

文章编号: 2095-2163(2021)02-0210-06

中图分类号: TP391.41

文献标志码: A

软件定义企业级无线网络平台设计与实现

林世荣, 姜守旭

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘要: 本文在企业级无线网络中引入软件定义网络协议并进行相应的采样和控制设计, 将网络划分为数据平面和逻辑平面, 并在这两个平面进行了相应的功能模块设计与实现。在数据平面, 本文实现了支持上层算法的实时数据采样以及支持上层算法的控制动作。在逻辑平面, 本文实现了实时的数据接收、预处理以及应用程序编程接口, 为采用高级语言进行计算的优化程序提供了支持。

关键词: 软件定义企业无线网络; 数据采样; 网络优化

Design and implementation of software defined enterprise wireless network platform

LIN Shirong, JIANG Shouxu

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] This paper introduces software-defined network protocol into enterprise wireless network and carries out corresponding sampling and control design. The network is divided into data plane and logic plane, and corresponding functional module design and implementation are carried out on these two planes. In the data plane, this paper implements real-time data sampling that supports upper-layer algorithms and control actions that support upper-layer algorithms. In the logic plane, this paper implements real-time data reception, preprocessing and application programming interface, and provides support for the optimization program that uses high-level language to execute.

[Key words] software defined enterprise wireless network; data sampling; network optimization

0 引言

WiFi 无线网络在生产生活实践中已经起到了不可替代的作用, 并正在呈现网络规模、数据量的爆炸式增长。在无线设备方面, 无线热点的数量已从 2017 年的 1.24×10^9 即将增长到 2022 年的 5.49×10^9 规模; 由此推知, 来自于 WiFi 网络数据的规模将在 2022 年超过所有 IP 数据量的 51%。WiFi 无线网络中最重要的网络组织形式是企业级无线网络。企业级无线网络指的是具有一个或者多个控制器、多个接入点的可配置可管理的无线局域网组织形式, 常见的校园无线网络、企事业单位无线网络、公共场所无线网络都是企业级无线网络的常见形式。企业级无线网络具有可中心配置管理、可软件定义、计算能力强、网络性能强、兼容性强等特点, 尤为重要是企业级无线网络可以让网络管理人员和研究从设备的弱兼容性、计算能力不足等限制中解脱出来, 专注于算法设计本身以提升网络的传输性能。正因为如此, 在软件定义网络协议^[1]出现后,

即迅速应用到企业级无线网络的组织管理中心^[2-5]。但是目前的企业级软件定义无线网络实现的出发点都是支撑各自的应用, 在本文涉及的项目中, 部分诸如链路调度、负载均衡的应用需要毫秒级的数据采样支持、需要特有的控制动作支持、需要毫秒级的数据集研究数据传输拥塞与 WiFi 特征变量之间的关系, 而这些支持从现有平台中是无法得到的, 所以本文开发了一个支持毫秒级数据传输优化的平台。

作为支撑优化算法计算和执行的平台基础, 数据采样一直是无线网络中数据传输优化研究的重点。考虑到本文平台与其他平台之间的差距主要在于数据采样, 为此拟着重介绍无线网络监控方面的相关研究。在文献[6]中, 通过 ping 方法来获取清华大学校园无线网络中的数据传输时延, 并从思科(该校网络设备提供商)提供的简单网络管理协议(simple network management protocol, SNMP)中提取粗粒度的网络状态统计信息, 从而研究 WiFi 时延与网络状态之间的量化关系。在文献[7-8]中, 采用

作者简介: 林世荣(1985-), 男, 博士研究生, 主要研究方向: 软件定义无线网络、拥塞控制; 姜守旭(1968-), 男, 博士, 教授, 博士生导师, 主要研究方向: P2P 网络、数据库、传感器网络。

收稿日期: 2020-10-29

哈尔滨工业大学主办 ◆ 科技创新与应用

TCPdump (<https://www.tcpdump.org/manpages/tcpdump.1.html>) 获取 TCP 通信的参数,收集系统信息 syslog 中的网络状态信息,使用 SNMP 获取并反馈数据流的流量信息;在某些时间点或者采样结束后,将数据收集到计算节点进行分析,从而得出典型网络特征变量的特征以及数据传输的一些特征。在文献[9]中,一个连接在康涅狄格大学网关路由器上的采样设备对所有进入和流出根网关的数据包进行采样,这些采样涵盖了应用层、传输层和网络层的特征变量。该采样用来研究校园网络中的数据来源、吞吐量、延迟、丢包、同时运行的流数量、往返时间(Round-trip time, RTT)、应用反应时间等。文献[10]中,研究了拥塞对网络吞吐量、信道繁忙占比(最近 1 s 内信道处于忙状态的时间比率)、RTS-CTS 机制、帧发送和接收以及接收延迟的影响。前述内容与本文的拥塞预测部分研究工作类似,但是其数据采集以及拥塞的定义存在问题:该研究受制于当时接入点(access point, AP)性能的限制,其数据包信息是通过一个叫做 tetheral 的软件进行采集的,而影响传输的因素使用一台搭载 sniffer(嗅探器)的 IBM R32 笔记本进行采集,这些相关参数包括发送速率、信道、信噪比和接收数据包量;但该研究并未严格区分拥塞丢包和非拥塞丢包,这也使得该研究的成果存在一定的争议。

综合上的研究工作都具有一个特征:采样频率低,采样不全面。本文旨在设计并实现了一个集细粒度且高频采样、实时预处理及优化算法应用程序编程接口(API)输入支持、速率控制为一体的软件定义无线网络平台,以支撑毫秒级网络优化,支撑传输控制协议(TCP)对应的数据拥塞与 WiFi 特征变量之间的关系研究。本文研发设计的平台旨在达到已有工作不能支撑的总体水平。

本文研究内容安排如下:第一部分给出平台的整体架构及功能模块;第二部分讨论了平台的数据采样;第三部分介绍了本文设计平台上的一个拥塞窗口回归应用,第四部分总结全文。

1 平台架构

本节立足于研究本文企业级软件定义无线网络平台的设计与实现,主要介绍平台的设计思想、采样模块、数据处理模块、以及相应的逻辑控制模块。对此可做阐释分述如下。

1.1 平台结构简介

在软件定义网络协议(通常为 OpenFlow)协议

中,网络被划分为数据传输平面和逻辑控制平面两个部分。其中,数据传输平面主要负责数据传输,而未能兼管逻辑相关的控制;逻辑平面主要负责逻辑的生成及管理。这是区别于传统数据传输网络的,在传统通信网络中并不区分逻辑平面与数据平面,不同厂家的数据设备设计并未形成统一的规范,各自算法固化在数据传输设备上,导致不同厂商的设备的协同工作调试极为困难,并且需要专业人员的配合参与。本文平台按照软件定义网络协议的逻辑进行开发,设计结构如图 1 所示。其中,无线接入点处于数据传输平面,与因特网连接,负责将用户请求的数据下行传输到移动设备上。无线接入点上运行的数据采样平面将数据实时采集并传送给控制器和数据采集器进行接收、存储和运算。一般计算能力强的服务器充当控制器角色,控制器上完成数据接收、处理并通过应用程序编程接口的形式传送给上层的优化算法进行数据传输优化计算,以提升数据传输的延迟和吞吐量性能。运行在控制器上的功能如数据传输性能与 WiFi 特征变量之间关系的研究、网络优化配置管理、负载均衡、链路调度以及传输协议优化。

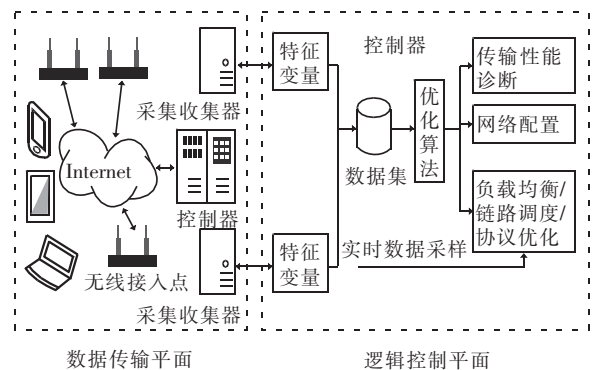


图 1 平台体系结构示意图

Fig. 1 The framework of the proposed platform

如上文所述,传统数据通信网络中并不区分数据传输平面和逻辑控制平面,传输控制逻辑直接固化在数据传输设备上,而软件定义网络中将数据传输逻辑剥离到控制器上运行,交换机只负责按照控制器生成的传输控制逻辑进行数据传输。所以构建软件定义无线网络的第一步是将无线接入点转化成为虚拟交换机^[11]。这一过程通常采用 Open vSwitch (www.openvswitch.org) 来实现,Open vSwitch 是一个软件定义网络套件,通过接管数据传输并创建相应流表的方式将传统网络设备虚拟为交换机,使其像真正的软件定义交换机一样配合控制器工作。总的来说,在无线接入点上配合 Open vSwitch 工作还

需要 OpenWRT, OpenWRT 是一个为无线接入设备开发的 Linux 操作系统。Open vSwitch 可以通过编程成为用户控件程序或者内核空间模块的模式运行在 OpenWRT 上,完成数据的传输控制。在无线接入点被虚拟为交换机后,本文通过编写内核模块以及用户空间程序的方式对企业级无线网络的特征变量(网络状态变量)进行采样、冗余消除和传输。并在控制器上实时接收处理的同时通过 API 的方式提供给网络优化算法使用。网络优化算法在数据采样基础上计算出网络优化策略后,通过编码、且加载到本文开发的消息中传输给接入点,接入点按照接收到的控制指令并在本文开发的控制模块的基础上进行网络优化控制。这里将针对整个控制过程中涉及的采样、数据接收及预处理、运算和控制等过程及对应的模块做出研讨论述如下。

1.2 数据平面模块设计

1.2.1 数据采样模块

如前文介绍,数据采样是本文区别于其他平台的主要原因。通常软件定义网络平台通过软件定义网络协议自带的状态消失进行数据采样。但是这种方式采样频率低、采样并不全面,不足以支撑数据传输优化算法的研究。本文数据采样的目的在于支撑毫秒级的链路调度研究以及 TCP 拥塞与 WiFi 特征变量之间的关系的研究。故本文在接入点内核开发了相应的模块,采样各种状态信息,如:

(1) 采样 WiFi 接入点的状态信息。如信道忙占比、信道接收占比、信道传输占比、信道扫描占比等,信道的状态占比是在过去 1 s 内处于该状态的时间比率。比如信道忙占比为 0.6,表示在过去的 1 s 内,有 600 ms 间信道处于工作状态。

(2) 链路状态信息。如链路传输/接收的数据包数量、传输失败、传输重试、链路信噪比等变量。

(3) 拓扑信息。从 Beacon 信标中可以得到接入点接收到的信标信息,可以显示邻居接入点的数量及信号强度信息。

(4) 队列信息。队列是数据传输过程中的一个重要结构体,数据进入设备后会被存储在相应的队列中等待被传输,TCP 中的拥塞丢包就发生在数据队列中,所以本文着重对接入点上的队列信息进行了采样。

(5) 数据包信息。本文在 Open vSwitch 的 datapath 中对数据包的 TCP 头部、IP 头部信息进行了采集,得到数据传输的端口、IP 地址、物理地址等。

本文采集的数据见表 1。

表 1 采样的内容
Tab. 1 The sample data items

采样项目	说 明
congestion drop	拥塞丢包数量
noncongestion drop	非拥塞丢包
time_busy	信道忙占比
time_rx	信道接收占比
time_tx	信道传输占比
time_scan	信道扫描占比
noise	噪声
packets_3 / packet_5	队列 3/队列 5 中接收数据包的速率
drops_3 / drops_5	队列 3/队列 5 中丢弃数据包的速率
requeues_3 / requeues_5	队列 3/队列 5 中重进队的数据速率
backlog_3 / backlog_5	队列 3/队列 5 中的队列余量
bytes_3 / bytes_5	队列 3/队列 5 的数据传输速率
inactive_time	非工作状态占比
expected_throughput	期望吞吐量
sta_count	终端数量

表 1 中,出现的“_3”表示来自于队列 3 的数据,下标为“_5”的数据项来自于队列 5 的数据。这是因为 OpenWRT 在网件公司的 NETGEAR WNDR3800/4300 无线路由器上会产生 5 个队列,标记为 1, ..., 5,但是通常数据采队列 3 中被传输,当且仅当数据压力大、数据突发产生的时候,部分数据才会在队列 5 中被传输。其他数据的解释说明已在表 1 中做出描述了,这些数据中如在拥塞丢包、非拥塞丢包、数据包数量等是在每个数据包都被采样,连接终端数量是按照 10 Hz 的频率进行采样,其他的数据则按照 250 Hz 的频率进行采样。

公共场所以及办公环境是企业级无线网络的典型场所,为了使得采集的数据集具有代表性,本文选择了 H 大学的图书馆来表征企业级无线网络中的普通公共场所,并选择 H 大学的实验室作为典型的办公场所代表。本文在 H 大学的图书馆布置了 12 个接入点,在该大学某实验楼的 4 楼布置了 8 个接入点。这些接入点并不设置接入密码,方便用户接入。采样持续了 2 个月,总共采集了 1 060.9 GB (600.9 GB TCP 头和 460 GB 的跨层统计数据)原始数据,该数据集的规模是 TB 级别的,并在与标准的企业级无线网络环境相同来采样的情况下,能支撑典型的企业无线网络中的优化研究。

1.2.2 负载监控

网络监控的目的在于在提醒网络管理员或者网

络优化算法在合适的时候发起网络优化配置或者选出数据传输优化算法以最大化网络传输性能。在一个典型的企业级软件定义无线网络中,通常有一个或者多个控制器、多个接入点和多个无线终端,控制器通过无线接入点采集的实时数据,判定网络当前所属的状态。当符合一定的特征的时时刻发起网络优化算法。

1.2.3 自定义消息和自定义动作

在软件定义网络协议(如 OpenFlow)中,通常有一套专门的消息用于在交换机和控制器之间传输采样消息和控制消息。比如在 OpenFlow 协议中,采用 6553 端口传输这些消息。但是因为系统实现的原因,这一消息机制并不能用于传输本文的采样信息和控制信息。所以在 OpenFlow 中的 Experimenter 消息的框架下,本文开发了 Open vSwitch 与控制器套件(RYU, <https://ryu-sdn.org/>)之间的自定义消息。这样在控制器 RYU 框架下计算出的控制消息可以通过用户自定义消息传送给交换机上的 Open vSwitch,后者在对应的用户空间协议 ofprotocol.c 中进行接收和处理。因为数据传输优化通常涉及到数据流之间的速率控制问题,同时要求数据传输的速率控制最好不要涉及丢包,因为丢包会造成额外的影响,所以本文开发了一种基于拥塞窗口的控制方式,其思路是通过修改 ACK 确认数据包中的广告串口限制发送窗口的方式进行速率控制。本次研究中即在 Open vSwitch 框架下开发了一个传输控制层动作 `set_window_to(x)` 用于改变当前数据传输数据流的速率,其中 x 是期望的网络速率。

1.3 逻辑平面模块

逻辑控制在使用软件定义网络协议的数据传输网络中扮演着“大脑”的角色,负责数据传输逻辑的生成和管理。如前文所述数据传输的逻辑被从数据传输设备剥离到控制器上进行计算,本文的逻辑控制建立在数据采样上层,负责对实时数据继续接收和处理,以及控制逻辑的运算。其中,数据接收和处理是有难度的,因为企业级无线网络中无论是设备规模、还是数据规模都是巨大的,但是只有对采样数据进行统计后才能得到最终可用的数据。比如信道的各种状态占比需要统计各个接入点在过去 1 s 的数据信息。而处理后的统计数据也不能直接供给优化算法使用,还要整合成应用程序编程接口的形式供优化算法进行调用。为此,对数据的接收过程可做如下分析详述。

1.3.1 数据实时接收及预处理

数据接收预处理的作用在于实时地处理接收到的来自于不同无线接入点的采样数据,并统计得到可以用于网络优化算法的输入变量。因为企业级无线网络中具有设备密度大、设备及数据规模大、设备与网络设备之间的连接关系复杂(及链路动态变化)等特征。所以数据接收处理的难度较大。并且采样到的元数据需要转化成为统计数据才具有价值,如前文论述的信道状态占比,以及队列信息、链路信息等等都需要进行统计分析,所以应该在数据流之间维护大量的采样数据缓存以计算网络状态变量的统计信息。在得到统计数据后,为了配合优化算法的执行,还要开发应用程序编程接口。所以本文采用 C 语言进行多线程编程以支持数据的接收,并且通过混合编程的方式开发应用程序编程接口来支撑网络优化算法的执行。

仍需一提的是 TCP 窗口的获取。因为历史原因,嵌套在数据的 TCP 头部的接收窗口并不是该数据流对应的真实接收窗口。接收窗口在数据传输的过程中表示当前数据接收方的接收能力,也是数据发送方数据发送的上限,发送方的发送窗口应该小于拥塞窗口和接收窗口大小。因此可以通过修改接收窗口的方式来限制当前数据流的速率。但是 TCP 头部中的窗口并不是当前数据流的真正窗口大小。这是因为在数据传输网络的初期,网络传输介质的能力相对较弱,此时采用 8 位大小就可以表征数据速率。但是随着数据通信技术的发展,数据传输介质的能力得到了提升,传统 8 位大小的接收窗口不足以支撑数据的高速率传输,随后就在 RFC1323 (<https://tools.ietf.org/html/rfc1323>)中提出了窗口放缩选项位,额外增加 8 位配合原有 8 位表征数据的接收窗口。但是该 8 位不在数据传输过程中发送,只在数据流的建立过程中相互通知。所以要得到一个数据流的接收窗口大小就要获得该窗口放缩选项以及当前 TCP 头部中的窗口大小。同时,本文设计中在接入点上的 Open vSwitch 上为每个数据流建立了相应的状态机,保存数据流建立过程中的各种协调信息。并在 TCP 数据流的各个生命过程中维护状态机的状态变化,从而获取到数据流的当前窗口大小,配合存储的窗口放缩选项,可以计算出当前数据流的接收窗口大小。

1.3.2 负载均衡

负载均衡用来管理客户端设备与无线接入点之间的连接关系,从而达到均衡网络负载以最大化数据传输的延迟和吞吐量性能的目的。该功能是建立

在数据采样基础上的,通常可以将其形式化为一个混合整数规划问题,求解并得到相应的连接关系,执行该连接关系,负载均衡则得以实现。

1.3.3 过载诊断

过载诊断通过研究网络负载与网络特征变量之间的关系,采用合适的回归模型可以得到当前的网络负载情况,也可以采用预测模型得到未来的网络负载状况,从而支撑各种网络优化算法的运行。

2 企业级无线网络中拥塞窗口回归

在本节中,介绍本文支撑网络优化的一个典型应用:拥塞窗口回归。拥塞窗口是 TCP 数据传输的一个重要参数,表示在未接收到下一个 ACK 的情况下最多可以注入网络的数据包的数量。通常 TCP 的发送窗口控制着发送速率,但是 TCP 的发送窗口必须同时小于拥塞窗口和接收窗口。已有研究表明在一般情况下,接收窗口大于拥塞窗口,所以如果在 TCP 传输的中间节点上得到拥塞窗口,则可以通过此拥塞窗口来控制 TCP 的速率以规避不必要的拥塞产生。文献[12]给出了一种通过未确认数据包量(unacknowledged, UNA)来进行拥塞窗口回归的方法,据此测算,该方法在有线网络中能够达到 0.9 以上的精确度,只是同时在研究中也发现,仅仅使用 UNA 在企业级无线网络中只能达到 0.291 1 的准确度。

既然只采用 UNA 解决不了企业级无线网络中的中间节点上的拥塞窗口回归问题,那么在企业级无线网络中,中间节点上的拥塞窗口回归的性能极限到底是多少,应该采用什么样的特征变量进行回归?本文开展了相应的回归研究。该回归研究拟欲设计 TCP 的拥塞窗口,限于篇幅,这里将重点评述拥塞窗口的采样工作。

为了采集作为回归对象的拥塞窗口,本文在 TCP 发送端进行了相应的采样实现,特开发了一个类似于 Linux 下“ss”命令的应用程序。该应用程序通过 NetLink 消息从内核获取当前数据流的统计状态信息,包括拥塞窗口、延迟等数据。同时再依据表 1 中的特征数据采样,共同组成回归所使用的数据集。研究中需特别关注的就是时间同步,本文在局域网内部架设专有的时间服务器以给采样的设备提供共同的时间,从而统一了采样点的时间,使得数据集具有共同的时钟。

在采样数据集的基础上,研究了表 1 中的特征变量与拥塞窗口之间的关系,从而得出所选特征变

量以及使用 seaborn. heatmap 计算的相关性见表 2。由表 2 可知,其中的数据项与拥塞窗口有显著的关系,即显示了 UNA、所有数据链路的数据速率、信道传输占比、信号强度、信道忙占比、重传率以及噪声与拥塞窗口具有显著的关系。依据这些特征变量,本文采用标准的机器学习回归方法研究了将其用于拥塞窗口的回归精度,最终结果见表 3。

表 2 所选特征变量以及使用 seaborn. heatmap 计算的相关性
Tab. 2 The selected characteristic variable and their relations obtained by utilizing seaborn. heatmap

特征变量	说明及数据来源	相关性
UNA	未确认的数据报的数量,从数据报信息中得出	0.52
Data rate for all the links	所有链路的速率和,从链路信息中得出	0.50
Channel transmit Utility	过去 1000 ms 中处于传输状态的时间,信道信息	0.49
Signal strength	信号强度,链路信息	0.27
Channel busy utility	过去 1000 ms 信道忙的时间,信道信息	0.23
Retransmission ratio	数据报重传率,丢包信息和数据报信息	0.20
Noise	噪声,信道信息	-0.13

表 3 仅使用 UNA 和使用本文的特征变量的回归算法及其性能
Tab. 3 The regression performance when utilizing UNA only and utilizing the variables in the paper

回归算法	仅仅使用 UNA	使用本文的特征变量
Gradient Boosting Regressor	0.291 1	0.536 2
Random Forest Regressor	0.290 2	0.513 5
Linear Regression	0.271 3	0.435 4
Voting Regressor	0.289 2	0.524 6
Ada Boost Regressor	0.271 5	0.390 0
Bagging Regressor	0.290 2	0.468 5
Extra Trees Regressor	0.290 2	0.504 5
XGB Regressor	0.291 1	0.537 4

回归结果显示,如果仅仅使用 UNA 进行回归,最好也只能达到 0.291 1 的精度,但是如果选用本文选择的特征变量,该回归精度可以提升到 0.537 4。基于选择的特征变量,本文采用回归表现最好的 XGBRegressor 算法对拥塞窗口进行回归,并截取了部分回归结果如图 2 所示。图 2 中,蓝色是实际窗口,红色是通过 XGBRegressor 获得的窗口。仿真结果显示回归窗口与实际窗口拟合程度可以接受。