北京交通大學

文献阅读报告——应用层

班级：通信 1903

学号：19211152

姓名：韦国轩

2022 年 05 月 23 日

电子信息工程学院

目录

一、文献信息

作者：Qizhen Zhang, Kelvin K.W. Ng, Charles W. Kazer, Shen Yan, João Sedoc, and Vincent Liu
论文题目：MimicNet: Fast Performance Estimates for Data Center Networks with Machine Learning
发表途径：SIGCOM 2021 Virtual Event, USA
发表时间：August 23-27, 2021

二、问题意义

多年来，人们提出了许多新的协议和系统来提高数据中心网络的性能。尽管这些提案在方法上具有创新性，结果也很充满前景，但它们面临着一个持续的挑战：难以对系统进行大规模评估。高度互连且充满依赖性的网络在这方面尤其具有挑战性，因为网络某个部分的微小变化可能会对其他部分产生较大的性能影响。对相关实验平台而言也确实如此，很少有人能够负担得起数据中心的专用、全尺寸副本。模拟也有着相似的处境，虽然最初设计正是为了评估相关的网络系统，但它很难适应当今网络的规模。

这篇论文提出了一种快速获得大型数据中心网络准确性能估计的方法。作者团队提出了一个系统——MimicNet，将网络的某一部分抽象为用户熟悉的 Packet-level 模拟——指的是主要关注数据包（packet）的特征及其到达过程，如数据包大小分布、数据包到达时间间隔的分布等——同时利用冗余和机器学习的最新进展，快速、准确地近似网络中无法直接看到的部分。与数据中心的常规模拟——通常具有数千台服务器相比，MimicNet 可以提供超过两个数量级的提速，即使在这种规模下，MimicNet 对尾部 FCT、吞吐量和 RTT 的估计也在真实结果的 5%以内。

即，本文针对大型数据中心网络的性能估计提出了一种快速，同时拥有着较高准确性的网络模拟系统，为相关领域的研究提供了低成本、高便捷的新方案。

三、思路方法

本文共划分为十一个部分， 在第一部分的简介中，向读者们介绍了数据中心网络领域的的发展现状，以及难以对系统进行大规模评估的现状，由此作者团队提出了一种用于快速评估大规模数据中心网络性能的工具——MimicNet。

在第二部分，作者团队从网络模拟工具的背景、当下模拟器的可扩展性两方面讲述了他们开发该工具的初衷。

该论文的之后正文部分结构如下：

第四部分标题为“OVERVIEW”，是 MimicNet 整体特性的总览，每个 MimicNet 模拟都包含一个“可观察”集群，以完全逼真的方式进行了模拟；未观察集群的各种行为都由经过训练的模型近似表示。并且介绍了 MimicNet 应用时的一些限制条件。

第五部分“INTERNAL MODELS”主要讲述了 MimicNet 的内部模型，即首先使用一个小尺度模型来收集训练/测试数据，并介绍了这一方案的建模目标、可扩展功能选择、数据通信网友好损失函数以及通用模型选择等相关内容。

虽然内部模型可以对集群的队列、路由器和内部流量的行为进行建模，但仍然需要对外部流量进行完整跟踪才能生成准确的结果，因此在第六部分“FEEDER MODELS”中，介绍了 MimicNet 通过引入一个馈线模型，估计 MimicNet 内部的流量到达率，并将其注入内部模型。

在第七部分“TUNING AND FINAL SIMULATION”中，作者团队向读者介绍了如何对 MimicNet 进行调整以及最终的模拟，包括如何构造一个 MimicNet，以及允许用户定义自己的优化功能。

在第八部分“PROTOTYPE IMPLEMENTATION”中，作者团队用 C++和 Python 实现了完整的 MimicNet 工作流原型，介绍了其仿真框架、并行执行能力以及所应用的机器学期框架。

第九部分“EVALUATION”包括了作者团队对 MimicNet 几个重要特性的评估，将在本文献报告的“四、实验结论”中详细说明。

在最后的第十、第十一部分中，作者团队对全文以及自己的研究成果进行了总结，介绍了 MimicNet 优势的同时，也表示在使过程更简单、更准确方面仍有许多工作要做，且这一设计为使用机器学习和问题分解来逼近大型网络提供了概念证明。

四、实验结论

作者团队的评估侧重于 MimicNet 的几个重要特性，包括：

（1）近似数据中心网络性能的准确性；

（2）其精度对大型网络的可扩展性；

（3）近似模拟的速度；

（4）其用于比较配置的实用性。

作者团队的模拟都采用了 FatTree 拓扑，并将链路速度配置为 100 Mbps，延迟为 500μs， 为了向上、向下扩展数据中心，还调整了每个集群中机架/交换机的数量以及数据中心中集群的数量。

由于传统的预测指标在此环境中参考价值不大，因此，作者团队利用了三个端到端指标：

（1）流量完成时间 FCT；

（2）以 100 毫秒为间隔的均服务器吞吐量；

（3）往返时延 RTT。

在下一节中检验更大规模的配置之前，首先评估 MimicNet 在用模仿者替换单个集群时的准确性。图 1 显示了本测试的三个指标的 CDF。如图所示，MimicNet 在所有指标上都达到了非常高的准确性。

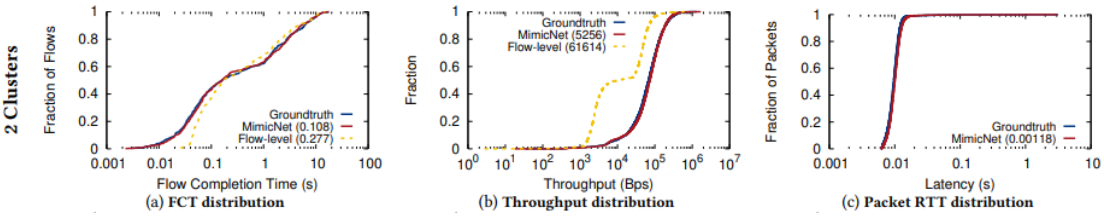


图 1—MimicNet 在 2 集群基线配置中的准确性

一个关键问题是，在流量交互变得更加复杂且添加了馈线的情况下，准确性是否会转化为更大的组成。作者团队使用由 128 个集群组成的模拟来回答这个问题（对于较大的集群，无法进行全保真度模拟）。在 MimicNet 中，127 个集群被替换为与前一小节相同的模仿者，图 2 显示了结果精度。

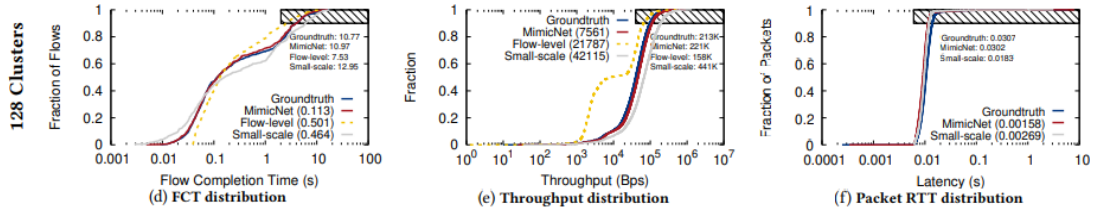


图 2—MimicNet 在 128 集群基线配置中的准确性

虽然 MimicNet 估计的准确性下降了，但这种下降是正常的。，MimicNet 的 FCT、吞吐量和 RTT 的 99%概率分布值分别在真实结果的 1.8%、3.3%和 2%以内。

与精确度同样重要的是，MimicNet 可以非常快速地评估性能。MimicNet 小规模仿真、模型训练、超参数调整和大规模合成的多个阶段都需要时间，但总而言之，仍然比直接运行全保真仿真更快。

图 3 显示了全尺寸仿真和 MimicNet 的运行时间明细，图 1 中 128 个集群、1024 个主机模拟分为三个阶段。对于 20 秒的仿真时间，全保真度模拟器几乎需要一星期加上五天；相比之下，MimicNet 总共只需要 8 小时 38 分钟，加速比为 34 倍。

	Factor	Time
MimicNet	Small-scale simulation	1h 3m
	Training and hyper-param tuning	7h 10m
	Large-scale simulation	**25m**
Full	Simulation	**1w 4d 22h 25m**

图 3—128 集群、1024 主机数据中心的 20s 仿真用时

图 4 显示了不同网络大小下的吞吐量结果。总的来说，不管网络大小如何，MimicNet 都能保持高吞吐量。另一方面，随着网络规模的增长，单次模拟的速度大大降低，在 128 集群中，完全模拟几乎比实时模拟慢五个数量级。较大的并行化实例开始遭受上述内存问题的困扰，但即使内存不受限制，MimicNet 在 128 个集群上的性能仍可能比并行化模拟高出 2-3 个数量级。

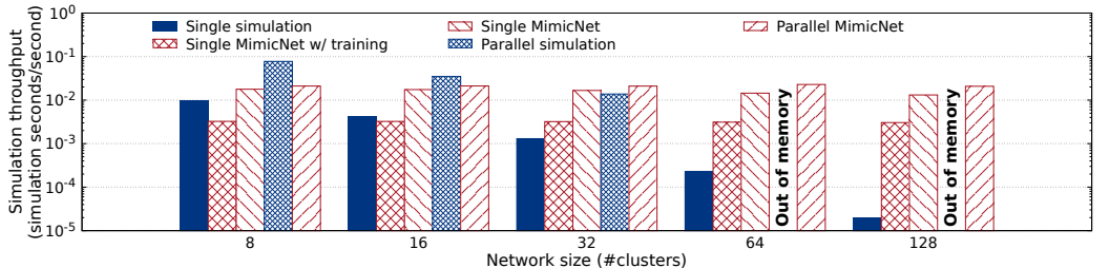


图 4—不同网络大小下的吞吐量

最后，MimicNet 可以近似各种协议，并为每种协议提供可操作的配置。在下文中将介绍两个潜在的用例：（1）DCTCP 配置的优化方法；（2）几种数据中心网络协议的性能比较。

DCTCP 利用来自网络的 ECN 反馈来调整拥塞窗口，一个重要配置参数是 ECN 标记阈值 K, 这会影响协议的延迟和吞吐量。本质上，一个较低的 K 更积极地发出拥塞信号，确保更低的延迟；但是 K 过低可能会导致网络带宽利用不足，从而限制吞吐量。FCT 同时受到这两个方面的影响：短流量受益于较低的延迟，而长流量有利于较高的吞吐量。

图 5 比较了不同的 K 值，若是只看 2 集群模拟，读者可能会认为工作负载的最佳设置是 K= 60；但是对于更大的 32 集群模拟，会发现 K= 60 几乎是最差配置之一，K = 20 成了最佳选择，MimicNet 成功得出了正确的结论。

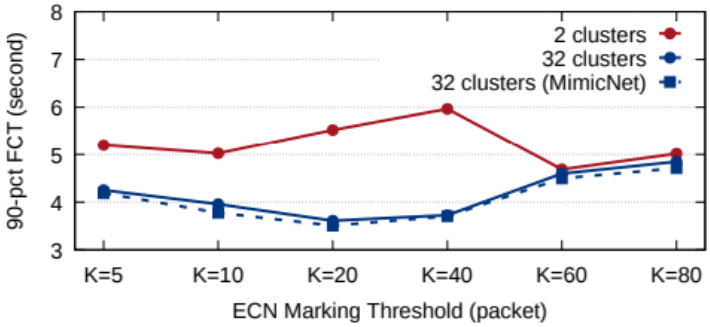


图 5—不同 K 值下的流量完成时间(FCT)

MimicNet 足够精确，可用于比较不同的传输协议。作者团队实现了四个协议，每个协议都以不同的方式体现出 MimicNet 的建模。

Homa 是一种利用优先级队列的低延迟数据中心网络协议，对于 MimicNet 来说，这是一项具有挑战性的额外功能，因为数据包可以重新排序。

TCP Vegas 是一种基于延迟的传输协议，它代表了对延迟的微小变化非常敏感的协议

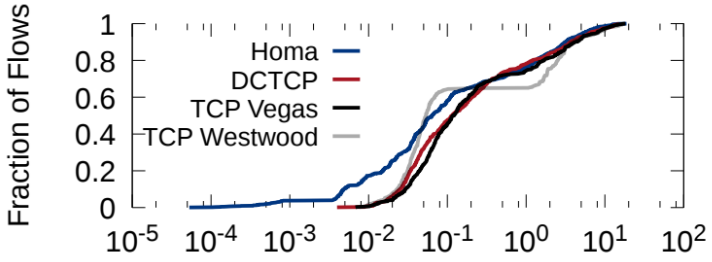
的最新趋势。

TCP Westwood 是一种发送方优化的 TCP，它测量端到端连接速率，以最大限度地提高吞吐量并避免拥塞。

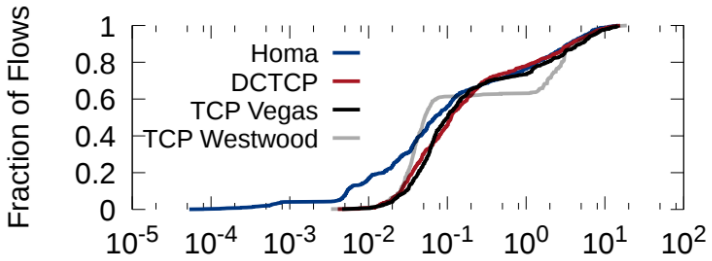
DCTCP(K=20)，使用 ECN 位，与其他协议相比，它增加了额外的功能和预测输出。

作者团队为每个协议运行完整的 MimicNet 模拟，训练单独的模型，随后对它们在相同相同工作负载下的性能进行了比较，并评估了 MimicNet 的准确性和速度。

图 5 为各协议的 FCT 结果，与基本配置一样，对于所有协议，MimicNet 都可以与全保真度模拟的 FCT 紧密匹配。事实上，平均而言，MimicNet 估计的 90%和 99%的尾部与真实值相差不到 5%。由于这种准确性，MimicNet 性能评估可以用来衡量这些协议的粗略相对性能。



(a) **Ground truth of the comparison**



(b) **The approximation of MimicNet.**

图 5—MimicNet 在各协议下的流量完成时间(FCT)与真实值的对比

五、启发思考

这是一篇发表于 SIGCOMM 2021 的论文，SIGCOMM 作为国际通信研究领域的顶级会议，能够通过其组委会审核，那么久表示这一文章得到了通信领域前沿研究者们的认可，而我作为一名通信工程专业的本科生，经过这一次的论文研读，对于文章的内涵只能说有了粗略的理解，但这依旧让我领略到了通信领域最前沿、最先进的科学思想，本文的作者十分巧妙地运用了以小见大的思想，利用了机器学习，来对小规模网络进行模拟，通过对训练模型的应用实现了对大规模网络的抽象。提前接触一下这些与我在未来的学习，或是科研生涯息息相关的研究成果，如同为依旧处于茫茫学海起步阶段的我点亮了一盏指路明灯，在以后的成长、进步中也有了正面的榜样。