Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心

高准确率的鲁棒加密恶意流量实时检测方法

博士研究生　张钊

2022年03月20日

内容提要

- 背景简介
- 历史发展
- 基本概念
  - 流量数据
  - 检测流程
- 算法原理
  - Whisper
  - ACID
- 应用总结
- 参考文献

预期收获

- 预期收获
  - 1. 了解网络流量和恶意流量
  - 2.了解恶意流量检测的问题和挑战
  - 3. 理解恶意流量机器学习检测方法
  - 4. 了解应用领域和发展方向等

背景简介

- 恶意流量检测
  - 互联网的快速发展，人们对网络各种服务的依赖增强
  - 网络安全形式严峻，恶意软件数量不断增加，以网络为主要传播途径的恶意软件不断地利用网络开展攻击
  - 通过分析网络流量特征来识别恶意流量，保护合法互联网用户免受网络攻击

背景简介

- 传统恶意流量检测面临挑战
  - 基于规则的检测无法检测到零日攻击
  - 由于机器学习算法的处理开销，现有的检测方法检测吞吐量较低，无法处理高速率的网络流量
  - 规避攻击可以轻松绕过使用粗粒度网络流统计信息的传统检测
  - 分析每个数据包特征序列的数据包检测无法实现稳健的检测
  - 对网络流量个别特征的微小变化极为敏感，在软件不断更新与流量环境不断变化情况下，有效性降低

历史发展

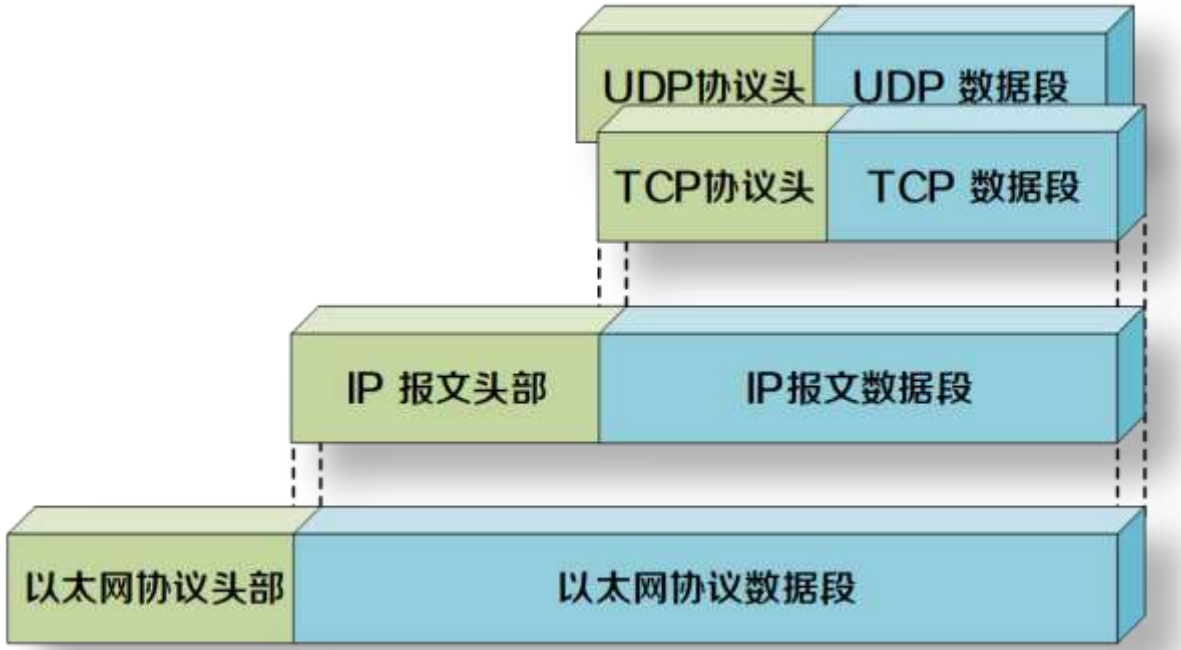
历史发展

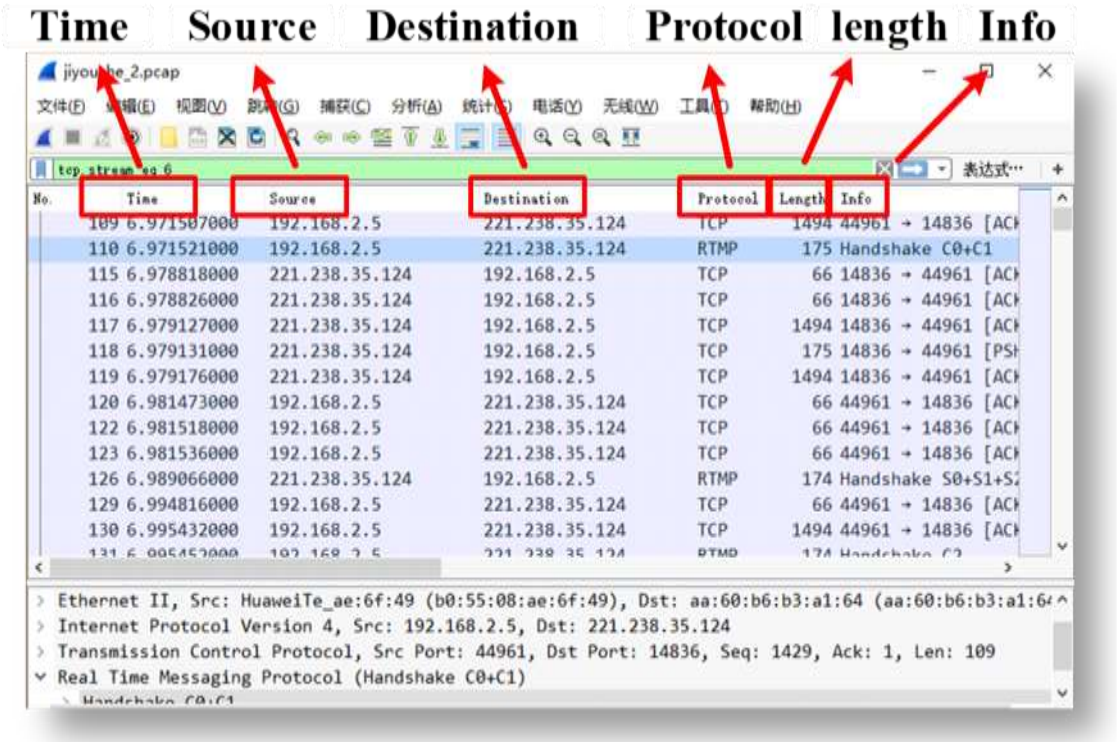
- 根据检测方法的不同性质，恶意流量检测方法可主要分为两类
  - 基于误用(Misuse)的恶意流量检测
    - 根据预定义的签名(Signature)来识别恶意流量，能有效过滤已知的恶意流量，但在识别未知恶意流量方面适用性较差
  - 基于异常(Anomaly)的恶意流量检测
    - 依靠启发式方法来捕获恶意流量，多为基于监督学习的机器学习方法
    - 基于机器学习的恶意流量检测主要步骤为数据预处理、特征选择以及检测模型训练

基本概念

基本概念

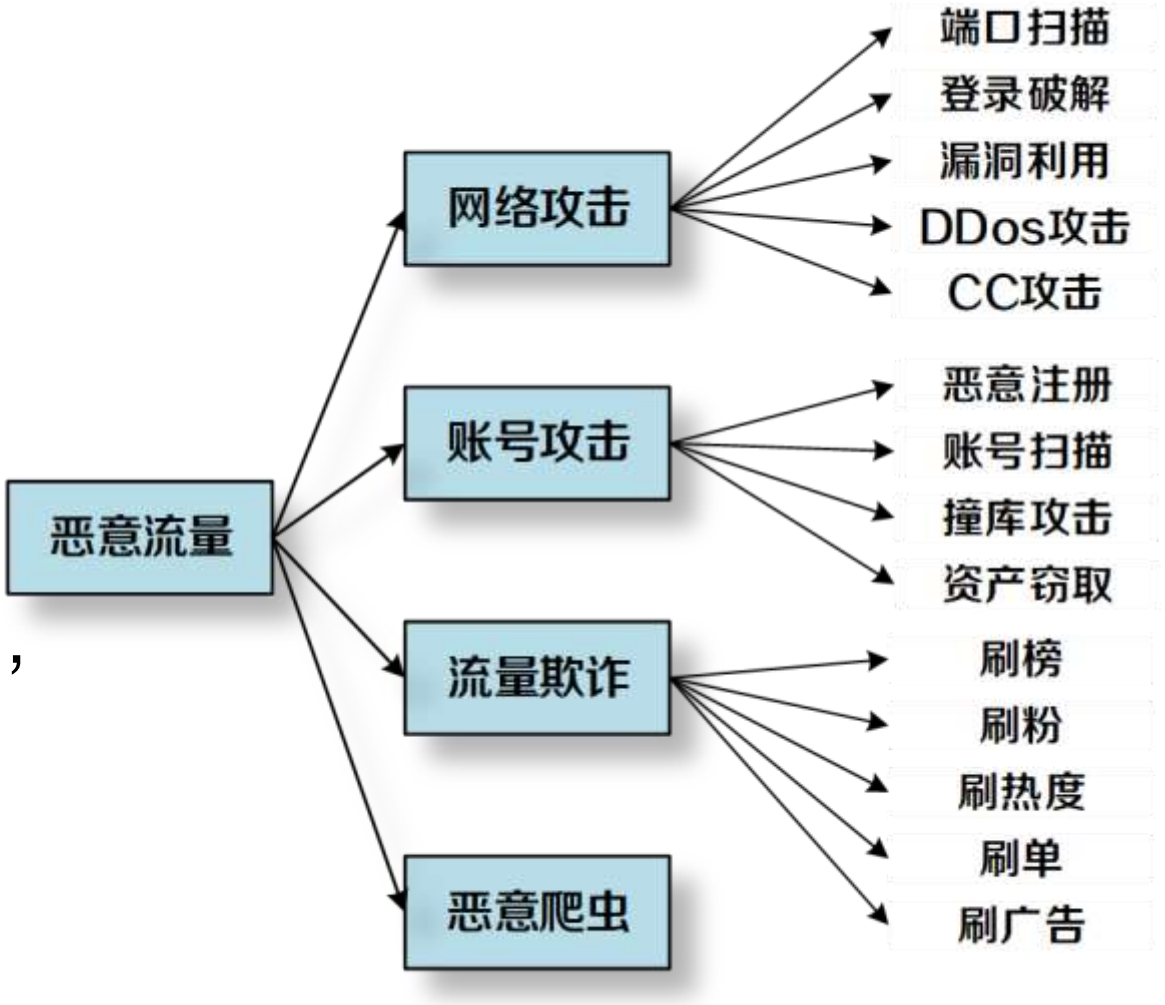
- 网络流量
  - 通过特定网络节点的数据包和网络请求数量
  - 作为网络攻击中的一种重要载体，对其进行研究具有极其重要意义





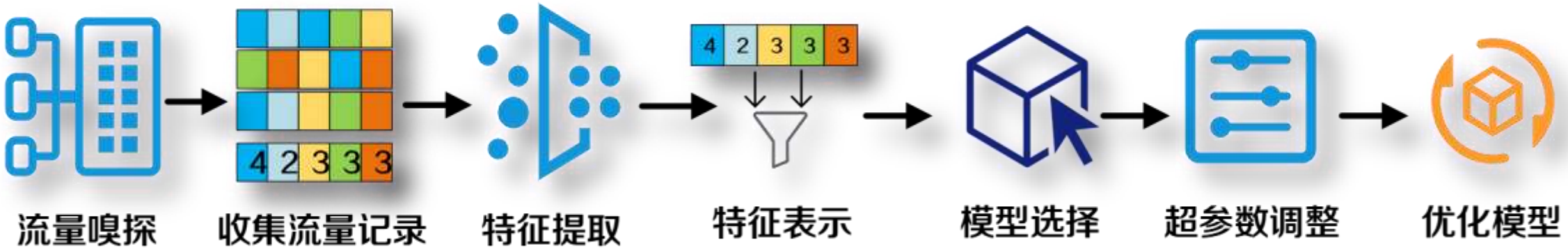
基本概念

- 恶意流量
  - 网络流量中属于“恶意”的部分，包括网络攻击、业务攻击、恶意爬虫等
  - 绝大部分都来自自动化程序，通常通过未经许可的方式侵入、干扰、抓取他方业务或数据
- 恶意流量检测器
  - 主要功能：监视流经设备的网络流量，捕获潜在恶意活动并阻止来自攻击者的恶意流量
  - 主要目标：从规模巨大的网络流量中识别恶意流量



基本概念

- 恶意流量检测基本流程
  - 收集流量（Collecting Flow Records）
  - 特征提取（Feature Extraction）
  - 特征表示（Feature Representation）
  - 模型选择（Model Selection）
  - 超参数调整（Hyperparameter Tuning）
  - 优化模型（Optimized Model）



算法原理

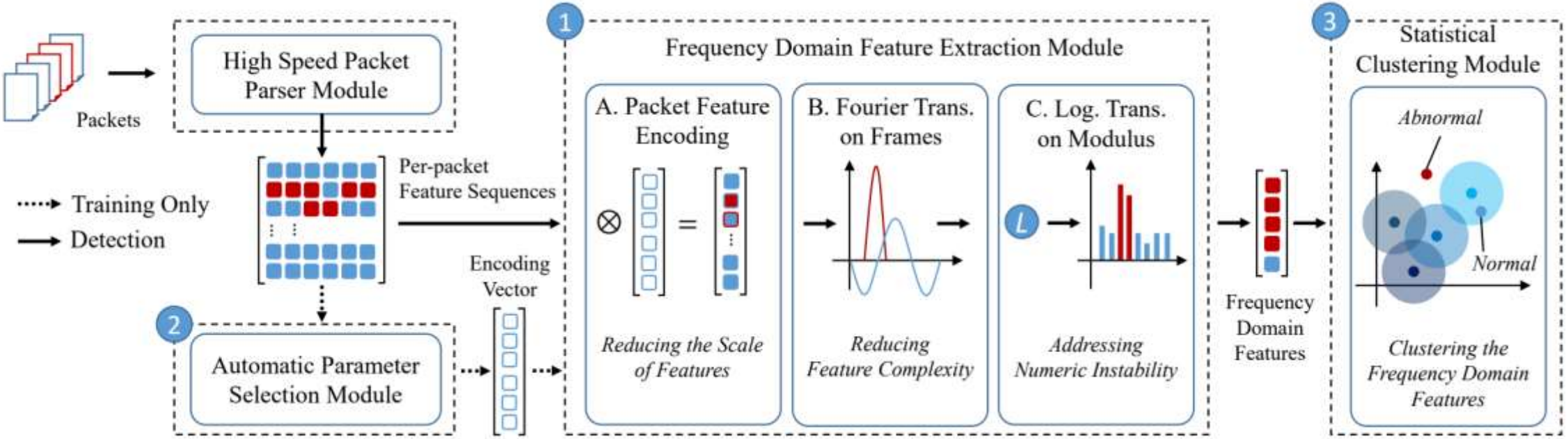
算法原理 Whisper

T	检测恶意流量
I	系统/应用运行的网络流量数据
P	1. 高速数据包解析，提取数据包特征 2. 轮询数据包解析器，提取频域特征 3. 自动参数选择，确定编码向量 4. 特征聚类训练，获取平均训练损失 5. 根据频域特征与聚类中心距离，检测恶意流量
O	待检测流量是/否为恶意流量

P	基于机器学习的实时鲁棒恶意流量检测
C	高速数据包解析
D	恶意流量的鲁棒精准检测以及高吞吐量的实时检测
L	CCF-A会议 (CCS 2021)

算法原理 Whisper

- 总体框架
  - 频域特征提取模块，获取特征并编码，通过频域提取序列信息
  - 自动参数选择模块，计算特征提取模块的编码向量
  - 统计聚类模块，用轻量级的统计聚类算法学习频域模式



算法原理 Whisper

- 频域特征提取模块
  - 特征提取：轮询高速数据包解析器，获取N个数据包的特征
  - 包特征编码：减少特征规模，降低处理开销

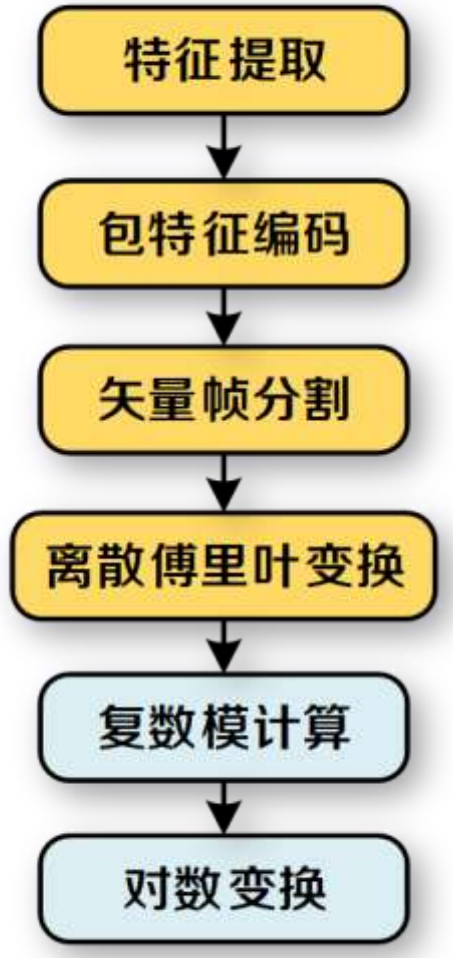
$$v = Sw = \left[v_1, \ldots, v_i, \ldots, v_N\right]^{\mathrm{T}}, v_i = \sum_{k=1}^{M} s_{ik} w_k$$

  - 矢量帧分割：限制数据包之间的长期依赖性

$$f_i = v\left[\!\left[(i-1)\times W_{seg} : i\times W_{seg}\right]\!\right] \quad \left(1 \le i \le N_f\right), \quad N_f = \left[\frac{N}{W_{seg}}\right]$$

  - 离散傅里叶变换：通过频域提取序列信息，减少信息损失

$$F_{ik} = \sum_{n=1}^{W_{seg}} f_{in} e^{-j\frac{2\pi(n-1)(k-1)}{W_{seg}}} \qquad (1 \le k \le W_{seg})$$



算法原理 Whisper

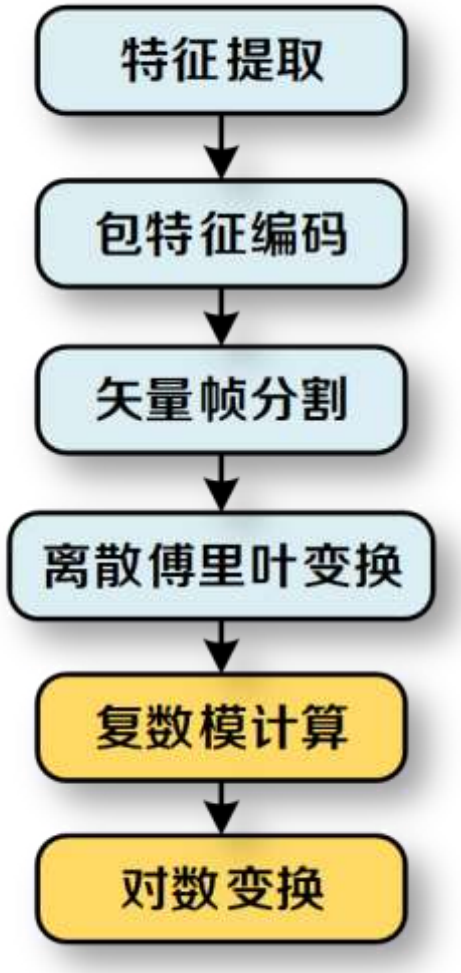
- 频域特征提取模块
  - 复数模计算：将频域表示的复数特征，转换为实数

$$F_{ik} = a_{ik} + jb_{ik}$$

$$\begin{cases} a_{ik} = \sum_{n=1}^{W_{seg}} f_{in} \cos \dfrac{2\pi(n-1)(k-1)}{W_{seg}} \\ b_{ik} = \sum_{n=1}^{W_{seg}} -f_{in} \sin \dfrac{2\pi(n-1)(k-1)}{W_{seg}} \end{cases} \Rightarrow \begin{cases} p_{ik} = a_{ik}^2 + b_{ik}^2 (1 \le k \le W_{seg}) \\ P_i = \left[p_{i1}, \ldots, p_{iK_f} \right]^T \left(K_f = \left[\dfrac{W_{seg}}{2} \right] + 1\right) \\ F_{ik} = F^*_{i(W_{seg}-k)} \Rightarrow p_{ik} = p_i(W_{seg} - k) \end{cases}$$

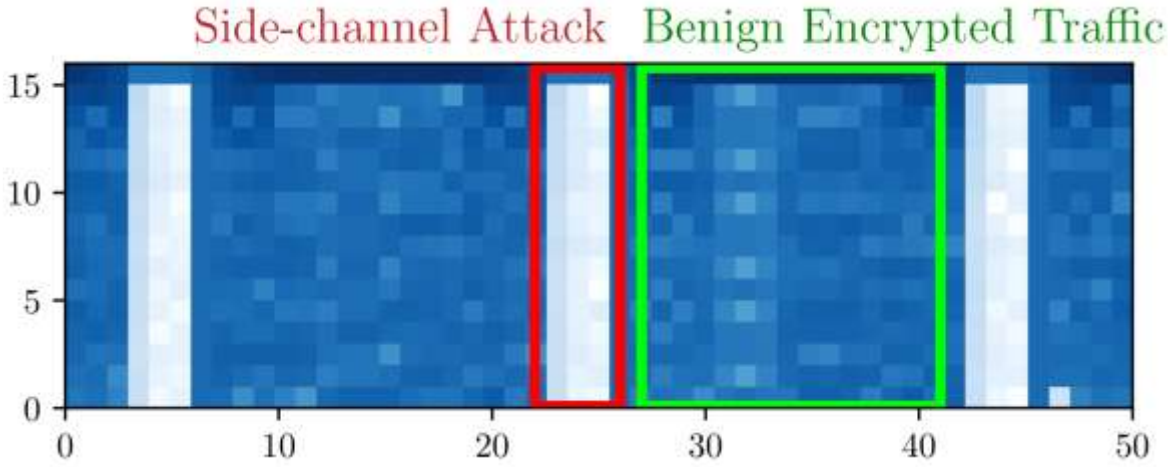
  - 对数变换：稳定频域特征数值，防止机器学习模型训练期间浮点溢出

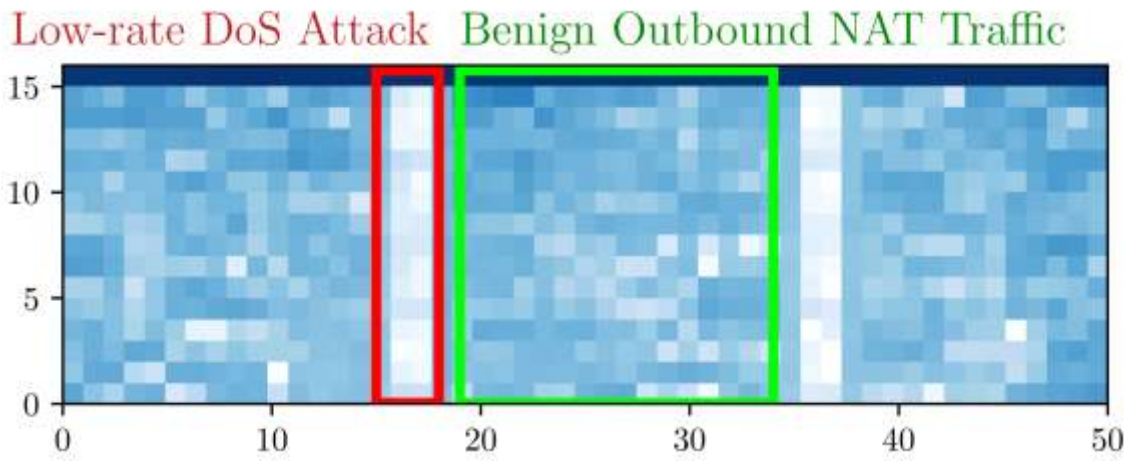
$$R_i = \frac{\ln(P_i + 1)}{C} (1 \le i \le N_f)$$

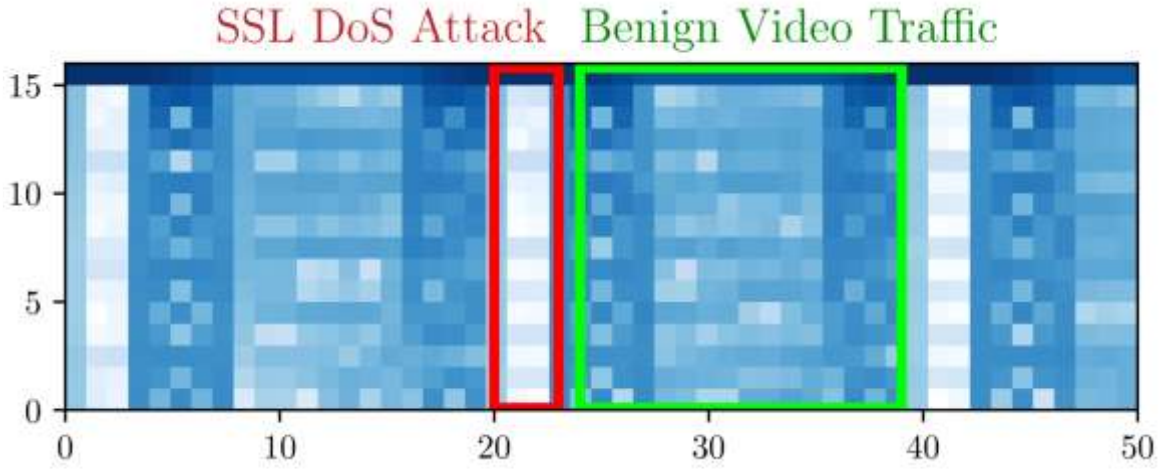


算法原理 Whisper

- 频域特征提取模块
  - 将具有三种典型恶意流量中提取的频域特征映射到RGB空间，并观察到少量恶意流量数据包的频域特征发生了显著变化（恶意流量攻击颜色更浅）







算法原理 Whisper

- 自动参数选择模块
  - 计算特征提取模块的编码向量，通过解决一个约束优化问题确定编码向量w，用来减少不同数据包特征之间的相互干扰
  - 通过求解可满足性模理论（SMT）问题，逼近原问题最优解
  - 减少手动选择参数的工作量
  - 可以在检测阶段准确配置编码向量

$$\tilde{w} = \arg\max \sum_{k=1}^{N} w_M n_{Mk} - w_1 n_{1k} - \sum_{i=2}^{M-1} 2w_i n_{ik} - w_{i-1} h_{(i-1)k} - w_{i+1} n_{(i+1)k}$$

$$\begin{cases} w_i \in [W_{\min}, W_{\max}] \\ \sum_{i=1}^{M} w_i n_{ik} \le B \\ w_i n_{ik} \le w_{i+1} n_{(i+1)k} \\ 2w_i n_{ik} \le w_{i-1} n_{(i-1)k} + w_{i+1} n_{(i+1)k} \end{cases}$$

算法原理 Whisper

- 统计聚类模块
  - 利用统计聚类算法学习频域特征模式
  - 通过训练良性流量的聚类算法，提高Whisper的鲁棒性
  - 将频域特征片段的均值作为聚类算法输入

$$r_i = mean(R[\![l : l + W_{win}]\!])$$

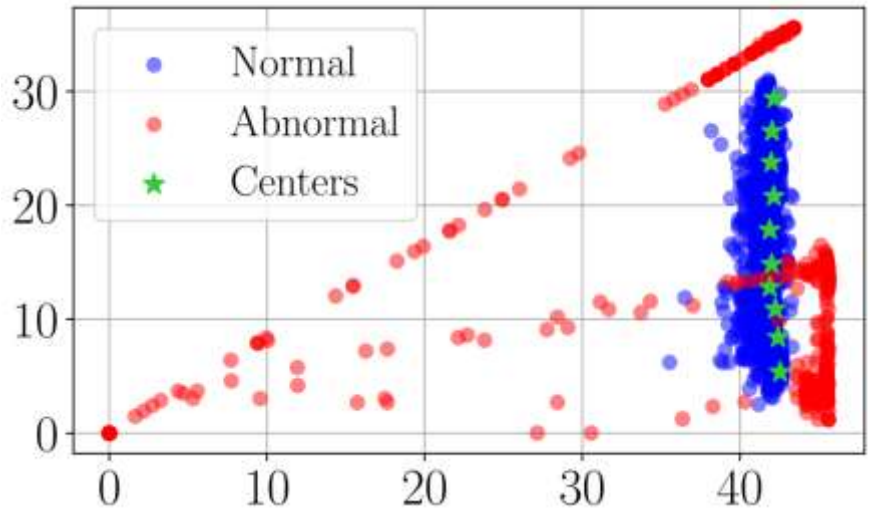
  - 获取良性流量的聚类中心，计算平均训练损失

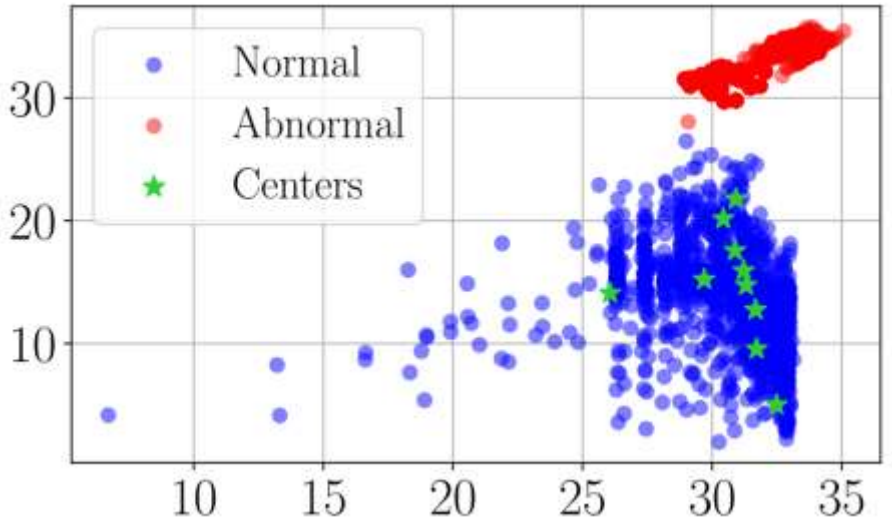
$$\hat{C}_i = \arg\min_{C_k} \left\| C_k - r_i \right\|_2 (1 \le i \le N_t)$$

$$train\_loss = \frac{1}{N_t} \sum_{t=1}^{N_t} \left\| r_i - \hat{C}_i \right\|_2$$

  - 输出检测损失

$$loss_i = \min(\left\| r_i - C_k \right\|_2)(1 \le k \le K_C)$$





- 数据集
  - WIDE MAWI 流量数据集结构

Group	Label	Attack Description	Benign Traffic[1]	Benign Flow Rate	Malicious Flow Rate	Ratio of Malicious[2]
Traditional Attacks	SYN DoS	TCP SYN flooding Deny-of-Service attack.	2020.6.10	5.276 Gbps	23.04 Mbps	0.0858
	Fuzz Scan	Scanning for vulnerabilities in protocols.	2020.6.10	5.276 Gbps	27.92 Mbps	0.0089
	OS Scan	Scanning for active hosts with vulnerable operating systems.	2019.1.2	4.827 Gbps	0.960 Mbps	0.0045
	SSL DoS	SSL renegotiation messages flooding Deny-of-Service attack.	2020.1.1	7.666 Gbps	21.60 Mbps	0.0128
	SSDP DoS	SSDP flooding Deny-of-Service attack.	2020.1.1	7.666 Gbps	27.20 Mbps	0.0321
	UDP DoS	High-rate UDP traffic blocks bottleneck links.	2019.1.2	4.827 Gbps	2.422 Gbps	0.4712
Multi-stage TCP Attacks	IPID SC	Side-channel attack via IPID assignments, disclosed in 2020 [17].	2020.6.10	5.276 Gbps	0.138 Mbps	0.0007
	ACK SC	ACK rate limit side-channel attack, disclosed in 2016 [10].	2019.1.2	4.827 Gbps	1.728 Mbps	0.0091
	TLS Oracle	TLS padding oracle attack [67].	2020.1.1	7.666 Gbps	1.626 Mbps	0.0031
Stealthy TCP Attacks	LRDoS 0.2	UDP burst triggers TCP retransmissions (burst interval 0.2s).	2019.1.2	4.827 Gbps	0.115 Gbps	0.0228
	LRDoS 0.5	UDP burst triggers TCP retransmissions (burst interval 0.5s).	2019.1.2	4.827 Gbps	0.046 Gbps	0.0112
	LRDoS 1.0	UDP burst triggers TCP retransmissions (burst interval 1.0s).	2019.1.2	4.827 Gbps	0.023 Gbps	0.0055
	IPID Scan	Prerequisite scanning of the IPID side-channel attack [17].	2020.6.10	5.276 Gbps	0.214 Mbps	0.0010
	TLS Scan	TLS vulnerabilities scanning [38].	2020.6.10	5.276 Gbps	0.046 Gbps	0.0071

1. 良性流量采集日期　2. 良性流量与恶意流量之比

算法原理 实验结果

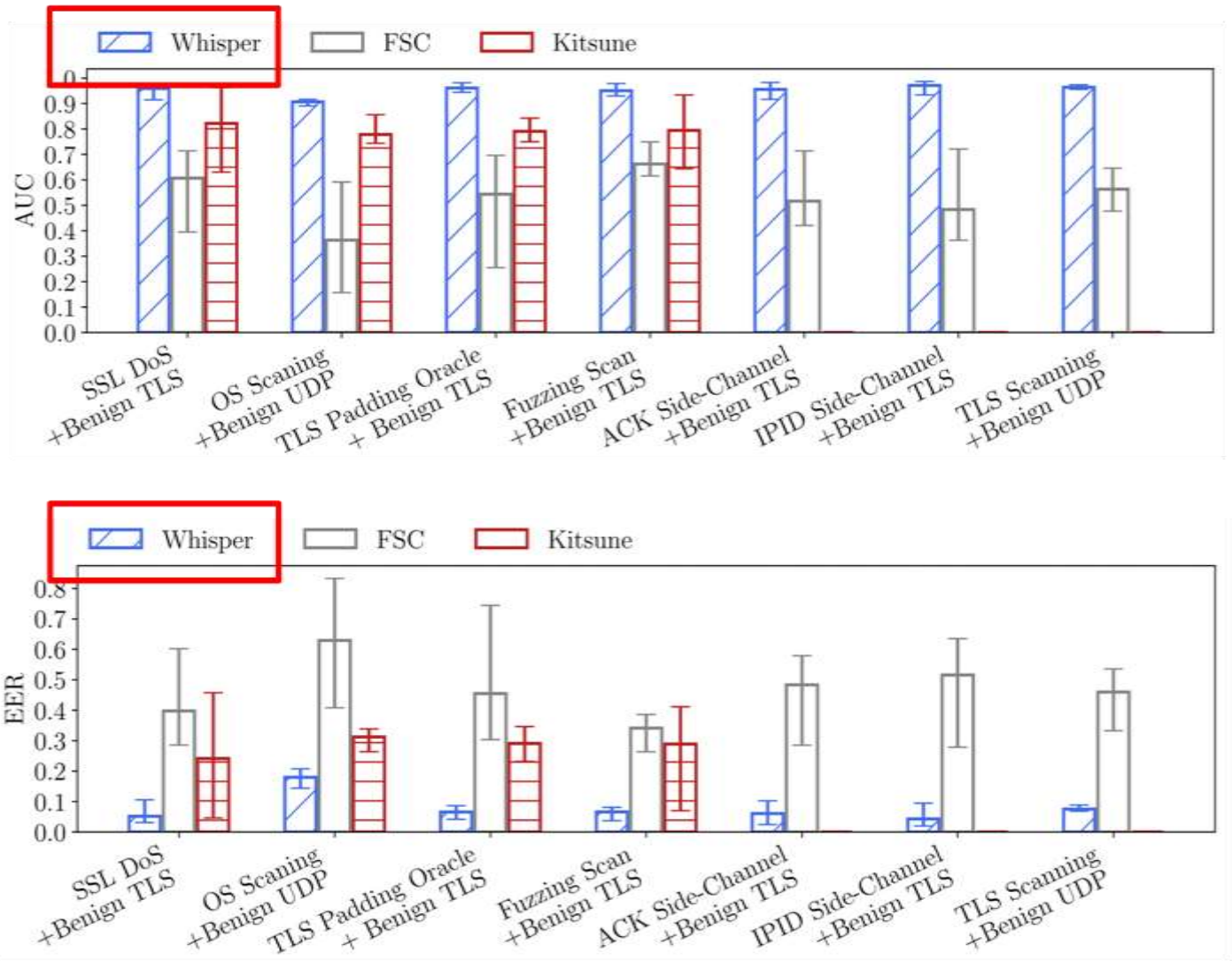
- **Whisper和基线算法检测14种攻击流量的准确率**

Methods	Kitsune				FSC				FAE				Whisper			
Metrics	TPR	FPR	AUC	EER	TPR	FPR	AUC	EER	TPR	FPR	AUC	EER	TPR	FPR	AUC	EER
SYN DoS	0.9801	0.0910	0.9562	0.0919	0.9999	0.0396	0.9603	0.0396	0.9813	0.0033	0.9840	0.0186	0.9924	0.0329	0.9870	0.0512
Fuzz Scan	0.9982	0.0015	0.9978	0.0336	0.0000	0.4007	0.6028	0.3964	0.0000	0.4111	0.6134	0.3954	0.9999	0.0046	0.9962	0.0047
OS Scan	0.9997	0.0786	0.9615	0.0800	0.0000	0.1114	0.8885	0.1114	0.9999	0.0069	0.9907	0.0075	0.9999	0.0106	0.9951	0.0111
SSL DoS	0.9417	0.0035	0.9781	0.0574	0.9992	0.0519	0.9732	0.0519	0.0000	0.1271	0.8774	0.1271	0.9699	0.0796	0.9391	0.0798
SSDP DoS	0.9901	0.0132	0.9955	0.0168	0.9999	0.0014	0.9986	0.0014	0.0003	0.1233	0.8770	0.1233	0.9969	0.0117	0.9902	0.0172
UDP DoS	0.4485	0.1811	0.8993	0.1433	0.9999	0.0173	0.9826	0.0173	0.9999	0.0068	0.9942	0.0071	0.9999	0.0083	0.9922	0.0093
IPID SC	/	/	/	/	0.0000	0.2716	0.7702	0.2716	0.8913	0.1001	0.9739	0.1001	0.6900	0.2324	0.9322	0.2014
ACK SC	/	/	/	/	0.0000	0.3090	0.6909	0.3090	-	-	-	-	0.9999	0.0001	0.9999	0.0001
TLS Oracle	0.9973	0.0335	0.9722	0.0392	-	-	-	-	-	-	-	-	0.9999	0.0121	0.9885	0.0124
LRDoS 0.2	0.6397	0.1270	0.9202	0.1239	0.9999	0.0254	0.9740	0.0254	0.9999	0.0254	0.9925	0.0088	0.9999	0.0109	0.9915	0.0123
LRDoS 0.5	0.0208	0.1882	0.8480	0.1835	0.9999	0.0551	0.9448	0.0551	0.9999	0.0078	0.9925	0.0081	0.9999	0.0101	0.9916	0.0114
LRDoS 1.0	0.0015	0.1774	0.8373	0.1758	0.9999	0.0940	0.9059	0.0940	0.9999	0.0074	0.9935	0.0074	0.9999	0.0115	0.9910	0.0122
IPID Scan	-	-	-	-	0.9999	0.0801	0.9255	0.0801	0.9999	0.0155	0.9934	0.0179	0.7964	0.1601	0.9579	0.1259
TLS Scan	-	-	-	-	-	-	-	-	0.0000	0.4014	0.6033	0.3973	0.9999	0.0091	0.9905	0.0095

绿色——准确率最高 红色——准确率最低 “/”——AUC<0.5 “–”——未在两小时内完成检测

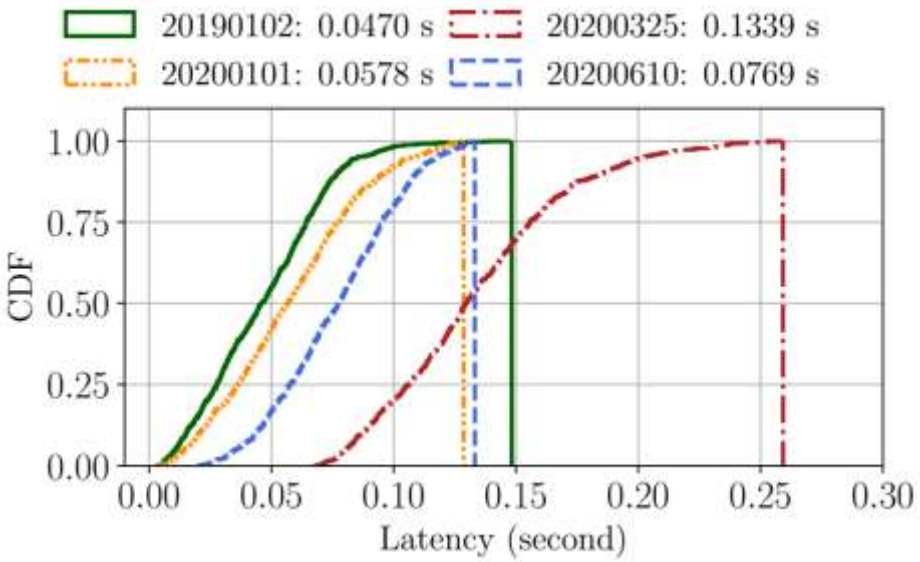
算法原理 实验结果

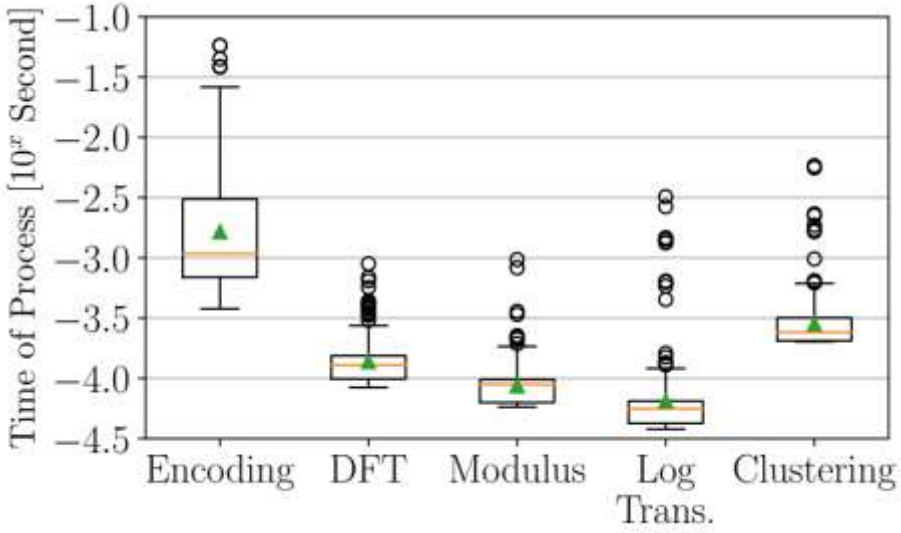
- 不同流量匹配下的平均AUC和等错误概率EER
  - Whisper使用频域特征代表了稳健的细粒度流量序列信息，伪装成良性流量的恶意流量不会引起网络数据流统计数据的显著变化
  - Whisper在不同流量匹配下具有稳定的检测精度



算法原理 实验结果

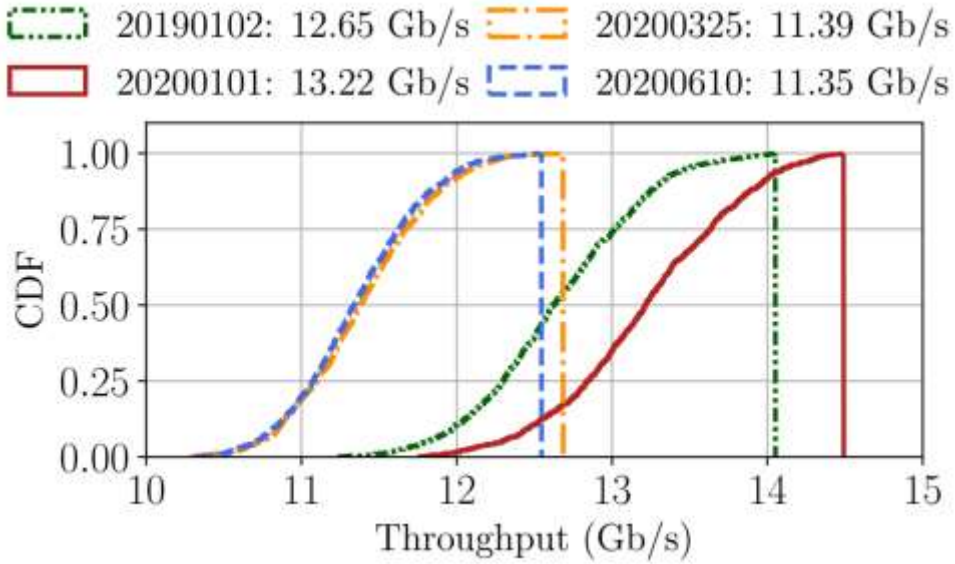
- Whisper的检测延迟
  - 总延迟（处理和排队）：通过四个数据集，Whisper的检测延迟在0.047到0.133秒之间，这表明Whisper在高通量网络中实现实时检测
  - 不同步骤的处理延迟：DFT、模计算和对数变换具有相似的计算复杂度，并产生相似的处理延迟。数据包编码的延迟最大。



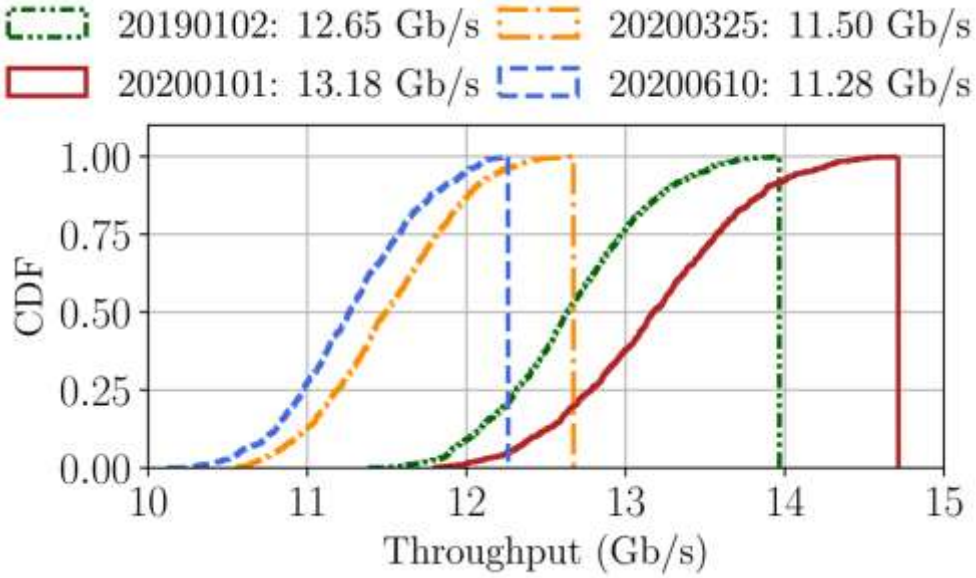


算法原理 实验结果

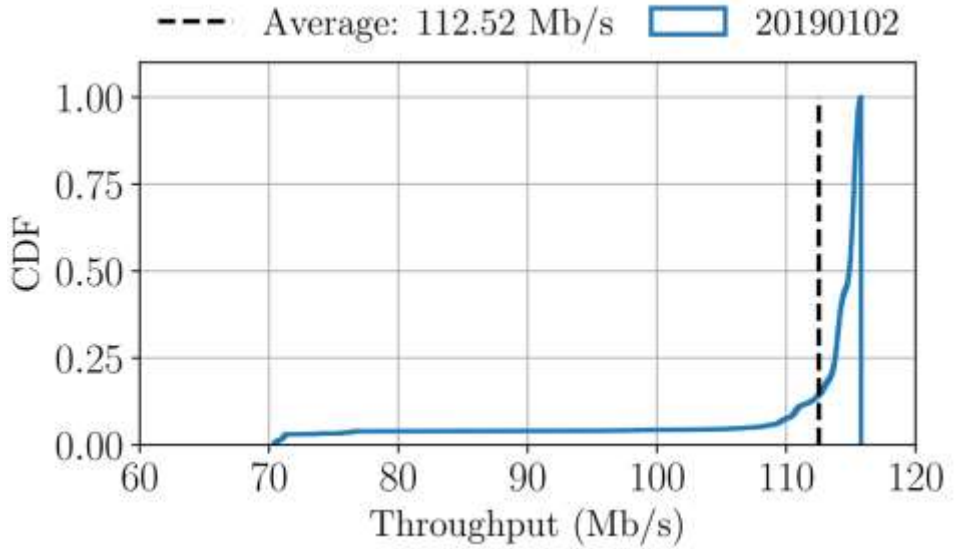
- 累计分布函数CDF和平均吞吐量
  - 四个MAWI主干网流量数据集用来测量吞吐量
  - (1)Whisper、(2)FAE、(3)Kitsune
  - Whisper和FAE的吞吐量最高



(1)



(2)



(3)

算法原理 优劣分析

- 优势：
  - Whisper能够实现高准确率和高吞吐量的检测
  - 细粒度频域特征表示数据包序列的顺序信息，确保了鲁棒检测，并防止攻击者逃避检测
  - 使用轻量级聚类算法实现了高效的攻击检测

- 局限性：
  - Whisper的编码向量的求解计算成本高

算法原理

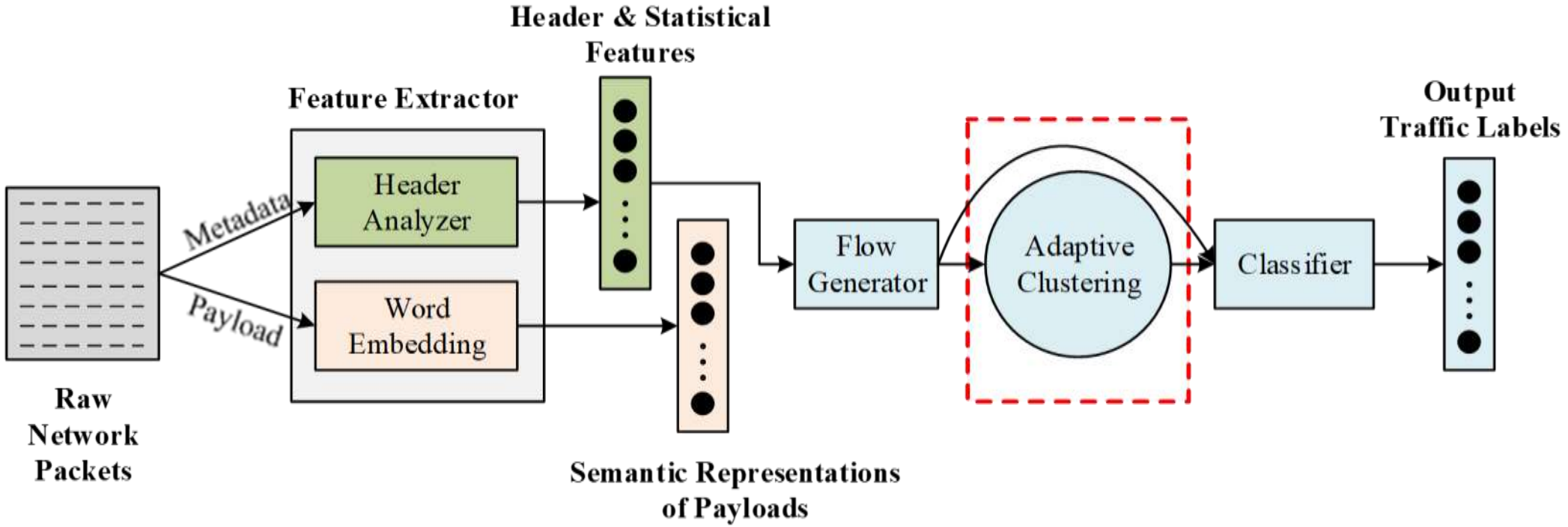
算法原理 ACID

T	恶意流量分类
I	系统/应用运行的网络流量数据
P	1. 特征提取，提取统计特征 2. 自适应聚类，产生聚类中心 3. 恶意流量分类
O	恶意流量类别

P	提高恶意流量的检测准确率，最大程度降低误报率
C	高速数据包解析
D	有效载荷的特征提取成本高
L	CCF-A会议 (INFOCOM 2021)

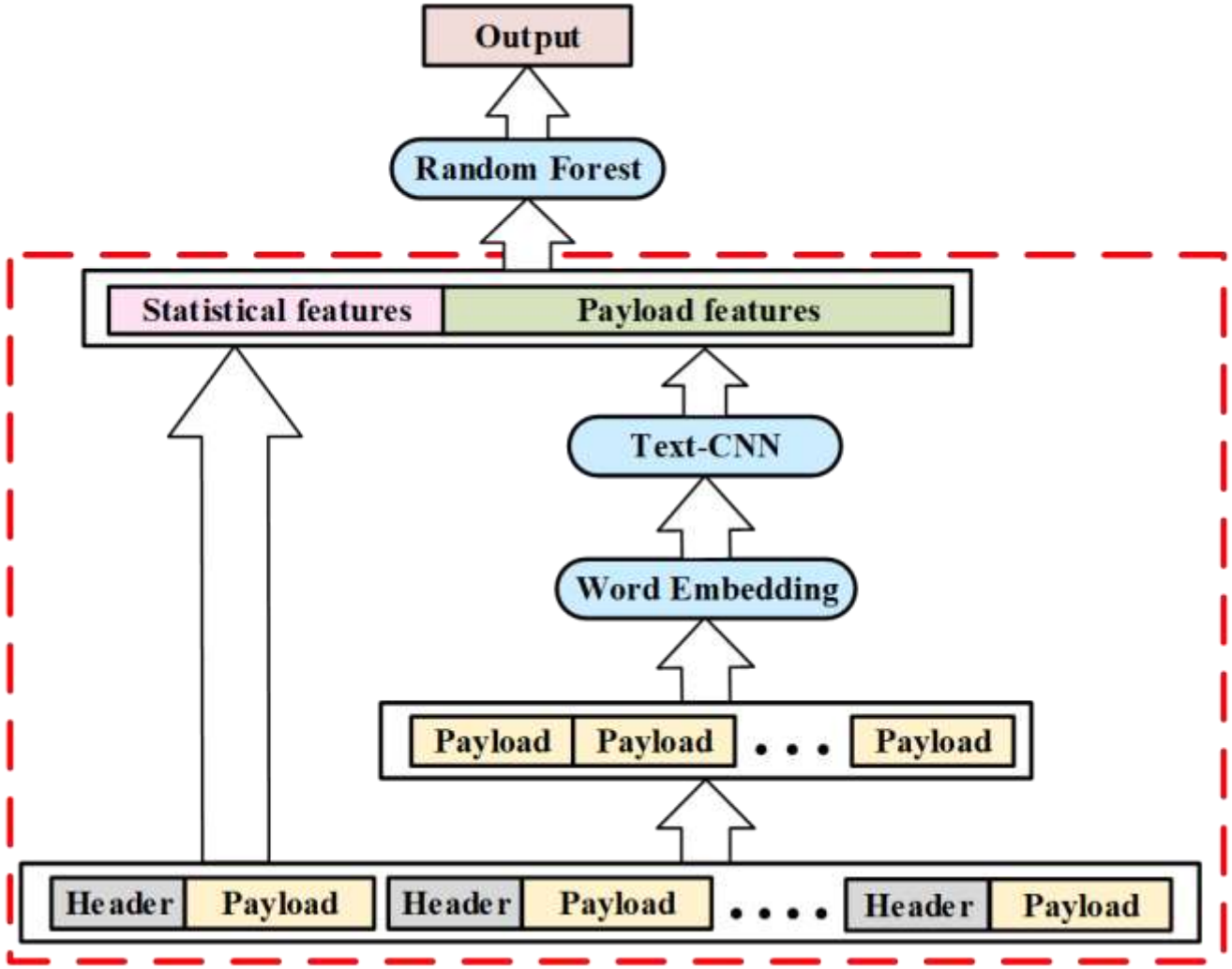
算法原理 ACID

- 总体框架
  - 特征提取器模块
  - 自适应聚类模块
  - 分类模块



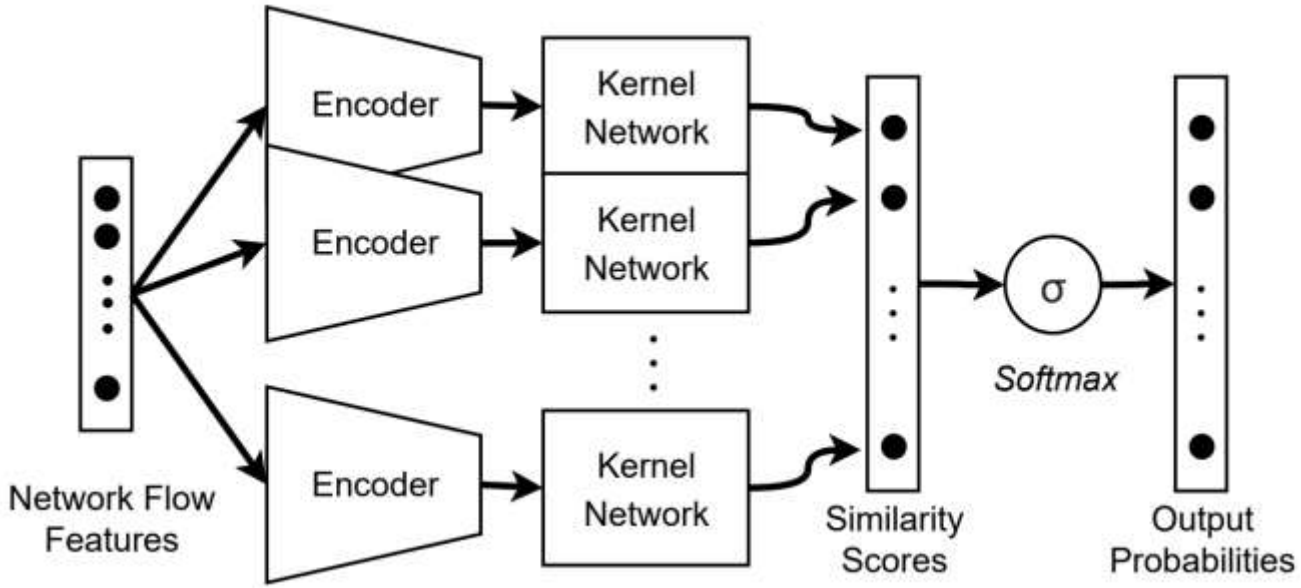
算法原理 ACID

- 特征提取器
  - 报头分析器：构建一组报头和统计特征（包括源/目标端口号、数据包到达时间、网络数据流中的数据包总数等），提供流量行为的紧凑表示
  - 可选的词嵌入：通过word2vec和Text-CNN技术构建有效负载的语义表示向量



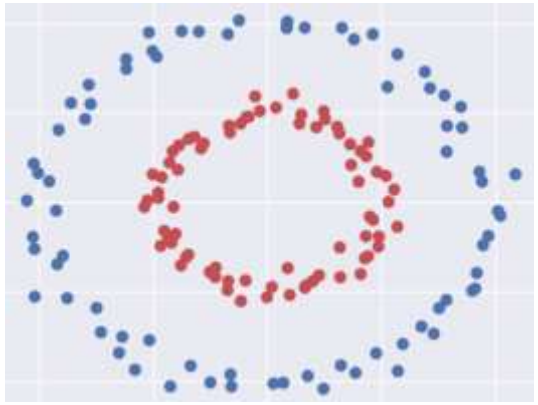
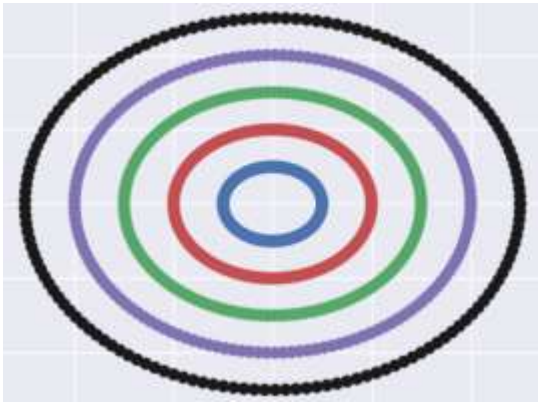
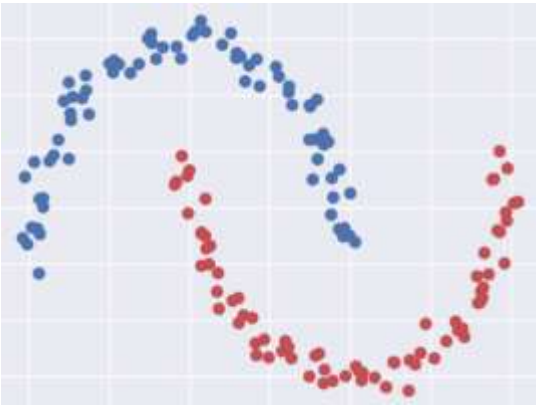
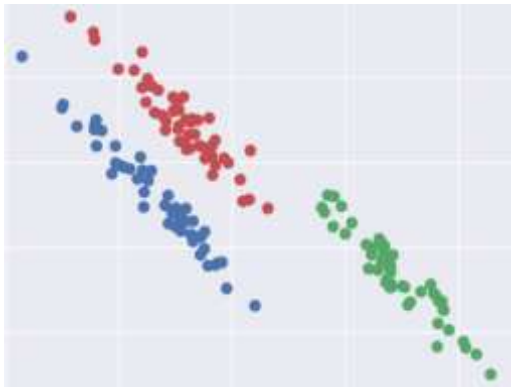
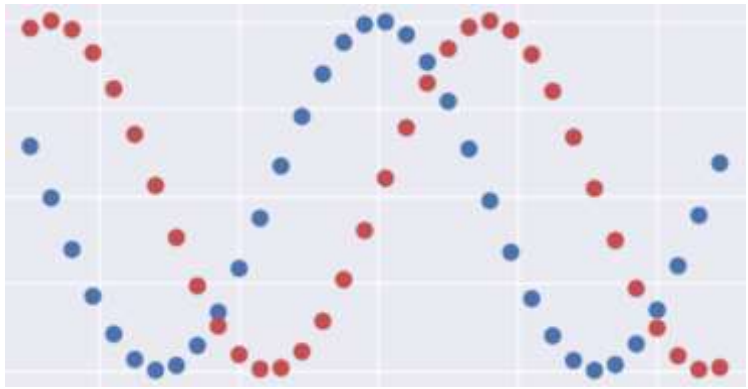
算法原理 ACID

- 自适应聚类网络
  - 采用全连接前馈神经网络，编码器将输入特征的维数降低到期望的维数，使用正弦函数作为激活函数，使网络更快学习并适应复杂的数据结构



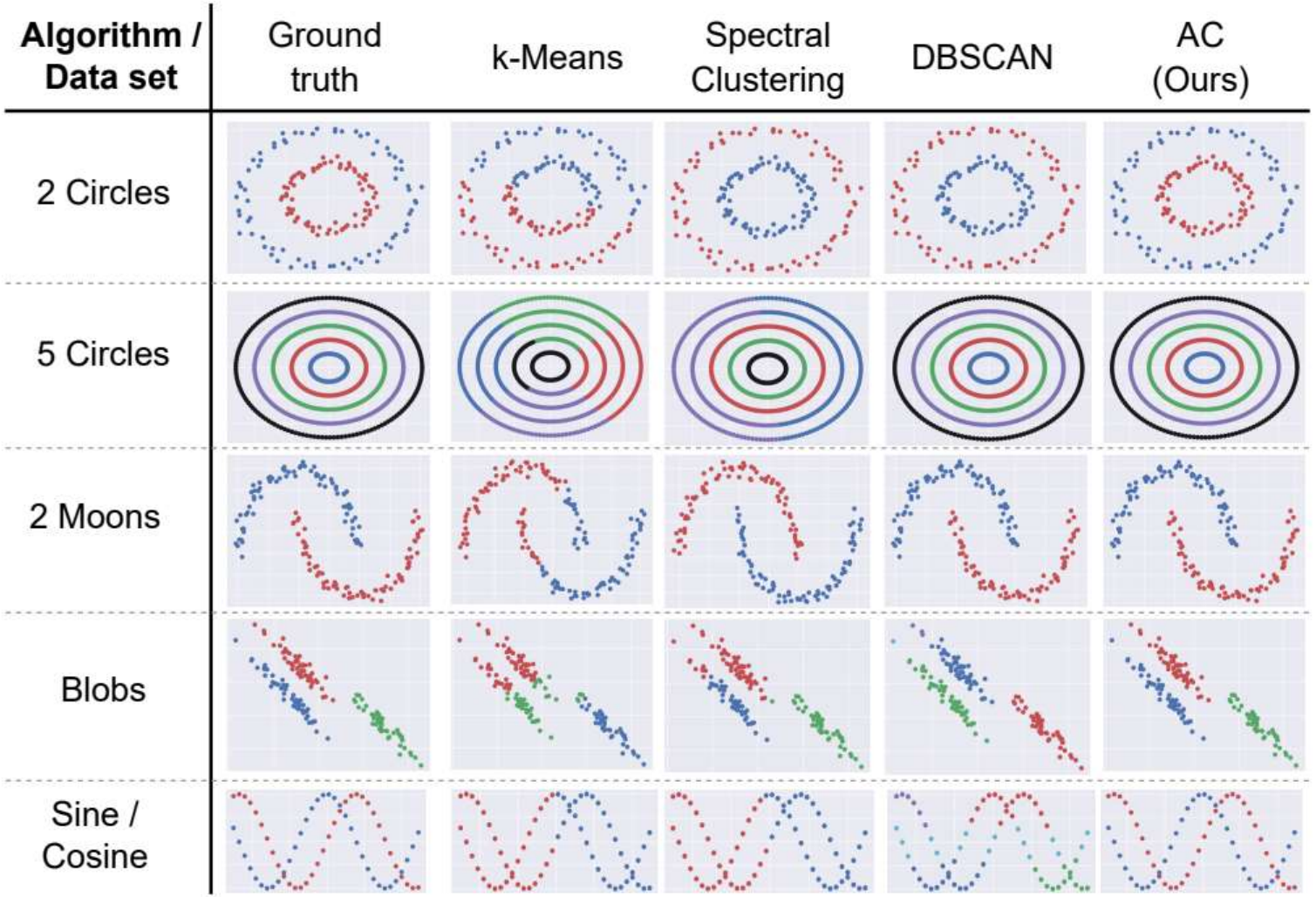
- 分类模块
  - 输入提取特征和聚类中心，输出该数据流所属推断类别
  - 自适应聚类算法减少了同一类样本之间的差异，无论分类器具体情况如何，均可提高分类性能

- 数据集
  - 合成数据集
    - 两个圆
    - 五个圆
    - 双月
    - 高斯分布点
    - 正弦和余弦
  - 入侵检测数据集
    - KDD Cup’99
    - ISCX-IDS 2012
    - CSECIC-IDS 2018

算法原理 实验结果

- 一般性的聚类结果
  - ACID将各类数据集中的数据点进行聚类，而不考虑形状、分布或复杂性。证明使用核网络识别聚类中心，并从分类的角度用这些中心的信息扩充特征集的关键优势



- 多标签分类任务数据集上的性能
  - 在进行二分类和多标签分类时，无论类别的数量如何，ACID都能达到100%的准确率、0%的FAR和100%的F1分数
  - 原因在于ACID对数据的处理方式，第一阶段对应于通过聚类对网络流量进行分类，第二阶段通过分类器进一步纠正错误分类的样本

Metric Dataset	Accuracy (%)	FAR (%)	F_1 (%)	Classes	Samples
KDD CUP'99	100.0	0.00	100.0	23	43,510
ISCX-IDS 2012	100.0	0.00	100.0	5	10,547
CSE-CIC-IDS 2018	100.0	0.00	100.0	15	144,772

- **ISCX-IDS 2012数据集上的性能比较**

Approach	Payload-based Features	Accuracy (%)	FAR (%)	F_1 (%)
DAGMM [13]	No	62.91	30.65	53.07
N-BaIoT [14]	No	89.19	10.80	89.19
Deep NN [15]	No	88.14	7.41	70.35
TR-IDS [24]	Yes	98.88	1.12	98.87
ACID (ours)	No	99.78	0.23	99.44
ACID (ours)	Yes	**100.0**	**0.00**	**100.0**

- 优势：
  - 最大限度地提高了恶意流量的检测准确率，最大限度地降低了误报率
  - 提高了对异常值的鲁棒性和分类模型的泛化能力
  - 降低模型对流量特征敏感的影响

- 局限性：
  - 采用了word2vec和Text CNN构建有效负载的语义表示向量，增加了计算成本

应用总结

应用总结

- 应用
  - 辅助运维/安全人员确定恶意流量类别与来源
  - 恶意流量攻击的趋势分析，有利于网络维护与运营
  - 在军事、医学、交通、物联网等领域的广泛应用

- 未来方向
  - 进一步构建全面、类别平衡、现实的公共数据集，研究所用数据集的质量直接影响其模型的最终性能
  - 研究加密恶意流量检测的泛化模型。随着加密协议和专有协议类型的增长，仅对一种或两种加密协议的加密恶意流量检测必将在未来发挥很小的作用
  - 目前的方法主要是二分类，可以探索多分类以获得更细粒度的检测结果

参考文献

[1] Fu C, Li Q, Shen M, et al. Realtime Robust Malicious Traffic Detection via Frequency Domain Analysis[C]. Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS 2021). 2021: 3431-3446.

[2] Diallo A F, Patras P. Adaptive Clustering-based Malicious Traffic Classification at the Network Edge[C]. IEEE Conference on Computer Communications (INFOCOM 2021). IEEE, 2021: 1-10.

[3] 腾讯安全云鼎实验室. 从恶意流量看2018十大互联网安全趋势[EB/OL]. 2018.

大成若缺，其用不弊。
大盈若冲，其用不穷。
大直若屈。大巧若拙。
大辩若讷。静胜躁，寒
胜热。清静为天下正。

谢谢！