VAS网络流分析

报告人：夏鸣轩

日期：2019年3月18日

内容简介

- 项目背景
- 项目理解与分析
- 项目进展
- 项目展望

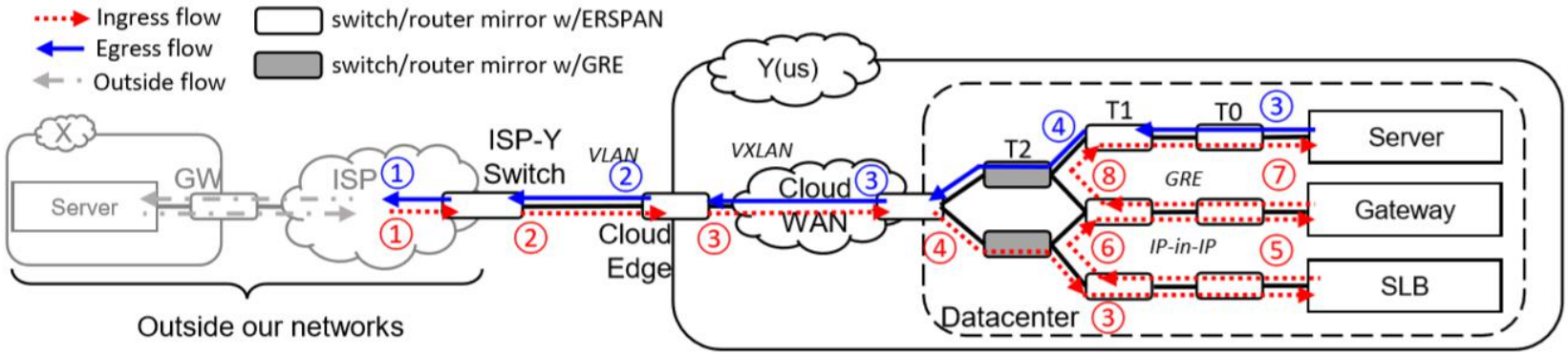
背景

- VAS网络定义
  - 基于NFV技术，提供除通信等基本服务以外，包括网络流量监控、网络故障管理等功能的网络
- 网络基本业务
  - 网络节点接入、网络节点通信
- 网络增值业务
  - 网络流量监控、网络故障管理、网络安全防护

问题

- 问题一：VAS网络功能组件修改报文内容
  - 其根据需要会修改报文头部或并合并不同报文的荷载，导致报文特征难以有效提取，从而难以准确拼接流信息。
- 问题二：VAS网络异构且复杂
  - VAS网络中网络功能异构复杂，包含来自于不同的第三方的物理设备和软件功能，无法直接采集各网络服务功能的配置信息。
- 问题三：VAS网络吞吐量大
  - 高速网络中网络吞吐量大，且网络功能规则动态性强，无法对报文进行逐包快速分析并实时识别流。

例子



编号	报文内容								网络设备
①	ETHERNET						IPV4	TCP	
②	ETHERNET					802.1Q	IPV4	TCP	ISP-Y Switch
③	ETHERNET		IPV4	UDP	VXLAN	ETHERNET	IPV4	TCP	Cloud Edge
④			IPV4	UDP	VXLAN	ETHERNET	IPV4	TCP	Datacenter Border
⑤	ETHERNET	IPV4	IPV4	UDP	VXLAN	ETHERNET	IPV4	TCP	SLB
⑥		IPV4	IPV4	UDP	VXLAN	ETHERNET	IPV4	TCP	SLB
⑦	ETHERNET		IPV4	GRE		ETHERNET	IPV4	TCP	VPN Gateway
⑧			IPV4	GRE		ETHERNET	IPV4	TCP	VPN Gateway

[1] Yu D, Zhu Y, Arzani B, et al. dShark: A General, Easy to Program and Scalable Framework for Analyzing In-network Packet Traces[C]//16th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 19). 2019.

分析

- 问题定义
  - VAS网络流串接技术→网络流量关联、分类
- 工作流程
  - 训练过程

    特征工程→样本构建→模型训练
  - 识别过程
- 提高效率
  - 抽样提取
  - 高速报文处理技术（Netmap，DPDK）

项目规划


3月中旬
4月中旬
5月中旬
6月中旬
7月中旬
平台搭建
平台
特征工程
模型设计
模型验证
模型训练优化
模型
初步采集
形成数据集
分布式采集
采集

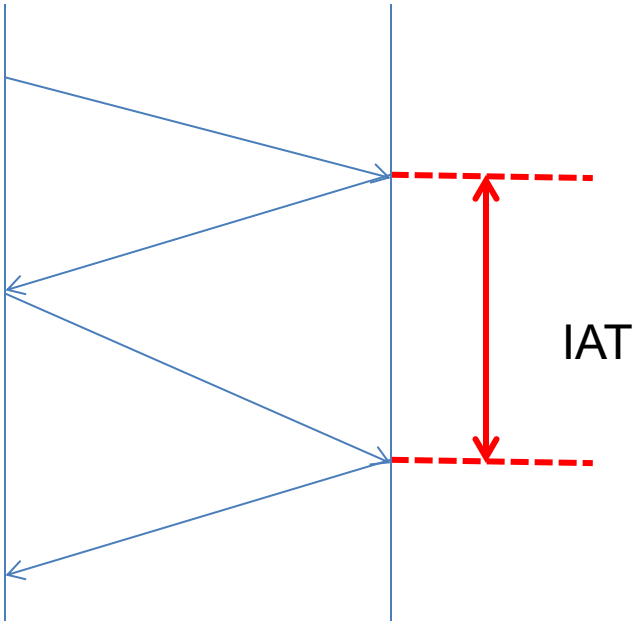
特征选取

- 流量特征定义
  - 表示网络流量行为的一系列度量，包括时间特征、规模特征等
  - 时间特征
    - ✓和数据包发送、接收时间相关的特征，比如达到间隔时间、空闲时间、链接持续时间等
  - 规模特征
    - ✓和数据包自身大小相关的特征，如数据包大小、有效负载大小、控制信息长度等

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Inter-arrival Time

- 定义
  - 到达间隔时间(IAT)
    - 两个连续数据包到达同一个节点的间隔时间
  - IAT = D(i+1) - D(i)，其中D(i)为第i个数据包到达的时间
- 适用场景
  - NAT
  - 隧道
- 开销分析
  - 存储开销大
  - 时间开销与数据流量大小相关



[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Connection duration

- 定义
  - 链接持续时间（CD）
    - 数据流在链路上保持传输的时间
  - CD＝链接空闲时间+链接活动时间
- 适用场景
  - NAT
  - 隧道
- 开销分析
  - 存储开销较小
  - 时间开销随流量大小变化

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Time Spent Idle

- 定义
  - 空闲时间（TSI）
    - 一条流的状态由空闲变为活跃之前的持续时间
  - TSI＝当前发送数据时间-上次发送数据时间
- 适用场景
  - NAT
  - 隧道
- 开销分析
  - 存储开销小
  - 时间开销和空闲的时长有关

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

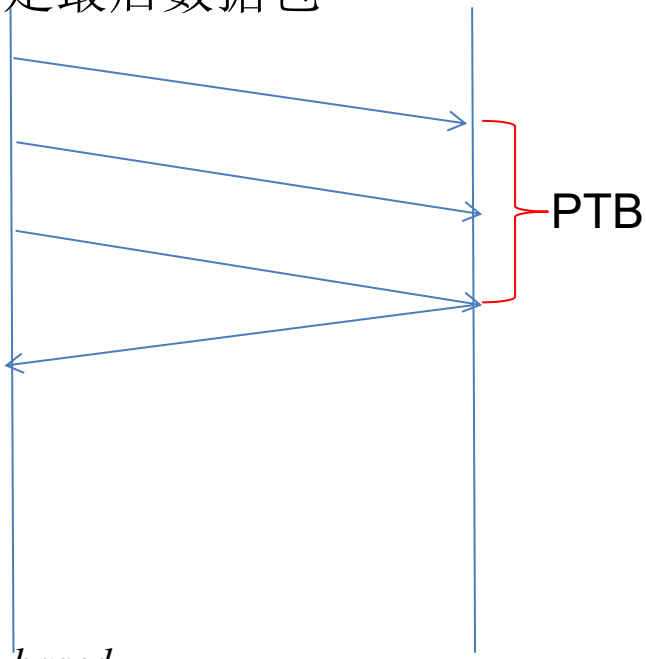
Time Spent Active

- 定义
  - 流活跃时间（TSA）
    - 一条流的状态由活动变为空闲之前的持续时间
  - TSA=最后数据包发送的时间-第一个数据包发送时间
- 适用场景
  - 隧道
  - NAT
- 开销分析
  - 存储开销小
  - 时间开销和流传输的时间有关

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Percent of Time Spent in Bulk Transfer

- 定义
  - 块传输时间（PTB）
    - 指在多个数据段连续发送而不需要等待对端应答的传输时间
  - PTB＝最后数据包到达时间－第一个数据包到达时间
    - 发送应答前接收到的最后一个数据包就是最后数据包
- 适用场合
  - NAT
  - 隧道
- 开销分析
  - 存储开销较大
  - 时间开销与块传输数据量有关



[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Packet Length

- 定义
  - 数据包长度（PL）
    - 数据包头部（控制信息）和数据包有效负载的整体大小
  - PL＝控制信息长度（SPH）＋有效负载大小（PS）
- 适用场景
  - 防火墙
  - NAT
- 开销分析
  - 存储开销较大
  - 时间开销与数据包数目有关

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Size of Packet Header

- 定义
  - 数据包头长度（SPH）
    - 指数据包中控制信息字段的长度，如IP首部信息
  - SPH＝数据包长度（PL）－有效负载大小（PS）
- 适用场景
  - 防火墙
  - NAT
- 开销分析
  - 存储开销较大
  - 时间开销与数据包数目有关

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Payload Size

- 定义
  - 有效载荷大小（PS）
    - 数据包中除头部控制信息外净荷的大小
  - PS=数据包大小（PL）—控制信息大小（SPH）
- 适用场景
  - 防火墙
  - NAT
- 开销分析
  - 存储开销较大
  - 时间开销与数据包数目有关

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Initial Window Size

- 定义
  -
  -
- 适
  -
  -
- 于
  -
  -

total segments	IW=3	IW=10
3	1	1
6	2	1
10	3	1
12	3	2
21	4	2
25	5	2
33	5	3
46	6	3
51	6	4
78	7	4
79	8	4
120	8	5
127	9	5

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

Number of Unique Bytes Sent

- 定义
  - 发送的唯一字节数（NUB）
    - 除重传数据包和窗口探针以外的数据字节数目
  - NUB＝接收数据字节总数－（接收探针字节数＋重传字节数）
- 适用场景
  - NAT
- 开销分析
  - 存储开销小
  - 时间开销与数据量相关

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

其他特征

- number of packets(bytes) in forward/backward direction 前/后向转发的数据包/字节数
- Time since the last connection between these hosts 上次链接后再次链接的时间差
- burst network traffic 突发网络流量
- periodic traffic 网络流量周期
- count of all the window probe packets 窗口探针数目

[1] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[R]. 2013.

未来工作

- 在网络流中实现对所提出的流量特征进行有效提取
- 实现特征学习的机器学习算法，并进行验证
- 提升特征提取以及机器学习算法的性能