ICS 点击此处添加 ICS 号
CCS 点击此处添加 CCS 号

T/

团 体 标 准

T/XXX XXXX—XXXX

面向 5G 通信网络的异常数据生成与信流检测技术规范

Technical specification for abnormal data generation and signaling flow detection in 5G communication networks

（征求意见稿）

XXXX－XX－XX 发布 XXXX－XX－XX 实施

发 布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利，本文件的发布机构不承担识别这些专利的责任。

本文件由××××提出。

本文件由××××归口。

本文件起草单位：重庆邮电大学、深圳北理莫斯科大学、香港大学、中国检验检疫科学研究院粤港澳大湾区研究院、广州波奇亚标准及检测技术有限公司、澳门科技大学、OrionAI Limited、中山市信裕科技有限公司、中山市政达企业管理服务有限公司。

本文件主要起草人：宁兆龙、王小洁、熊炫睿、胡希平、邱云翔、邹道远、陈源、巩勋、王杰、欧阳瑞琦、吴宇、谢晓刚、Edith C. H. Ngai、田晋宇、Terry H. S. Chu、王云帆、关乐淳、吴锡标、罗丙燕、钟晓雨。

面向5G通信网络的异常数据生成与信流检测技术规范

1 范围

本文件规定了通信网络的异常数据生成与信流检测的要求，数据的获取，生成原理，实验方法，信流分析机制，传输，重组等内容。

本文件适用于面向5G通信网各类通信协议的异常数据生成与信流检测技术。包括面向5G用户面流量生成任务的异常数据模拟生成系统，也适用于基于信令流量的信令网异常检测技术研究。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 25069-2022 信息安全技术术语

GB/T 22239-2020 信息安全技术 网络安全等级保护基本要求

GB/T 37988-2019 云计算服务安全能力要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

信令流量重组 signaling traffic reassembly

在移动通信网络中，根据协议规范对捕获的分散信令消息进行解析、分类和重新组合，恢复完整的信令交互过程。该过程通常用于分析通信设备间的交互行为、检测网络异常或优化网络性能。

3.2

特征计算 feature computation

从原始数据中提取、转换和生成能够表征数据本质或关键模式的数值或描述性指标的过程。通过特征计算，原始数据被转换为结构化的特征集合，用于模型训练、预测或分析。

3.3

强化学习 reinforcement learning

一种机器学习方法，通过让智能体 (Agent) 在与环境 (Environment) 的交互中学习策略(Policy)，以最大化其长期累积奖励 (Reward)。智能体根据环境的反馈（奖励或惩罚）调整自身的决策行为，以达到优化目标。

3.4

深度学习 deep learning

一种机器学习方法，通过构建和训练多层神经网络（通常包括输入层、隐藏层和输出层），从大量数据中自动学习特征和模式，解决复杂的感知和决策问题。它是人工神经网络（ANN）的扩展，具有更高的表现力和学习能力。

3.5

Free5GC网络 Free5GC network

一款基于 3GPP 标准的 5G 核心网开源软件，它支持 5G 独立组网和非独立组网。Free5GC 提供了一系列的核心网节点，实现了 5G 核心网的控制平面和用户平面的各个功能，同时使用其提供的可视化界

面可以管理和监控网络，此外还虚拟了 IP 地址，自定义了环回地址，完成各个网元之间的功能接口，完成了接入和移动管理功能、认证服务功能、网络开放功能、网络存储库功能、网络切片选择功能、控制策略功能、会话管理功能、统一数据管理功能等。

3.6

XPRO仪表 XPRO instrument

一种用于5G通信网络部署、优化和维护的专业测试设备。它可以执行与5G相关的关键参数测量，如射频性能、频谱分析、信号质量、网络流量、基站同步、传输延迟等,确保网络运行的高效性和可靠性。

3.7

模糊聚类 fuzzy clustering

一种基于模糊集合理论的聚类分析方法，用于将数据集划分为多个类（簇），但与传统的硬聚类不同，每个数据点可以以不同的隶属度同时属于多个类。隶属度通常用介于0和1之间的值表示。模糊聚类适用于数据集存在模糊性、不确定性或边界不清晰的场景，常见算法包括模糊C均值（FCM）算法。

3.8

异常检测 anomaly detection

一种数据分析技术，用于识别数据集中显著偏离正常模式或行为的数据点、事件或观测值（称为异常或异常值）。异常可能代表错误、欺诈行为、故障或其他非正常状态。异常检测通过分析数据的分布、趋势或统计特性，确定哪些数据点不符合预期模式。

3.9

生成对抗网络 generative adversarial network

一种基于深度学习的生成模型，由两个神经网络组成：生成器（Generator）和判别器（Discriminator）。生成器的目标是通过学习输入数据的分布生成尽可能真实的伪造数据，而判别器的目标是区分真实数据和生成数据。这两个网络在对抗训练过程中相互优化，生成器逐步提升生成数据的质量，直到难以被判别器识别。

3.10

入侵检测系统 intrusion detection system

用于监测计算机网络或系统中恶意活动、不正常行为或安全威胁的安全设备或软件。它通过分析网络流量、系统日志或用户行为，检测潜在的攻击或违规行为，并发出警报以便及时响应。

3.11

注意力机制 attention mechanism

深度学习技术，模仿人类注意力集中和分配的过程，用于增强神经网络在处理复杂任务（如序列建模）时的性能。它通过赋予输入数据中关键部分更高的权重，忽略或弱化不重要的部分，从而提高模型的学习效率和效果。

4 程序

面向5G通信网络的异常数据生成与信流检测技术规范包括以下阶段:异常数据生成与信流检测装置的设计与加工；5G网络信令流量的监测与采集要求；异常信流检测的实验流程与数据处理；信号处理与异常流量识别；信流检测分析工具与算法流程；相关数据库与云平台的应用。考虑到目前技术发展水平和未来预期，对面向5G通信网络的异常数据生成与信流检测技术进行了分别的说明（如图1所示）。

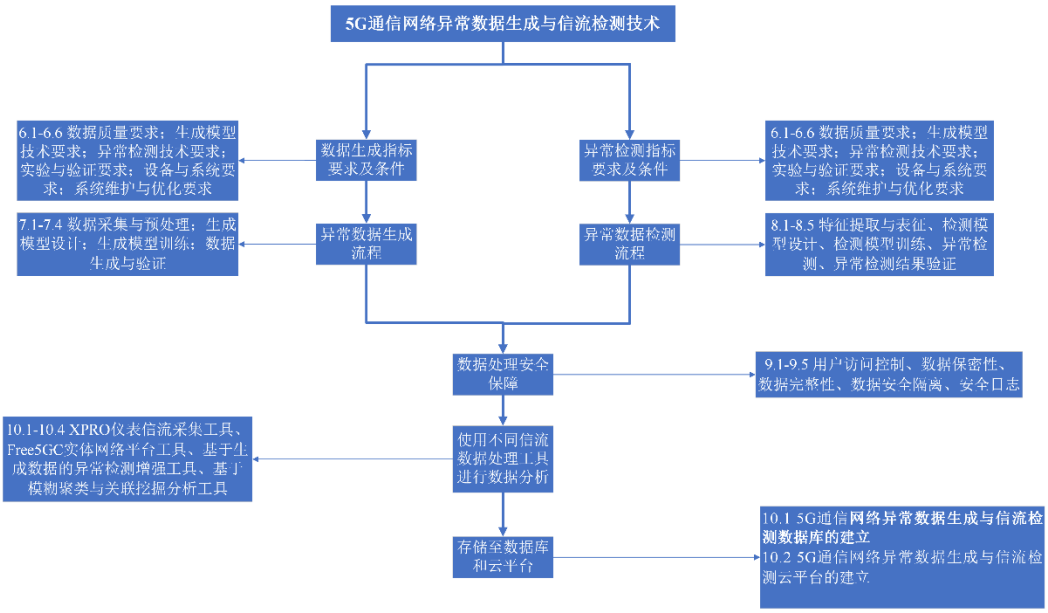

图1　5G 通信网络的异常数据生成与信流检测技术标准流程示意图

5　生成与检测原理

5.1　异常数据生成原理

在5G通信网络的异常数据生成中，采用生成对抗网络（GAN）进行数据生成。生成器通过输入随机噪声生成流量样本，并根据生成数据与真实数据的对比，不断优化生成策略。判别器则评估生成的数据是否接近真实流量，通过反馈机制指导生成器改进生成策略。随着博弈过程的进行，生成器能够生成更加真实且具有异常行为特征的流量数据。通过此过程，系统能够模拟各种潜在的网络攻击流量，为异常检测系统提供多样化的训练数据。异常数据生成原理如图2所示。

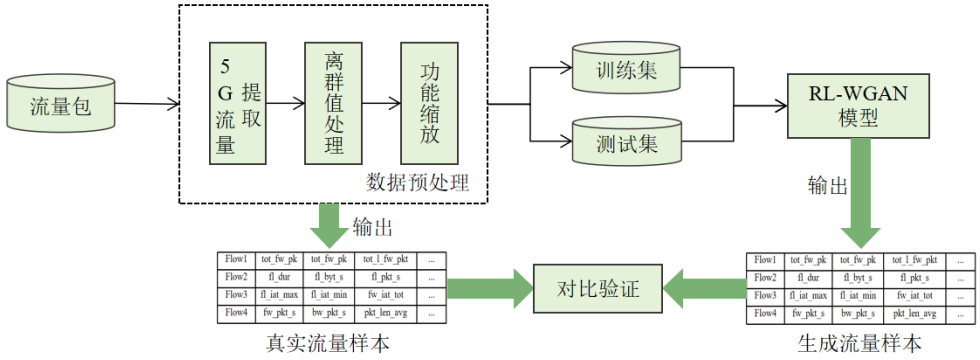

图2　异常数据生成原理示意图

5.2　异常信流检测原理

异常信流检测结合了深度学习与传统机器学习技术，使用卷积神经网络（CNN）和长短时记忆网络（LSTM）来分析信令流中的时间序列和协议特征。LSTM网络能够有效捕捉信令流中的长期依赖关系，而CNN则擅长提取信号中的局部特征。这些特征帮助检测系统区分正常流量和异常流量，识别如DDoS攻击、协议滥用等复杂的异常行为。通过训练深度学习模型，系统能够对实时数据流进行高效分析，及时发现潜在的异常信流。异常检测算法如图3所示。

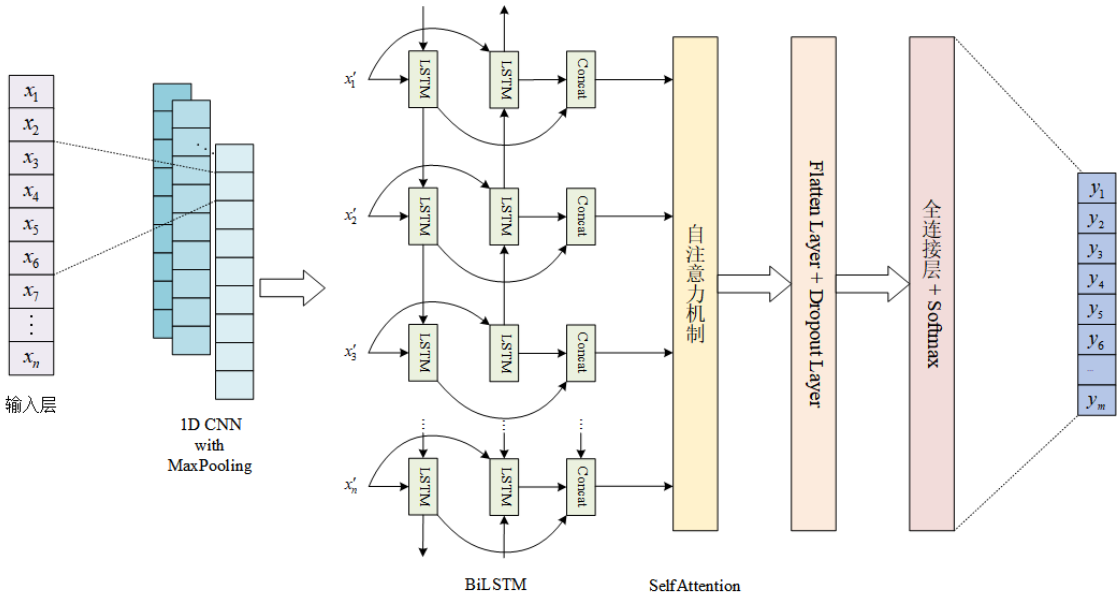

图3　异常检测算法示意图

6　生成与检测的指标要求与条件

6.1　数据质量要求

生成的异常数据应符合5G通信网络协议的标准，确保数据在协议结构、通信流程上与实际流量一致。生成的数据应覆盖多种攻击类型（如DDoS攻击、流量窃取、协议滥用等），以增强数据集的多样性，并避免数据不平衡问题。每种异常流量模式的生成数据应具备合理的攻击强度、时序模式和数据流量分布，确保模拟的攻击能有效反映真实世界中可能出现的安全威胁。

生成的数据应具有较高的真实性，并能够与实际网络流量特征高度相似。生成过程需保证每类流量的特征值（如数据包大小、流量速率等）与真实流量分布相符，且能在实际环境中体现其攻击性质。为了验证生成数据的质量，应使用与真实数据的相似性分析（如Wasserstein距离）和一致性检验方法，确保生成数据能够在不同场景中有效支撑异常检测任务。

数据生成过程中，所有实验应在严格控制的环境下进行。生成算法应能够根据实时反馈动态调整生成策略，减少人为干扰，并确保生成的数据具有高可靠性。生成过程应避免外部环境因素的干扰，例如网络带宽限制、硬件故障等，以确保生成数据的一致性和高质量。

6.2　生成模型技术要求

生成模型应具备高效生成异常流量数据的能力，生成时间必须在可接受范围内，避免延迟对实时检测产生影响。生成器的训练应在合理时间内完成，生成的流量数据应满足质量标准，且能够实时为检测模型提供训练数据。

6.3　异常检测技术要求

异常流量检测系统必须能够在复杂的网络环境中精准区分正常流量和异常流量，尤其在检测少数类异常时应保持较高的准确率。系统应优化其算法，减少误报率，避免对正常流量的错误识别，确保检测结果的可靠性。检测精度应通过定期评估和优化，确保高效准确地发现网络中的异常流量。

检测系统应能够实时处理5G网络中的海量流量数据，并在短时间内完成对流量的分析与判断。系统响应时间应小于预定的阈值，确保异常流量被及时识别，并触发报警。实时检测的能力至关重要，尤其在应对实时攻击（如DDoS）时，能够保证网络安全性。

检测模型应结合深度学习（如CNN、LSTM）与传统机器学习算法（如SVM、决策树等），自动提取流量中的时序特征和协议特征。深度学习模型能够从复杂的网络流量中提取高维度特征，提升检测系统对复杂攻击的识别能力。传统算法则用于提升模型的准确性和稳定性，从而有效应对不同类型的异常信流。

6.4 实验与验证要求

实验应在真实或仿真的5G通信网络环境中进行。推荐使用如Free5GC等开源平台进行流量数据采集和分析，确保生成数据的真实性与可靠性。实验中应考虑网络拓扑结构、带宽限制等因素，模拟真实的通信场景，确保结果的代表性。

系统应定期进行性能评估，采用准确率、召回率、F1-score等评估指标，评估生成模型与检测系统的表现。生成数据的质量应通过与实际数据的相似性进行验证，检测模型应通过对比测试与现实网络攻击案例进行验证，确保模型的有效性和适用性。

6.5 设备与系统要求

生成与检测系统应具备强大的计算和存储资源，能够应对生成大规模异常流量数据和实时检测任务的需求。硬件设备应支持高性能计算，如GPU加速和大规模并行处理能力，以保证系统的高效运行。

系统应能够在长时间运行过程中保持稳定，尤其是在高负载情况下，检测系统应具备自我恢复能力。设备应支持冗余设计和故障恢复机制，确保在出现故障时，系统能够快速恢复并继续提供正常服务。

6.6 系统维护与优化要求

生成与检测系统应具备持续优化机制，通过定期更新生成模型和检测算法，使其适应新的攻击模式和流量特征。定期的优化能够确保检测系统始终保持高效并具有较强的适应性。

系统设计应具有较强的扩展性，以便后续加入新的攻击模式、生成算法和检测方法。模块化设计和开放接口是系统扩展的基础，能够根据需求快速调整和升级。

7 异常数据生成流程

7.1 数据采集与预处理

7.1.1 数据采集

数据采集过程应通过5G通信网的核心网（5GCN）及用户面（UPF）网元收集流量数据。采集的数据应包括正常流量数据和已知的攻击流量数据（如DDoS攻击、协议滥用等）。数据应以PCAP格式存储，并通过专用通道传输到远程服务器进行后续处理。采集的原始数据应具有代表性，能够覆盖不同类型的流量和攻击场景。

7.1.2 数据预处理

采集到的原始数据应经过数据清洗、特征提取和标准化等预处理步骤。特征提取包括从原始流量中提取关键特征，如数据包大小、传输速率、时序特征等。所有特征应标准化，确保生成模型能够处理具有相同尺度的输入数据。特征选择应依赖于5G通信网的流量协议和攻击类型，重点提取有助于识别异常流量的关键指标。用户面流量特征预处理流程如图4所示。

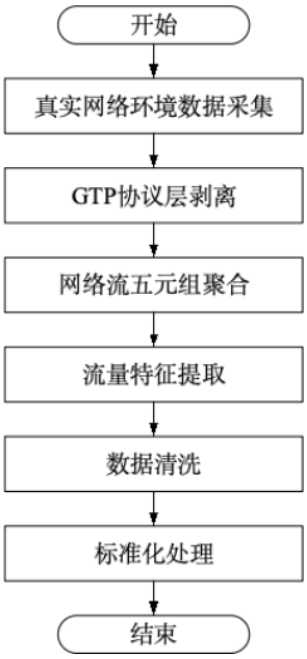

图4　用户面流量特征预处理流程

7.2　生成模型设计

7.2.1　强化学习与生成对抗网络结合

生成模型应基于强化学习与生成对抗网络（GAN）的结合。强化生成对抗网络（RL-WGAN）模型通过在生成器中引入强化学习的动态奖励机制，确保生成的数据不仅符合协议标准，而且具备高质量的异常特征。生成器和判别器通过博弈式训练，不断优化生成的数据，使其能够接近真实流量数据分布。

7.2.2　生成器与判别器设计

生成器应基于双向门控循环单元（BiGRU），并结合协议标准对生成的流量样本进行约束。判别器采用改进的多层感知机（MLP），用于区分真实数据和生成数据。生成器的奖励机制应根据协议特征和生成数据的质量进行优化，判别器通过评估生成样本与真实数据的相似性，不断提升其辨别能力。生成器整体结构如图5所示。

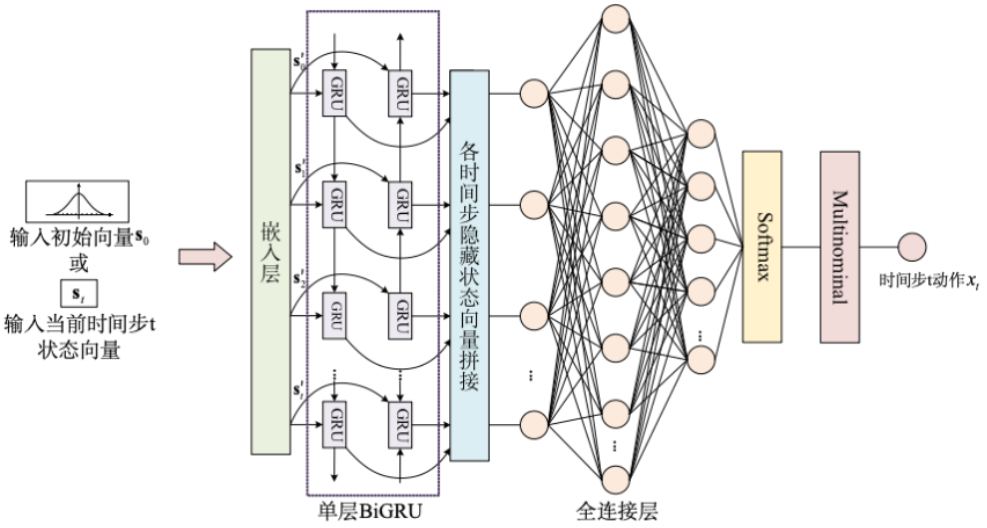

图5　生成器模型结构示意图

7.2.3　数据生成流程

包括以下步骤：

a) 初始化模型：根据 5G 通信网用户面流量的特征构建并初始化生成器与判别器。
b) 生成数据：生成器接收噪声作为输入，通过强化学习机制生成符合协议标准的异常流量数据。
c) 评估与优化：判别器对生成数据与真实数据进行对比评估，反馈生成器优化策略，通过对抗训练不断提高生成数据的质量。
d) 输出生成数据：经过多轮训练后，生成器输出高质量的异常数据样本，用于异常检测系统的训练。

7.3 生成模型训练

在训练阶段，RL-WGAN模型需要在标注数据集上进行训练，算法细节如表1所示。训练数据包括正常流量数据和不同类型的攻击流量数据。训练过程的目标是让生成器能够根据输入的正常流量数据生成异常流量，同时让判别器学会区分真实数据与生成的数据。通过多轮的训练，生成器和判别器逐渐优化，最终生成器能够生成具有较高真实性的异常流量。具体步骤包括：

a) 生成异常流量数据：生成器根据输入的正常流量数据生成新的一批异常数据。
b) 判别异常流量数据：判别器判断生成的数据是否为真实的流量数据。如果判别器判断为假，生成器根据反馈调整其生成策略。

表1 RL-WGAN 算法

算法 RL-WGAN 训练与更新
1: **输入**：初始化参数的生成器 G_θ 、判别器 D_w ;奖励折扣因子 α ;专家样本数据集 X;
2: **输出**：更新参数后的 G'_θ ， D'_w ;
3: **function** Reward(x_g)
4: 判别奖励 $Q_{D_w}^{G_\theta} = D_w\left(x_g\right)$;
5: 协议奖励 $R_p = \text{Calculate\_Protocol\_Reward}\left(x_g\right)$;
6: 总奖励 $R = Q_{D_w}^{G_\theta} - \alpha \cdot R_p$;
7: **return** R
8: **end function**
9: **function** Calculate_Protocol_Reward(x_g)
10: 协议特征奖励 C_1 = 核查协议特征字段(x_g);
11: 端口映射奖励 C_2 = 核查端口号到应用层服务的映射(x_g);
12: 标志字段一致性奖励 C_3 = 核查协议对应的特征标志字段(x_g);
13: 服务与协议映射奖励 C_4 = 核查服务与协议的对应关系(x_g);
14: 总奖励 $R_p = C_1 \cdot C_2 \cdot C_3 \cdot C_4$;
15: **return** R_p
16: **end function**
17: **for** episodes = 0, 1, …, E **do**
18: **for** each discriminator_iteration **do**
19: 获取真实数据样本 X_r ;
20: 获取生成数据样本 X_g ;

21: 计算判别损失 $L_w\left(D_w\left(X_r, X_g\right)\right)$;
22: 更新判别器参数 D_w;
23: **end for**
24: **for** each generator_iteration **do**
25: **for** t = 0, 1, …, T **do**
26: 获取当前状态 s_t;
27: 生成网络流序列时间步 t 的动作 $x_t = G_\theta\left(s_t\right)$;
28: 时间步序列拼接 $x_g = \left(x_1, x_2, \ldots, x_t, \mathrm{MCSampling}(x_{1:T}, t)\right)$;
29: 计算奖励 $R = \mathrm{Reward}\left(x_g\right)$;
30: 更新状态 $s_{t+1} = \left(x_1, x_2, \ldots, x_t\right)$;
31: **end for**
32: 计算回合总奖励 R_{total};
33: 计算生成损失 $L_G\left(R_{total}\right)$;
34: 更新生成器参数 G'_θ;
35: **end for**
36: **end for**

7.4 数据生成与验证

7.4.1 数据生成

完成训练后，RL-WGAN模型可以用于生成符合5G网络协议的异常流量数据。生成的数据应具有真实攻击特征，如拒绝服务（DoS）攻击、恶意中间人攻击等。这些异常数据将用于增强异常检测系统，提升其在实际网络环境中的效果。

7.4.2 验证过程

生成的数据必须经过验证，以确保其与实际网络攻击流量的统计特征一致。通常使用Wasserstein距离等度量方法来评估生成数据的相似性。通过这种方法，可以确认生成的异常流量与真实攻击流量之间的差异，确保生成的异常数据真实可信。

8 异常数据检测流程

8.1 特征提取与表征

8.1.1 特征提取

特征提取是信令流量分析的核心步骤，特征表如下表2所示，流程如下表3所示，旨在从原始信令流中提取与流量行为密切相关的关键特征。提取的特征应能够全面表征信令流的状态，并涵盖不同类型的流量，包括正常流量与攻击流量。常见的特征包括：

a) 包计数：在一定时间内统计信令包的数量，用于分析流量密集度，尤其在 DDoS 攻击中具有显著变化。
b) 流持续时间：记录信令会话的时长，能够检测到异常连接持续时间。

c） 协议分布：分析不同协议（如 NGAP、SCTP 等）在信令流量中的比例，帮助识别协议滥用等异常行为。
d） 数据包大小：统计数据包的大小，异常包大小可能表示数据包注入攻击。
e） 时序特征：包括数据包到达的时间间隔、会话的建立与释放等，帮助揭示时序异常，如洪水攻击。

表2 信令特征表

通用流量特征	描述	信令流特有特征	描述
fl_dur	流持续时间	fw_pdum_flag	C2SPDUM 标志次数
tot_fw_pk	C2S 包数量	fw_pduc_flag	S2CPDUC 标志次数
tot_bw_pk	S2C 包数量	fw_up_flag	C2S 包中 UP 标志次数
tot_fw_pkt	C2S 包总大小	fw_do_flag	S2C 包中 DOWN 次数
fl_byt_sv	每秒传输包字节数	tot_nas_flag	包中 NAS 次数
fl_pkt_s	秒传输的数据包数	fw_init_flag	C2S 包中 INIT 次数
fw_pkt_s	每秒 C2S 包的数量	tot_heart_flag	HEARTBEAT 次数
bw_pkt_s	每秒 S2C 包的数量	tot_shut_flag	SHUTDOWN 次数
pkt_len_min	流的最小长度	fw_ngre_flag	NG 连接请求次数
pkt_len_max	流的最大长度	bw_ngre_flag	NG 请求响应次数
bw_iat_tot	S2C 最大时间差	Flow_id	信令流唯一标识符
bw_iat_min	S2C 最小时间差		
fl_iat_max	流之间的最大时间差		
fl_iat_min	流之间的最小时间差		
fw_iat_tot	C2S 流最大时间差		
fw_iat_min	C2S 流最小时间差		
pkt_len_avg	流的平均长度		
pkt_size_avg	数据包的平均大小		
fw_seg_avg	C2S 包平均大小		
bw_seg_avg	S2C 包平均大小		

表3 特征提取流程

算法 特征提取过程
1: **Input：** Pcap信流数据包
2: **Output：** csv数据集
3: 初始化信令流处理模块
4: **for each** packet in captured_signaling_flow_packets **do**
5: 提取源、目标IP、端口及协议信息

6: **if**存在相同的源、目标主机、端口和协议 **then**
7: 将数据包添加到相应的流缓存区
8: **eles**
9: 创建新的流缓存区并将数据包添加
10: **end if**
11:**end for**
12:遍历流缓存区模块
13**: each** flow_buffer in flow_buffers **do**
14: 对每个缓存区信令流进行双向流归并
15: **for each** packet in flow_buffer **do**
16: **if** protocol_info == SCTP **then**
17: 执行SCTP流重组方法
18**: end for**
19:信令流特征提取模块
20: **for each** reassembled_packet in reassembled_signaling_flow_packets **do**
21: 初始化统计特征变量
22: **for each** reassembled_packet in reassembled_signaling_flow_packets **do**
23: 计算各条信令流特征值
24: **end for**
25: **end for**

8.1.2 特征表征：

在特征提取后，需对数据进行标准化处理，以确保不同特征的数值范围一致，避免在后续模型中产生偏差。特征表征的关键步骤包括：

a) 标准化：将所有提取特征按相同尺度标准化，以消除不同特征量纲上的差异，便于后续算法的处理。
b) 降维：通过主成分分析（PCA）或线性判别分析（LDA）等方法减少特征维度，去除冗余信息，提高计算效率。
c) 特征选择：使用算法筛选最具代表性的特征，去除不相关或冗余的特征，以提高检测模型的精度和效率。

8.2 检测模型设计

基于深度学习的异常流量检测：采用了生成对抗网络（GAN）和强化学习（RL）技术相结合的方法。检测模型的设计包括两个主要组成部分：

a) 生成对抗网络（GAN）：
   1) 生成器：生成器负责通过输入的正常流量数据，生成可疑的异常流量数据。生成器的目标是生成能够通过判别器检测的异常流量。生成器不仅要模拟攻击行为，还要符合网络流量的协议标准。

2) 判别器：判别器的任务是判断输入的流量数据是否为真实数据或由生成器生成的假数据。通过与生成器的对抗训练，判别器可以提高对异常流量的识别能力。
3) 优化过程：生成器和判别器通过博弈的方式互相优化。生成器不断改进以生成更具欺骗性的流量，而判别器则变得越来越强大以区分真实和虚假的流量。

b) 强化学习机制：强化学习机制引入了动态奖励机制，优化生成器的学习过程。生成器根据奖励反馈调整其策略，从而生成更符合实际攻击流量特征的异常数据。具体来说，生成器根据奖励的大小来调整其生成的数据样本，使得生成的数据更加逼真。生成器的优化过程不仅考虑数据生成的正确性，还考虑到协议的复杂性和攻击行为的多样性。

8.3 检测模型训练

在训练阶段，模型使用标注数据集进行训练，数据集包括正常流量数据和各种类型的攻击数据。训练的目的是使生成器能够生成接近真实攻击数据的流量，而判别器则学习如何正确区分真实流量和生成的异常流量。训练过程包括以下步骤：

a) 生成异常流量：生成器从正常流量中生成新的异常流量数据样本。这些样本模拟各种攻击行为，如 DoS 攻击、流量洪水、恶意中间人攻击等。
b) 判别异常流量：判别器对生成的数据进行判断，评估数据是否符合正常流量的特征。根据判别结果，生成器调整生成策略，尽可能生成更接近真实攻击数据的样本。
c) 迭代优化：生成器和判别器经过多轮对抗训练后，生成器的能力逐步提高，能够生成更加真实的异常流量数据。

8.4 异常检测

经过训练后，模型可以投入实际的异常流量检测任务。通过输入实时网络流量数据，检测模型能够识别出其中的异常流量。模型的检测过程包括：

a) 实时流量输入：检测系统接收来自网络中的流量数据，将其输入到经过训练的生成对抗模型中。
b) 异常检测：模型将流量与训练数据进行对比，判断其是否属于异常流量类别（如 DoS 攻击、恶意中间人攻击等）。

8.5 8.5 异常检测结果验证

为了验证模型的有效性，生成的异常数据需要与实际的攻击流量进行对比，确保其具有较高的相似性。验证方法包括：

a) Wasserstein 距离：该方法用于衡量生成数据与真实攻击数据之间的分布差异。Wasserstein 距离越小，表明生成的数据越接近真实数据，模型的生成能力越强。
b) 2)混淆矩阵：通过混淆矩阵评估检测系统的性能，帮助计算模型的准确率、召回率、F1 值等。混淆矩阵能够显示系统在检测异常流量时的表现，包括真正例、假正例、假负例和真负例的数量。

9 信流数据处理的安全性

9.1 用户访问控制

用户注册时应与其约定访问权限，管理员应可以随时收回该用户权限。用户访问时应通过保护手段校验其身份（如验证码或口令验证机制等）。当未授权用户访问本地存储空间时，系统应阻止未授权用户的访问行为同时向管理员提供可查的告警信息。

9.2 数据保密性

应对本地测序数据进行加密存储，对数据的机密性和完整性进行保护，加解密操作应采用国家密码行政主管部门认可的密码算法。

对于加密的数据，未授权用户无法获得或操作明文，授权用户可以通过解密算法获得明文保证数据的可用性。

9.3 数据完整性

应能够对本地测序数据进行完整性校验，校验数据的完整性是否被破坏，防止数据被刻意的篡改删除或插入。当检测到数据完整性被破坏时，应向管理员提供可查的告警信息。

9.4 数据安全隔离

应对本地测序数据进行安全隔离，为不同用户分配独立的内存空间，防止用户对其他用户数据的非法访问。

9.5 安全日志

用户的所有操作均应形成完整且不可篡改的操作日志，管理员有权限查看所有用户的操作日志。

10 流数据处理工具

10.1 XPRO 仪表信令流采集工具

5G核心网信令风暴仿真平台（XPRO仪表）以高性能5G核心网仿真测试仪为核心，通过40000TPS以上单节点业务过程、200Gbps单节点的媒体流量带宽、20Gbps单UE接入吞吐量、65536个最大仿真节点数实现高度逼近现网规模的仿真。通过XPRO仪表，可以仿照现实环境进行相应的网元配置，并根据具体需求进行信令负载配置，达到仿真5GC环境的目的，仪表架构如下图6所示。

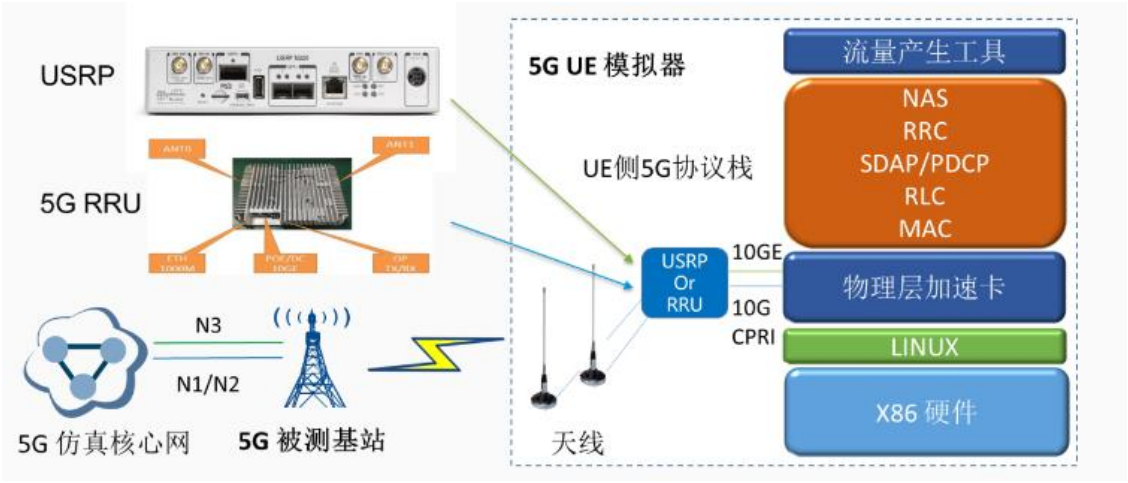

图6 XPRO 仪表架构图

XPRO仪表信令流采集工具的核心功能包括：

a) 信令仿真：平台能够精确模拟 5G 核心网中的信令流程，实现了对 5GC 主要信令网元 AMF、SMF、AUSF、UDM、NSSF、NRF、PCF 和 AF 以及各网元之间接口协议的仿真，包括但不限于用户设备注册、鉴权过程、接入控制决策以及会话管理等。这些仿真信令能够复现真实世界中的网络行为，为测试提供真实的场景。

b) 风暴测试：平台具备在短时间内生成大量信令的能力，这种高强度的信令生成可以模拟网络攻击、设备故障或异常流量等极端情况，从而测试核心网在极端压力下的响应能力和恢复能力。

c) 性能评估：通过实时监控网络的关键性能指标，如端到端延迟、数据包丢失率、网络吞吐量等，平台能够全面评估核心网在不同负载条件下的性能表现。这有助于发现性能瓶颈，为网络优化提供数据支持。

10.2 Free5GC 实体网络平台工具

Free5GC是一款基于3GPP标准的5G核心网开源软件，架构图下图7所示，它支持5G独立组网和非独立组网。Free5GC提供了一系列的核心网节点，实现了5G核心网的控制平面和用户平面的各个功能，同时使用其提供的可视化界面可以管理和监控网络，此外还虚拟了IP 地址，自定义了环回地址，完成各个网元之间的功能接口，完成了接入和移动管理功能、认证服务功能、网络开放功能、网络存储库功能、网络切片选择功能、控制策略功能、会话管理功能、统一数据管理功能等。

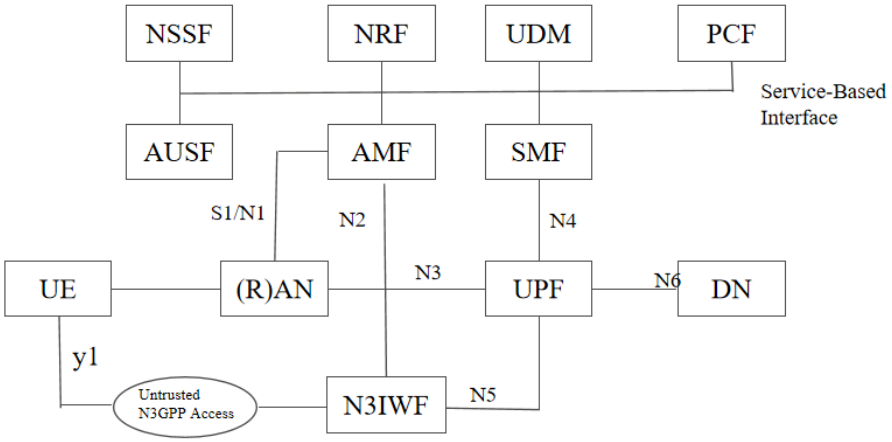

图7 Free5GC 架构图

Free5GC实体网络平台工具包含了以下主要组件：

a) AMF（Access and Mobility Management Function）：负责用户的接入和移动管理。AMF 处理与用户设备的接入过程，包括鉴权、安全性协商和移动性管理。
b) SMF（Session Management Function）：负责会话管理和策略控制。SMF 处理用户会话的创建、终止和修改，同时负责策略控制，包括服务质量（Quality of Service，QoS）的管理和网络切片的支持。
c) UPF（User Plane Function）：负责用户数据的传输。UPF 处理用户数据包的路由和转发，包括数据的分类、转换和加密。
d) UDM（Unified Data Management）：负责用户数据的管理。UDM 存储和管理与用户相关的信息，例如用户配置数据、认证数据和服务订阅数据。
e) UDR（Unified Data Management）：作为 UDM 的后端存储，负责持久化和管理用户数据。
f) f)NRF：负责网络功能 NF 的注册和发现。NRF 维护一个 NF 注册表，允许 NF 进行注册，并提供查询接口以发现已注册的 NF。
g) NSSF（Network Slice Selection Function）：负责网络切片的选择。NSSF 根据用户需求和策略，选择合适的网络切片并分配资源。
h) AUSF（Authentication Server Function）：负责用户设备的认证和安全性管理。AUSF 验证用户设备的身份，并生成用于加密和安全性保护的密钥。
i) PCF（Policy Control Function）：PCF 是负责管理 UE 服务质量 QoS 策略的核心网节点。它根据网络的负载和资源需求，为 UE 提供最佳的 QoS 策略，以及支持网络切片策略的管理。

10.3 基于生成数据的异常检测增强工具

10.3.1 工具实现

异常检测增强工具结合了生成对抗网络（GAN）生成的数据与先进的深度学习模型，如卷积神经网络（CNN）、双向长短时记忆网络（BiLSTM）、自注意力机制（Attention Mechanism）等，通过多层次融合的方式显著提升异常流量检测的准确性与鲁棒性。该工具通过引入基于生成的数据增强模型（RL-WGAN）及检测增强模块（CBA-Net），充分利用生成的异常数据扩展训练样本空间，从而提升检测系统的能力，特别是在少数类异常流量检测中的效果，结构如下图8所示。

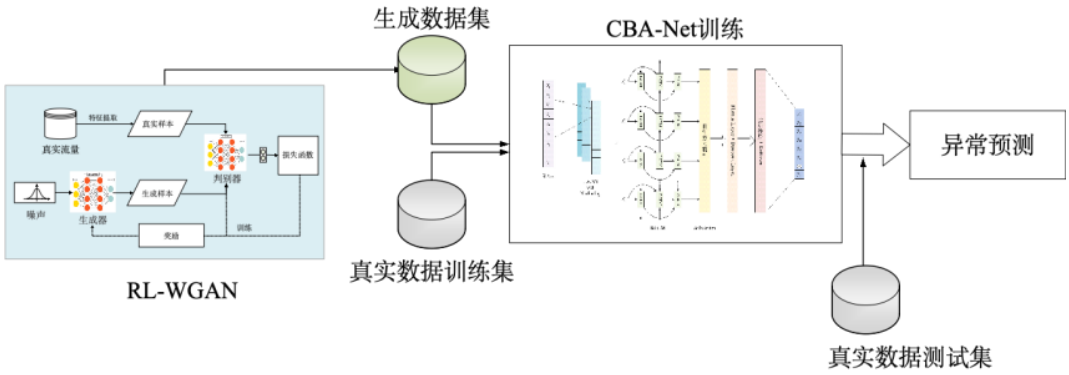

图8　eCBA-Net 结构图

10.3.2　数据生成与增强

在面对5G通信网络中流量数据稀缺的情况下，传统的数据收集与生成方法无法有效支持异常检测任务，尤其是在异常样本不平衡的场景下。为此，通过集成基于生成对抗网络（GAN）和强化学习（RL）优化的生成模型（RL-WGAN），增强工具能够自动生成符合5G通信协议和攻击特征的异常流量数据。这一生成模型不仅能够模拟各种攻击行为，如DDoS攻击、协议滥用等，还能够适应多种协议标准，生成高质量、多样化的异常流量数据，补充了原始数据集中的稀缺样本。通过生成这些异常数据，增强工具能够有效平衡训练数据集中的类别不平衡问题，增强模型对少数类异常流量的识别能力。工具个层级运作关系时序图如图9所示。

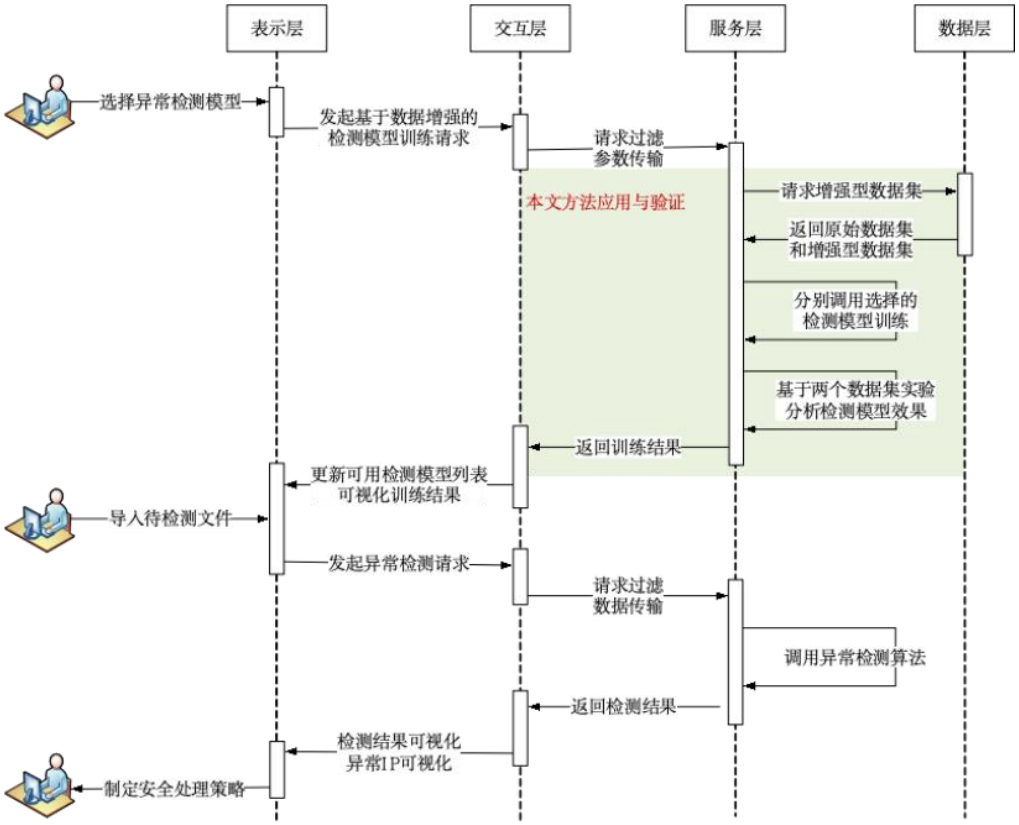

图9　基于数据增强的异常检测时序图

10.3.3　异常检测

异常检测模块采用了融合深度学习模型的增强算法，特别是在特征提取与序列模式识别中，结合卷积神经网络（CNN）与长短期记忆网络（BiLSTM）的优势。CNN用于从流量数据中提取局部特征，而BiLSTM

则处理时序数据，捕捉长距离依赖关系。通过自注意力机制（Attention Mechanism），该模块能够对异常流量中的关键信息进行动态加权，确保在检测过程中对重要特征的关注，提高了检测系统的准确性。

10.4 基于模糊聚类与关联挖掘的异常信流分析工具

10.4.1 研究过程

威胁信流分析的具体研究过程如下图10所示。

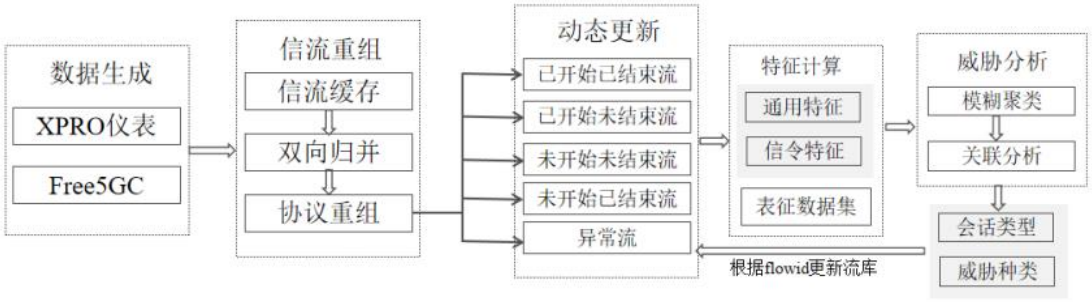

图10 威胁信流分析过程

10.4.2 数据采集与预处理

在进行异常信流分析之前，工具首先从5G核心网中采集信令流数据，包括GTP-U协议流量、SCTP协议数据等。由于信令流量具有较高的实时性与多样性，数据预处理阶段非常关键。预处理过程中，工具会进行数据清洗，去除无效数据和噪声，并通过特征提取方法，从每条信令流中提取包括数据包大小、流量速率、消息类型、会话时长等多个统计特性。

在特征提取的过程中，特别针对NGAP、SCTP等协议的信令数据，通过解析每一条信令消息中的字段，提取出各个信令协议的关键参数，如源地址、目标地址、协议类型、消息体长度、会话持续时间等。这些特征为后续的异常检测提供了基础。

10.4.3 模糊聚类分析

信令数据本身存在一定的模糊性，同一会话中可能包含不同类型的信令行为，导致其在特征空间中呈现模糊的分布。为了应对这种模糊性，工具采用了模糊C均值（FCM）聚类算法，将信令流数据进行模糊划分。模糊聚类是将每个信令数据点划分到多个聚类中，并为其分配一个隶属度，表示该数据点属于某一类别的程度。与传统硬聚类方法不同，模糊聚类能够处理数据的不确定性，使得工具能够在复杂的信令行为中，更灵活地识别出可能的异常模式。

表4 模糊 C 均值聚类算法

算法 模糊 C 均值聚类算法过程
1: **input：** 聚类数 *C*，初始聚类中心 $X=\{x_1,x_2,...x_n\}$，模糊指标 *m*，终止误差
2: **output：** 聚类中心 V$=[v_1,v_2,...v_n]$，隶属矩阵u_{ij}
3: **Step 1:** 设置聚类数目 *C*
4: **Step 2:** 初始化模糊因子 *m*、迭代允许的误差 ε、迭代次数 t=0 和隶属度矩阵 U(0)
5: **Step 3:** 将根据公计算或更新隶属度矩阵 U，并更新聚类中心
6: 加到相应的流缓存区。如果不存在创建新的流缓存区并更新流标识符
7: **Step 4:** 比较和$J(t)$和$J(t-1)$;若$\|Jt-J(t-1)\|\leq\varepsilon$,则满足迭代停止条件，迭代停止，否则
8: 返回 Step3，继续迭代

10.4.4 关联挖掘与威胁识别

在信令流量分析过程中，工具通过关联挖掘技术，分析信令数据之间的潜在关系。信令流的数据特征往往是相互关联的，某些特征的异常可能会与其他特征的变化相联系。通过关联规则挖掘，系统能够识别出特征之间的依赖关系，从而揭示出潜在的攻击模式。通过对不同会话特征（如信令次数、消息类型、源IP与目标IP等）的关联分析，工具能够准确区分不同类型的攻击威胁，如DoS攻击、DDoS攻击、伪基站攻击等。工具通过识别会话特征之间的关联性，能够有效地从大量信令流中识别出符合特定攻击特征的异常信流。

11 数据库及云平台

11.1 建立 5G 通信网络异常数据生成与信流检测数据库

11.1.1 存储引擎与事务支持

对于5G通信网络异常数据生成与信流检测的数据库，我们强烈推荐使用InnoDB存储引擎。其支持事务（ACID属性）、行级锁定和多版本并发控制（MVCC），能够提供高并发的数据处理能力，确保在5G通信网络中大量并发读写请求时，数据库依然能够稳定高效地工作。此外，InnoDB支持良好的崩溃恢复机制，能在断电或系统故障时自动恢复事务，提高系统的容错能力。特别是在涉及到异常流量的实时检测和大数据存储时，InnoDB的这些特性尤为重要。

11.1.2 字符集

为了确保全球范围内的兼容性和稳定性，建议使用UTF-8字符集。UTF-8可以兼容所有语言，避免了因字符编码不一致而导致的乱码问题。在5G通信网络中，由于网络数据流量来自全球各地，采用UTF-8字符集能够保证数据在不同地域、不同语言环境下的正确存储和传输。此外，UTF-8字符集具有存储优化特性，减少了空间浪费，特别是在数据库中存储大量实时信流数据时，能够提高存储效率。

11.1.3 数据表及字段注释

建议在所有数据表和字段中加入中文注释，这样不仅有助于项目组成员理解数据库结构，也便于将来其他研究人员理解和维护代码。在5G通信网络的异常流量检测系统中，数据表和字段的名称和含义较为复杂，因此，为每个表和字段提供详细注释，能够显著提高系统的可维护性和扩展性。

11.1.4 禁用存储过程与触发器

建议禁用存储过程、视图、触发器和Event等数据库功能，尤其是在高并发的5G通信网络应用中，这些功能可能导致性能瓶颈。尽管这些功能可以简化数据库操作，但它们可能在数据量巨大时影响数据库的处理速度。在异常数据生成与信流检测场景中，业务逻辑和数据操作更适合在应用层处理，而非依赖数据库内部复杂的机制。

11.1.5 大文件存储

数据库应避免存储大文件如图片、视频等。5G网络流量数据通常包括实时的网络信流数据或日志数据，建议将这些大数据存储在专门的文件系统中，而在数据库中仅存储指向这些大文件的URI或路径。这种做法能显著减轻数据库的负担，提升查询效率，并能够确保数据存储的灵活性和扩展性。

11.1.6 数据库连接与命名规范

数据库连接应使用内网域名，而非IP地址，增加安全性并降低维护复杂度。数据库命名规范的设计有助于保证5G通信网络中不同环境（如生产环境、开发环境、测试环境）数据库的管理一致性。例如：线上环境：dj.xxx.db；开发环境：dj.xxx.rdb；测试环境：dj.xxx.tdb；从库：dj.xxx-s.db；备库：dj.xxx-sss.db。

11.1.7 命名规范

为了提升代码和数据库结构的可读性，建议库名、表名、字段名均采用小写字母，并使用下划线命名风格。例如，表名应使用user_data、traffic_flow等命名方式，避免使用拼音或混合命名方式。此举有助于保证数据库的跨团队协作和维护的一致性。

11.1.8 主键与外键

每张数据表应具有主键，例如采用自增主键。这样能够保证数据的一致性，并提高查询性能。特别是在5G通信网络的高并发环境下，主键自增可以有效避免page分裂，减少内存和磁盘的资源消耗。外键约束应当禁用，因为外键约束会增加表之间的耦合度，影响插入和更新操作的效率，甚至可能导致死锁问题。

11.1.9 字段设计

字段设计时，应尽量避免使用TEXT和BLOB类型，因为它们会占用大量的磁盘空间和内存，导致性能下降。对于字段的设计，建议将所有字段定义为NOT NULL，并为必要字段提供默认值。这样不仅简化了数据库的优化，也能提升数据库性能。

11.1.10 索引设计

合理的索引设计对于提高查询性能至关重要。建议创建组合索引，并将区分度较高的字段放在索引前面。对于常被查询的字段，创建适当的索引能够显著提高检索速度。但对于更新频繁且区分度较低的字段，应该避免创建索引，因为这会增加数据库的负担。

11.1.11 查询优化

查询时应避免使用SELECT *，仅查询必要的字段，减少不必要的CPU、内存和I/O消耗。在WHERE条件的属性上，应避免使用函数或表达式，避免引起全表扫描，影响数据库性能。避免在查询中使用不必要的复杂条件，如NOT LIKE或!=，这些会导致全表扫描并且增加查询的复杂度。

11.2 建立 5G 通信网络异常数据生成与信流检测云平台

11.2.1 云平台架构选择

在5G通信网络的云平台设计中，推荐采用Unix服务器架构，尤其是需要处理大规模并发数据和高稳定性要求的应用。Unix系统在高负载、高并发场景下表现稳定，能够有效支持5G网络流量数据的实时处理。针对存储系统，建议使用光纤的SAN存储，适合大规模数据的高性能存储。

11.2.2 数据存储与处理

建议采用高密度计算系统来提升数据处理能力，减少空间占用。对于流量数据量较大的情况，采用基于SCSI存储的方案，在不需要高复杂度应用时具有更高的适配性。高密度计算和存储能够为5G网络中实时流量监测提供更强的支持。

11.2.3 云平台兼容性

云平台应具备强大的兼容性，能够支持多种应用接入，尤其是在5G网络中，不同的服务和应用程序会不断增加，平台必须能够无缝接入新类型的应用。建议选择支持开放硬件架构和操作系统的方案，确保平台在未来能够灵活扩展并支持不同类型的5G应用。

11.2.4 系统迁移与平滑过渡

在进行5G通信网络平台的升级或迁移时，云平台应确保系统的迁移不会影响现有服务的正常运行。平台应支持平滑过渡，避免系统大规模的调整和修改，确保现有系统的数据和功能能够无缝过渡。

11.2.5 节能与降耗

为了减少运营成本，云平台需要优化能耗。采用低功耗硬件和优化供电系统，结合高效的散热设计，能够显著减少平台的整体能耗，提高运维效率。特别是在5G通信网络中，数据量庞大，云平台的能源效率对长期运营至关重要。

11.2.6 数据计算分析功能

云平台应具备高效的数据计算和分析功能，特别是对于5G通信网络中实时生成的异常数据进行快速检测。平台应提供以下模块：

a) 数据管理模块：支持数据上传、存储和分类管理，确保 5G 网络中的大数据流量可以高效存储与访问。
b) 信流检测与分析模块：实时分析和处理流量数据，检测异常行为，生成检测报告。
c) 任务管理模块：支持任务分配、执行和监控，确保数据分析的及时性。
d) 用户权限管理模块：细粒度的权限管理，保障数据的安全性。