或者离线分析控制器的性能;其二,相关的研究工作对控制器性能的研究指标较多的都是控制器的吞吐量和平均响应时延,而本文定义了一个性能衡量参数Lpdex,可更加准确地反映控制器性能状态。

5 结束语

本文基于软件定义网络控制器性能,设计了一个在线的SDN控制器性能监控器,并定义了一个比平均响应时延更能反映群体响应时延性能的指标Lpdex。为了克服性能监控器的扩展问题和对原控制器最小化性能的影响,设计的监控器采用分布式采集模块和集中式性能信息处理模块,用以收集控制器性能信息。根据这个框架,本文在FloodLight控制器中实现了一个简化的性能监控器,并在监控器中定义了Lpdex用于在线衡量控制器时延性能。最后,在实现的监控器中,经过多个实验评估分析,验证了设计的控制器性能监控器和定义的Lpdex具有实际意义,针对不同的负载压力,可在线准确地衡量控制器的响应时延性能。

参考文献

- McKeown N, Anderson T, Balakrishnan H, *et al.* OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 2008, 38(2): 69~74
- Mendonca M, Nunes B A A, Nguyen X N, *et al.* A survey of software-defined networking: past, present, and future of programmable networks. *IEEE Communications Society, Institute of Electrical and Electronics Engineers (IEEE)*, 2014, 16(3): 1617~1634
- Kandula S, Sengupta S, Greenberg A, *et al.* The nature of data center traffic: measurements & analysis. *Proceedings of Internet Measurement Conference*, Chicago, Illinois, 2009
- Benson T, Akella A, Maltz D A. Network traffic characteristics of data centers in the wild. *Proceedings of Internet Measurement Conference*, Melbourne, Australia, 2010
- Tootoonchian A, Gorbunov S, Ganjali Y, *et al.* On controller performance in software-defined networks. *Proceedings of USENIX Workshop on Hot Topics in Management of Internet, Cloud,*

- and Enterprise Networks and Services (Hot-ICE)*, San Jose, CA, 2012
- FloodLight. FloodLight controller. <http://floodlight.openflowhub.org/>
- Apdex. <http://en.wikipedia.org/wiki/Apdex>
- Github code. <https://github.com/rainmetersjtu/floodlight>
- Cbench. <http://archive.openflow.org/wk/index.php/Oflows>
- Handigol N, Heller B, Jeyakumar V, *et al.* Reproducible network experiments using container-based emulation. *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, Guilin, China, 2012
- Jarschel M, Oechsner S, Schlosser D, *et al.* Modeling and performance evaluation of an OpenFlow architecture. *Proceedings of the 23rd International Teletraffic Congress International Teletraffic Congress*, San Francisco, USA, 2011
- Shah S A, Faiz J, Farooq M, *et al.* An architectural evaluation of SDN controllers. *Proceedings of 2013 IEEE International Conference on Communications (ICC)*, London, UK, 2013
- Azodolmolky S, Wieder P, Yahyapour R. Performance evaluation of a scalable software-defined networking deployment. *Proceedings of 2013 Second European Workshop on Software Defined Networks (EWSDN)*, Singapore, 2013
- Jarschel M, Lehrieder F, Magyari Z, *et al.* A flexible OpenFlow-controller benchmark. *Proceedings of 2012 European Workshop on Software Defined Networking (EWSDN)*, San Diego, CA, USA, 2012

[作者简介]



黄序富,男,上海交通大学硕士生,主要研究方向为云计算和软件定义网络。

刘川,男,现就职于中国电力科学研究院,主要研究方向为软件定义网络。

林亦雷,男,现就职于国网上海市电力公司信息通信公司,主要研究方向为软件定义网络。

金耀辉,男,上海交通大学教授、网络信息中心副主任,主要研究方向为云计算架构、软件定义网络、数据中心及宽带网络和海量流式数据分析。

罗萱,男,博士,现就职于上海交通大学网络信息中心,负责学校数据中心网络的规划和建设工作,主要研究方向为云计算、软件定义网络、网络虚拟化和数据挖掘。