ICS 35.240
L 60

团　　体　　标　　准

T/××× ××××—××××

大数据平台网络攻击日志的交叉认证要求

The cross-certification requirements for network attack logs of the big data platform

××××－××－××发布　　××××－××－××实施

X X X X X X X X 发 布

目　　次

目次 I
前言 II
引言 III
大数据平台网络攻击日志的交叉认证要求 1
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
3.1 交叉认证 cross-certification 1
3.2 差分攻击 differential attack 1
3.3 重标识攻击 re-identify addack 1
3.4 统计推断攻击 statistical inference attack 1
3.5 原始日志数据 source log data 2
3.6 证据 evidence 2
3.7 关联证据 correlative evidence 2
3.8 关联规则 association rules 2
4 交叉认证概述 2
5 交叉认证工作流程规范 2
5.1 证据生成 3
5.2 关联规则设计 3
5.3 构建关联证据库 3
5.4 证据间交叉认证流程 3
5.5 证据间交叉认证要求 4
6 交叉认证要求 4
6.1 原始日志收集范围 4
6.2 日志要求 4
6.3 证据要求 4
6.4 关联规则要求 5
6.5 关联规则维护要求 5
6.6 后期维护要求 5
6.7 更新管理要求 5
附录 A 6
A．1.1 差分攻击交叉认证 6
A．1.2 重标识攻击交叉认证 6
A．1.3 统计推断攻击交叉认证 6

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。

本文件由中国通信标准化协会提出并归口。

本文件起草单位：华中科技大学网络空间安全学院、西安电子科技大学网络与信息安全学院、国网湖北省电力有限公司武汉供电公司。

本文件主要起草人：戴子城、韩兰胜、李新、孙海丽、秦泽青、朱东君、陈鹏、廖伟、尤伟、张行柯、杨帆、马铭芮、冯铭希。

引　言

本文件由国家重点研发计划子课题大数据平台安全审计与攻击溯源关键技术（2022YFB3103403）支持起草。由于攻击行为可能跨多个应用系统，攻击数据存在于多个系统，每个系统数据格式不一致，攻击行为特征不明显，单源证据证明力不足，针对单一证据证明力不足进行跨系统多来源证据之间的关联挖掘，提出基于多方风险数据关联性证据的交叉认证方法，支撑大数据平台攻击证据的跨域交叉认证，提升证据证明的效力及可信度。

目前在网络仿真中，仍缺少证据交叉认证要求标准，异构证据交叉认证无法通用，因此，亟需制定相关标准予以规范。

大数据平台网络攻击日志的交叉认证要求

1 范围

本文件提供了大数据环境下实施证据交叉认证的要求，简要概述交叉认证使用场景、所需要素、认证流程，并给出交叉认证实现必需的数据和规则信息。

本文件适用于数据监管者实现证据交叉认证过程中使用。

2 规范性引用文件

下列文件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，凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型

ISO/IEC 27037:2012 信息技术-安全技术-数字证据的识别、收集、获取和保存指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

交叉认证 cross-certification

通过多个独立来源的证据数据集之间的相互验证，实现对网络攻击行为的协同印证过程，确保攻击特征在不同维度的证据链中具有一致性表现。

3.2

差分攻击 differential attack

给定两个最多一个元素不同的数据集，攻击者可以通过计算两个数据集之间的差异来获得用户的隐私数据。

3.3

重标识攻击 re-identify attack

使用一个或多个去标识数据集并结合辅助信息，通过跨数据集的数据链接寻找数据重叠，或是结合已有信息进行推断，来将去标识化数据集中的标识信息还原出来，实现数据的重标识的过程。

3.4

统计推断攻击 statistical inference attack

攻击者能够通过统计用户提交的搜索和访问操作，以及其他一些信息，与公开或已知的数据库进行比对，从中恢复出一些关键信息。

3.5

原始日志数据 source log data

可用于证明攻击行为的记录行为等描述信息的日志文件。

3.6

证据 evidence

基于多重校验规则生成的复合证据实体，包含规则匹配标记与置信度评分

3.7

关联证据 correlative evidence

通过关联规则推导产生的逻辑证据，具有行为链标识与上下文关联特征

3.8

关联规则 association rule

将单个证据与其他证据进行匹配的确定模式。

4 交叉认证概述

交叉认证是一种基于多源信息数据的分析方法，其本质是将来自不同数据源的异构数据进行交叉，从而获得更全面、更准确的信息，进而做出更为精准的判断。在大数据安全领域，交叉认证可以被应用于攻击溯源等场景中，利用已有的证据和规则对多个数据源进行交叉验证，以此来推导出更多的相关证据和攻击行为信息。

在构建关联规则数据库的场景中，通过将原始日志数据与对应的关联规则进行交叉认证，可生成合成证据与关联证据，形成多层次证据体系，进而构建关联性证据库。进一步实现安全分析和威胁情报研究，从而帮助安全团队更好地了解攻击者的行为模式和攻击手段，提高安全防范能力。

在攻击检测场景中，由于同一攻击行为可能会留下多种证据，因此通过将大量关联性的证据进行交叉认证，可以由交集得到确定的攻击行为。这种方法可以帮助安全团队快速准确地确认攻击行为，并采取相应的应对措施，从而缩短安全事件的响应时间，降低安全风险。

5 交叉认证工作流程规范

证据的交叉认证工作流程应包含证据生成、关联规则设计、构建关联证据数据库、证据间交叉认证四个部分。

5.1 能源日志采集要求

日志采集要求如下：

a) 必须保存原始日志载体；
b) 电子数据副本需与原始载体哈希值一致；
c) 日志需包含毫秒级时间戳、操作主体、操作对象等元数据，确保行为可回溯；
d) 修改记录必须划改留痕，禁止直接覆盖；

e) 强制采集网络流量、主机事件、应用操作、安全告警四类日志（如电力系统需记录电压/电流时序数据）；
f) 关键操作需记录完整上下文。

5.2 证据生成

为确保能够有效实现关联证据的交叉认证，证据生成要求如下：

a) 证据的基本信息应包含时间、IP地址、用户身份信息、具体系统或服务器名称或编号、攻击类型等；
b) 在生成证据过程中应适当保留日志与证据的关联关系；
c) 应保证生成证据的可靠性和保密性；
d) 数据处理应包含去噪清洗、格式标准化、数据脱敏三个环节。

5.3 关联规则设计

交叉认证关联规则应包含专家设计和算法生成两种，专家通过分析源IP、目的IP、时间、操作行为等关联关系，构建基础关联规则；而算法生成通过将同一事件在不同的层级留下的日志数据关联起来 ，并保存日志数据和关联规则。此外，应根据具体攻击类型具体分析，构建具有差异性的关联规则，便于实现较高准确率的交叉认证。其中专家设计的基础关联规则应包含以下几种：

a) 通过日志属性值之间的关联关系构建的属性关联规则，基于日志属性间的相似性，如:IP地址相同、访问时间临近、目的IP一致等；
b) 通过访问模式特征构建的关联规则，如：访问来源一致、访问频率固定等；
c) 通过用户行为特征构建的关联规则，如：多次登录失败、测试访问权限等；
d) 专家规则设计维度，如：定义时间窗口阈值及地理围栏范围的时空关联、预定义高危行为模式库的行为序列、设置网络标识权重、时间敏感系数的属性权重等；
e) 算法生成规则要求，如基于支持度阈值和置信度阈值的频繁项集挖掘的关联性算法，基于大规模样本训练的自动化规则生成的深度学习模型，以及采用多规则权重投票策略的冲突解决机制。

5.4 构建关联证据库

在构建初始证据集之后，需要结合专家规则与算法规则进行多维度比对，并将日志属性与规则库逐项比对，触发匹配判定，针对每一个初始证据集的需求包括：

a) 需利用事先设计并确定的基础关联规则，以及通过算法生成的关联规则，对其进行对比分析；
b) 应将初始证据集中的待认证证据的属性，与关联规则库中的关联规则进行逐一的比对；
c) 若待认证证据的日志属性中的某个属性项，与关联规则库中的某一关联规则的项集完全吻合，则需判定该待认证证据与该关联规则所对应的初始证据集相匹配；
d) 建立规则匹配度量化指标，如：完全匹配（所有属性符合）、强匹配（关键属性符合数量阈值）、以及弱匹配（基本属性符合数量阈值）；
e) 设置规则冲突解决机制，比如：专家规则优先级高于算法生成规则，在出现冲突时启动三方会审流程。

5.5 证据间交叉认证流程

证据间交叉认证流程可以分为以下三步：

a) 从原始日志中提取攻击特征指纹，包括网络会话哈希、进程行为签名、异常操作模式编码；

b) 在关联规则库中执行多级匹配，优先匹配专家规则中的时空约束条件，其次应用机器学习规则进行行为模式扩展匹配；
c) 构建攻击行为置信度模型，关联证据作为直接证据赋予高权重，关联证据作为间接证据赋予低权重。

5.6 证据间交叉认证要求

对于已确定相匹配的初始关联证据集，需执行以下分析认证要求：
a) 攻击行为一致性检查：深入分析该相匹配的初始关联证据集中的所有证据，确认是否存在与待认证证据所指向的同一攻击行为的证据；
b) 认证结果判定：若存在与待认证证据指向相同攻击行为的证据，则判定为认证成功，若不存在与待认证证据指向相同攻击行为的证据，则判定为认证失败；
c) 交叉认证过程：根据待认证证据所指证的攻击行为，全面遍历初始关联证据集中的每一项证据；
d) 在遍历过程中，若发现有某一关联证据所指向的攻击行为与待认证证据所指证的攻击行为完全一致，即视为交叉认证成功。

6 交叉认证要求

6.1 原始日志收集范围

大数据平台原始日志收集范围应包含多个系统，且各系统数据间存在关联关系，收集数据类型主要为系统运行的日志数据，其中每条日志数据中应记录该操作的IP、时间、设备类型等可以用与交叉认证的信息。具体日志收集要求：
a) 数据类型：多源日志（网络流量、系统事件、应用审计、安全设备日志）；
b) 采集时间精度要求：记录时间、网络标识符、用户身份、操作参数等核心属性；
c) 数据质量保障：完整性保护、可靠性验证、空值处理机制。

6.2 日志要求

对交叉认证的日志数据的要求包括：
a) 日志数据应记录时间、IP地址、用户身份信息、攻击发生的系统或服务器等设施、攻击涉及的参数和数据等信息；
b) 应保证日志数据的完整性、可靠性和保密性；
c) 对于未收集到的属性应以空值存储；
d) 应构建日志数据库。

6.3 证据要求

对证据的要求包括：
a) 证据应对原始日志进行清洗、提纯、标准化等处理和转换，保留高信息熵的属性值，形成可靠的证据实体，如保留时间戳、主题IP、客体IP、操作、所在域、攻击类型、调用进程等可用于交叉认证的属性；
b) 证据库中的证据应随着原始日志的增加同步进行更新；
c) 证据库应存储海量的证据实体；
d) 应构建证据库；
e) 需要对证据进行标准化处理，如：清洗、提纯、属性保留；
f) 并且需要动态更新机制，将证据库与日志库同步更新。

6.4 关联规则要求

对关联规则的要求包括：

a) 关联规则的制定应根据不同攻击类型有所变化；
b) 关联规则应将同一事件在不同的层级留下的日志数据关联起来；
c) 关联规则应根据证据属性而分析；
d) 构建关联规则应符合攻击特性；
e) 关联规则应采取聚类、深度学习等方法构建；
f) 应构建关联规则数据库用于交叉认证。

6.5 关联规则维护要求

关联规则维护机制：

a) 规则有效性周期检测；
b) 规则库更新比例阈值；
c) 规则生命周期管理。

6.6 后期维护要求

应在交叉认证流程完成后，将构建的关联规则定期上传至服务端，并由专业人员定期进行安全检查、规则维护、数据备份等。

版本控制要求：

a) 规则库版本号格式标准化；
b) 历史版本保留策略。

6.7 更新管理要求

更新管理流程：

a) 测试环境验证周期；
b) 低负载窗口部署；
c) 分阶段发布策略。

附录 A

（规范性）

特定隐私攻击交叉认证标准示例

A．1.1 差分攻击交叉认证

检测特征：

a） 查询模式相似性检测：通过相似度算法分析连续查询的结构特征；

b） 响应数据分布异常监测：监控数据发布前后的信息熵变化；

c） 隐私保护机制失效检测：基于隐私保护算法输出的安全状态评估。

关联规则：

a） 专家规则：定义高频相似查询的时空模式特征；

b） 动态规则：采用时序模式分析模型识别查询序列的递进关联性。

验证流程：

a） 提取查询语义特征生成逻辑结构树；

b） 匹配历史攻击模式库中的敏感信息探测特征；

c） 结合身份验证信息与数据分布异常特征综合判定。

A．1.2 重标识攻击交叉认证

检测特征：

a） 准标识符组合风险评估：基于唯一性计算模型分析字段组合强度；

b） 外部数据关联性检测：通过数据链接算法评估跨源匹配风险；

c） 匿名化机制有效性验证：基于去标识化算法的安全参数输出。

关联规则：

a） 专家规则：预定义高危字段组合模式库；

b） 动态规则：应用图结构分析模型构建身份关联概率网络。

验证流程：

d） 执行模拟攻击测试验证数据防护能力；

e） 监测数据使用中的跨源关联操作特征；

f） 基于实际攻击成功率触发防御机制。

A．1.3 统计推断攻击交叉认证

检检测特征：

a） 敏感属性关联性分析：通过多维关联算法评估数据维度相关性；

b） 统计单元安全检测：基于统计归并算法的最小单元保护验证；

c） 逆向推断风险监测：采用概率推断模型评估属性推测风险。

关联规则：

a） 专家规则：定义统计查询的维度叠加模式特征；

b） 动态规则：应用对抗分析模型模拟潜在推断路径。

验证流程：

a） 在数据发布环节植入可追踪特征标记；

b） 分析查询行为中的维度组合演进规律；

c） 基于标记特征还原概率触发主动防御。