威胁行为分析工具概要设计

# 1.开发目标

面向网络威胁行为实时检测和深度检测的需求，研究基于主机日志、网络设备日志和应用系统报警日志的多源异构数据汇聚技术和关联分析方法，通过构建攻击行为分类模型和攻击阶段特征模型，实现对攻击行为的属性定义和过滤识别，设计可扩展的检测分析引擎，开发实现一款配置方便、部署灵活的网络威胁行为分析工具软件。

# 2.总体设计

## 2.1需求规定

该软件应用于具备多种网络安全设施的网络环境，以黑盒方式部署，通过授权和自身配置获得已有网络防护软件的安全监测数据（包括报警记录和日志记录等），并在多类型、多层次检测和分析结果基础上，提取威胁行为的阶段过程、典型特征、危害分类等技术特征，挖掘攻击者的攻击路径、攻击目标、攻击能力和攻击习惯等属性信息，从而为企业用户提供安全威胁关键信息参考，为信息安全防护提供支持。

在应用功能方面要求该软件具备良好的现有安全设施或安全防护软件的兼容能力，并通过简单配置建立对已有安全监测数据的实时采集和汇聚能力；具备针对网络攻击行为的技术属性分析识别，包括攻击方法、攻击阶段、攻击危害等；具备对攻击者层次属性信息的关联挖掘能力，包括攻击目标、攻击范围、攻击路径、攻击能力、攻击习惯等信息的统计分析；具备支持分析方法的插件式扩展的框架结构。

在用户基本使用体验方面要求具备威胁报警分级控制能力，关键报警数应小于1000条/天，报警信息按照由低到高可划分为多个数据层次（系统层、动作层、行为层、攻击者层）对威胁报警分级，逐层向上支撑，对下抽取，形成多层报警的协同管理；能够根据威胁模型标示出不同层次信息的支撑关系和组织关系；攻击模型发生修改和升级时，报警信息可以重新生成，更新对威胁行为的分析识别结果。

在安装部署方面要求该软件工具经过配置能以黑盒方式部署在可正常通信的任何内网位置；支持通用安全防护设备和系统的日志配置接收；支持通用安全防护监测数据主动采集；支持对特定系统监测数据的主动采集；支持Web方式管理。

## 2.2基本设计概念

### 2.2.1技术架构

本软件由分析服务器和管理界面组成，分析服务器包括数据汇聚模块和关联分析模块。这个结构包含了日志分析的完整过程，包括数据的采集、归一化、分析、输出和展示。

软件整体分为三个层次，分别是：数据汇聚层、关联分析层、管理界面层。

数据汇聚层实现多源数据的汇聚功能，包括数据采集、数据识别、归一化。用户通过编写数据汇聚规则，接入要监测的数据源，采集监测数据，对数据进行识别，通过威胁信息交换协议实现数据的归一化，为关联分析层提供数据支持。

关联分析层，是软件的核心层，实现数据的分析过程，组织建立关联分析算法调度和执行框架。分析算法以插件形式加载，通过实现不同的关联分析算法插件，实现对关联分析能力的扩展。根据业务需求，选择加载相应的关联分析插件，建立关联分析任务，完成业务数据分析。用户通过管理界面的菜单选项实现分析任务的调度，分析任务是针对具体业务的一次分析计算执行过程，对应一种具体的分析模型，插件是分析模型的具体实现。

管理界面层实现用户与软件的交互功能，主要包括：综合监控、行为分析、配置管理。综合监控集成实时报警、历史统计和威胁趋势评估，提供全局展示；行为分析辅助进行局部行为分析和还原；配置管理作为管理支撑，实现用户对软件的配置输入，包括数据汇聚规则信息配置、插件管理配置信息等。



图2-2-1 技术架构图

### 2.2.2业务处理流程

本工具的业务流程分为三个模块，分别是多源数据汇聚模块、关联分析模块、管理界面模块。多源数据汇聚模块通过数据汇聚规则和威胁信息交换协议，采集多源异构数据并进行归一化，为关联分析模块提供数据支持。关联分析模块通过插件管理和任务调度，接收数据汇聚模块传输来的数据，进行具体的关联分析任务，得到分析结果，保存到数据库中。管理界面模块实现数据的展示、分析和配置管理，其主要功能是作为关联分析引模块析结果的输出，读取数据库数据，展示在用户界面中。本软件的业务处理流程如图2-2-2所示。

图2-2-2 业务处理流程图

多源数据汇聚模块的逻辑流程如图2-2-3所示，首先初始化：加载数据汇聚规则；然后执行任务：根据数据汇聚规则，对要监听的数据源进行数据采集；采集到数据之后要对数据进行归一化处理，定义一个威胁信息交换协议，把采集数据按照层次进行划分，转换为统一格式；最后发送到威胁信息库中，等待分析任务调度。

图2-2-3 多源数据汇聚模块逻辑流程图

关联分析引擎通过任务调度加载相应的插件，进行关联分析操作。任务调度的处理流程如图2-2-4所示。首先初始化：加载插件；然后根据插件的具体要求选择数据池中的数据，不同需求的插件分析的数据不同；获取数据之后，执行插件处理操作，进行关联分析，得到分析结果，保存到数据库中。

管理分析模块实现分析数据的界面展示和用户的配置管理。管理界面的主要功能包括：综合监控、实时报警和配置管理；此外，用户还可以通过配置插件列表，对插件进行增删改等操作，实现更多的业务功能，如威胁趋势分析，历史统计，行为分析，攻击者分析等。



图2-2-4 关联分析引擎任务调度处理流程

## 2.3总体结构与模块设计

### 2.3.1总体结构设计

如图2-3-1所示，本工具的总体结构包括三个部分：数据汇聚模块、关联分析模块、管理界面模块。

数据汇聚层主要为分析工具提供数据支持，通过加载数据汇聚规则，采集数据，进行数据解析，然后根据威胁信息交换协议对数据进行归一化处理，对处理过的数据进行归并，减少数据量。

关联分析层是对采集到的数据进行分析，是本工具的核心层。关联分析引擎基于插件实现多种关联分析方法的并行操作，通过任务调度加载关联分析插件，执行关联分析任务，得到分析结果，保存到数据库中。

管理界面层，提供用户接口，实现用户操作，包括综合监控、行为分析、配置管理。



图2-3-1总体结构设计

### 2.3.2数据汇聚模块

数据汇聚模块，实现多源异构数据的采集、识别、归一化和归并功能。设计一套多源数据接口协议，通过定义数据采集接口规则实现对多源数据的并行采集，兼容通用安全监测系统数据接口并支持数据接口的动态扩展。

设计威胁信息交换协议，对采集到的安全监测数据进行转换和协议封装，实现多层次监测分析数据的归一化，为进一步的威胁行为分析提供数据支持。

数据汇聚模块的结构如图2-3-2所示，输入为来自不同数据源的日志数据，向关联分析引擎输出数据。



图2-3-2多源数据汇聚模块结构图

数据汇聚模块包括四部分：数据采集、数据识别、归一化、数据归并。

数据采集：通过加载多源数据汇聚规则，得到采集任务。当有采集任务时，轮询数据源的数据，进行数据采集，对采集到的数据进行格式化，转换为内存数据。

数据识别：对采集到的数据识别，得出数据源的类型（主机日志、网络设备日志、应用系统报警日志）以及数据所处威胁信息交换协议的层次，为数据归一化提供依据。

数据归一化，是把采集到的数据通过协议封装，转换为统一的格式，利用威胁信息交换协议实现。威胁信息交换协议把数据分为四个层次：系统层、动作层、行为层、攻击者层。每个层次的数据对应不同的数据源，四个层次表示的数据抽象层次递增。系统层表示主机日志和网络设备日志数据，动作层表示应用系统的报警数据和经过一级分析的攻击动作属性信息，行为层表示关联分析得到的攻击行为特征信息，攻击者层表示个人或个体表现的特征信息，如攻击者的习惯、熟悉的工具和技术、以及其国籍、宗教、民族等属性。威胁信息交换协议中还包括数据解码规则和一级分析规则。数据归一化实际实现的是对底层数据的预处理。

数据归并：利用属性相似度算法，判断相似事件，对相似数据进行聚类归并，从而减少关联分析的数据量。

### 2.3.3关联分析模块

关联分析模块的核心是任务调度，通过任务调度加载插件库，读取数据池中的数据，执行对应插件的处理操作，最终返回结果。在执行的过程中，还可以产生进度报告和程序报错信息。任务调度通过用户界面实现，不同的分析菜单选项对应不同的插件。在分析过程中，凡是加载到插件库中的插件，都自动执行，得到分析结果保存到数据库中，用户根据具体业务需求去数据库中检索数据。

图2-3-3 关联分析引擎结构图

关联分析引擎通过任务调度，实现具体的关联分析操作。任务调度的实现流程如图2-3-4所示，通过任务调度接口，加载任务需要用到的插件；并根据分析插件的具体要求，选择数据池中的数据。插件和数据都准备好之后，开始执行插件操作，调用插件的关联分析算法，对数据过滤器中的数据进行分析；分析完成后，返回结果，输出到数据库中。



图2-3-4 任务调度实现流程图

### 2.3.4特征序列关联分析插件

特征序列关联分析插件通过匹配同一个攻击主体行为的特征序列，来判断攻击者的攻击阶段。基于特征序列的关联分析技术的实现，主要是对关联规则进行匹配。把同一个攻击主体的所有相关数据进行统计，然后对这些数据进行特征序列关联规则匹配，当匹配成功时，发出告警信息，如果不能完全匹配，则发送到交互界面，用户来研判是否属于攻击行为。

实现流程如下图所示。



图2-3-8 特征序列关联分析流程图

关联规则匹配流程：



图2-3-9关联规则匹配流程

匹配一条关联规则时，首先匹配初始特征，只有初始特征匹配成功，才继续匹配下一层的规则，直到最后一层，如果全部匹配完成，则关联规则匹配成功，否则匹配失败。

# 3.接口设计

## 3.1内部接口：数据结构（ossec）

内部接口表示软件内部实现的数据结构，包括：威胁信息结构体、解码规则、特征提取规则。

### 威胁信息结构体

|  |  |  |
| --- | --- | --- |
| **属性名称** | **类型** | **含义** |
| log | char\* | 原始数据的日志内容：不包含主机名+进程名 |
| full\_log | char\* | 原始数据的整条日志内容：包含主机名和进程名 |
| hostname | char\* | 主机名 |
| location | char\* | 数据源：发送数据的机制（目录、进程、模块） |
| program\_name | char\* | 进程名 |
| srcip | char\* | 以上为预处理得到，以下为解码得到的日志内容信息 |
| dstip | char\* |  |
| srcport | char\* |  |
| dstport | char\* |  |
| protocol | char\* |  |
| action | char\* |  |
| srcuser | char\* |  |
| dstuser | char\* |  |
| id | char\* |  |
| status | char\* |  |
| command | char\* |  |
| url | char\* |  |
| data | char\* |  |
| systemname | char\* |  |
| generated\_rule | RuleInfo\* | 匹配到的特征提取规则 |
| decoder\_info | OSDecoderInfo | 匹配到的解码规则 |
| sid\_node\_to\_delete | OSListNode\* | 要释放的节点 |
| size | int | 日志大小 |
| p\_name\_size | int | 进程名大小 |
| matched | int | 匹配标志 |
| time | int | 时间 |
| day | int | 天 |
| year | int | 年 |
| hour[10] | char | 小时 |
| mon[4] | char | 月份 |
| filename | char\* | 以下为SYSCHECK要用到的属性 |
| perm\_before | int |  |
| perm\_after | int |  |
| md5\_before | char\* |  |
| md5\_after | char\* |  |
| sha1\_before | char\* |  |
| sha1\_after | char\* |  |
| size\_before | char\* |  |
| size\_after | char\* |  |
| owner\_before | char\* |  |
| owner\_after | char\* |  |
| gowner\_before | char\* |  |
| gowner\_after | char\* |  |

### 解码规则

|  |  |  |
| --- | --- | --- |
| **属性名称** | **类型** | **含义** |
| get\_next | u\_int8\_t | 下一条规则 |
| type | u\_int8\_t | 解码类型 |
| use\_own\_name | u\_int8\_t | 标识：是否独立编号 |
| id | u\_int16\_t | 规则编号 |
| regex\_offset | u\_int16\_t | 正则表达式的匹配选项值 |
| prematch\_offset | u\_int16\_t | 字符串匹配的的匹配选项值 |
| fts | int | 第一次判断的标识 |
| parent | char \* | 父节点 |
| name | char \* | 名称 |
| ftscomment | char \* | ftc描述 |
| regex | OSRegex | 匹配的正则表达式 |
| prematch | OSRegex | 匹配的字符串 |
| program\_name | OSRegex | 进程名称 |
| plugindecoder | void | 插件解码 |
| order | void | 字段的顺序 |

### 特征提取规则

|  |  |  |
| --- | --- | --- |
| **属性名称** | **类型** | **含义** |
| sigid | int | 规则编号 |
| level | int | 报警级别 |
| maxsize | int | 日志大小 |
| frequency | int | 频率 |
| timeframe | int | 时间范围 |
| context | u\_int8\_t | 内容 |
| firedtimes | int | 匹配成功的时间 |
| time\_ignored | int | 忽略时间 |
| ignore\_time | int |  |
| ignore | int |  |
| ckignore | int |  |
| group\_prev\_matched\_sz | int |  |
| \_\_frequency | int |  |
| last\_events | char \*\* |  |
| alert\_opts | u\_int16\_t |  |
| context\_opts | u\_int16\_t |  |
| category | u\_int8\_t |  |
| decoded\_as | u\_int16\_t |  |
| sid\_prev\_matched | OSList \* |  |
| sid\_search | OSList \* |  |
| group\_prev\_matched | OSList \*\* |  |
| group\_search | OSList \* |  |
| event\_search | void |  |
| group | char \* |  |
| match | OSMatch \* |  |
| regex | OSMatch \* |  |
| day\_time | char \* |  |
| week\_day | char \* |  |
| srcip | os\_ip \*\* |  |
| dstip | os\_ip \*\* |  |
| srcport | OSMatch \* |  |
| dstport | OSMatch \* |  |
| user | OSMatch \* |  |
| url | OSMatch \* |  |
| id | OSMatch \* |  |
| status | OSMatch \* |  |
| hostname | OSMatch \* |  |
| program\_name | OSMatch \* |  |
| extra\_data | OSMatch \* |  |
| action | char \* |  |
| comment | char \* |  |
| info | char \* |  |
| cve | char \* |  |
| info\_details | RuleInfoDetail \* |  |
| lists | ListRule \* |  |
| if\_sid | char \* |  |
| if\_level | char \* |  |
| if\_group | char \* |  |
| if\_matched\_regex | OSMatch \* |  |
| if\_matched\_group | OSMatch \* |  |
| if\_matched\_sid | int |  |
| compiled\_rule | void \* |  |
| ar | active\_response \*\* |  |

## 3.2外部接口：输出的数据库表结构

外部接口表示软件输出的数据的存储接口，本软件中数据存储到数据库中，设计数据库表结构。

# 4.模型设计

## 4.1攻击行为模型设计

攻击行为模型分为三个层次：系统层、动作层、行为层。三个层次分别在不同的维度对攻击行为进行分类描述。

攻击行为模型最底层的是系统层，原始数据表现出的数据属性能反映出动作层次的攻击动作。位于中间层的动作层次的分类树，可以根据分类依据找到攻击的属性，包括攻击阶段段、攻击手段、攻击动作等。分析攻击层次的动作属性进行关联分析，得到个行为特征，与阶段模型中的特征序列进行匹配，可以得到攻击行为目前所处的攻击阶段，通过特征序列可以得到攻击的过程。

三个层次分别在数据属性、动作属性、行为特征三个不同的维度对攻击行为进行定义，比较全面的分解攻击行为，使得原始日志数据通过一层层的映射，反映出目前网络中的威胁行为，以及网络安全趋势。攻击行为模型如图4-1所示。



图4-1攻击行为模型

定义三个层次的属性，位于最底层的是数据属性，中间层的为动作属性，位于最高层的是行为特征。数据属性表示底层系统数据的基本属性，定义如下表：

|  |  |
| --- | --- |
| **数据属性** | **说明** |
| 数据来源 | 数据的来源，表示数据的产生机制，如主机、网络、监测系统 |
| 主机名 | 数据标记的主机名 |
| 进程名 | 数据记录的进程名，说明是哪个进程产生的数据 |
| 源IP | 日志内容中的源IP地址 |
| 目的IP | 日志内容中的目的IP地址 |
| 源端口号 | 日志内容中的源端口号 |
| 目的端口号 | 日志内容中的目的端口号 |
| 协议 | 日志内容中包含的协议名称 |
| 动作 | 日志内容中包含的动作内容 |
| 源用户名 | 日志内容中的源用户名信息 |
| 目的用户名 | 日志内容中的目的用户名信息 |
| ID | 日志内容中的编号信息 |
| 状态 | 日志内容中记录某些进程、服务等的状态 |
| 命令 | 日志内容中记录的命令 |
| url链接 | 日志内容中记录的url链接 |
| 数据 | 日志内容中包含的数据 |
| 系统名称 | 日志内容中包含的系统名称 |

动作属性，是根据数据属性提取出的比数据属性层次高一级的属性，能表示出攻击的信息，定义如下表所示：

|  |  |
| --- | --- |
| **动作属性** | **说明** |
| 攻击阶段 | 动作属性所处的攻击阶段：探测、入侵、控制、操作 |
| 攻击手段 | 攻击使用的攻击手段：ping探测、协议连接探测、 |
| 动作描述 | 属于哪种攻击动作，具体的攻击名称：如SQL溢出 |
| 威胁级别 | 用数字表示威胁级别 |
| 频率 | 统计一段时间内，相同记录的条数 |

行为特征，表示动作属性继续提取出的表示攻击行为的高层特征，定义如下表所示：

|  |  |
| --- | --- |
| **行为特征** | **说明** |
| 方式 | 采用何种方式，不同的阶段有不同的攻击方式，如探测阶段包括：在线探测、服务探测、漏洞探测 |
| 范围 | 攻击的广度，一般为主机数：1个主机、多个主机、全部主机 |
| 频度 | 攻击发生的次数：高、低 |
| 程度 | 攻击的深度，深、浅、一般 |
| 顺序 | 攻击多个主机、端口、或服务时，是否有顺序性 |
| 结果 | 攻击的结果：成功、失败 |

行为特征中的范围、频度、程度，都可以通过利用统计方法实现。

最后行为特征的序列，对应事件序列规则，进行匹配，重构攻击场景，可以得到攻击目前所处阶段、攻击程度、威胁级别等信息。

## 4.2关联分析模型设计

关联分析的目的是得到目前攻击的程度和所达到的阶段，以及威胁趋势。关联分析方法主要分为两个方面：横向的数据关联分析和纵向的阶段关联。

数据关联，是把来自主机、网络设备、应用的不同数据进行交叉关联，来实现对某一个攻击行为属性的分析，得到数据表示的攻击阶段。

阶段关联，是通过攻击行为模型，分析攻击进行到哪一个阶段，以及达到了什么样的程度和产生的影响。



图4-2 关联分析模型

此关联分析模型可以得到目前攻击者攻击的程度和深度，实现实时报警和预警功能。数据之间的横向关联分析通过威胁信息交换协议实现，在4.4中有详细介绍。

## 4.3多源数据接口协议模型

多源数据接口协议通过定义数据采集接口规则实现对多源数据的并行采集，兼容通用安全监测系统数据接口并支持数据接口的动态扩展。

数据采集规则包括三部分：数据源接口、数据连接信息、数据存储格式。具体规则设计如下表所示。

|  |  |  |  |
| --- | --- | --- | --- |
| **规则** | **关键字** | **说明** | |
| 数据源接口 | enablename | 需要轮询的数据源名称：SpecTrojan、DataLeak | |
| type | 数据源类型：系统层日志、报警日志 | |
| 数据连接信息 | name | 连接的名称 | |
| type | 连接的类型：mysql、file、sock、（oracle、sqlserver） | |
| db | hostname | 数据库所在主机IP地址 |
| dbname | 数据库名称 |
| dbuser | 用户名 |
| password | 密码 |
| file | path | 文件路径 |
| sock | domain | 协议域：AF\_INET、AF\_INET6、AF\_LOCAL |
| socktype | socket类型：SOCK\_STREAM、SOCK\_DGRAM |
| port | 端口号 |
| protocol | 协议：UDP、TCP |
| 数据存储格式 | dataname | 数据类型：SpecTrojan、DataLeak | |
| connection | 连接名称，用来指定使用哪个连接 | |
| tablename | 表名 | |
| keyfield | 主键ID，用来查询到最大ID号 判断是否有新数据 | |
| fieldname | 数据表中的字段名（可以有很多个fieldname） | |

## 4.4威胁信息交换协议模型

威胁信息交换协议模型包括四层：系统层、动作层、行为层、攻击者层。

系统层数据位于协议的最底层，表示主机和网络设备的原始日志，是最底层的日志信息，只包含数据属性。

动作层数据位于协议的第二层，表示监测系统产生的报警数据和系统层数据经过一级分析得到的数据，这些数据比系统层数据高级，表示一次攻击的动作。

行为层数据，是经过关联分析之后，得到的攻击特征序列，对应攻击分类树中各层的攻击行为属性。

攻击者层次位于协议的最高层，表示攻击者的属性，包括攻击者个体的特征信息，如攻击能力、攻击习惯、常用工具及技术、国籍、宗教等等。



图4-3威胁信息交换协议模型

威胁信息交换协议元素：

|  |  |  |  |
| --- | --- | --- | --- |
| **元素** | | **数据类型** | **说明** |
| 数据协议类型 | | int | 表示数据所处信息交换协议的层次：0、1、2、3分为代表日志、动作、行为、攻击者 |
| 攻击主体 | | char | 定义的供给主体名称（搜集总结得到） |
| 时间 | | Time | 数据记录的发生时间 |
| 源IP地址 | | long | 攻击者主机的IP地址 |
| 目的IP地址 | | long | 被攻击者主机的IP地址 |
| 日志信息 | 源端口号 | int | 日志内容中的源端口号 |
| 目的端口号 | int | 日志内容中的目的端口号 |
| 协议 | char | 日志内容中包含的协议名称 |
| 动作 | char | 日志内容中包含的动作内容 |
| 源用户名 | char | 日志内容中的源用户名信息 |
| 目的用户名 | char | 日志内容中的目的用户名信息 |
| ID | int | 日志内容中的编号信息 |
| 状态 | char | 日志内容中记录某些进程、服务等的状态 |
| 命令 | char | 日志内容中记录的命令 |
| url链接 | char | 日志内容中记录的url链接 |
| 数据 | char | 日志内容中包含的数据 |
| 系统名称 | char | 日志内容中包含的系统名称 |
| 动作信息 | 攻击阶段 | char | 处于攻击的哪一阶段 |
| 攻击手段 | char | 使用的攻击手段 |
| 动作分类 | char | 属于哪种攻击行为 |
| 威胁级别 | int | 威胁的报警级别 |
| 频率 | int | 统计相同记录的条数 |
| 行为信息 | 方式 | char | 使用的攻击方式 |
| 范围 | char | 攻击的广度 |
| 频度 | int | 统计相同攻击的次数 |
| 程度 | char | 攻击的深度 |
| 顺序 | char | 是否有顺序性 |
| 结果 | char | 攻击结果：成功、失败 |
| 攻击者 | 攻击能力 | int | 攻击者的威胁程度 |
| 针对性 | char | 攻击习惯，是否有针对性 |
| 常用工具 | char | 攻击者常用的工具 |
| 常用技术 | char | 攻击者常用的技术 |
| 国籍 | char | 攻击者的国籍（位置） |

威胁信息的数据格式解释：

1. 数据协议类型：日志、动作、行为、攻击者。分别代表威胁信息交换协议模型中的四个层次。在数据报中用数字表示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 协议类型 | 日志 | 动作 | 行为 | 攻击者 |
| 协议域数值 | 0 | 1 | 2 | 3 |

2. 攻击主体：攻击者的标识名称。

一个攻击者可以对应一个地区的多个IP地址，通过搜集总结，我们得到攻击者和源IP地址的对应关系。如果无法对应，则攻击主体为空。维护统计攻击者信息如下表所示：

|  |  |
| --- | --- |
| **攻击主体** | **IP地址** |
| 001 | 10.10.10.1 |
| 10.10.10.2 |
| …… |
| 002 | …… |
|  |
| …… | …… |

3 时间：威胁信息记录的发生时间。

4. 源IP地址：表示攻击者发出攻击的源主机IP地址。

5. 目的IP地址：表示被攻击主机的IP地址。

源IP地址和目的IP地址表现出攻击的方向。

6. 日志信息：

如果协议类型为0，则数据包中只有日志信息。日志信息包括源端口号、目的端口号、源用户、目的用户、日志记录。这些信息都是从原始日志数据中提取出来的对应字段的信息。

7. 动作信息：日志信息经过一级分析，得到动作信息，动作信息包括动作名称、动作在攻击属性中的分类、威胁级别。

8. 行为信息：动作信息经过关联分析，得到行为信息，行为信息包括行为名称、行为在攻击特征模型中的分类、威胁级别。

9. 攻击者信息：攻击者信息是通过大量数据的搜集、分析、总结得到的对攻击者的认识。此处把攻击方看作一个个体或集体，描述攻击者的特征信息，如攻击能力、攻击习惯、常用工具、常用技术、国籍。攻击者信息是威胁信息交换协议中最高层的数据，是分析的最终目的，本模型中暂时只分析到第三层的行为信息，攻击者信息作为后期研究工作。

威胁信息交换协议中的数据进行关联分析时，包括同层间数据的横向关联和上下层之间的纵向关联，关联模型如图4-4所示。



图4-4 信息交换协议数据关联模型

# 5.附录：分析实例

## 5.1 日志数据：数据属性

ossec: output: 'netstat -tan |grep LISTEN |grep -v 127.0.0.1 | sort':  
tcp 0 0 0.0.0.0:111 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:21 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:22 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:3306 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:36975 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:4567 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:4730 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:48869 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:5672 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:5901 0.0.0.0:\* LISTEN   
tcp 0 0 0.0.0.0:6001 0.0.0.0:\* LISTEN   
tcp 0 0 192.168.122.1:53 0.0.0.0:\* LISTEN   
tcp 0 0 ::1:631 :::\* LISTEN   
tcp 0 0 :::111   
数据属性：

|  |  |
| --- | --- |
| 目的端口号 | 111、21、22、3306、36975、4567、4730、48869、5672、5901、6001 |
| 协议 | tcp |
| 动作 | connection |
| ID | 11 |
| 状态 | LISTEN |
| 命令 | netstat |

Oct 20 14:50:02 dlpserver81 sshd[16882]: Failed password for root from 192.168.1.65 port 39157 ssh2

Oct 20 14:49:58 dlpserver81 sshd[16882]: Failed password for root from 192.168.1.65 port 39157 ssh2

Oct 20 14:49:55 dlpserver81 sshd[16882]: Failed password for root from 192.168.1.65 port 39157 ssh2

Oct 20 14:49:38 dlpserver81 sshd[16869]: Failed password for root from 192.168.1.65 port 39154 ssh2

Oct 20 14:49:35 dlpserver81 sshd[16869]: Failed password for root from 192.168.1.65 port 39154 ssh2

Oct 20 14:49:32 dlpserver81 sshd[16869]: Failed password for root from 192.168.1.65 port 39154 ssh2

Oct 20 14:49:26 dlpserver81 sshd[16862]: Failed password for root from 192.168.1.65 port 39153 ssh2

Oct 20 14:49:19 dlpserver81 sshd[16862]: Failed password for root from 192.168.1.65 port 39153 ssh2

Oct 20 14:51:02 dlpserver81 sshd[16882]: Accepted password for root from 192.168.1.65 port 39157 ssh2

数据属性：

|  |  |
| --- | --- |
| 源端口号 | 39157、39154、39153、 |
| 进程名 | sshd |
| 动作 | login |
| 目的用户名 | root |
| 状态 | LISTEN |

Oct 20 15:47:11 dlpserver81 sudo: root : TTY=pts/1 ; PWD=/root ; USER=root ; COMMAND=validate

数据属性：

|  |  |
| --- | --- |
| 源端口号 | 39157、39154、39153、 |
| 目的用户名 | root |
| 状态 | LISTEN |
| 命令 | sudo |

## 5.2 ossec报警数据：动作属性

上述日志数据经过ossec分析得到的报警数据：

\*\* Alert 1371796268.67442: mail - ossec,

2013 Jun 21 14:31:08 (agent81) 192.168.1.81->netstat -tan |grep LISTEN |grep -v 127.0.0.1 | sort

Rule: 533 (level 7) -> 'Listened ports status (netstat) changed (new port opened or closed).'

动作属性：

|  |  |
| --- | --- |
| 攻击阶段 | 探测 |
| 攻击手段 | tcp连接 |
| 动作描述 | tcp连接探测端口 |
| 威胁级别 | 3 |
| 频率 | 高 |

\*\* Alert 1382253232.47590: mail - syslog,sshd,authentication\_failures,

2013 Oct 20 15:13:52 dlpserver81->datasource

Rule: 5720 (level 10) -> 'Multiple SSHD authentication failures.'

Src IP: 192.168.1.65

User: root

动作属性：

|  |  |
| --- | --- |
| 攻击阶段 | 入侵 |
| 攻击手段 | 破解口令 |
| 动作描述 | 暴力破解 |
| 威胁级别 | 5 |
| 频率 | 高 |

\*\* Alert 1382253232.48606: mail - syslog,attacks,

2013 Oct 20 15:13:52 dlpserver81->datasource

Rule: 40112 (level 12) -> 'Multiple authentication failures followed by a success.'

Src IP: 192.168.1.65

User: root

动作属性：

|  |  |
| --- | --- |
| 攻击阶段 | 入侵 |
| 攻击手段 | 破解口令 |
| 动作描述 | 暴力破解成功 |
| 威胁级别 | 7 |
| 频率 | 低 |

\*\* Alert 1382254406.51720: - syslog,sudo

2013 Oct 20 15:33:26 (agent81) 192.168.1.81->/var/log/secure

Rule: 5402 (level 7) -> 'Successful sudo to ROOT executed'

User: root

动作属性：

|  |  |
| --- | --- |
| 攻击阶段 | 入侵获取权限 |
| 攻击手段 | 权限 |
| 动作描述 | 获取最高权限 |
| 威胁级别 | 7 |
| 频率 | 低 |

## 5.3 提取行为特征

1. tcp探测

行为特征：

|  |  |
| --- | --- |
| 方式 | 端口在线探测 |
| 范围 | 1台主机 |
| 频度 | 高 |
| 程度 | 多个端口 |
| 顺序 | 无序 |
| 结果 | 成功 |

2. 暴力入侵

行为特征：

|  |  |
| --- | --- |
| 方式 | 口令 |
| 范围 | 1台主机 |
| 频度 | 高 |
| 程度 | 多个端口 |
| 顺序 | 部分有序 |
| 结果 | 多次失败、1次成功 |

3. 获取最高权限：

行为特征：

|  |  |
| --- | --- |
| 方式 | 口令 |
| 范围 | 一个主机 |
| 频度 | 低 |
| 程度 | 一个主机 |
| 顺序 | 无序 |
| 结果 | 成功 |

三个特征组成一个序列：

tcp探测 =>暴力入侵 =>获取最高权限

匹配规则成功！

得到分析结果：目前的攻击已经获取了系统最高权限。怀疑攻击者IP：192.168.1.65

接下来的动作有待于更多数据的进一步分析。