基于FPGA的高速校园网全流量采集处理方案设计与实现

姚仁龙，赵琼，何海涛，韦雨君，黎恩磊
（中山大学网络与信息中心，广东 广州 510275）

摘　要：通过SDN架构采集校园网传输的流量数据完整码流，使用FPGA技术进行解析处理后，生成各种级别应用级日志数据。通过大数据技术进行分析和挖掘后，为校园网管理和安全提供数据支撑，发现网络中存在的潜在风险。

关键词：FPGA；SDN；流量采集；安全分析
中图分类号：TP393
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024257

Design and realization of a high-speed traffic collection and processing scheme for campus networks using FPGA

YAO Renlong, ZHAO Qiong, HE Haitao, WEI Yujun, LI Enlei
Network and Information Center, Sun Yat-sen University, Guangzhou 510275, China

Abstract: The study captures the entire traffic data stream of the campus network using the SDN architecture, followed by parsing and processing with FPGA technology to generate application-level log data at different levels. Subsequent analysis and mining with big data techniques provide data support for campus network management and security, revealing potential risks within the network.

Keywords: FPGA, SDN, traffic collection, security analysis

0　引言

随着互联网技术的飞速发展，校园网作为高校日常管理的重要基础设施，承载的流量日益增长，在办公、教学和科研方面的应用程度也越来越广泛[1]，校园网流量的快速增长给网络安全、数据分析等方面带来了新的挑战。同时校园网由于存在规模大、计算机系统复杂、用户活跃等特点，常常成为黑客重点攻击的目标[2]。因此，如何高效、准确地采集和处理高速校园网流量，成为当前研究的热点问题。

目前流量采集主要有Netflow方案和Sflow方案，但是随着网络速度的提高，在设备上开启流量采集功能会降低设备转发性能，与此同时，基于flow的流量采集使用抽样方案，在需要深度流量分析的场景中并不适用[3]。而使用软件定义网络（SDN）技术能够在不牺牲设备性能的情况下实现全流量实时采集需求。

传统的基于软件的流量处理方案在处理速度、扩展性和灵活性方面存在一定的局限性[4]。相比之下，现场可编程门阵列（FPGA）作为一种硬件加速技术，具有并行处理能力强、功耗低、可编程性高等优点，在高速流量处理领域具有广泛的应用前景。本文使用FPGA解析处理网络流量，并以传输控制协议（TCP）、安全套接层协议（SSL）和域名系统（DNS）为例介绍了应用层日志的生成和应用。

本文旨在设计并实现一种基于FPGA的高速校园网流量处理方案，通过SDN架构对校园网流量

收稿日期：2024–11–07

进行采集分发，利用FPGA实现高速、高效的流量处理，生成应用层可用日志，并为大数据平台等提供数据支撑。

1　相关技术与概念

1.1　SDN架构介绍

SDN是一种新兴的网络架构，其核心思想是将网络控制层（控制平面）与数据转发层（数据平面）分离，实现网络的可编程化和自动化管理。SDN架构主要包括3个部分：SDN控制器、SDN交换机和SDN应用。SDN控制器负责整个网络的控制逻辑，SDN交换机负责数据转发，SDN应用则提供丰富的网络功能和服务。SDN架构具有灵活性高、扩展性强、管理简便等优点，在高速网络环境中具有广泛的应用前景[5]。

1.2　FPGA技术基础

FPGA是一种基于可编程逻辑器件的数字集成电路，内部包含大量的逻辑单元、计算资源和存储器，可以根据实际需求进行编程和配置。FPGA具有并行处理能力强、功耗低、可编程性高等优点，在高速流量处理领域具有广泛的应用。FPGA的主要特点包括以下几点。

1) 并行处理能力：FPGA内部包含大量的逻辑单元和计算资源，可以实现高度并行处理，提高流量处理速度。

2) 低功耗：FPGA采用静态功耗控制技术，相较于传统CPU和GPU，具有更低的功耗。

3) 可编程性：FPGA可以根据实际需求进行编程和配置，实现定制化的流量处理功能。

4) 灵活性：FPGA支持在线重构，可以根据不同的流量处理需求动态调整硬件资源。

1.3　TCP、SSL和DNS

应用层网络流量日志作为网络管理和安全分析的基础，在网络安全态势感知中发挥了重要的作用[6]。本文主要以TCP、SSL和DNS流量解析为例阐述FPGA流量处理过程。

TCP是一种面向连接的、可靠的、基于字节流的传输层通信协议。TCP流量处理主要包括TCP连接建立、数据传输、连接终止等过程。随着网络攻击的隐匿性日渐加深，深度TCP流量分析相较于传统的数据包分析能够提高恶意数据流检测的准确性。

SSL是一种安全协议，用于在互联网上提供加密通信。SSL流量处理主要包括SSL握手、密钥交换、数据加密解密等过程。当前诸多恶意软件使用SSL加密通信协议进行通信，例如木马、蠕虫、下载器、翻墙软件和洋葱网络等[7]。

DNS是一种用于将域名解析为IP地址的系统。DNS流量处理主要包括DNS查询、DNS响应、DNS缓存等过程。DNS流量在恶意域名、僵尸网络通信检测等方面发挥重要作用[8]。

2　系统设计

为实现校园网主要链路的完整数据包采集需求，本文使用端口镜像技术捕获校园网流量。同时基于SDN技术实现捕获点的统一接入管理和对流量的复制、汇聚以及流量出口的负载均衡等功能。流量处理模块接收到流量后，将端口数据合并、分类、过滤和分流，添加标准描述符和扩展描述符后进入主机相应的Hostbuffer中，后续根据hash值定位独立的flow以识别不同的应用层协议日志。最后应用日志通过flume采集到大数据平台。全流量处理平台拓扑如图1所示。

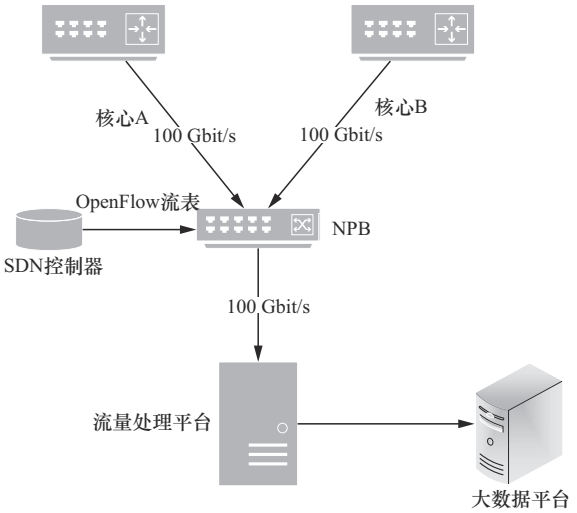


图1　全流量处理平台拓扑

3　系统实现

3.1　流量采集与分发

根据全流量采集的需求，结合现网资源条件，实施的网络拓扑如图2所示。两台OpenFlow交换机承担捕获层、过滤（NPB）层和分发层功能，基于ODL（Opendaylight）的SDN控制器下发OpenFlow流表到OpenFlow交换机后，经过二次处理之

后的流量，由交换机分发给后端的流量处理模块。根据具体需求，可以对入站端口流量进行过滤，同时为了避免出站端口流量过载，也可对出站端口流量进行负载均衡[9]。

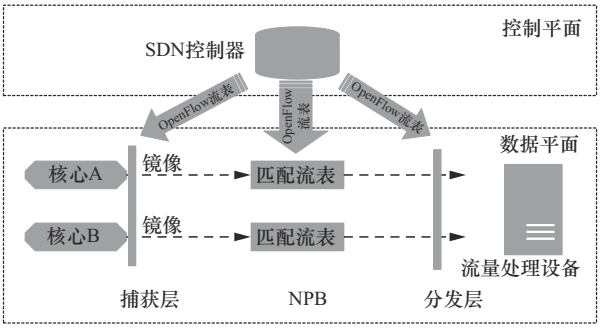


图2　全流量采集拓扑

3.2　FPGA流量处理流程

为满足大数据平台的网络数据需求，需要构建高性能的流量处理模块。现有加速卡具备处理100Gbit/s流量能力。加速卡主要包括硬件，内核驱动、软件服务及用户应用程序接口（API）构成。可通过NTPL（Napatech Programming Language）或API对设备进行配置，并使用软件开发工具包（SDK）提供的API进行数据的处理，实现灵活方便的数据聚合编排功能。

流量处理模块使用加速卡将原始流量转换为应用日志。到达加速卡网口的数据包会按到达顺序添加时间戳，分别经过合并、分类、过滤和分流，添加标准描述符和扩展描述符后进入主机相应的Hostbuffer中。从各个Hostbuffer中循环取出数据包，根据加速卡添加的标准描符和扩展描述符对数据包进行进一步解析处理，如图3所示。

同时支持将多个点捕获的数据帧合并为单一流，例如通过“5-tuple sorted”hash算法生成的Hash值，可以定位出一条独立的flow：源IP+源端口号+协议类型+目的IP+目的端口号。定位出独立的flow后，根据不同的应用层协议的特征，识别出各应用层协议，并存储到数据结构中，待后期处理。

3.3　TCP流重组算法实现

TCP流包含由硬件在帧处理流程中生成的帧描述符，其中包括TCP校验和、哈希值、时间戳和IP校验以及帧长等信息。之后对帧进行解析，通过对帧描述符的解析，判断是否为有效帧、TCP头部信息、哈希值是否一致和校验和是否有效。解析完成后生成一条由五元组确定的TCP流。flow开始于TCP的3次握手，结束于TCP的FIN或RESET（或者15 min内无新的连接）。日志定义格式如下

时间|持续时长|协议|应用类型|源IP|源端口号|目的IP|目的端口号|发送最小包长|发送最大包长|发送平均包长|接收最小包长|接收最大包长|接收平均包长|发送包数|接收包数|发送字节数|接收字节数|发送速率|接收速率|最长发送应答|最短发送应答|平均发送应答|最长接收应答|最短接收应答|平均接收应答|发送重传|接收重传|建立时间|发送SACK|接收SACK|TCP状态|发送首包负荷|接收首包负荷

TCP流重组流程如图4所示。

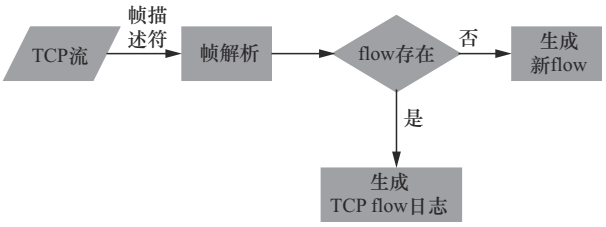


图4　TCP流重组流程

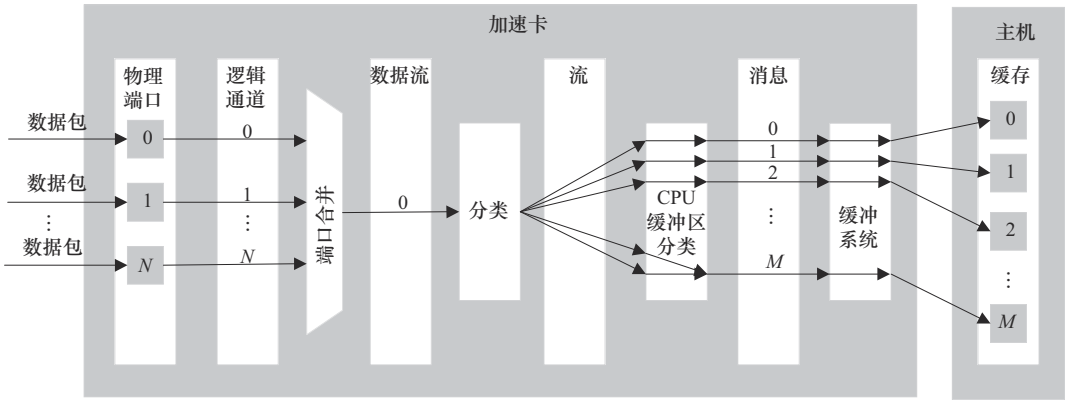


图3　加速卡包流程处理框架

3.4　SSL日志解析算法

协议解析流程中判断TCP流是否使用SSL加密，对于使用SSL加密的TCP流会判断SSL的Hello信息，并最终从certificate包中提取证书信息。certificate包中包含证书持有者及颁发者的相关信息，如表1所示。

表1　　certificate包内容

参数	含义
C（Country Name）	所在国家字母简称，如中国CN
ST（State or Province Name)	所在省份简称，如Beijing
L（Locality Name）	所在城市
O（Organization Name）	公司或者机构名称
OU（Organizational Unit Name）	部门简称
CN（Common Name）	公用名称

日志格式定义如下

时间|源IP|源端口号|目的IP|目的端口号|公用名称|组织机构名称

3.5　DNS日志解析算法

协议解析流程中判断用户数据报协议（UDP）流是否为DNS协议，针对DNS协议分析DNS请求报文及应答报文。从DNS应答报文中提取日志，并根据UDP的五元组生成UDP流日志。DNS应答报文采用了TLV（tag-length-value）格式，依据此规则依次提取DNS的关键信息。本例中DNS日志提取了CNAME（Canonical NAME）及IN（Internet）信息，出于后续数据处理的需要，CNMAE及IN各存储4条信息，日志定义格式如下

时间|DNS服务器地址|请求域名|CNAME|IN|TTL|CNAME1|CNAME|IN|TTL|CNAME2|CNAME|IN|TTL|CNAME3|CNAME|IN|TTL|CNAME4|A|IN|TTL|IP1|A|IN|TTL|IP2|A|IN|TTL|IP3|A|IN|TTL|IP4|请求DNS解析客户端地址

4　应用与效果展示

本文使用由43个节点组成的集群搭建的cloudera大数据管理平台进行验证。大数据管理平台通过flume采集流量处理模块的流式日志存储到Hadoop分布式文件系统（HDFS）。在前端可通过spark、impala等引擎进行数据查询，同时也可以建立周期任务，实现异常检测，监测告警等功能。

通过对校园网流量的分析，发现了诸多安全事件，其中一个例子为论文过量下载行为。根据信息显示，某数据库认为学校某个出口IP存在过量下载行为，并进行封禁。根据网络地址转换（NAT）日志和TCP流量显示，校内某IP在时间和TCP流特征上都符合该过量下载行为，如图5所示。另外，通过在登录节点搭建API服务，在第三方服务器通过调用API实现SQL语句获取数据，从而实现监控告警或者数据展示等诸多功能，如图6所示。

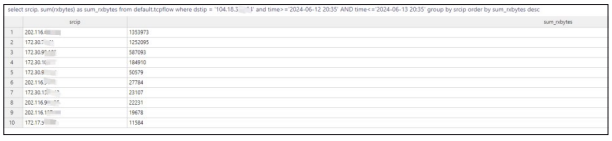
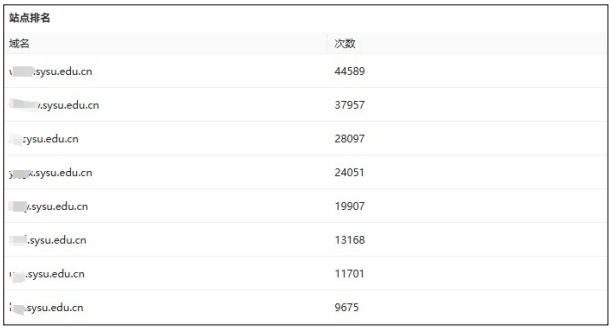


图5　TCP流表溯源过量下载



站点排名

域名	次数
.sysu.edu.cn	44589
.sysu.edu.cn	37957
sysu.edu.cn	28097
.sysu.edu.cn	24051
.sysu.edu.cn	19907
.sysu.edu.cn	13168
.sysu.edu.cn	11701
.sysu.edu.cn	9675

图6　站点访问趋势排名

5　结束语

本文针对高速校园网流量处理的需求，设计并实现了一种基于FPGA的流量处理方案。通过SDN架构，实现了校园网流量的高效采集与分发。利用FPGA的高速处理能力，对TCP、SSL和DNS流量进行了深度处理，生成了应用层可用的日志信息。此外，通过设计并实现的TCP帧合并算法，支持将多个TCP帧合并为同一个流，进一步提高了流量处理的效率。通过对海量数据的异常捕获和分析，发现网络中潜在的风险和漏洞，弥补安全体系检测的不足。

参考文献：

[1] 周康乐. 基于数据挖掘的校园流量监测系统设计[J]. 现代电子技术, 2020, 43(21): 59-63.
ZHOU K L. Design of campus traffic monitoring system based on data mining[J]. Modern Electronics Technique, 2020, 43(21): 59-63.

[2] 孔令浩. 浅谈校园网网络安全问题的分析与对策[J]. 网络安全技术与应用, 2018(5): 56-57.

KONG L H. Analysis and countermeasures of network security problems in campus network[J]. Network Security Technology & Application, 2018(5): 56-57.

[3] 赵琼, 何海涛, 杨敏. 基于SDN的大规模校园网流量采集方案研究与实现[J]. 中国教育网络, 2018(10): 31-35.
ZHAO Q, HE H T, YANG M. Research and implementation of large-scale campus network traffic collection scheme based on SDN[J]. China Education Network, 2018(10): 31-35.

[4] 任菲梵. 高速流量威胁检测方法研究[D]. 长沙: 国防科技大学, 2022.
REN F F. Research on threat detection method of high-speed traffic[D]. Changsha: National University of Defense Technology, 2022.

[5] 张俊茸. 软件定义网络(SDN)技术分析[J]. 数字通信世界, 2024(6): 115-117.
ZHANG J R. Analysis of software defined network (SDN) technology [J]. Digital Communication World, 2024(6): 115-117.

[6] 方鹏. 基于TCP流特征提取技术的网络流量识别应用研究[D]. 合肥: 中国科学技术大学, 2018.
FANG P. Application research of network traffic identification based on TCP flow feature extraction technology[D]. Hefei: University of Science and Technology of China, 2018.

[7] 陈子嘉. 基于机器学习的SSL/TLS恶意加密流量识别技术研究与实现[D]. 北京: 北京邮电大学, 2023.
CHEN Z J. Research and implementation of SSL/TLS malicious encrypted traffic identification technology based on machine learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2023.

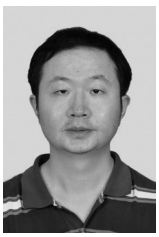
[8] 周婧莹, 黎宇, 黄坤, 等. 基于DNS流量分析识别加密货币矿工的研究和实现[J]. 邮电设计技术, 2023(8): 48-52.
ZHOU J Y, LI Y, HUANG K, et al. Research and implementation of cryptocurrency miner detection based on DNS traffic analysis[J]. Designing Techniques of Posts and Telecommunications, 2023(8): 48-52.

[9] 杨敏, 何海涛, 赵琼. 流量大数据安全分析平台的设计与实现[J]. 通信学报, 2018, 39(S1): 104-109.
YANG M, HE H T, ZHAO Q. Design and implementation of traffic big data security analysis platform[J]. Journal on Communications, 2018, 39 (S1): 104-109.

[作者简介]

姚仁龙（1992– ），男，安徽安庆人，中山大学助理工程师，主要研究方向为人工智能、流量分析。

赵琼（1983– ），男，湖北黄冈人，中山大学工程师，主要研究方向为计算机网络技术。



何海涛（1975– ），男，安徽淮北人，中山大学高级工程师，主要研究方向为因特网流量行为、大数据等。

韦雨君（1997– ），女，贵州贵阳人，中山大学助理工程师，主要研究方向为网络安全、流量分析。

黎恩磊（1995– ），男，河南漯河人，中山大学助理工程师，主要研究方向为数字取证、流量分析。