Beijing Forest Studio

北京理工大学信息系统及安全对抗实验中心

基于深度学习的恶意流量检测方法

杨若晗

导师：罗森林

2020年12月27日

内容提要

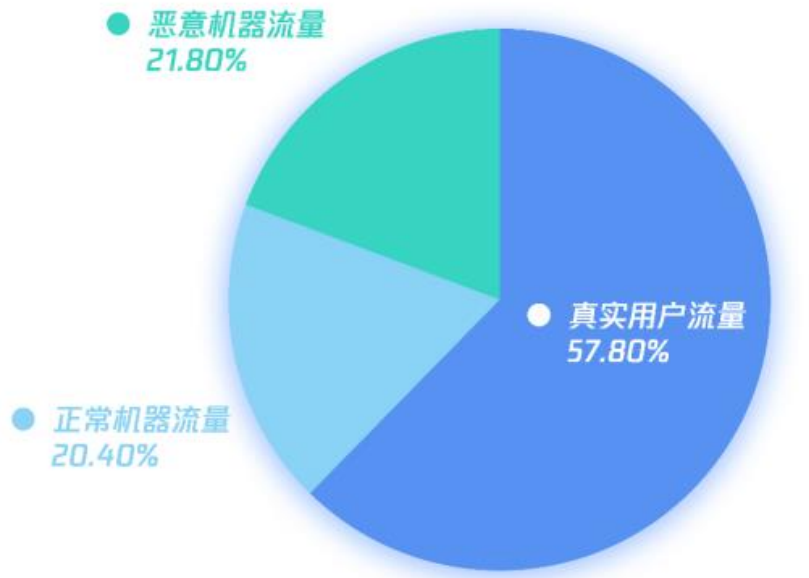
- 背景简介
- 基本概念
- 解决方法
- 算法原理
- 应用总结
- 参考文献

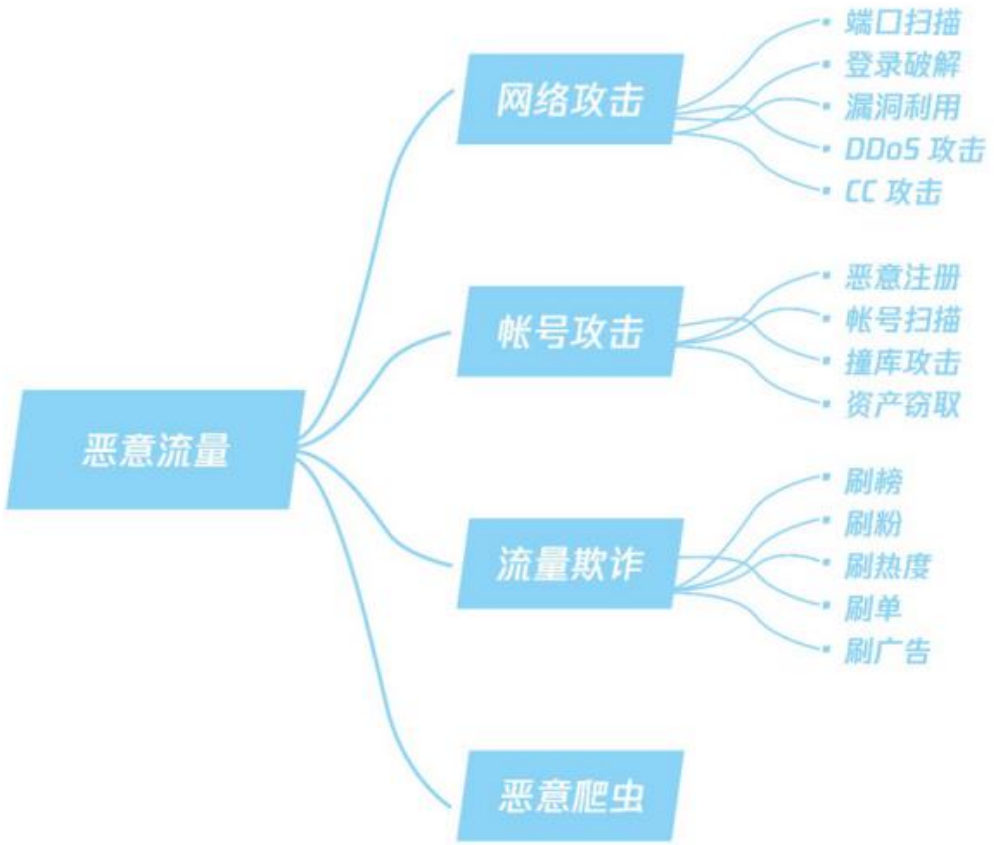
预期收获

- 预期收获
  - 1.了解恶意流量检测概念
  - 2.了解恶意流量检测方法
  - 3.了解基于深度学习的恶意流量检测算法

背景介绍

- 恶意流量检测已经成为网络安全领域常见的问题
- 随着大数据时代的来临，传统的检测方法面临挑战





基本概念

基本概念

- 流量
  - 手机流量：每个月给运营商付费获得若干G上网流量。
  - 网站流量：网站访问量，用来描述一个网站的用户数和页面访问次数。
  - 网络流量：通过特定网络节点的数据包和网络请求数量。
- 恶意流量
  - 定义：在安全领域，主要是指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
  - 来源：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
  - 特点：攻击普遍性、攻击持续性、攻击目的性

- Accuracy（准确率：被正确预测的样本数占所有样本数的比例）

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- Precision（精度：被正确预测为正样本的数量占所有被预测为正样本数的比例）

$$precision = \frac{TP}{TP+FP}$$

- Recall（召回率：被正确预测为正样本的数量占实际为正样本数的比例）

$$recall = \frac{TP}{TP+FN}$$

- F_1 score

$$F_1 = 2 * \frac{precision*recall}{precision+recall}$$

真实情况	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN

- ROC曲线
  - 横轴：假阳性率（False Positive Rate, FPR）

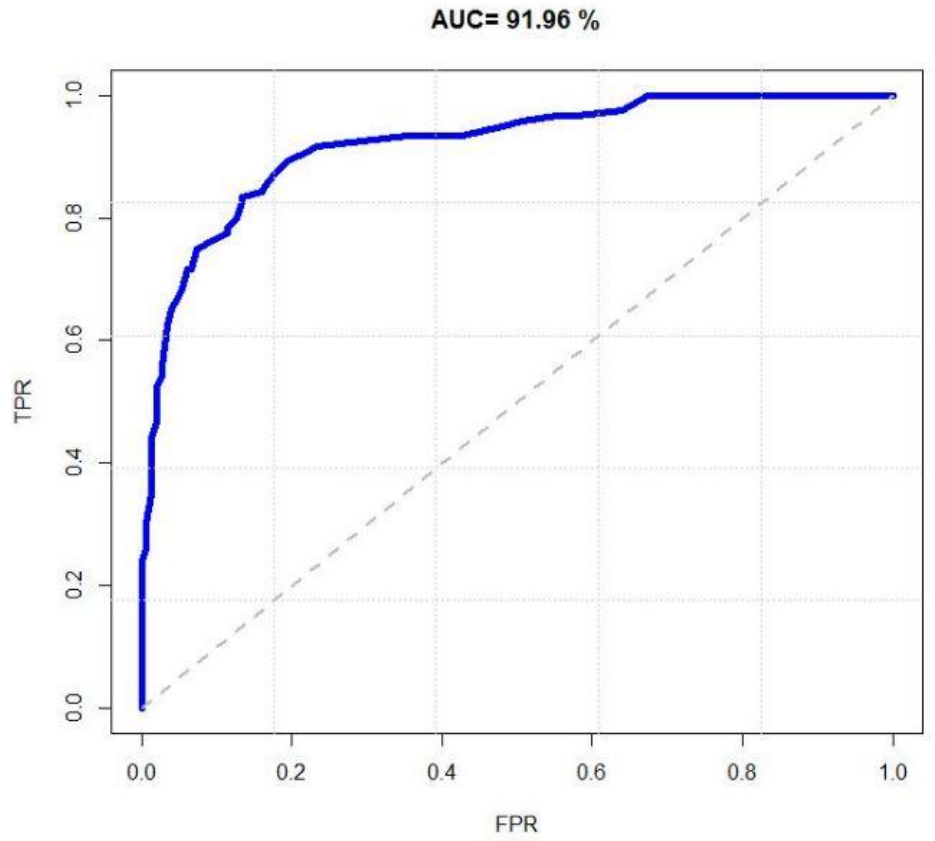
$$FPR = \frac{FP}{TN + FP}$$

  - 纵轴：真阳性率（True Positive Rate, TPR）

$$TPR = \frac{TP}{TP + FN}$$

  - AUC：曲线下面积，值越大，分类效果越好

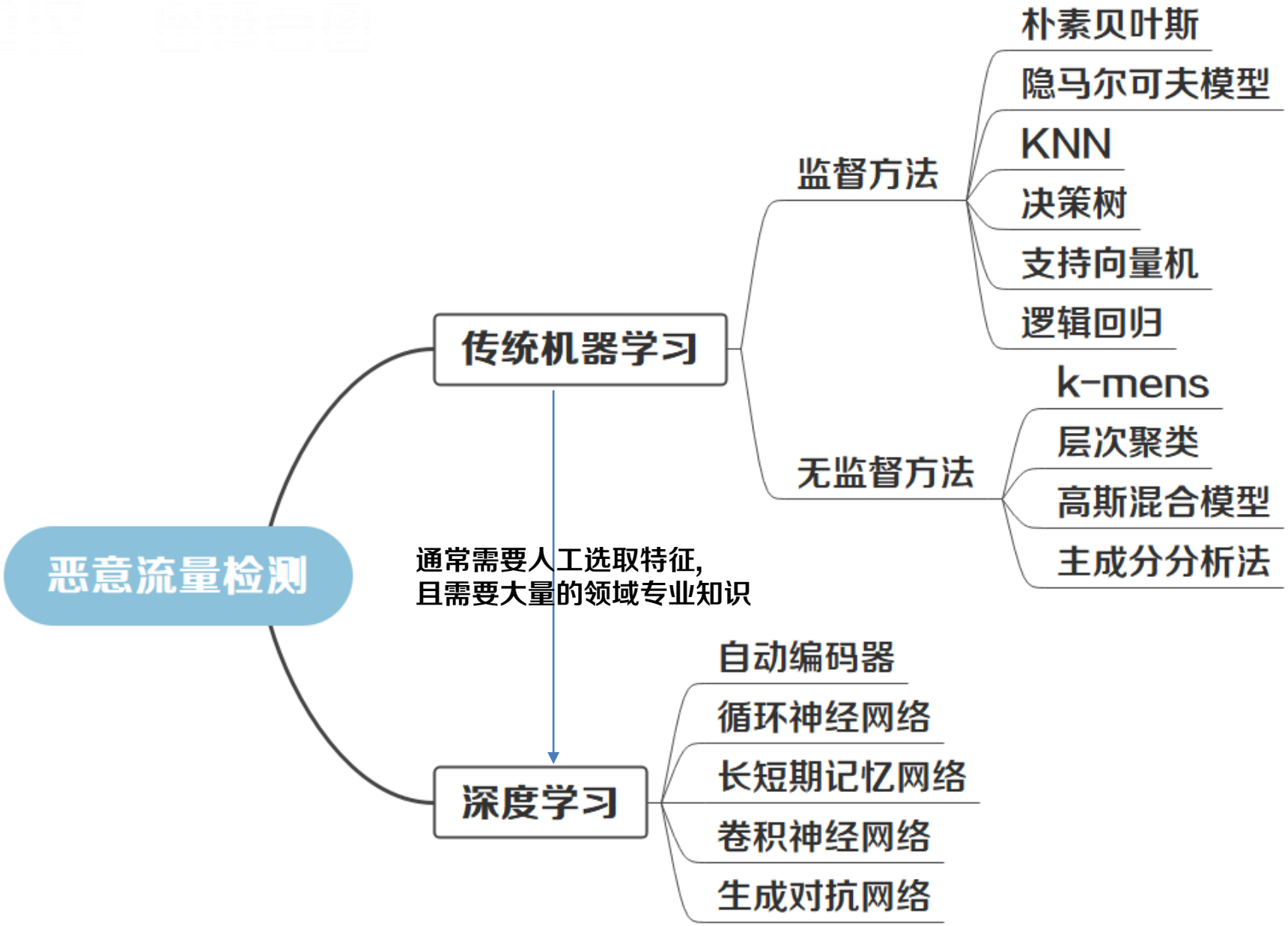
真实情况	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN



恶意流量检测方法

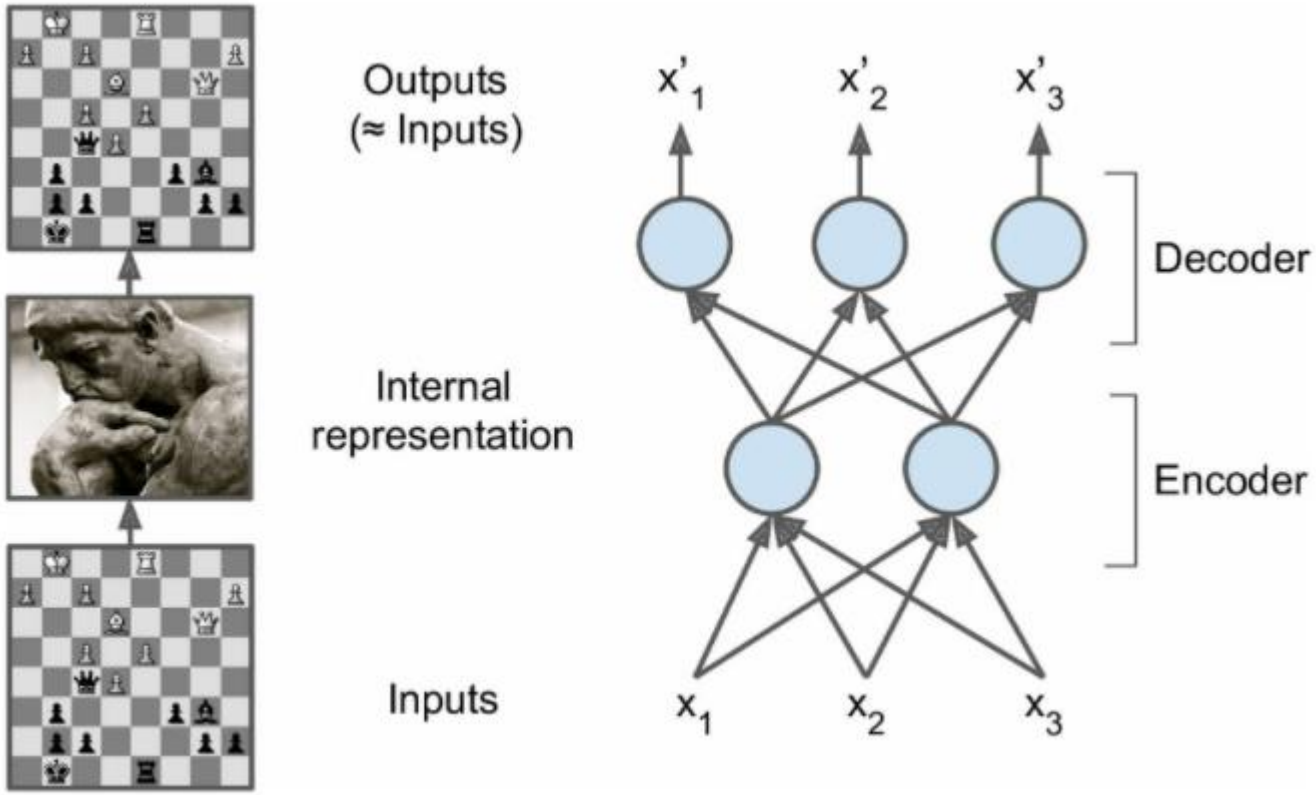
解决方法

解决方法　思维导图


恶意流量检测
传统机器学习
监督方法
朴素贝叶斯
隐马尔可夫模型
KNN
决策树
支持向量机
逻辑回归
无监督方法
k-mens
层次聚类
高斯混合模型
主成分分析法
通常需要人工选取特征,
且需要大量的领域专业知识
深度学习
自动编码器
循环神经网络
长短期记忆网络
卷积神经网络
生成对抗网络

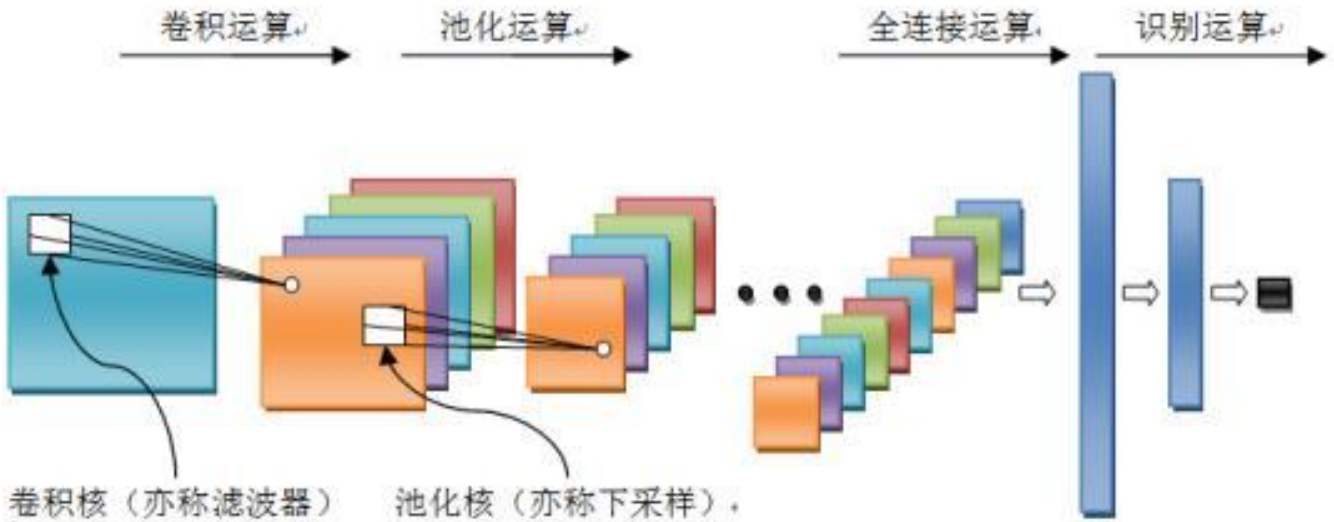
解决方法

- 基于深度学习的方法
  - 自动编码器：由编码器和解码器组成，包含输入层、编码层、解码层
  - 优点：能够表征线性变换和非线性变换，广泛应用于降维任务



解决方法

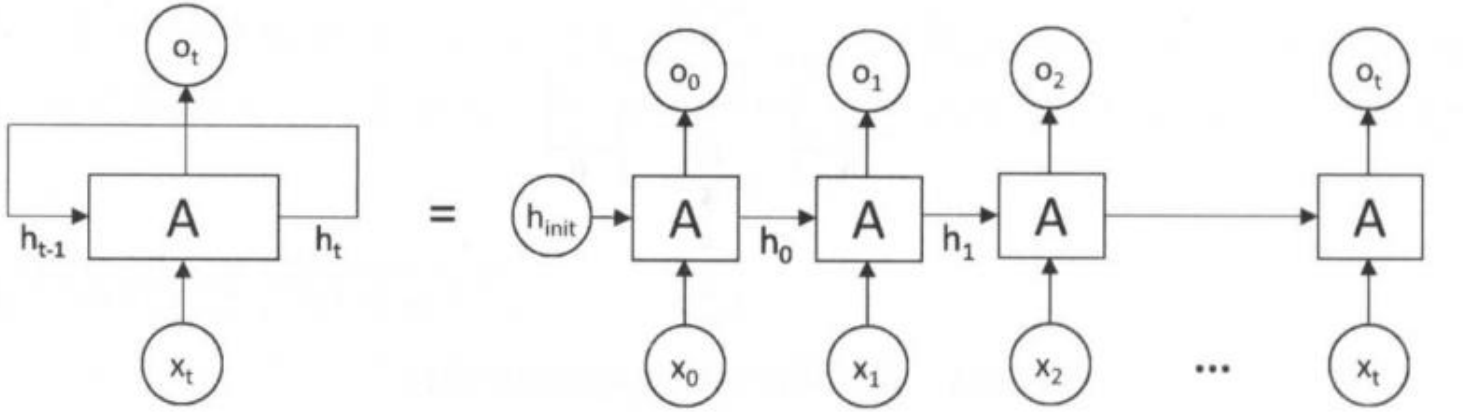
- 基于深度学习的方法
  - 卷积神经网络：包含了输入层、隐藏层（卷积层、池化层、全连接层）、输出层，具有深度结构的前馈神经网络
  - 优点：能够更准确且高效地提取特征，特别当输入是图像时表现更为明显

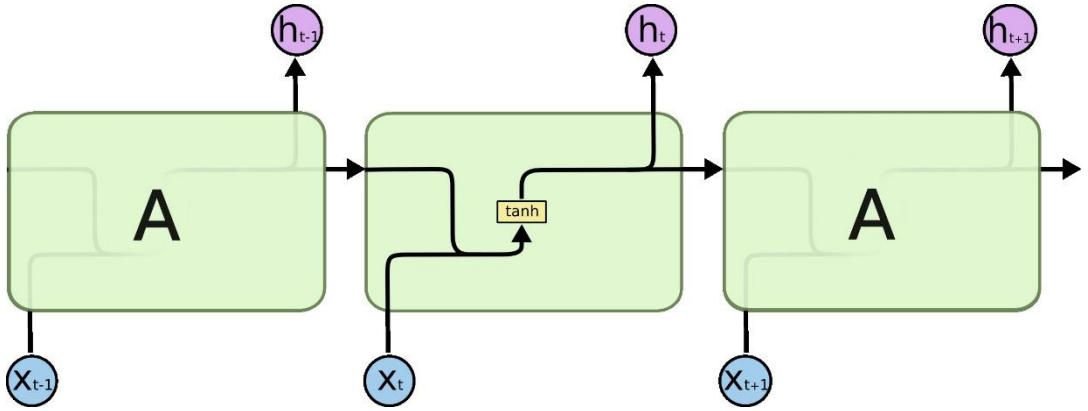


解决方法

- 基于深度学习的方法
  - 循环神经网络：以序列数据为输入，在序列演进方向进行递归
  - 优点：能够深度挖掘数据中的时序信息和语义信息，广泛用于序列相关的任务

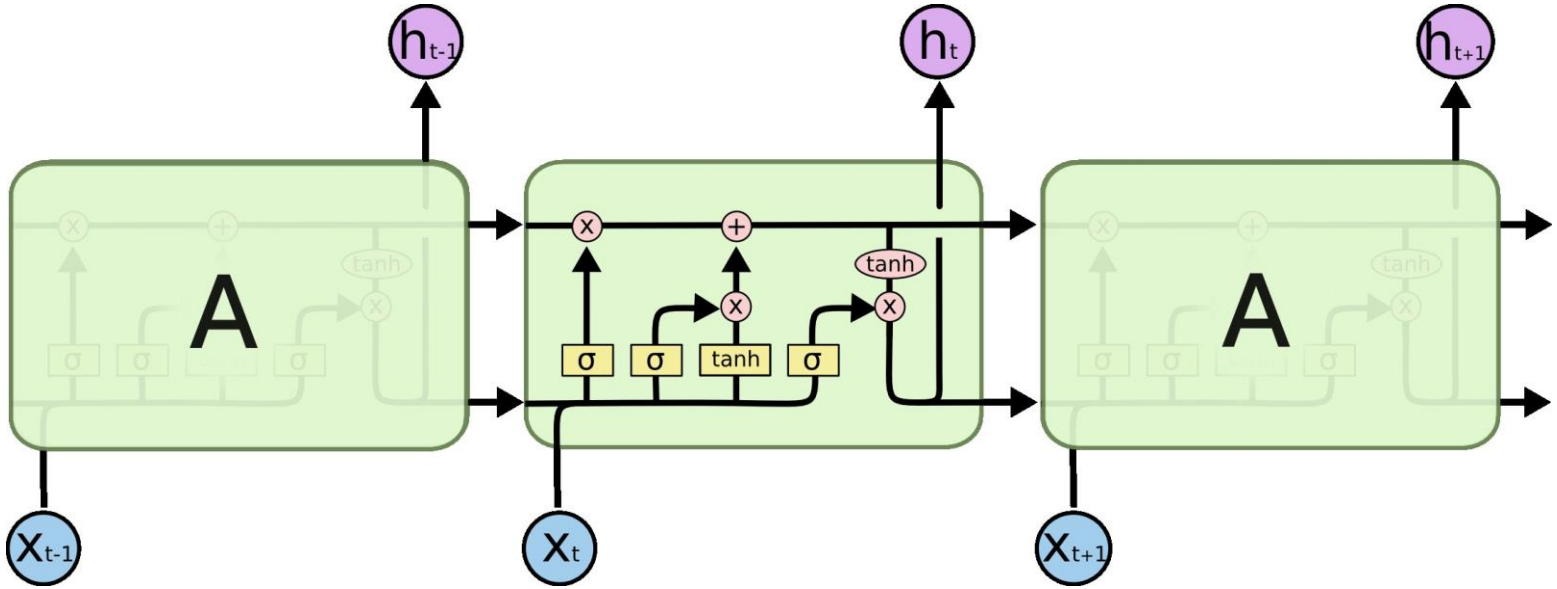
“我→吃→苹果”
n. v. n.





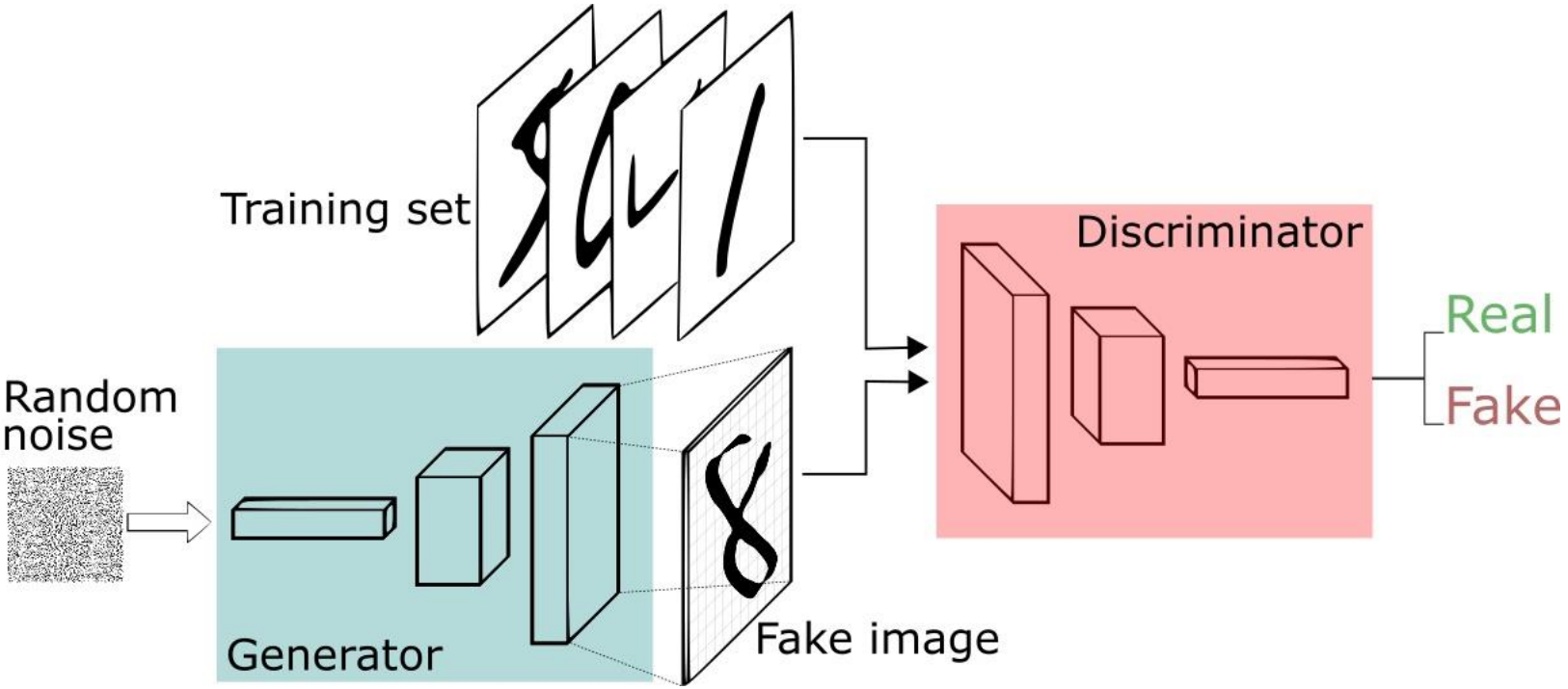
解决方法

- 基于深度学习的方法
  - 长短期记忆网络：是一种特殊的RNN，与一般的RNN区别是中间的更新状态的方式不同。
  - 优点：通过“门”结构来实现对重要内容的保留和对不重要内容的去除



解决方法

- 基于深度学习的方法
  - 生成对抗网络：通过生成模型和判别模型的相互博弈学习产生高质量输出
  - 优点：在处理数据类别不平衡问题上表现良好



算法原理

算法原理

T	检测和分类恶意网络流量
I	基于字节流的原始数据
P	1.按照固定尺寸截取原始数据 2.将处理后的数据输入模型进行训练
O	网络流量的分类

P	需要专家知识构建的特征在不同场景下分类准确率可能不高
C	带有流量类别标签的网络流量数据
D	直接将原始字节流作为输入进行模型训练
L	S&P2019

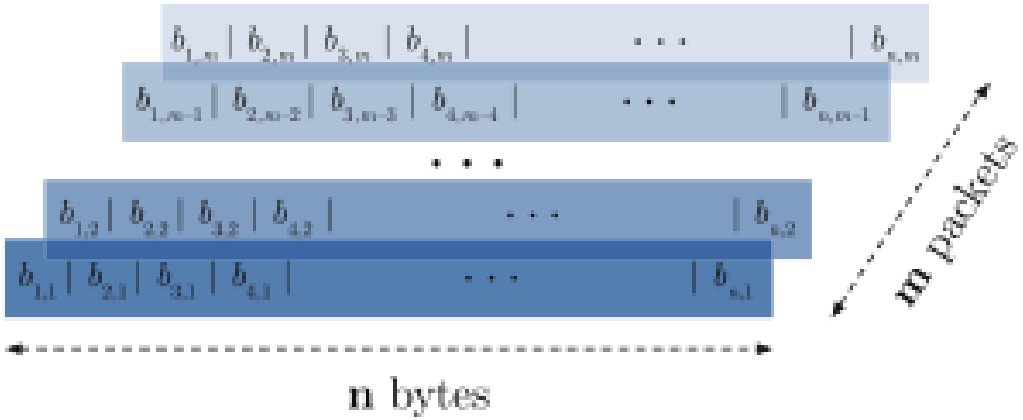
算法原理 输入数据表示

- 数据包表示法
  - 输入数据表示为（N，n）
  - N：数据包数量
  - n：字节数量

- 将每个数据包的有效负载作为分析和构建数据集的关键信息
  - 设定n值，截取有效负载的前n个字节

算法原理 输入数据表示

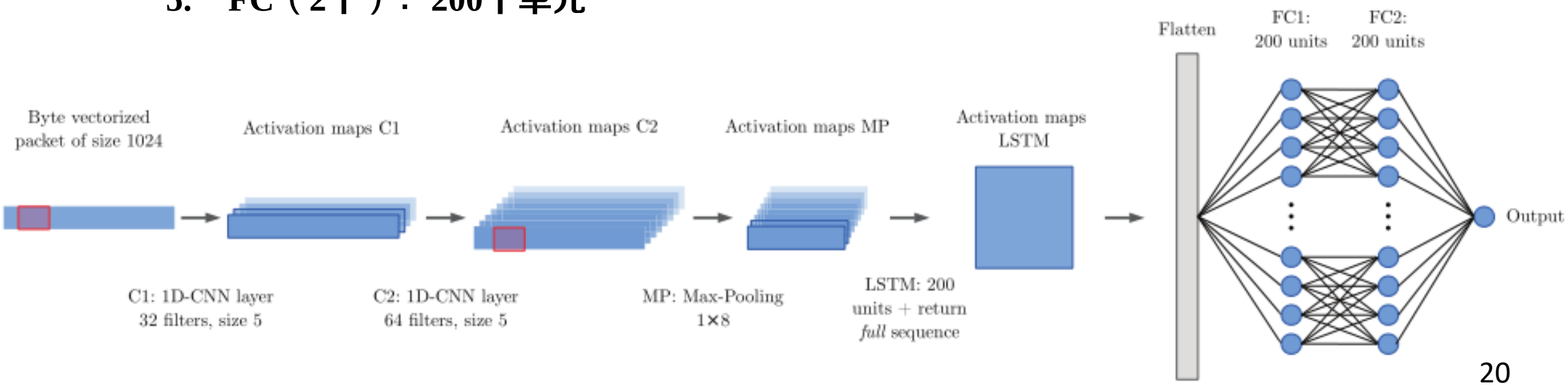
- 数据流表示法
  - 输入数据表示为（N，m，n）
  - N：数据流数量
  - m：数据包数量
  - n：字节数量



- 将每个数据包的有效负载作为分析和构建数据集的关键信息
  - 设定m：截取前m个数据包
  - 设定n：截取有效负载的前n个字节

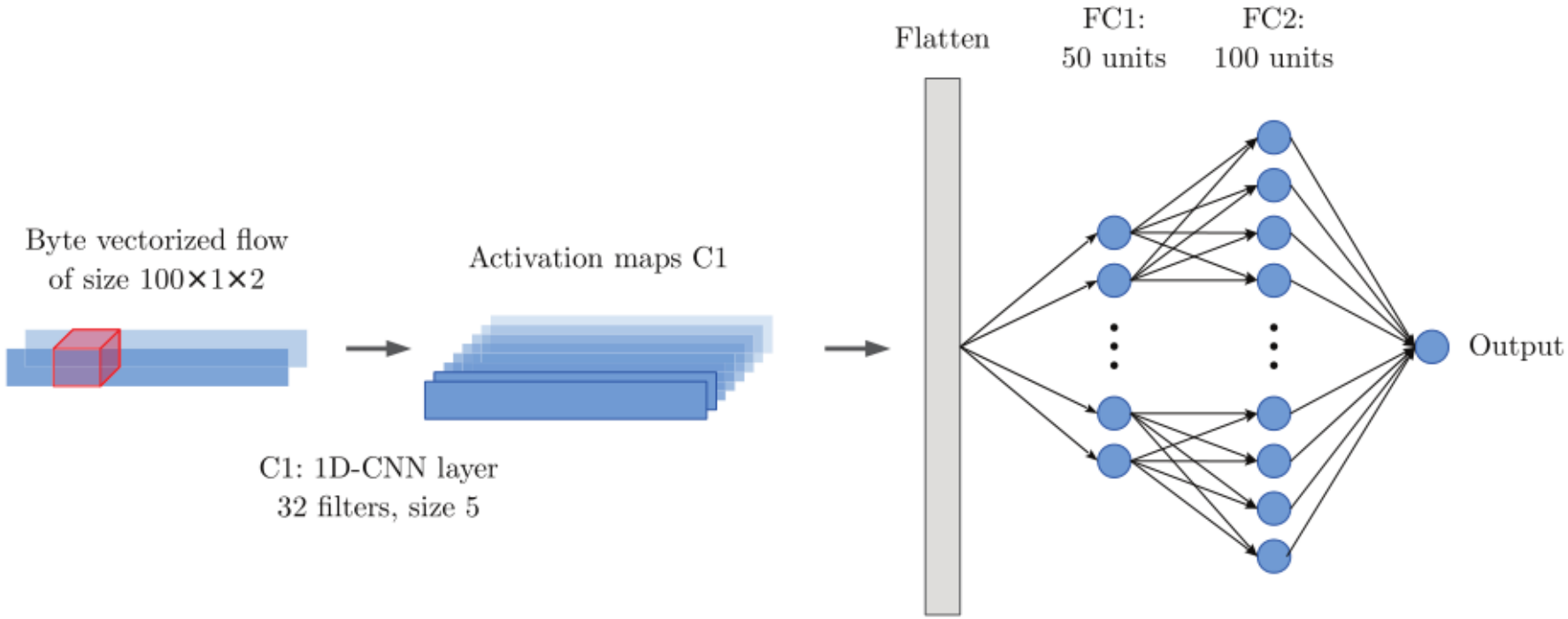
算法原理　算法原理图

- 用原始数据包作为输入的DL算法原理
  - 包含两个1D-CNN层、一个LSTM、两个全连接层
    1. C1：32个大小为5的卷积核
    2. C2：64个大小为5的卷积核
    3. MP：大小为1*8
    4. LSTM：200个单元
    5. FC（2个）：200个单元



算法原理　算法原理图

- 用原始数据流作为输入的DL算法原理
  - 包含一个1D-CNN层、两个全连接层
    1. C1：32个大小为5的卷积核
    2. FC（2个）：分别为50个、100个单元



实验分析

序号	类型	数据集	模型	输入数据
对比实验1	二分类	USTCTFC2016 （异常流量10种，良性流量10种）	DL	原始数据包
			RF	
对比实验2	二分类	扩充的USTCTFC2016 （异常流量10种，良性流量16种）	DL	原始数据包
			RF	
			DL	原始数据流
			RF	
对比实验3	二分类	扩充的USTCTFC2016 （异常流量10种，良性流量16种）	DL	原始数据流
			RF	自己构建的具有200维特征的数据
对比实验4	四分类	Neris、Rbot、Virut、良性流量 各40000个样本	DL	原始数据包
			RF	

- 数据集（ USTCTFC2016 ）
  - 简介：该数据集由两组来自恶意软件和良性应用程序的标记pcap文件组成

TABLE I. USTC-TFC2016 PART I (MALWARE TRAFFIC)

Name	CTU num	Binary MD5	process
Cridex	108-1	25b8631afeea279ac00b2da70fffe18a	original
Geodo	119-2	306573e52008779a0801a25fafb18101	part
Htbot	110-1	e515267ba19417974a63b51e4f7dd9e9	original
Miuref	127-1	a41d395286deb113e17bd3f4b69ec182	original
Neris	42,43	bf08e6b02e00d2bc6dd493e93e69872f	merged
Nsis-ay	53	eaf85db9898d3c9101fd5fcfa4ac80e4	original
Shifu	142-1	b9bc3f1b2aace824482c10ffa422f78b	part
Tinba	150-1	e9718e38e35ca31c6bc0281cb4ecfae8	part
Virut	54	85f9a5247afbe51e64794193f1dd72eb	original
Zeus	116-2	8df6603d7cbc2fd5862b14377582d46	original

恶意流量：从真实的网络环境中收集到的10种恶意软件流量

TABLE II. USTC-TFC2016 PART II (NORMAL TRAFFIC)

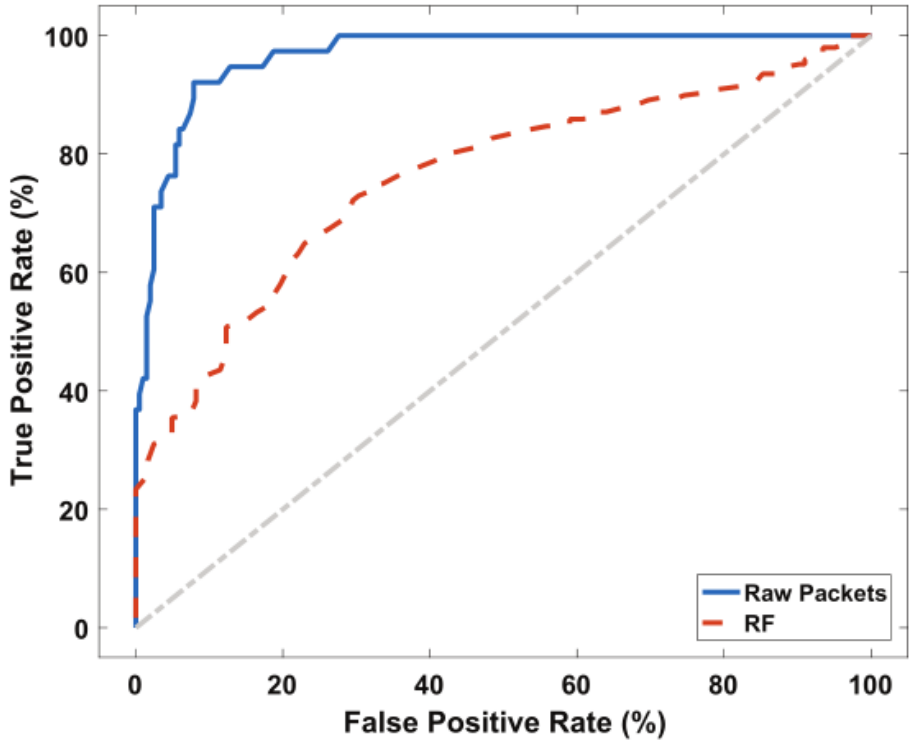
Name	Class	Name	Class
BitTorrent	P2P	Outlook	Email/WebMail
Facetime	Voice/Video	Skype	Chat/IM
FTP	Data Transfer	SMB	Data Transfer
Gmail	Email/WebMail	Weibo	Social NetWork
MySQL	Database	WorldOfWarcraft	Game

良性流量：使用网络流量虚拟平台Ixia Breaking Point采集的10种正常流量

实验分析　对比实验1

序号	类型	数据集	模型	输入数据
对比实验1	二分类	USTCTFC2016	DL RF	原始数据包

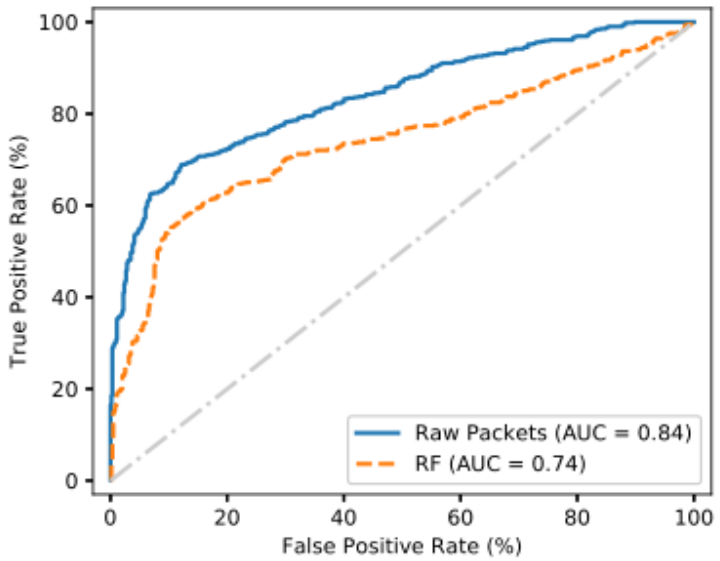
- 输入数据处理
  - 打标签：将来自pcap文件的每个数据包打标签为“良性”或“恶意”
  - 数据集划分：比例为训练集（80%）、验证集（10%）、测试集（10%）
  - 设n=1024
- 实验结果
  - 可以检测到70%以上的恶意软件流量数据包
  - 误报率低于3%
  - ROC对比曲线见右图



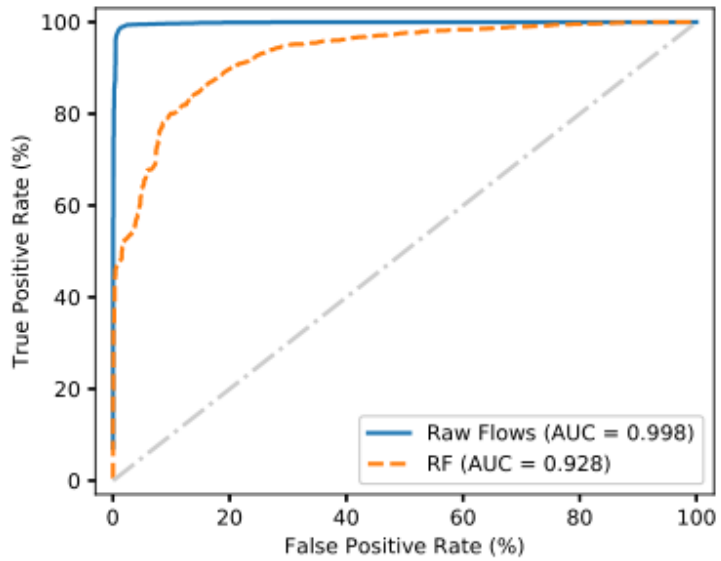
实验分析　对比实验2

序号	类型	数据集	模型	输入数据
对比实验2	二分类	扩充的 USTCTFC2016	DL	原始数据包
			RF	
			DL	原始数据流
			RF	

- 输入数据处理
  - 原始数据
  - 原始数据
- 实验结果



(a) *Raw Packets* representation.

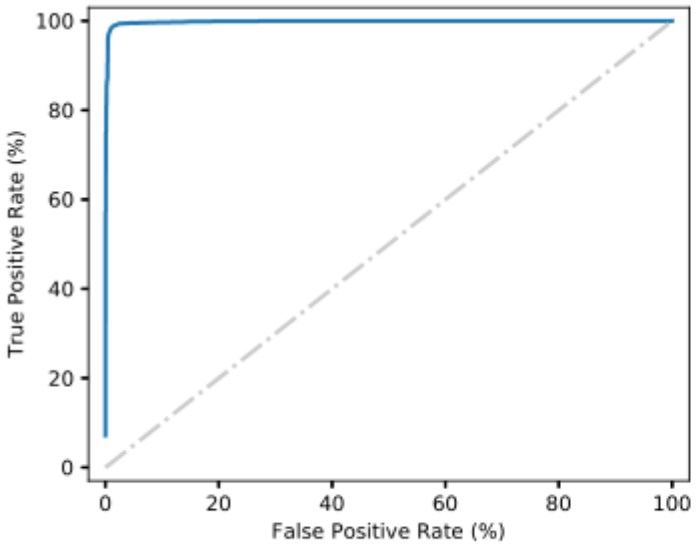


(b) *Raw Flows* representation.

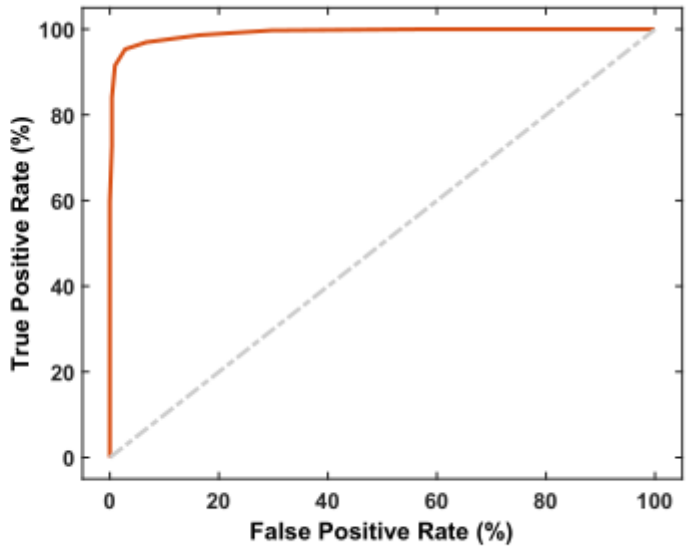
实验分析　对比实验3

序号	类型	数据集	模型	输入数据
对比实验3	二分类	扩充的 USTCTFC2016	DL	原始数据流
			RF	自己构建的具有200维特征的数据

- 输入数据处理
  - 原始数据流的处理：同对比实验2
  - RF模型输入数据：自己构建的具有近200维特征的数据
- 实验结果



(a) *Raw Flows* representation.



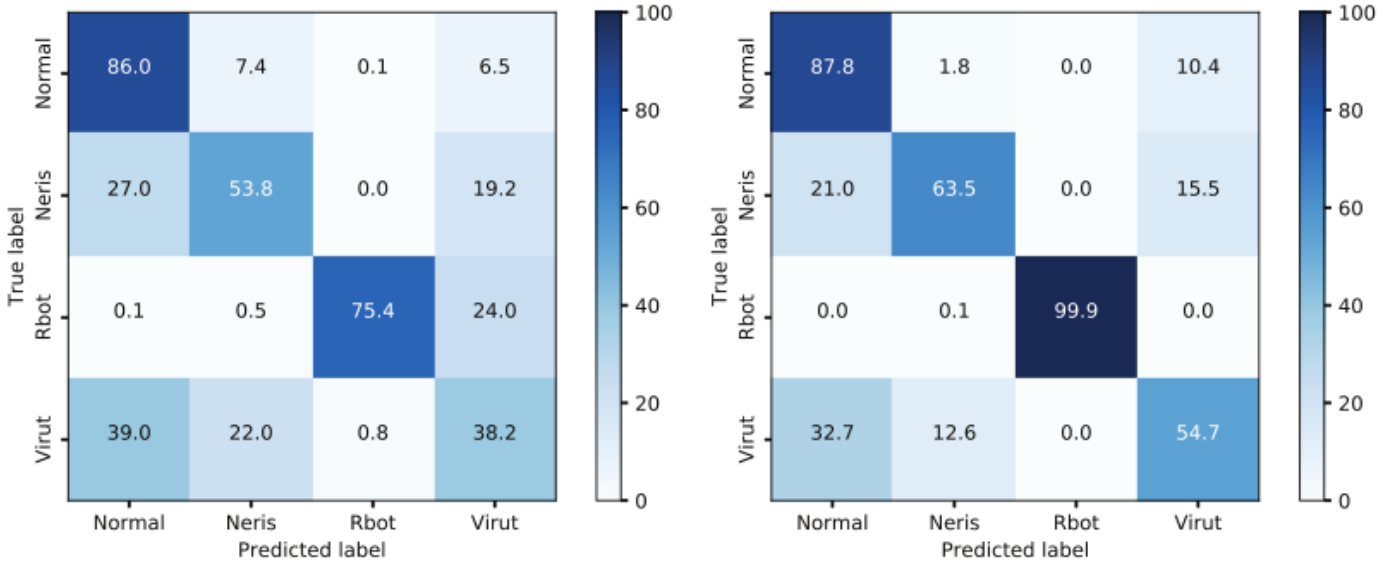
(b) Expert-Knowledge based input features.

实验分析　对比实验4

序号	类型	数据集	模型	输入数据
对比实验4	四分类	Neris、Rbot、Virut、良性流量各40000个样本	DL RF	原始数据包

- 输入数据处理
  - 同对比实验1
- 实验结果

Class	Accuracy	Precision	Recall	F_1 score
Normal	0.878	0.621	0.878	0.727
Neris	0.635	0.814	0.635	0.714
Rbot	0.999	1.000	0.999	1.000
Virut	0.547	0.679	0.547	0.606



(a) Random Forest.　(b) DL Multi-Class.

实验分析

- 总结
  - 使用原始数据流作为输入比使用原始数据包能产生更好的性能；
  - 无论是使用原始数据流还是原始数据包作为输入，DL模型性能都优于RF模型；
  - 使用原始数据流作为输入的DL模型性能跟使用专家知识构建特征作为输入的RF模型性能相当甚至更优，但使用原始数据流作为输入的DL模型省去了构建特征等繁琐的操作。

应用总结

应用总结

- 算法的应用领域
  - 互联网
  - 军事领域
  - 医学领域
  - 交通领域
  - 物联网领域
  - 工业控制领域
  - …
- 发展方向
  - 提高原始数据包作为输入的分类性能
  - 提升原始数据流作为输入的多分类性能

参考文献

[1] Marin G , Casas P , Capdehourat G . Deep in the Dark - Deep Learning-Based Malware Traffic Detection Without Expert Knowledge[C]// 2019 IEEE Security and Privacy Workshops. IEEE, 2019.

[2] Ali W A , Manasa K N , Aljunid M F , et al. A Review of Current Machine Learning Approaches for Anomaly Detection in Network Traffic[J]. Journal of Telecommunications and the Digital Economy, 2020, 8(4):64-95.

知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者，寿。

谢谢！