文章编号:1001 - 9081(2007)09 - 2177 - 03

攻击案例综合学习系统研究

咎 鑫,郑庆华,范宇倩,韩九强
(西安交通大学 电子与信息工程学院,西安 710049)
(zanxin@ mail. xjtu. edu. cn)

摘 要:随着入侵检测系统在安全领域的广泛应用,入侵报警学习和分析已经成为一个研究热点。针对目前入侵报警泛滥和知识贫乏等问题,设计了一个完整的攻击案例学习系统框架。该学习系统分为两个阶段:入侵报警精简和典型攻击案例挖掘。前者利用改进的密度聚类方法实现相似报警聚合以及报警聚类的自动精简表示,后者利用序列模式挖掘方法挖掘频繁入侵事件序列。进一步提出一种基于入侵执行顺序约束关系的攻击案例评估算法实现典型攻击案例的自动筛选。最后,利用真实入侵报警数据测试了该攻击案例学习系统,结果表明该系统能够实现高效报警精简和典型攻击案例的准确学习。

关键词:入侵检测; 密度聚类算法; 序列模式挖掘; 攻击案例
中图分类号: TP393.3 **文献标志码**:A

Study on a comprehensive attack case learning system

ZAN Xin, ZHENG Qing-hua, FAN Yu-qian, HAN Jiu-qiang
(*School of Electronic & Information Engineering, Xi' an Jiaotong University, Xi' an Shaanxi* 710049, *China*)

Abstract: With the widespread deployment of Intrusion Detection Systems (IDS) in network security community, intrusion alert learning and analysis has increasingly become an active research area. Due to some problems such as alert flooding and lack of knowledge about attack scenario etc, a comprehensive attack case learning system composed of two learning phases: similar alerts aggregation and typical attack instance learning was presented. Firstly, an improved density-based clustering algorithm was introduced to aggregate huge volume of similar alerts to numbers of alert clusters. Secondly, some representative alerts were chosen to represent the overall alert clusters according to some reduction rules. Eventually, sequence pattern mining approach is used to mine frequent intrusive incidents. Furthermore, an evaluation approach based on execution ordering of attacks was proposed to identify valuable attack instances from frequent sequences of intrusive incidents. A real intrusion alert dataset was used to test our learning system. The experimental results show that our learning system can not only effectively reduce the large amount of alerts but also correctly learn the valuable attack cases.

Key words: intrusion detection; density-based clustering algorithm; sequence pattern mining; attack case

0 引言

随着入侵检测系统在安全领域的广泛应用,入侵报警学习和分析已经成为一个研究热点。一些数据挖掘方法被应用于入侵检测领域并取得了一定的进展。文献[1]指出每个入侵报警产生都是有"根源"的。基于这个观点,提出一个入侵报警处理框架:利用概念聚类方法实现相似报警聚,合并由专家分析报警根源,从而消除报警泛滥的根源。文献[2]设计了一个自适应学习器(ALAC)用于解决误报识别问题。ALAC 首先给出真实报警和误报的初始分类结果。然后经过安全分析员的修正,作为标定的训练数据再由 ALAC 进行学习,得到更新的误报识别规则。这一过程不断循环,直到获得较高的分类准确度。文献[3]设计了一个 CRIM 系统实现了入侵报警的聚类、合并、关联等功能。CRIM 系统采用聚类方法实现原始入侵报警的聚合,利用 LAMBDA 语言描述攻击行为并提出了两种有效报警关联方法。

入侵报警学习按照学习目标可以分为:误报规则学习、报警精简学习、典型攻击事件序列学习。第一类问题一般根据误报警与真实报警的特征差异,从中提取规则来识别误报,这类学习基本上属于无监督学习;第二类问题通过选择有效的机器学习方法实现大量相似报警聚合;第三类问题一般利用序列模式挖掘方法挖掘典型、有价值的频繁入侵事件序列。

本文研究目标主要有两点:1)入侵报警精简,即利用一种高效的密度聚类方法实现入侵报警聚合;2)典型攻击案例挖掘。利用序列模式挖掘方法挖掘频繁入侵事件序列,经过案例评估及筛选,转换成经典攻击案例,用于进一步的入侵预测及入侵误报识别。

1 攻击案例综合学习系统框架

原始入侵报警是包含一定噪声数据、属性之间存在复杂关联关系的高维海量数据集。单一学习方法很难获得理想的学习效果。考虑到报警数据的复杂性及相关性,我们提出一个典型攻击案例学习系统框架(如图 1 所示)。整个案例学习系统分为两个模块:入侵报警精简和典型攻击案例挖掘。

收稿日期:2007 - 03 - 13;**修回日期**:2007 - 06 - 01。 **基金项目**:国家 863 计划项目(2003AA142060)。

作者简介:咎鑫(1974 -),男,安徽安庆人,讲师,博士研究生,主要研究方向:计算机网络安全、机器学习; 郑庆华(1969 -),男,浙江埏州人,教授,博士生导师,博士,主要研究方向:智能网络学习环境、计算机网络安全; 范宇倩(1982 -),女,安徽合肥人,工程师,硕士,主要研究方向:计算机网络安全; 韩九强(1947 -),男,陕西西安人,教授,博士生导师,主要研究方向:智能检测、软测量与多传感信息融合、无接触图像测量、模式识别。

前者解决报警泛滥问题,主要包括数据预处理、数据精简、数据解释三部分;后者实现从原始入侵报警中提取典型攻击案例,主要包括入侵事件序列挖掘、入侵事件序列评估两个部分。

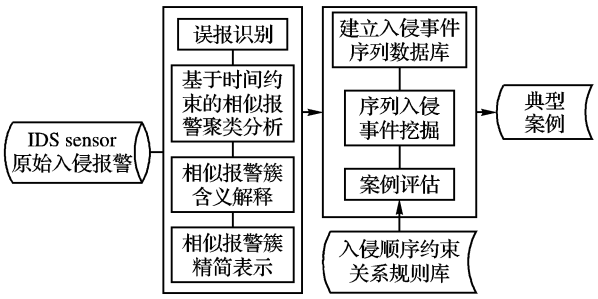

图1 典型攻击案例学习系统

2 入侵报警精简

2.1 简单误报识别

误报是指将用户正常行为误认为黑客的入侵行为。原始入侵报警中的误报警相当于噪声数据。为了获得较好的学习效果必须首先消除误报。下面给出几种简单识别误报策略:

1)利用网络拓扑信息及主机配置信息识别误报。比如:主机安装 apache 作为 Web 服务器则针对 IIS 服务器的 unicode 攻击报警就可以通过主机配置信息识别为误报。

2)基于入侵场景完整性原则识别误报。一般来说,一次成功的黑客入侵是通过多个相关入侵攻击活动组成的。相反,孤立的单一入侵报警则很有可能是误报或是失败的入侵企图。因此,基于此假设,我们可以有效识别部分误报。

3)利用漏洞扫描器的检测结果验证入侵报警是否为误报。

2.2 入侵报警的属性约简及形式化描述

原始入侵报警数据包含十几个属性,如果直接利用这些属性进行聚类分析,不仅计算时间长而且无关属性可能影响聚类结果。根据经验我们选择下列属性进行攻击案例学习:攻击时间(*time*),攻击源(*sip*),攻击目标地址(*dip*),目的端口(*dport*),攻击名称(*alert*),攻击类型(*type*)。入侵报警形式化描述为:$A = \{time, sip, dip, dport, alert, type\}$。

2.3 入侵报警的距离计算

聚类分析的基本思想是将本质上相似的数据聚成一簇。簇内数据具有很高的相似性而簇间数据具有很大的差异性。因此,区分数据的关键在于定义能够反映数据差异性的距离。根据报警属性在攻击过程中所起的作用,给出下面的距离公式。

1) 攻击源距离计算:

$$\mathrm{dis}(sip_1, sip_2) = \begin{cases} 0, & sip_1 = sip_2 \\ 0.4, & sip_1, sip_2 \in samenet \\ 0.8, & \text{其他} \end{cases}$$

2) 攻击目标 IP 地址距离计算:

$$\mathrm{dis}(dip_1, dip_2) = \begin{cases} 0, & dip_1 = dip_2 \\ 1, & dip_1 \neq dip_2 \end{cases}$$

3) 目的端口距离计算:

$$\mathrm{dis}(d_1, d_2) = \begin{cases} 0, & d_1 = d_2 \\ 0.4, & d_1 \neq d_2, d_1, d_2 \leqslant 1024 \\ 0.9, & \text{其他} \end{cases}$$

其中:d_1,d_2 分别表示两个报警的目的端口 *dport*。

4) 攻击名称距离计算:

$$\mathrm{dis}(alert_1, alert_2) = \begin{cases} 0, & alert_1 = alert_2 \\ 0.3, & type_1 = type_2 \\ 0.7, & \text{其他} \end{cases}$$

最终攻击报警的距离计算公式如下:

$$\mathrm{distance}(A_i, A_j) = \sum_{k=1}^{4} W_i \cdot \mathrm{dis}(a_i k, a_j k)$$

$\mathrm{distance}(A_i, A_j)$ 表示入侵报警 A_i,A_j 之间的距离;a_{ik},a_{jk}($k = 1, \cdots, m$)分别表示入侵报警 A_i,A_j 的第 k 个属性取值,W_i 是第 i 个属性在整个攻击报警距离中所占权值,其取值范围是[0,1],由安全分析员根据网络环境实际安全态势及安全策略给出。

2.4 基于时间约束的 dbscan 聚类算法

dbscan[3] 是一种高效密度聚类算法。其优点是不需要事先指定聚类数目,这一点非常适合入侵报警聚类分析要求。另外,dbscan 算法指定的参数只有两个:最小邻域半径 *eps*,即相似报警间的最小距离和 *minPts* 即报警聚类包含的最少报警数。dbscan 算法也存在一定的不足,由于要计算任意两个对象之间的距离,算法的平均时间复杂度达到了 $O(n \log_2 n)$。入侵报警数据量一般每天都在几十万条,直接利用 dbscan 算法计算相似报警聚类从时间上是不可行的。

通过统计分析我们发现大量相似报警都是在极短的时间内连续出现,这是使用自动化攻击工具的结果。根据相似报警的连续性特征,我们提出基于时间约束的 dbscan 算法聚合相似报警,其优点是不需要扫描整个数据集,只扫描在指定时间间隔内的候选数据,大大缩小了计算相似报警的数目,提高了计算效率。下面给出基于时间约束的 dbscan 算法步骤:

输入:相似报警最小距离 *eps*,报警聚类包含的最少报警数 *minPts*,时间间隔 $t = 1$ min

输出:相似报警聚类 C

1) 按照攻击时间对入侵报警数据进行排序,形成顺序报警数据集 D;

2) 在报警数据集 D 中选择当前时间最小的报警 $A_{\min}$,根据给定时间间隔 t,建立满足时间约束条件的候选相似报警集 $C_{\mathrm{can}} = \{A_i \mid t_i - t_{\min} < t\}$;

3) 计算当前报警 $A_{\min}$ 与候选相似报警集 C_{can} 的距离。如果 $A_{\min}$ 是核心点则将其添加到核心点集 D_{core}。从数据集 D 中删除记录 $A_{\min}$,清空候选相似报警集 C_{can};

4) 重复第 2 步和第 3 步,直到报警数据集 D 中没有数据为止;

5) 由于核心点之间密度可达满足传递性、对称性,在核心点集 D_{core} 中利用广度优先搜索算法搜索所有的最大密度相连点集,得到所有的相似报警聚类 C。

2.5 报警聚类的解释及表示

相似报警聚类根据攻击源及攻击目标可分为四类:单对单,单对多,多对单,多对多。其中单对单是指攻击源和攻击目标都是单一的,这是最常见的入侵场景。该类报警聚类中绝大部分报警的目标地址相同而源地址也相同或是属于同一个 C 类网。单对多报警聚类表示单一攻击源对多个攻击目标的相似报警集合,一般是大范围的网络扫描攻击场景。多对单报警聚类表示多个攻击源同时攻击同一个目标的相似报警集合,这是典型的分布式攻击场景。而多对多表示多个攻击源同时攻击多个目标的报警集合,这种情况较少出现。

为了达到精简报警的目的,我们考虑用少数有代表性的报警表示整个报警聚类。精简后的综合报警所包含信息与原始报警信息相同。报警的精简规则按照如下顺序应用:

规则一:若攻击名称相同和目的地址相同且源地址属于同一网段则合并报警,报警时间以第一条报警为准,攻击源及攻击目标等信息不变;

规则二:若报警聚类中来自同源地址的报警数目占整个报警簇一半以上则将该攻击源作为报警聚类的攻击源并将报警类型相同的报警合并;

规则三:若报警簇中同目标地址的报警数目占整个报警簇一半以上则将该攻击目标作为整个报警聚类的攻击目标并将报警类型相同的报警合并;

规则四:应用上面三条精简规则后,报警簇中剩余的报警作为精简报警保留。

3　典型攻击案例挖掘

3.1　频繁入侵事件序列模式挖掘

通常一个黑客入侵过程可以描述成由相关入侵活动所构成的一个入侵场景。而频繁的入侵事件序列就包含了典型的入侵场景。首先按照攻击源建立入侵事件序列数据库 $D = \{S_1, S_2, \cdots S_n\}$。其中 $S_i = \{a_1, \cdots a_n\}$,$S_i$ 表示第 i 个黑客在一定时间内产生的连续入侵事件序列,而 a_i 表示入侵事件(报警)。t_i 表示入侵时间且满足 $t_i < t_j$,当 $i < j$。

序列模式挖掘是数据挖掘领域中一个非常重要的研究方向。其核心思想是根据指定的频繁度提取原始序列数据中的最大频繁子序列集。本文引入高效序列模式挖掘算法 Prefixspan[4],用于发现频繁入侵事件序列。序列模式挖掘本质上是一种基于统计分析的数据挖掘技术,挖掘得到的频繁入侵序列模式中可能包含一些无意义的频繁模式。因此,必须通过安全专家或其他评估技术对频繁入侵序列模式进行评估。

3.2　典型攻击案例评估

一个多步骤攻击场景是由一系列相关的入侵事件组成的。这些入侵事件存在关联关系的同时也存在一定的约束关系。不同入侵事件之间存在一定的执行顺序约束关系,即首先需要执行某些入侵活动来获取一定的权限才能实施进一步入侵活动。我们定义了入侵执行顺序约束集合:

$R = \{before, late, simultaneous, non\text{-}simultaneous, irrelevant\}$

以 ARB 为例,上述取值分别表示报警 A 在报警 B 之前发生,报警 A 在报警 B 之后发生,报警 A 和报警 B 同时发生,报警 A 和报警 B 不能同时发生。报警 A 和报警 B 之间不存在时间关系。

根据上面定义的入侵执行顺序约束关系,我们定义了一个入侵执行顺序约束规则库(attack_ordering_ruleset)用于描述各种入侵事件之间的执行顺序约束关系。任意两种(类)攻击之间只能满足一种执行顺序约束关系。如果实际入侵事件序列与入侵执行顺序约束规则冲突则表明该序列可能是误报或是对整个入侵过程不起作用。入侵顺序约束规则包括不同攻击和不同类型攻击两种约束关系,前者优先级高于后者。规则形式如下:

规则一:{*service_probing before privilege_upgrading*}

规则二:{*install_backdoor after remote_attack*}

基于上述的入侵执行顺序约束规则,我们提出了一种基于序列相似度的攻击案例评估算法。给定的一个入侵序列模式,通过计算入侵序列中每个报警与序列其他报警之间顺序关系满足入侵顺序约束关系的程序,最终得到一个案例支持度并与给定阈值比较,来判断是否可以作为典型攻击案例。设 $S = \{a_1, \cdots a_n\}$ 为频繁入侵事件序列,$a_i (i = 1 \cdots n)$ 为入侵报警,$t_i (i = 1, \cdots, n)$ 为对应的攻击类型。参数 $W_j (j = 1, \cdots, 5)$。R 为类型约束关系,取值为:{*before*, *late*, *simultaneous*, *non-simultaneous*, *irrelevant*}。下面给出入侵序列评估算法:

```
Input: 入侵事件序列 S = {a1, ...an}
Output: 攻击案例支持度 CaseSupport
CaseSupport = 0
Count = 0
For (i = 0; i < S.length; i++){
  For(j = i+1; j <= S.length; j++){
    if a(i) match a(j) in ordering_attack_ruleset
      Count++;
    Else
      Count = 0;
    CaseSupport = CaseSupport + Count;
  }
}
```

说明:$a(i)$ match $a(j)$ 表示判断 $a(i)$ 与 $a(j)$ 实际顺序关系是否满足入侵顺序约束规则。如果满足规则,则连续计数器(Count)递增,否则清零。连续计数器(Count)反映了连续匹配子序列长度特性。最后计算得到的攻击案例支持度 CaseSupport 与给定的支持度阈值比较,来决定该入侵事件序列模式是否可以作为一个典型攻击案例。

4　攻击案例实验

我们在西安交通大学网络中心安装了网络入侵检测器 Snort 2.3 检测校园网内某 C 类网段的网络流量,检测时间从 2005 年 10 月 10 日至 2005 年 10 月 15 日,选择其中三天数据作为三个攻击案例学习和测试数据集。原始入侵报警记录为 414 460 条,经过简单误报识别,剔除 3 417 条明显误报警。

根据对入侵报警数据的统计分析,我们确定如下实验参数:密度聚类计算中的聚类半径 *eps* 取值为 0.12 和核心点包含最少元素个数 *minPts* 取值为 10。序列模式挖掘中的频繁序列支持度 $S = 3$。

表 1　入侵报警精简结果

数据集	原始报警	聚类	最小聚类	最大聚类	精简报警
1	182 833	171	32	5 146	11 618
2	92 781	73	47	4 567	5 079
3	135 429	107	29	4 221	6 123

表 2　典型攻击案例挖掘结果

数据集	入侵事件	攻击源	频繁入侵事件序列	攻击案例
1	7 619	73	47	19
2	3 217	49	23	9
3	3 836	45	25	11

表 1 是入侵报警精简实验结果。以数据集 1 为例,报警精简率达到 93.6%,其中最大报警聚类包含 5146 条报警,包含 8 种类型的攻击报警,攻击源地址有 3 个,精简后只有 257 条。实验结果表明大量原始入侵报警可以聚合成少数的报警聚类。

表 2 是典型攻击案例学习的实验结果。其中入侵事件是精简报警集剔除明显误报警聚类之后的数据集。我们按照攻击源建立入侵序列,数据集 1 包括 73 个入侵序列,序列模式挖掘发现了 47 个频繁入侵序列,经过攻击案例评估,最终得到 19 个攻击案例。攻击案例形式如下:

{scan _ null, FTP _ STOU _ overflow _ attempt, ATTACK-RESPONSES_id_check_returned_root}　(下转第 2183 页)

能通过计算 $k_T = H(y^{xT} \bmod p)$ 并解密 c 来验证明文内容。因此,Alice 不能否认签密及其内容。

3) 公平性。基于改进组签密协议,可充分利用签密的性质,进行验证和仲裁,维护收发双方的利益。依据验证的层次,可分为密文有效性验证、明文内容验证和明文仲裁。

(1) 通过可公开计算的 $y = (y_A g^v)^s \bmod p$ 和 $r_i = H(y, c_i, M)$ 可以验证签密的发送者及信息的完整性;

(2) 签密接收者可以通过解签密查看明文的内容和完整性;

(3) 若签密接收者对明文内容有疑议,可申请仲裁,由 *TTP* 判断明文的合法性等性质。

同时,由于签密的 UF-CMA 安全性,若 R_i 接收了 (c, c_i, r_i, v, s, M),并不能进行篡改。因此, R_i 不能修改一个已有签密的内容, 对 Alice 进行诬陷。因此,协议保证了收发双方的利益,是一个公平的协议。

4)防止中间者攻击。由签密的性质可知,签密的收发双方是确定的,且其身份信息包含在签密中。基于签密的 UF-CMA安全性,中间攻击者不能伪造 Alice 的一个合法签密,因此不能执行中间者攻击。

4.3　性能分析

1) 公开可验证性。由于加入了信息 M,并可通过任意一个签密求 $y = (y_A g^v)^s \bmod p$,则可通过 $r_i = H(y, c_i, M)$ 验证签密的发送方和有效性。验证有效性不需要明文信息 m 的参与,保护了明文的安全。

2) 不需要使用带钥的 hash 函数,提高了效率。

3) 附加通信量降低。文献[3] 方案中的附加通信量为 $(n+1)|H| + |q| + |KH|$, 改进方案中附加通信量为 $2|H| + 2|q|$, 节省了通信带宽。

4) 不考虑仲裁所需的计算(只有必要的时候 TTP 才被使用), 在改进方案中,若方案中有 n 个参与方,那么在改进的签密方案中签密时需要 $n+1$ 个指数运算,解签密时每个用户需要 3 次指数运算。与文献[3] 中的效率是相同的。

定义 Exp 为模数长度为 $|p|$,幂指数长度为 $|q|$ 的一次幂运算;l_c 为密文长度;并假设 $|KH|$ 和 $|H|$ 的长度相同。在 n 个用户的情况下,比较文献[4] 方案、文献[3] 方案与本方案的性能如表 1 所示。

改进方案中,每个用户不仅不需要接收其他用户的密文[3],而且不需要接收 $r_i (i = 1, 2, \cdots, n)$ 值。因此,通信量为常数,不随用户数量的增加而增长,具有较好的性能。

表 1　基于组签密的方案比较

方案	加密计算量	每个用户的解密计算量	每个用户的通信量	可验证性	可仲裁性
文献[4]方案	n Exp	3 Exp	$n(l_c+\|H\|+\|q\|)$	否	否
文献[3]方案	$(n+1)$ Exp	3 Exp	$2l_c+(n+2)\|H\|+\|q\|$	是	是
本文方案	$(n+1)$ Exp	3 Exp	$2l_c+2\|H\|+2\|q\|$	是	是

5　结语

本文分析了文献[3] 提出的组签密方案,指出其方案中存在冗余数据和不必要的计算。去掉冗余数据和计算,设计新的组可验证签密方案。并基于新的签密方案,设计了一个一对多的秘密传输协议。新协议不仅没有增加计算量,且与原方案一样具有可验证性和可仲裁性;新协议能保证信息的机密性和非否认性;能保证收发双方的公平性,并能抵抗中间者攻击。新方案中,每个用户的通信量降低了 $(n-1)|H|$;且通信量为常数,不随用户人数增加而增长,具有较好的性能。

参考文献:

[1] ZHENG Y. Digital signcryption or how to achieve cost(signature & encryption) < cost(siganture) + cost(encryption) [C]// Crypto' 97, LNCS 1294. Berlin: Springer-Verlag, 1997: 165 - 179.

[2] ZHENG Y, IMAI H. Using signcryption to build compact and unforgeable key establishment over an ATM network [C]// Proceedings of IEEE INFOCOM' 98. [S. l.]: IEEE Press, 1998: 411 - 418.

[3] 王彩芬, 贾爱库, 刘军龙. 基于签密的多方认证邮件协议[J]. 电子学报, 2005, 33(11): 2070 - 2073.

[4] SEO M, KIM K. Electronic funds transfer protocol using domain-verifiable signcryption scheme [C]// Proceeding of Information Security And Cryptology (ICISC' 99). Berlin: Springer-Verlag, 2000. 269 - 277.

[5] 陈伟东, 冯登国. 签密方案在分布式协议中的应用[J]. 计算机学报, 2005, 28(9): 1421 - 1430.

[6] GAMAGE C, LEIWO J, ZHENG Y. Encrpted message authentication by firewalls [C]// Proceeding of PKC' 99, LNCS 1560. Berlin: Springer-Verlag, 1999: 69 - 81.

[7] BACK J, STEINFELD R, ZHENG Y L. Formal proofs for the security of signcryption[C]// Proceeding of PKC '02. Berlin: Springer-Verlag, 2002: 81 - 98.

[8] SHIN J B, LEE K, SHIM K. New dsa - verifiable signcryption schemes [C]// Proceedings of ICISC 2002. Berlin: Springer-Verlag, 2003. 35 - 47.

(上接第 2179 页)

从实验结果来看,攻击案例主要分为三类:远程网络入侵、DDoS、蠕虫传播。大部分攻击案例是远程攻击者扫描系统漏洞,然后利用漏洞实现权限提升的攻击场景,这也反映了当前网络入侵的一般模式。

5　结语

本文提出了一个完整的攻击案例学习系统框架,详细描述了各模块的功能和实现流程。利用密度聚类方法实现了高效报警聚合,分析了入侵攻击执行顺序约束关系,并提出了基于入侵攻击执行顺序约束关系的典型攻击案例评估算法。最后利用真实入侵报警测试我们的攻击案例学习系统,实验数据表明本系统能够准确、有效的报警精简及典型攻击案例自动学习。下一步我们将继续完善报警精简规则及典型攻击案例评估准则,进一步提高案例学习系统的准确性。

参考文献:

[1] JULISCH K. Clustering intrusion detection alarms to support root cause analysis [J]. ACM Transactions on Information and System Security, 2003, 6(4): 443 - 471.

[2] PIETRASZEK T. Using adaptive alert classification to reduce false positives in intrusion detection [C]// Recent Advances in Intrusion Detection (RAID2004). Sophia Antipolis: [s. n.], 2004: 102 - 124.

[3] CUPPENS F, MIEGE A. Alert correlation in a cooperative intrusion detection framework [C]// Proceedings of the IEEE Symposium on Security and Privacy. [S. l.]: IEEE Press, 2002.

[4] ESTER M, KRIEGEL H, SANDER J, *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise [C]// KDD' 96. Portland: [s. n.], 1996.

[5] PEI J, HAN J, PINTO H, *et al.* PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth[C]// the IEEE 17th International Conference on Data Engineering. [S. l.]: IEEE Press, 2001: 215 - 226.

[6] ROESCH M. Snort[EB/OL]. [2005 - 10 - 20]. http://www.snort.org/dl/binaries/win32/old/.