TOGATHER

同道 KCon 2024

网络流量密态匿迹场景下的体系对抗

汇报人：强王

目录

CONTENT

介绍（ID：强王）

KCon

- 中国科学院 网络空间安全专业；
- 在CCS、TIFS、ToN等学术会议和期刊上发表学术论文多篇；
- 获强网杯智能赛道、Datacon等竞赛奖项，在多个网络防御主题论坛发表过演讲；
- 担任过TIFS/TDSC/Computer Networks审稿人；
- 长期在云边端网络流量监测一线攻研关键课题，主要研究领域包括网络流量识别、加密通信检测等防御课题。

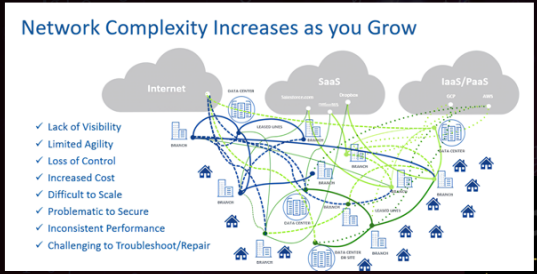
KCon

PART ONE

01

背景挑战

背景挑战



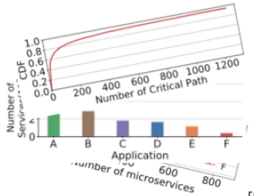
网络资产拓扑高复杂性

生产/办公/外协多域交织 [1]

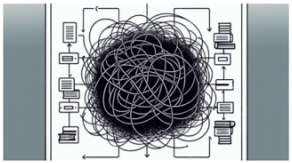
[1] CSOP 2024 深圳站_供应链、0day、黑产带来的办公网安全挑战

背景挑战

网络应用**调用关系复杂**，基础设施**规模增长**，供应链冗长导致**组件多样**



[8]



云原生应用与基础设施伴生式演进，规模持续增长，迭代速度惊人 → 原有运维技术不能直接应用

云原生应用与基础设施交互复杂，相互影响
云原生应用分散在基础设施不同位置 → 应用状态并非独立存在，需要同时监控基础设施

云原生应用种类繁多，运维需求各不相同 → 应用感知过程需要以应用为根本感知对象

[2]

[2] 以网络为中心的服务感知技术和工程实践，张晗
[3] An in-depth study of microservice call graph and runtime performance. Luo, Shutian, et al.

背景挑战

- 网络结构复杂多样
- 应用状态依赖各异
- 背景流量噪音巨大

通信网已成为社会运行的基础设施。海量数据通过网络流动，通信流量的传输基础呈现出异构性和复杂性。在该背景下，低频攻击流量因其隐蔽性，给识别工作带来了巨大挑战。

KCon

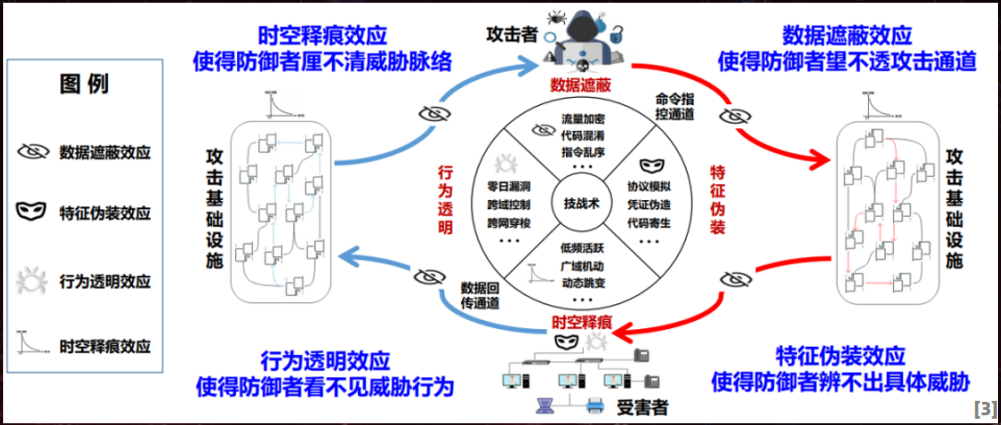
PART ONE

02

研究方法

高隐蔽攻击行为隐匿【定义】

高隐蔽攻击行为匿迹于正常信息服务、藏匿攻击流量于背景通信流量中



[3]

研究方法：流量分析

KCon

➢流量分析：

网络协议IP端口、深度包检测、统计行为分析、机器学习方法

1. 基于统计方法：从行为分析统计理论出发，关注流量分布异常等；
2. 机器学习方法：是目前应用误报较多的流量识别方法，能够实现一定的泛化分类任务；
3. 深度学习方法：注重从流量数据中提取潜在的表征特征。

流量分析是指使用技术手段对网络通信进行流量分析的工作，
一般用来识别流量承载的传输内容是否包含恶意信息、攻击载荷等

研究方法：目标

KCon

- 针对网络流量传输过程的**对抗性分析**与防御策略，在**实网攻防**中的重要性日益凸显，催生了对网络流量透视分析和构建有效防御机制的迫切需求；
- 亟需体系性开展对网络流量中的**密态匿迹行为**、**违规通联**或**攻击载荷的识别与拒止**技术研究。

网络流量作为攻击链中关键环节的对应载体：贯穿了从漏洞利用、载荷投递、指令控制到信息泄漏等多个作业阶段；
网络空间威胁行为体的匿迹动作，使其通信流量具有极高的隐蔽性，大量控制指令和窃密信息通过隐蔽的通信手段进行传输。

研究方法

KCon

➢网络流量层面实现密态匿迹效应的方法综述：

方向一：匿名通信技术研究

- 着重对网络流量传输隐私性进行保护，通过多种方法掩盖通信双方身份

方向二：流量混淆技术研究

- 着重对通信流量进行变形扰动，掩盖能够区分信息的多种流量侧信道特征

由于网络空间的虚拟化和匿名化等客观特性，及攻击者隐匿威胁行为的主观动机，信流中网络威胁一般具有高隐蔽性。网络流量密态匿迹是指网空威胁在通信过程隐匿恶意行为、藏匿自身存在的现象，造成防御视角的信息缺失。

研究方法：技术思路

KCon

➢密态匿迹技术研究进展

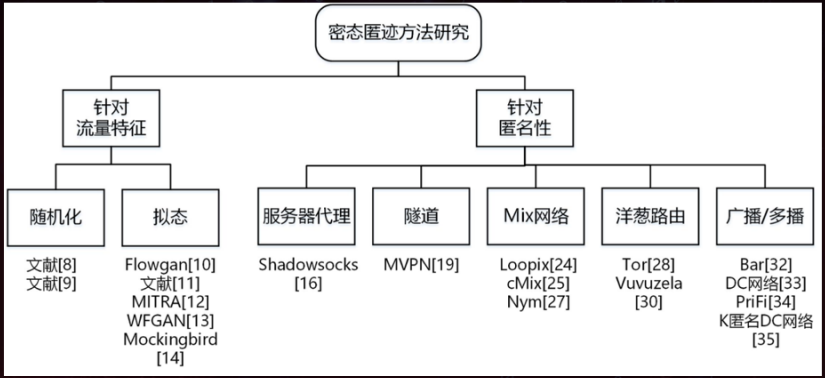
技术思路一：通联关系

- 通过修改流量传输方式，藏匿原始流量，此类技术按具体实施策略不同，分为通过服务器代理、隧道、Mix网络、洋葱路由以及匿名网络广播/多播的方法

技术思路二：流量特征

- 研究检测机制重点关注的流量特征信息，针对性修改敏感特征，此类技术可分为基于随机化的方法和基于拟态的方法

研究方法：技术脉络



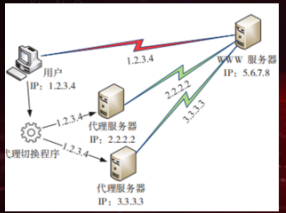
- 技术早期，多考虑针对传输方法进行改进，如VPN技术以及TOR等技术的改进发展和大规模应用；近年来针对流量特征的方法逐渐成为研究主流。
- 两种方法也常常结合使用，使兼具多种方法优点，同时保护通信双方的匿名性和通信内容的机密性。

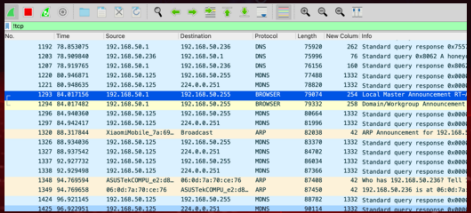
研究方法：针对通联关系的隐匿

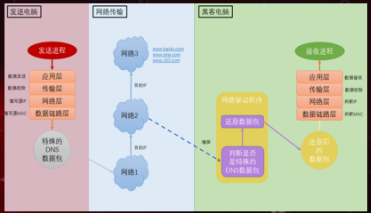
➢ 针对通联关系逃逸的方法核心在于对通信方匿名性的保护

1. 广播/多播方法、业务链路**嵌入隐匿**、**通用良性**信息基础设施借道等方法；

2. 改变原有流量传输方法，通过重新设计**传输体系结构**，隐匿通信方真实网络地址；







研究方法：针对流量特征的逃逸

➢基于随机化的流量混淆

1. Obfsproxy和ScrambleSuit：采用随机填充方法为单次通信生成新报文长度值，并分割变形剩余通信量。方法可抵御主动探测和其他指纹识别技术（如协议分类和正则表达式），实现极低开销条件下的应用层协议有效混淆
2. 针对缺乏先验知识的未知流量，优化添加独立于输入流量的“盲”扰动，设计了特殊的映射函数和正则化器在满足流量约束条件情况下生成对抗样本。

对流量关键特征进行修改，可使流量的通信协议、报文长度、通信模式等信息分散，更不具备明显特征，从而使高度依赖特征的流量分析体系难以对目标流量做出准确判断

研究方法：针对流量特征的逃逸

KCon

➢基于拟态理论的流量混淆

- ➢ 利用表达式转换、连接借用等方法，辅以加密、填充等技术；
- ➢ 将样本流量特征整形为目标流量特征，使流量难以从观测流量集识别出原始状态.

基于拟态的流量混淆技术开辟了一条新思路，即通过给流量赋予强特征，将其隐匿于具备相似特征的流量集合中，实现匿迹。

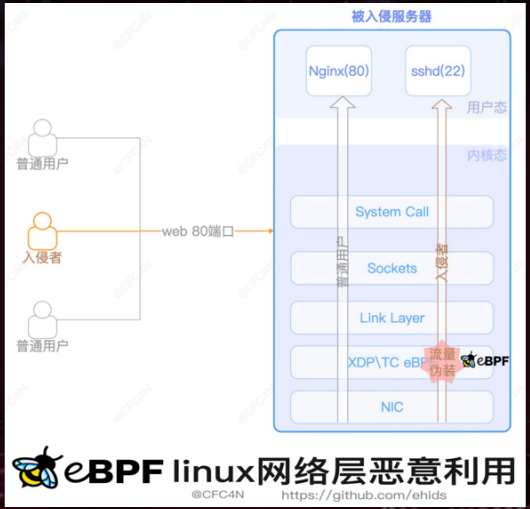
1. 针对现有流量变形/协议隧道技术主要依赖于学习特定流量的模式，缺乏动态性、被识别可能性高这一问题，Flowgan算法将流量特征动态变形为白流量，并提出了一种新的混淆结果评价算法，来测量GAN生成的网络流与目标网络流之间的不可区分性
2. 针对数据包大小这一关键特征进行混淆操作，对拟态目标的数据包长度概率分布进行建模，并将源应用程序的数据包长度突变为目标应用程序中具有相似二进制概率的数据包长度，对多种类型的应用流量都有较好的模仿效果。
3. 物联网设备低资源条件下的流量保护场景，MITRA提出物联网网络流量混淆方法，分析网络流量，并根据上下文动态生成不同级别的伪装流量，降低网络开销

研究方法：针对防御方案的逃逸

KCon

➢ 底层新方法：

- **业务链路**嵌入私有协议
- 开展**驱动层的流量伪装**



技术进展-New

➢GPT base方法：预训练支持的网络流量分析

1. 通用的编码方案，使模型具有较强的通用性和广泛任务支持性
2. 覆盖包/流两个层次，加入随机化头部字段和数据包分割，缓解小样本现象

NetGPT: Generative Pretrained Transformer for Network Traffic

Xuying Meng[1], Chungang Lin[1], Yequan Wang[2], Yujun Zhang[1]

[1]Institute of Computing Technology, Chinese Academy of Sciences, China
[2]Beijing Academy of Artificial Intelligence, China
{mengxuying, linchungang22s, nrcyujun}@ict.ac.cn, tshwangyequan@gmail.com

论文题目：NetGPT: Generative Pretrained Transformer for Network Traffic
论文作者：Xuying Meng, Chungang Lin, Yequan Wang, Yujun Zhang
发表会议/期刊：arXiv
发布时间：2023
主题类型：流量分析
笔记作者：JSY@Web 攻击检测与追踪课程
作者主页：孟绪颖 http://www.ict.ac.cn/sourcedb/cn/jssrck/202012/t20201204_5808220.html

KCon

PART ONE

03

研究成果

研究成果：效能评估方法应用

- ✓ 集合匿名性
- ✓ 不可观测性
- ✓ 不可关联性

类别	方法	匿名性	不可关联性	不可观测性	计算开销	延迟
代理	Shadowsocks	不稳定（仅具备关系匿名性）	无	无	低	低
隧道	MVPN	不稳定（仅具备关系匿名性）	无	无	低	低
Mix网络	Loopix	稳定	稳定	接收者不稳定，其他稳定	高	可改变
	cMix	稳定	稳定	无	部署中等，通信低	中
	Nym	稳定	稳定	稳定	N/A	N/A
洋葱路由	Tor	不稳定	无	无	高	低
	Vuvuzela	不稳定	不稳定	不稳定	高	高
广播/多播	BAR	稳定	不稳定	无	高	部署高，通信中等
	DC 网络	不稳定	不稳定	不稳定	中	高
	PriFi	不稳定	不稳定	不稳定	高	中
	K 匿名 DC 网络	稳定	稳定	稳定	中	中

类别	文献	方法特点	检测算法	数据集	评估
随机化	[8]	采用流量可视化的方法将网络流量转换为灰度图像，分别采用 3 种不同算法向图像中添加干扰噪声生成伪装流量，使分类器错误分类	LeNet-5 卷积神经网络模型	Moore 数据集	流量的应用类型被错误分类的概率大大提高，以 FGSM 方法为例，攻击者使用 LeNet-5 对生成的欺骗网络流量进行分类时错误率达到 99%。
	[9]	设计特殊的映射函数和正则化器来满足实时流量生成情况下的约束条件，通过改变数据包大小、信息或插入虚拟数据包的方式生成对抗样本。	DF	DeepCorr、DF 和 Var-CNN 论文所用数据集	该算法针对基于深度学习的 DF 等指纹检测方法具有较好的防御力，并且相对于之前的对抗算法更具有鲁棒性。
拟态	[10]	针对现有流量变形/协议隧道技术主要依赖于学习特定流量的模式，缺乏动态性、被识别可能性高这一问题，提出 flowgan 算法，将流量特征动态变形为白流量	SVM、NB 和其他论文提出的评价参数	自采	采用不可区分性来评价混淆有效性，该参数由算子特征曲线下的面积来确定（曲线由真阳性率与假阳性率构成）。实验证明 flowgan 的有效性。
	[11]	针对数据包大小这一关键特征进行混淆操作，对拟态目标的数据包长度概率分布进行建模，并将源应用程序的数据包长度突变为目标应用程序中具有相似二进制概率的数据包长度	SVM、决策树、KNN、随机森林	自采（来源其他论文）	以 Game 流量转为 Viber 类型为例，可以将 SVM 分类器的准确率从 76.7%降至 0.48%，将 Bagged Trees 分类器的准确率从 90%降至 0.19%，将 KNN 分类器的准确率从 83.9%降至 2.18%。
	[12]	针对物联网设备的流量保护问题，提出 MITRA 方法，根据上下文动态生成不同级别的伪装流量，避免不必要的额外开销	XGBoost、随机森林	自采	与其他文献方法相比检出率结果稍差，但网络开销较低。相比其他工作网络开销仅有百分之一甚至千分之一
		对抗样本实现指纹识别防御，流量突发特征作为生成器模型原始输入，梯度信息来优化对抗样本			获得了 90%的对抗成功率，优于 W-T 模型。
	[14]	基于对抗样本思想提出 Mockingbird 算法，不关注检测器的损失函数，产生的对抗样本具有随机性，使算法具有更好的鲁棒性。	DF、Var-CNN、CUMUL、k-FP 和 k-NN	自采	与 WTF-PAD 算法相比，DF 和 Var-CNN 检测方法的 Top-1 准确率至少低 28%，识别错误率提高了两倍

体系对抗

网络流量侧预处理引擎

框架支撑

基于网内通联基线恶意识别

+

小样本学习、迁移学习的检测方案

+

基于溯源图的广泛APT攻击检测

Buff

对抗

AI

缓解

■智能模型部署难度大

- 计算复杂度高
- 模型结构复杂
- 数据处理延迟高

■ 稳定特征难构造

- 特征分布常漂移
- 已知样本强相关

■ 标签获取困难

- APT恶意流量量少
- 高质量标签成本高

■长期潜伏运行

- 较难捕获跨周期的系统行为
- 零日漏洞利用需要多维关联

密态对抗与流量防御新阵地

第 1 卷第 3 期
2023 年 12 月

网络空间安全科学学报
Journal of Cybersecurity

Vol. 1 No. 3
Dec. 2023

加密流量实体行为内视

王 强[1,2]，尹 鹏[1,2,3*]，刘 畅[4]，乔可春[4]，胡春卉[5]

(1. 中国科学院信息工程研究所，北京 100093；
2. 中国科学院大学网络空间安全学院，北京 100049；
3. 军工保密资格审查认证中心，北京 100089；
4. 中国人民解放军 61035 部队，北京 100094；
5. 中关村实验室，北京 100094)

摘要：网络流量检测和识别是一个持久的话题，但是对网络流量行为与实体的关联研究较少，流量主体不明。由于加密流量的广泛应用，在安全运营中较难获取对网络通信传输的可见性，而检测网络攻击和异常行为通常需要获取加密流量对应的底层明文信息。在本文中，我们探索了通过进程行为实体运行时采集上下文信息的方式，对加密流量行为测量提供信息增益，实现加密流量的实体行为内视。我们的工作为加密流量的测量和可见性，增加了观测维度，提供精准的平行安全监测和高效的实体行为回溯分析能力，能够在一定程度上缓解网络空间防御视角下的遥测难题，从而有效提高加密流量隐蔽威胁监测的预警能力，并提升威胁实体恶意行为分析工作的回溯取证效率。

关键词：实体行为；投影测量；加密流量内视

中图分类号：TN915.08 **文献标识码**：A

图 5 端—网流量投影的实体插桩回溯技术

Fig. 5 Entity Stake Backtracking Technology for Endpoint-to-Network Traffic Projection

PART ONE

04

小结

KCon

红队

隐匿方法

- 混淆保护
- 匿名对抗
- 加密对抗

VS

- 内容
- 实体
- 过程
- 流量
- 熵值
- 通信

蓝队

对抗策略

- 特征检出
- 通联关系
- 行为识别

三点未来研究方向建议：

KCon

1.精细网格化的数据通道综合治理

- 未纳管、难纳管的设备单元，提供了大量隐蔽的攻击路径
- 可开展在**数据通道层面建立综合治理的旁路感知能力**的研究

2.面向身份的行为基线和攻击预防

- 信息化进程中数字身份成为访问资源的基础设施
- 可开展**以身份为核心的行为基线研究**，分析网络流量承载身份认证、授权行为模式和异常活动

3.跨时空域的威胁行为体痕迹关联

- 高等级对抗场景攻击方主动降低人机交互频次，构建更加低频且庞大的攻击基础设施，隐藏其攻击意图
- 网络安全防御任务的现实需求演进为**工程化和研究能力交织并进**
- 建立云边端**集体研判的协作机制**，成为必要选项

TONGDAO

KCon 2024

Q & A