ICS 35.110
CCS L 67

YD

中华人民共和国通信行业标准

YD/T XXXXX—202X
[代替 YD/T]

网络安全态势感知　数据采集要求

Cybersecurity situational awareness — Requirements for data collection

（报批稿）

XXXX - XX - XX 发布　　XXXX - XX - XX 实施

中华人民共和国工业和信息化部　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。

本文件是网络安全态势感知系列标准之一，与YD/T 2388-2022《网络脆弱性指数评估方法》、YD/T 2389-2022《网络威胁指数评估方法》共同构成网络安全态势感知的标准体系。该标准体系还计划发布《网络安全态势感知 态势分析要求》、《网络安全态势感知 资源管理要求》、《网络安全态势感知 数据描述要求》、《网络安全态势感知 事件检测要求》《网络安全态势感知 量化评估要求》等标准。

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专利的责任。

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：鹏城实验室、国家计算机网络应急技术处理协调中心、广州大学网络空间先进技术研究院、哈尔滨工业大学（深圳）、四川亿览态势科技有限公司、北京天融信网络安全技术有限公司、北京神州绿盟科技有限公司、郑州信大捷安信息技术股份有限公司、深信服科技股份有限公司、长扬科技（北京）股份有限公司。

本文件主要起草人：贾焰，王海燕，李树栋，韩伟红，李润恒，谢敏容，舒敏，贺敏，王龑，刘为华，赵华，陶莎，肖岩军，谭运强，钟广辉。

网络安全态势感知　数据采集要求

1　范围

本文件规定了网络安全态势感知的数据采集要求。

本文件适用于网络安全态势感知的事件检测、量化评估、态势分析等方面的业务应用。

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 29246-2017　信息技术　安全技术　信息安全管理体系　概述和词汇

3　术语和定义

GB/T 25069-2022、GB/T 29246-2017界定的以及下列术语和定义适用于本文件。

3.1

数据源　data source

网络安全平台中目标网络中的节点，提供节点状态信息、日志数据或流量数据的设备或系统。也称为采集对象。

3.2

采集器　collector

从数据源获取数据的组件或系统。

3.3

日志　log

各类型网络产品的运行记录，记录的内容是未经判断的客观事实。日志通常从网络流量中提取元数据信息而生成，分为系统日志和网络日志两大类。系统日志是由操作系统、应用程序自身生成，记录系统运行的情况。网络日志是网络上发生的行为的记录，通常是根据分析的需要，从网络流量中提取元数据信息生成的。

[GB/T 25068.1—2020，定义 3.5]

3.4

网络流量数据　network traffic data

计算机网络上传输的数据，包括网络协议封装格式信息，以及封装的净负载信息，或基于这些信息不同维度的统计数据。

不同协议层的网络流量数据分类组成不尽相同，IP网络的流量数据在传输层上包括源IP地址、目的IP地址、源端口、目的端口、传输协议类型，以及传输层封装的净负载信息。

3.5

安全告警 security alert

安全告警是安全产品根据对网络流量、日志、扫描探测返回信息等数据的分析结论或基于机器学习、引擎类产品、工具、组件关联分析生成的，描述异常网络情况、异常系统访问或系统脆弱性的信息。

4 缩略语

下列缩略语适用于本文件。

DDoS：分布式拒绝服务（Distributed denial of service）
DNS：域名系统（Domain Name System）
FTP：文件传输协议（File Transfer Protocol）
HTTP：超文本传输协议（HyperText Transfer Protocol）
HTTPS：超文本传输安全协议（Hypertext Transfer Protocol Secure）
IDS：入侵检测系统（Intrusion Detection Systems）
IPS：入侵防御系统（Intrusion Prevention System）
IMAP：交互信息访问协议（Internet Message Access Protocol），
JDBC：Java数据库连接(Java Database connect)
MQTT：消息队列遥测传输协议（Message Queuing Telemetry Transport）
ODBC：开放式数据库互连（Open DataBase Connectivity）
SFTP：安全文件传输协议（Secure File Transfer Protocol）
SMTP：简单邮件传输协议（Simple Mail Transfer Protocol）
SNMP：简单网络管理协议（Simple Network Management Protocol）
SSH：安全外壳协议（Secure Shell）
TCP：传输控制协议（Transmission Control Protocol）
TLS：传输层安全性协议（Transport Layer Security）
TLCP 传输层密码协议（Transport Layer Cryptography Protocol）
VPN：虚拟专用网络（Virtual Private Network)

5 概述

数据源提供的数据采集接口不尽相同，有的提供了专用日志接口，有的只是存放在文件或数据库中。采集器需要根据采集对象的采集接口、采集数据类别情况进行相应的适配处理，采取合适的采集方式。

网络安全态势感知中数据采集的对象为目标网络节点，包括交换机、路由器等网络设备、防火墙、IDS/IPS、VPN、蜜罐等安全设备、应用系统及操作站、服务器等。

按照数据传递方式，采集方式可分为被动采集方式或主动采集方式，被动采集方式是被动地从采集对象接收数据，主动采集方式是主动地向采集对象轮询获取数据。主动采集方式通常有FTP/SFTP主

动采集、ODBC/JDBC、HTTP/HTTPS、SNMP、SSH、Agent代理程序方式等；被动采集方式通常有Kafka、MQTT、Syslog、SNMP TRAP、FTP/SFTP被动方式等。

网络安全态势感知中数据采集的数据类别包括日志数据、网络流量数据、资产数据、安全告警数据以及业务数据等。

6 采集的数据类别要求

收集的类别可根据应用场景不同进行自定义。采集的数据类别应支持以下至少两种不同类型数据的收集：

a) 日志数据

本项要求包括：

1) 应支持对不同类型日志的收集，包括但不限于重要业务日志、操作日志、登录日志、系统日志和告警日志等；

2) 应支持对不同数据来源日志的收集，包括但不仅限于安全设备、网络设备、操作系统、数据库管理系统、终端、应用系统等。

b) 网络流量数据

本项要求包括：

1) 应支持对全网网络流量数据的实时收集，数据内容应包括但不限于时间、源IP、源端口、目的IP、目的端口、协议等；

2) 应支持对协议的识别、统计，包括但不限于TCP、HTTP、FTP、IMAP、SNMP、TELNET、DNS、SMTP等；

3) 应支持从原始数据包中还原不同文件格式的样本，包括压缩包、可执行文件等。

c) 资产数据

本项要求包括：

1) 应支持对网络中资产数据的收集，并进行定期更新，资产数据类型包括但不限于网站、信息系统等；

2) 应支持对资产数据多维度属性的收集，数据内容包括但不限于资产名称、资产类型、资产IP地址、资产IP 归属地、域名、资产重要程度、所属关系、数量等。

d) 安全告警数据

本项要求包括：

1) 应支持不同类型网络攻击事件（见附录A）的收集，包括但不限于DDOS 攻击、后门攻击、漏洞攻击、网络扫描窃听、干扰事件等；

2) 应支持不同类型有害程序事件的收集，包括但不限于计算机病毒、蠕虫、特洛伊木马、僵尸网络、混合攻击程序、网页内嵌恶意代码事件等；

e) 业务数据

本项要求包括：

1) 应支持应急处置业务场景数据的收集，包括但不限于处置启动时间、结束时间、处置类型、处置人员、处置事件描述、处置方式、处置结果等；

2) 应支持通报预警业务场景数据的收集，包括但不限于通报下发时间、通报对象名称、通报类型、通报人员、通报事件描述、通报反馈期限、通报反馈结果等。

7 数据采集总体要求

数据采集总体要求如下：

a) 数据采集的对象和采集内容应满足数据分析的要求，避免数据内容的缺失，但也无需过度采集，满足要求即可，避免对系统性能造成影响。
b) 数据采集传输时宜采用可靠安全的传输方式，如基于TLS/TLCP加密传输等。
c) 采集通道出现故障时，采集器应进行告警，故障恢复时采集工作应自动恢复。
d) 采集器应及时采集数据源产生的数据，满足及时性要求。
e) 采集的数据提取后，宜以JSON等格式提供给后续分析模块使用。

8 日志数据采集要求

8.1 采集对象

应支持包括交换机、路由器等网络设备、防火墙、IDS/IPS、VPN、蜜罐等安全设备、靶标应用系统、操作站、服务器等的数据采集。

8.2 采集方式

通常日志类数据采集方式有FTP/SFTP、ODBC/JDBC、Kafka、MQTT、HTTP/HTTPS、Syslog、SNMP TRAP、SNMP、SSH、Agent代理程序方式，采集方式至少支持一种被动式和一种主动式的日志采集方式，建议优先采用被动式日志采集方式。

采集方式应支持扩展。

8.2.1 主动式采集要求

主动式采集要求如下：

a) 主动式数据采集，应支持采集对象上的多个数据文件、或多个数据表等多个数据存放处的数据采集。
b) 主动采集时，应支持采集频率、采集数据量等的配置或自适应调整，以减轻对采集对象的性能影响，避免对网络造成拥堵，也需避免数据积压过多，影响数据采集的及时性（尤其是系统刚上线时存在大量历史数据的情况）。
c) 主动采集时应在采集端维持采集数据的读取指针，不能遗漏数据，或重复采集造成数据冗余。
d) 在采集方式选取上，根据网络情况，应当选择安全的采集方式，如SFTP、HTTPS、SSH，避免数据被窃听。
e) Agent代理程序采集方式，应限制日志采集代理程序使用的内存和CPU资源，避免对采集对象造成影响。CPU资源占用不宜超过5%，内存使用不宜超过10%。
f) Agent日志采集代理应当将采集的日志数据及时发送到采集器，发送通道需要安全可靠，当通道发生网络故障时，应将日志数据在本地缓存，网络恢复时继续传输，避免数据的丢失。

8.2.2 被动式采集要求

被动式数据采集时，应满足日志接收的性能要求，如采用缓存机制，避免日志源数据发送过快时造成日志数据丢失。

9 流量数据采集要求

9.1 采集方式

流量数据常用的采集接口有SNMP、Netflow、Netstream、SFlow、JFlow、镜像侦听等。

流量数据采集的采样方式采集和全量方式采集方式中应至少支持一种，采集方式的选择应满足数据分析的需求。

采样采集方式宜采用大多数设备厂商都支持的采集接口。

9.2 采集内容

针对采样采集方式，采集内容应包括协议五元组内容，即源IP地址、目的IP地址、源端口、目的端口和协议类型信息。还应包括网络流基础信息，如网络流起止时间、网络流中的总字节数及上下行字节数、网络流中的数据包数量及上下行数据包数量、网络流中的第一个和最后一个数据包时间戳等。

针对流量全量采集方式，应支持DNS、DHCP、HTTP、FTP、TELNET、SMTP、POP3、IMAP、SSH、TLS、TLCP 等常用协议内容的采集，应采集到这些协议网络层和应用层协议的内容。如针对电子邮件协议SMTP、POP3、IMAP等，应采集到邮件的发送人、接收人、邮件主题、邮件内容、邮件附件名称及内容等数据。

针对工业互联网的流量数据采集，应支持OPC、Modbus TCP、Profinet、BACnet、DNP3、IEC104等常用工业控制系统的协议内容的采集提取。

9.3 流量数据采集预处理

流量数据采集应支持按照采集策略进行筛选处理，如果按照源或目的IP地址，源或目的端口，或者协议类型等策略进行筛选处理。

应支持流量数据的去重处理。

10 资产数据采集要求

10.1 采集对象

包括硬件及型号、操作系统及版本号、软件及版本号等资产信息。其中，资产包括主机节点、网络设备节点、安全设备节点等，包括且不限于：服务器、计算机等；交换机、路由器等；防火墙、入侵检测、防病毒等。

10.2 采集内容

网站类数据、信息系统类数据、关键信息基础设施类数据、移动应用类数据、数据库类数据

等。

10.3 采集方式

根据采集数据来源分类：主动采集、被动采集。

10.3.1 主动采集

主动采集是通过主动向目标网络资产发送构造的数据包，并从返回数据包的相关信息(包括各层协议内容、包重传时间等)中提取目标指纹，与指纹进库中的指纹进行比对，实现对开放端口、操作系统、服务及应用类型的数据进行探测。

10.3.2 被动采集

被动采集是采集目标网络的流量，对流量中应用层 HTTP、 FTP、SMTP 等协议数据包中的特殊字段 banner 或 IP、TCP 三次握手、DHCP 等协议数据包的指纹特征进行分析，实现对网络资产信息的被动探测。

附 录 A
（资料性）
数据采集功能架构示例

数据采集架构由采集器构成，主要功能模块包括：数据提取、数据预处理、采集配置管理、数据缓存等部分，如图A.1所示。

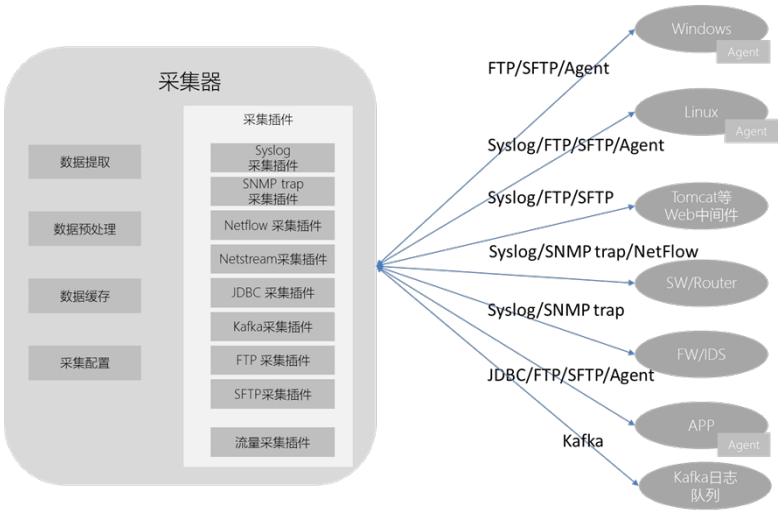
采集器需考虑到数据采集的灵活性和扩展性，可采用插件的方式。采集插件根据采集对象、采集内容和采集接口等进行相应的设计，实现数据提取和预处理的功能。采集插件可以是独立的采集代理程序，也可以是不同的采集插件。对于数据量较大的网络流量数据，宜采用单独的流量采集插件。

数据提取模块，将采集的原始数据进行抽取和格式化处理，解析出具有明确意义的字段。数据提取的粒度粗细取决于数据分析使用的需要。

数据预处理模块，将采集的重复数据进行清洗、数据过滤处理、或者数据补全处理（如添加标签、补充来源名称等）等。

采集配置管理模块，主要配置采集的数据源采集方式及其接口信息（数据源IP地址、认证方式等），采集的策略，数据提取的配置等，并对采集插件等进行管理。

数据缓存，针对采集提取到的数据进行缓存处理，提供给后续的数据分析模块使用。通常可采用内存数据库或Kafka等消息队列方式缓存。



图A.1 数据采集架构

参 考 文 献

[1] 《网络安全态势感知技术标准化白皮书》，（2020 版）.
[2] 《网络安全态势感知》，贾焰，方滨兴等编著，中国工信出版集团，2020.
[3] GB/T 20278-2013 信息安全技术 网络脆弱性扫描产品安全技术要求
[4] GB/T 20280-2006 信息安全技术 网络脆弱性扫描产品测试评价方法
[5] GB/T 20984-2007 信息安全技术 信息安全风险评估规范
[6] GB/T 31509-2015 信息安全技术 信息安全风险评估实施指南
[7] GB/T 31722-2015 信息技术 安全技术 信息安全风险管理
[8] YD/T 2252-2011 网络与信息安全风险评估服务能力评估方法
[9] YD/T 2707-2014 互联网主机网络安全属性描述格式
[10] YD/T 3153-2016 Web 应用安全评估系统技术要求
[11] YD/T 3463-2019 漏洞扫描系统通用技术要求
[12] YD/T 2388-2022 网络脆弱性指数评估方法
[13] YD/T 2389-2022 网络威胁指数评估方法