一种基于 Choquet 模糊积分的入侵检测警报关联方法

努尔布力[1,2],柴 胜[1,3],李红炜[1,3],胡 亮[1,3]
(1.吉林大学计算机科学与技术学院,吉林长春 130012; 2.新疆大学信息科学与工程学院,新疆乌鲁木齐 830046;
3.吉林大学符号计算与知识工程教育部重点实验室,吉林长春 130012)

摘 要: 文章研究了警报关联方法,模糊积分和模糊认知图基本理论,提出了一种基于 Choquet 模糊积分的入侵检测警报关联方法,设计并实现了一个能够识别多步攻击的警报关联引擎.通过 DRDOS 和 LLDOS 实验表明,该引擎能够有效的检测网络中存在的大规模分布式多步攻击.

关键词: 入侵检测; 模糊积分; 模糊测度; 警报关联; 告警关联
中图分类号: TP302.7 **文献标识码:** A **文章编号:** 0372-2112 (2011) 12-2741-07

Intrusion Detection Alert Correlation Based on Choquet Fuzzy Integral

Nurbol[1,2], CHAI Sheng[1,3], LI Hong-wei[1,3], HU Liang[1,3]
(1. *College of Computer Science and Technology Institute, Jilin University, Changchun, Jilin* 130012, *China*;
2. *College of Information Science and Engineering, XinJiang University, Urumqi, Xinjiang* 830046, *China*;
3. *Key Laboratory for Symbol Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun, Jilin* 130012, *China*)

Abstract: The alert correlation, choquet fuzzy integral and fuzzy cognitive maps was analyzed, the correlation of IDS alerts based choquet fuzzy integral was proposed and the correlation engine of intrusion detection system was designed. Though experiences of the DRDOS attack and LLDOS attack, it is proved that the alert correlation in the paper could correlate the alerts with high feasibility.

Key words: intrusion detection; fuzzy integral; fuzzy measure; alert correlation; alarm correlation

1 引言

赛门铁克在 2010 年 7 月的报告表明[1],现在的网络安全状况由原先“单纯的”木马、病毒问题,已经转化为“互联网”+“攻击团队”+“病毒”+“商业利益”合成到一起,形成了复杂的攻击集体.然而,目前一般的入侵检测系统只能检测出单步攻击,无法对网络中的大规模协同式多步攻击进行检测.针对入侵检测系统的这一不足,本文提出了一种基于模糊积分的警报关联分析技术.警报关联能够将入侵检测系统产生的单步攻击进行分析、关联,最终识别一个完整的多步攻击,并及时阻断攻击或者将攻击造成的损失降到最低.

本文将现有的警报关联分析技术作为研究对象,进行了基于模糊积分的入侵检测警报关联分析方法的研究.该方法引入模糊积分,顺应自然人的知识逻辑,并能进行多步攻击的智能化关联分析.实验证明,本文提出的基于模糊积分的入侵检测警报关联方法,相对 PENG NING 研究小组[2]提出的基于贝叶斯网络的概率统计方法,有更高的警报关联完备率和有效率.

2 相关工作

2.1 模糊积分

由于生产生活中的大部分问题都是非可加的,而模糊积分[3]恰恰是通过定义的一组非负非可加的单调集函数作为模糊测度,度量模糊程度,来表示不同分类器之间的相互影响,模糊积分已经成为信息融合、信息关联领域普遍而又重要的合成算子.

2.1.1 模糊测度

设 $X=\{x_1, x_2, x_3, \cdots, x_n\}$ 是一个有限集合,幂集 $P(X)$ 是 X 上的 δ 代数,定义在 $P(X)$ 上的集函数 $\mu: P(X)\to[0,1]$ 称为模糊测度[4,5],满足以下条件:

(1) $\mu(\phi)=0, \mu(X)=1$.

收稿日期:2010-12-28;修回日期:2011-04-23
基金项目:国家重点基础研究发展计划(973 计划)(No.2009CB320706);国家自然科学基金(No.61073009);吉林省科技发展计划资助项目(No.20090111);长春国际合作项目(No.11GH12)

(2) $\forall A, B \in p(X)$,若 $A \subseteq B$ 则,$\mu(A) \leqslant \mu(B)$(单调性).

2.1.2 Choquet 模糊积分

设 μ 是 X 上的模糊测度,$f: X \in [0, +\infty]$为定义在 X 上的非负实值函数,f 关于 μ 的 Choquet 积分公式[4]定义为:

$$(c)\int f\mathrm{d}\mu = \int_0^{\infty} \mu(F_\alpha)\mathrm{d}\alpha$$

式中的 $F_\alpha = \{x \mid f(x) \geqslant \alpha, x \in X\}$,当 $X = \{x_1, x_2, x_3, \cdots, x_n\}$为一个有限集合且 $f: X \in [0,1]$时,Choquet 积分的计算公式可以简化为:

$$(c)\int f\mathrm{d}\mu = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\mu(A_i)$$

其中 $A_i = \{x_i, x_{i+1}, \cdots, x_n\}$,计算时,如果不能满足 $0 \leqslant f(x_1) \leqslant f(x_2) \leqslant \cdots \leqslant f(x_n) \leqslant 1$,则需要对有限集合 X 中的元素进行重新排列 $\{x_1^*, x_2^*, \cdots, x_n^*\}$,以使排列后的元素满足 $0 \leqslant f(x_1^*) \leqslant f(x_2^*) \leqslant \cdots \leqslant f(x_n^*) \leqslant 1$.

2.2. 模糊认知图

认知图主要包括两大类:古典认知图和模糊认知图.认知图主要由节点和带箭头的弧线构成,节点表示概念,表示系统的一些属性或者状态,弧线表示这些概念节点之间的因果关联关系,本文使用基于概率的模糊认知图[6],如图 1、2 所示.

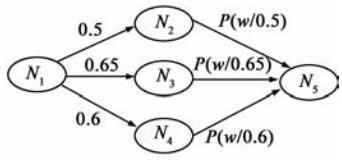

图1 基于概率模糊认知图的结构

$$\begin{bmatrix} 0 & 0.5 & 0.65 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & P(w/0.5) \\ 0 & 0 & 0 & 0 & P(w/0.65) \\ 0 & 0 & 0 & 0 & P(w/0.6) \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

图 2 连接矩阵

概念节点的状态值计算公式为[7]:

$$V_{cj(t+1)} = f\Big(\sum_{i=1}^{n} V_{ci(t)} P(W_{ij(t)}/V_{ci(t)}, V_{cm(t)}, \cdots) + \gamma V_{cj(t)}\Big)_{j=1,2,\cdots,n}$$

其中,n 为概念节点的总数,$V_{cj(t+1)}$表示为概念节点 C_j 在 $t+1$ 时刻的状态值,$V_{ci(t)}$、$V_{cm(t)}$分别表示结果、原因概念节点 C_i 在 t 时刻的状态值、相关概念节点 C_m 在 t 时刻的状态值,$W_{ij(t)}$表示在 t 时刻概念节点 C_i 对概念节点 C_j 的因果关系强度,$P(W_{ij(t)}/V_{ci(t)}, V_{cm(t)}, \cdots)$表示概率测度,$\gamma$ 表示上一时刻状态值对下一时刻状态值的影响因子,$f(x)$为概念 C_j 的阈值函数[8].

2.3 警报关联方法

目前典型的警报关联方法包括:基于属性相似度的警报关联、基于已知场景的警报关联、基于因果关系的警报关联、基于关联规则的警报关联、基于统计分析的警报关联.

2.3.1 基于属性相似度的警报关联

这种方法是基于警报的某些属性的相似度进行关联[9]的,比如源 IP 地址、目的 IP 地址、端口号等.具有高相似度的警报将被关联在一起,但是这种方法的一个缺点是不能很好地发现关联警报之间的因果关系.

2.3.2 基于已知场景的警报关联

攻击场景的表述方法包括使用攻击场景描述语言(STATL, LAMDBA)和使用数据挖掘的方法来训练数据集[10].这种方法能够揭示警报之间的因果关系,但是受限于已知的场景.

2.3.3 基于因果关系的警报关联

这种方法基于对于大多数非独立的警报的关联,这些警报产生于攻击的不同阶段,前一个阶段为后一个阶段做准备.这种方法需要精确的攻击知识来识别前提和结果警报[11].通过匹配前提警报的结果和结果警报的前提来进行警报关联.但是这种方法的一个局限是无法关联未知攻击,因为攻击的前提和结果没有被定义,并且很难定义前提以及产生的后果.

2.3.4 基于关联规则的警报关联

利用基于关联规则的方法进行警报的关联,首先需要寻找警报之间各种重要的联系,然后将这些重要的联系制定成各种规则[12],可分为两步:(1)将警报按照某种方法进行分类,保证每类警报具有某些相同的特点;(2)将这些分好类的警报按照某种规则进行关联.

由于针对传统的基于关联规则的方法没有考虑警报之间的时间顺序关系,因此目前又提出了一种动态的基于时间序列的关联规则模型[13].

2.3.5 基于统计分析的警报关联

基于统计分析的方法最典型的方法是基于时间序列[14]的分析方法,其主要思想是通过将某特定时间段分割成 N 份,利用贝叶斯网络获取发生在时间段 i 内的具有高优先级的警报集合,通过所制造的时间序列,统计各个时间段内不同类警报在数量上是否满足格兰格尔因果关系,进而发现这些警报是否存在某种关联关系.

本文将基于相似度、基于因果关系以及基于关联规则的警报关联方法相结合,进行基于模糊积分的警报关联方法研究,能够准确地发现多步骤攻击之间的关联关系.

3 基于 Choquet 模糊积分的警报关联

3.1 警报关联框架

本文将警报关联引擎设计为以下 5 部分:输入模

块、预处理模块、关联引擎、规则数据库、输出模块.

3.1.1 输入模块

为警报关联引擎提供分析数据,这些数据为入侵检测系统产生的单步攻击的警报,警报格式采用基于 XML 的 IDMEF 格式[15],警报类型分为两大类,即 Alert 类型和 HeartBeat 类型,其中 Alert 是入侵检测系统产生的各种警报,而 HeartBeat 定时地获取当前入侵检测器的状态.

3.1.2 预处理模块

将输入模块的数据进行预处理,包括将输入数据转换为关联引擎能够使用的格式、将警报进行分类、剔除重复冗余的警报等.

3.1.3 关联引擎

对经过预处理模块处理的警报关联,按照规则数据库中设定的规则进行关联,一旦警报满足了规则库的某种多步攻击的规则,则认定发生了多步攻击,将关联的结果传送给输出模块进行相应的处理.

3.1.4 规则数据库

主要包括两部分:否定规则数据库、预定规则数据库.关联引擎在对警报关联的过程中,首先查找否定规则数据库中的关联规则,如果满足否定规则,则认为此警报不是多步攻击的某个步骤,舍弃此警报,不再与预定规则数据库中的关联规则进行计算;如果不满足否定规则数据库中的规则,则与预定规则数据库中的关联规则进行关联.

3.1.5 输出模块

接收关联引擎传送来的多步攻击,对这些多步攻击进行相应的响应,比如发出多步攻击警告、根据攻击多步攻击的危害程度关闭相应的端口或者停止相应的服务.

3.2 警报关联引擎

为了更好说明警报关联引擎规则的定义,在本小节给我们出警报关联引擎相关的定义.

定义 1 在一次攻击中,如果包含 2 个或 2 个攻击步骤以上的攻击,那么这次攻击可以被称为多步攻击.

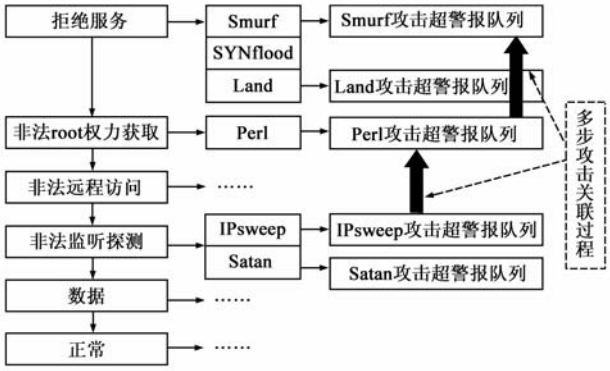

图3 警报关联引擎

定义 2 多步攻击的每个攻击步骤,按照攻击分类规则进行分类.

在本文研究中,我们将攻击分为 6 大类[16],将每个攻击步骤分成对应成相应的攻击类别.如图 3 所示,每一个大类又被分为若干小类,其中黑色的箭头表示一个多步攻击的一次完整的关联过程.

图 3 中显示了一个具有三个攻击步骤的多步攻击,首先进行 IPsweep 地址扫描,其次通过 Perl 攻击进行 root 权限的获取,最后通过 Smurf 攻击对目标主机进行攻击,最终完成多步攻击.

定义 3 寻找多步攻击中完整攻击步骤的过程,可以被称作多步攻击关联过程.

定义 4 存在攻击步骤 Attack_i 和 Attack_j,$\text{Caculate\_Correlate}(\text{Attack}_i, \text{Attack}_j)$ 表示 Attack_i 和 Attack_j 之间的因果关联度,通过模糊积分来计算 Attack_i 和 Attack_j 之间的因果关联度,$P(\text{Attack}_i \mid P(\text{Attack}_j)) = \text{Caculate\_Correlate}(\text{Attack}_i, \text{Attack}_j)$,其中 $P(\text{Attack}_j) >= 0.5$,其中 $P(\text{Attack}_i)$ 和 $P(\text{Attack}_j)$ 分别表示攻击 Attack_i 和 Attack_j 发生的概率.

这里,对于每一个攻击,都可以抽象出攻击的重要属性特征,根据每类攻击的特点,为每一个属性赋予不同的模糊测度值.假设攻击 Attack_1 属性 1 到属性 n 的模糊测度值为 $\{\text{Attack}_1\_\text{Property}_1, \text{Attack}_1\_\text{Property}_2, \cdots, \text{Attack}_1\_\text{Property}_n\}$,攻击 Attack_2 属性 1 到属性 n 的模糊测度值为 $\{\text{Attack}_2\_\text{Property}_1, \text{Attack}_2\_\text{Property}_2, \cdots, \text{Attack}_2\_\text{Property}_n\}$,…攻击 Attack_n 属性 1 到属性 n 的模糊测度值为 $\{\text{Attack}_n\_\text{Property}_1, \text{Attack}_n\_\text{Property}_2, \cdots, \text{Attack}_n\_\text{Property}_n\}$,他们之间的关联度可以表示为 $I = P(\text{Attack\_Relation}_{(1,2)}), P(\text{Attack\_Relation}_{(2,3)}), \cdots, P(\text{Attack\_Relation}_{(n-1,n)}) = \{P(\text{Attack\_Relation}_2) \mid P(\text{Attack\_Relation}_1), P(\text{Attack\_Relation}_3 \mid P(\text{Attack\_Relation}_2)), \cdots, P(\text{Attack\_Relation}_n \mid P(\text{Attack\_Relation}_{n-1}))\}$.

为了说明以上定义的过程,给出一次多步攻击的关联过程的分析.如图 4,多步攻击共包括三个步骤的多步攻击,分别假设为:A、B、C,攻击 C 为攻击者的最后一步攻击.攻击 A 的属性 1 的模糊测度值为 0.3,属性 n 的模糊测度值为 0.5.由此,$P(B/A)$、$P(C/D)$ 分别表示 A 与 B 之间、B 与 C 之间的关联度(为了便于表示,这里用 D 来作为 $P(B/A)$ 简写标识),其定义分别为:

$$P(B/A) = \begin{cases} \text{Caculate\_Correlate}(A, B), & P(A) \geqslant 0.5 \\ 0, & P(A) < 0.5 \end{cases}$$

其中 $\text{Caculate\_Correlate}(A, B)$ 表示 A、B 之间的因果关联度,通过模糊积分进行计算,$P(A)$ 表示攻击 A 发生的概率.

$$P(C/D)=\begin{cases}\text{Caculate\_Correlate}(B,C), & D=P(B/A)\geqslant 0.5\\ 0, & D=P(B/A)<0.5\end{cases}$$

其中 Caculate_Correlate(B, C)表示 B 与 C 之间的因果关联度,通过模糊积分进行计算,$D=P(B/A)$表示在攻击 A 发生的条件下,攻击 B 发生的概率.当 $P(C/D)$ 的关联值满足阈值要求时,认为对于一个具有三个攻击步骤的多步攻击关联成功.

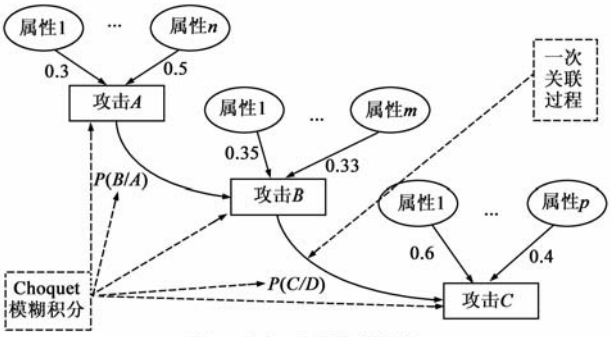

图4 多步攻击关联过程

根据定义 1 和 2,我们分析得到定理 1.

定理 1 对于一个完整的多步攻击关联过程,至少包含 2 个或 2 个步骤以上的攻击步骤.

在一个多步攻击发生时,必然至少包含 2 个或 2 个以上的攻击步骤,如果发现一个多步攻击的关键步骤,那么必然存在相对应的其他攻击过程.当然,在真实环境中,网络攻击数据的获取可能会存在不完整的情况,这里暂不讨论此内容.

3.3 警报关联规则

需要进行警报关联规则的设计,首先要进行数据结构的定义,以便更为科学合理的进行资源空间的利用.

3.3.1 数据结构设计

根据规则数据库中的关联规则,制定了相应的警报、超警报的数据结构,如表 1、表 2.

表 1 警报数据结构

源 IP 地址	源端口号	目的 IP 地址	目的端口号	警报时间戳
Source_IP	Source_Port	Desti_IP	Desti_Port	Alert_Time
协议类型	标志	源 MAC 地址	目的 MAC 地址	警报 ID
Alert_Protocol	Alert_Flag	Source_MAC	Desti_MAC	Alert_ID

表 2 超警报数据结构

超警报 ID	超警报名称	超警报时间戳	融合警报数量	融合积分阈值
Halert_ID	Halert_Name	Halert_Time	Alert_Number	Halert_Value
成熟度	时间计数器	成熟标识	后项指针	警报指针
Halert_Mature	Halert_Timer	Halert_Flag	Halert_Next	Alert_Pointer
积分值数组	最小积分值标号			
Halert_Array	Min_Halert_Value_Number			

3.3.2 警报关联规则设计原则

这里的警报关联规则设计原则主要为以下九个方面[17]:

(1)具有相同目的 IP 地址、相同攻击类别的警报,可能是多个攻击者共同协作,对同一目标节点实行的分布式攻击,如分布式拒绝服务攻击.

(2)具有相同源 IP 地址的警报,可能是同一攻击者针对不同目标节点所发起的不同攻击.

(3)具有相同源 IP 地址、相同目的 IP 地址、相同攻击类别的警报,可能是攻击者针对同一节点发起的一系列相同、相似的攻击或者多次尝试攻击同一个节点,如攻击者试图对 Web 服务器发起一系列的 Web 攻击.

(4)具有相同源 IP 地址、相同攻击类别的警报,可能是同一攻击者针对某一网络发起的相同攻击,如同一攻击者针对不同的域名服务器发起的攻击.

(5)具有相同源 IP 地址、相同目的 IP 地址的警报,可能是同一攻击者针对同一节点上的不同服务发起的一系列攻击.

(6)具有相同目的 IP 地址的警报,可能是不同攻击者针对系统的不同漏洞发起的攻击,或者是针对同一目标节点发起的分布式协同攻击,如 SYNflood 攻击.

(7)目的 IP 地址、端口号相同,且为某类漏洞的端口号的警报,可能是针对某一服务发起的攻击,如 23 为 Telnet、514 为 RPC 后门端口等.

(8)如果在时间戳上出现规律性变化的警报,可能是攻击者定时地向受害者发起的攻击.

(9)MAC 地址与 IP 不同的警报,可能是与 IP、MAC 相关的欺骗类攻击.

3.3.3 关联规则设定方法

通过选取警报的重要特征属性,并根据属性的重要程度,设定不同模糊测度值,来对两个超警报进行关联.首先需要对各个属性设定属性值转换方法,得到相应的属性值,然后依据 Choquet 模糊积分的计算过程进行警报关联值的计算.

(1)IP 地址转换函数

设警报 $A1$、$A2$ 的源 IP 地址分别为 $A1$.Source_IP、$A2$.Source_IP,从最高位开始取其相同的位数 r,由于 IP 地址是 32 位,设定函数值为 $r/32$,是一个在[0,1]区间内的数值,函数的运算方法为:

(a) $s=A1.\text{Source\_IP}\oplus A2.\text{Source\_IP}$.

(b)获取数值 s 的位数,赋值于 s_number.

(c) $r=32-s\_\text{number}$, $f=r/32$.

(2)协议类型转换函数

设警报 $A1$、$A2$ 的协议类型分别为 $A1$.Alert_protocol、$A2$.Alert_protocol,其转换函数定义如下:

$$f=\begin{cases}1, & A1.\text{Alert\_Protocol}=A2.\text{Alert\_Protocol}\\ 0, & A1.\text{Alert\_Protocol}=A2.\text{Alert\_Protocol}\end{cases}$$

在计算时,将协议类型依次编号,0 表示 TCP,1 表

示 UDP,2 表示 ICMP,3 表示 ARP 等.

(3)时间戳转换函数

设警报 $A1$、$A2$ 的时间戳分别为 $A1.\text{Alert\_time}$、$A2.\text{Alert\_time}$,设定时间窗口大小为 TIME,表明两个警报在一定间隔时间内时间属性的关联关系,如果 $A1.\text{Alert\_time}$ 与 $A2.\text{Alert\_time}$ 的间隔时间超过 TIME,函数值取值为 0,如果 $A1.\text{Alert\_time}$ 与 $A2.\text{Alert\_time}$ 的间隔没有超过 TIME,则认为两个警报在时间上是有一定关联的,公式定义如下:

$$f=\begin{cases}0, |A1.\text{Alert\_Time}-A2.\text{Alert\_Time}|>\text{TIME}\\ (\text{TIME}-|A1.\text{Alert\_Time}-A2.\text{Alert\_Time}|)/\text{TIME},\\ \qquad |A1.\text{Alert\_Time}-A2.\text{Alert\_Time}|\leqslant\text{TIME}\end{cases}$$

(4)端口转换函数

设警报 $A1$、$A2$ 的源端口号分别为 $A1.\text{Source\_Port}$、$A2.\text{Source\_Port}$,选择端口属性的主要目的是判断数据包的发送端与接收端分别运行何种服务,因此,如果两个警报的端口一致,并且取值为设定的特定数值,认为警报是针对同一服务类型的攻击,函数值为 1;如果两个警报的端口值不等或者与设定的特殊值不同,定义函数值为 0:

$$f=\begin{cases}1, A1.\text{Source\_Port}=A2.\text{Source\_Port}=\text{设定值}\\ 0, A1.\text{Source\_Port}\neq A2.\text{Source\_Port}\end{cases}$$

3.3.4 Choquet 模糊积分计算过程

对于警报 $A1$、$A2$ 进行超警报的融合训练,判定他们是否可以属于同一个超警报,选取警报的源 IP 地址值、协议类型值、以及时间戳为特征属性,分别设为 x_1、x_2、x_3,并根据经验制定模糊测度表,如表 3、表 4、表 5.

表 3　模糊测度表

$\mu(\phi)=0$	0	$\mu(\{x_2\})$	0.5	$\mu(\{x_1,x_2\})$	0.8	$\mu(\{x_2,x_3\})$	0.53
$\mu(\{x_1\})$	0.7	$\mu(\{x_3\})$	0.4	$\mu(\{x_1,x_3\})$	0.75	$\mu(\{x_1,x_2,x_3\})$	1

表 4　警报 1

源 IP 地址	协议类型值	时间戳
202.198.16.3	TCP	20091205143625

表 5　警报 2

源 IP 地址	协议类型值	时间戳
202.198.16.20	TCP	20091205143630

由以上的转换函数计算公式得到,IP 地址的转换函数值为 0.84375,协议类型的转换函数值为 1,假设 TIME 的转换函数值为 0.5,利用 Choque 模糊积分计算公式得到两个警报之间的关联度为:$0.5*1+(0.84375-0.5)*0.8+(1-0.84375)*0.5=0.853125$.

4　实验评估

利用基于 Choquet 模糊积分的入侵检测警报关联方法,通过实验,关联了两种多步攻击,即 DRDOS[18] 攻击及 LLDOS1.0 攻击.实验采用的数据集和评估标准根据测试要求进行分段的描述,实验结果与算法的比较为以后的研究提供了理论上的指导,同时为今后的研究工作提供了一定的实验参考数据.

4.1　警报关联评估方法

常用的入侵检测系统警报关联的评估标准包括[19]:完备性和有效性.

(1)完备性,实际正确关联的警报数量在正确的警报中所占的比例.

$$\text{完备率}=\frac{\text{实际正确关联的警报数量}}{\text{正确的警报数量}}$$

(2)有效性,实际正确关联的警报数量在实际关联的警报中所占的比例.

$$\text{有效率}=\frac{\text{实际正确关联的警报数量}}{\text{实际关联的警报数量}}$$

4.2　测试数据集

(1)DRDOS 攻击仿真

如图 5 所示,攻击发起者为 10.60.55.104,受害者为 10.60.55.108,傀儡机为 10.60.55.106 和 10.60.55.107,其中 10.60.55.106 装有 Snort 入侵检测系统,能够产生单步攻击的警报,警报关联引擎接收这些警报,最终识别 DRDOS 攻击.主要分为 5 个步骤:网络带宽设置、运行 Snort 入侵检测系统、运行 DRDOS 攻击程序、获取 DRDOS 攻击的警报、进行 DRDOS 的检测.

(2)LLDOS1.0 攻击仿真

采用 DARPA 2000 数据集中的 LLDOS 1.0 攻击进行警报关联测试,数据集的搜集时间从 2000 年 3 月 7 日上午 9 点 25 分至 12 点 35 分,共 3 小时 10 分钟.

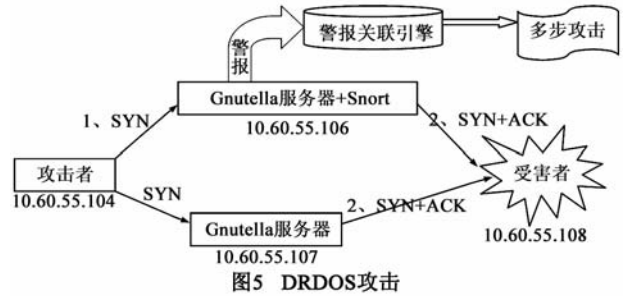

图5　DRDOS攻击

4.3　关联结果

(1)DRDOS 攻击

图 6 为采用本文设计的警报关联引擎,关联 DRDOS 攻击的结果截图.

```
2010-1-27 16:49:2
10.60.55.108 are suffering DRDOS Attack !
**********************************************************
2010-1-27 18:45:5
10.60.55.109 are suffering DRDOS Attack !
**********************************************************
```

图6　DRDOS攻击关联结果

(2)LLDOS1.0 攻击

图 7 为采用本文设计的警报关联引擎,关联 LL-

DOS1.0 攻击的结果截图.

```
131.84.1.31 are suffering LLDOS1.0 Attack !
Sadmind_Ping:
2000-3-7 10:08:07  202.77.162.213->172.16.115.20 (1)
2000-3-7 10:15:10  202.77.162.213->172.16.112.50 (1)
2000-3-7 10:15:10  202.77.162.213->172.16.112.10 (1)
Sadmind_Overflow:
2000-3-7 10:33:29  202.77.162.213->172.16.115.20 (6)
2000-3-7 10:34:46  202.77.162.213->172.16.112.10 (4)
2000-3-7 10:34:59  202.77.162.213->172.16.112.50 (4)
Rsh:
2000-3-7 10:50:37  202.77.162.213->172.16.112.10 (3)
2000-3-7 10:50:53  202.77.162.213->172.16.112.50 (3)
2000-3-7 10:50:21  202.77.162.213->172.16.115.20 (6)
Mstream_Zombie:
2000-3-7 10:50:37  172.16.112.10->255.255.255.255 (1)
2000-3-7 10:50:53  172.16.112.50->255.255.255.255 (1)
2000-3-7 10:50:21  172.16.115.20->255.255.255.255 (2)
2000-3-7 11:27:51  172.16.115.20->172.16.112.50 (1)
2000-3-7 11:27:51  172.16.115.20->172.16.112.10 (1)
Stream_DOS:
2000-3-7 11:27:51  8.138.161.2->131.84.1.31 (33751)
daying@daying-laptop:~/paper$
```

图7 LLDOS1.0攻击关联结果

根据表 6 所示数据,得出表 7 的警报关联的各项指标.

最后,我们进行了本文算法的比较实验.针对 LLDOS1.0 攻击进行关联,将本文算法与 TIAA[2] 进行比较的结果如表 8 所示.

表 6 LLDOS1.0 警报、超警报训练、超警报关联的警报数量关系表

关联指标	步骤 1	步骤 2	步骤 3	步骤 4	步骤 5
各阶段原始警报总数量	3	14	12	6	33782
超警报融合的警报数量	3	14	12	6	33782
超警报的数量	3	3	3	5	1
正确的警报数量	3	14	12	6	33751
实际正确关联的警报数量	3	14	12	6	33751

表 7 DRDOS 警报关联指标

关联指标	步骤 1	步骤 2	步骤 3	步骤 4	步骤 5	总体
完备率	100%	100%	100%	100%	100%	100%
有效率	100%	100%	100%	100%	99.91%	99.98%

表 8 与 TIAA 关联结果对照表

关联指标	本文算法	TIAA
警报关联有效率	99.98%	93.18%
警报关联完备率	100%	93.18%

通过以上两个实验,证明了基于 Choquet 模糊积分的警报关联的可行性、可用性及具有较高的完备性和有效性.

表 8 的实验结果证明,在当前测试环境下,本文的方法相对于 TIAA 有 6% 左右的效果提升,这说明我们的研究是存在一定研究意义的.但是在本文的实验中,并未对时间复杂度进行测试,因为 TIAA 是一个离线的非开源警报关联测试工具,不能够精准的获取到关联的时间和算法消耗的资源.针对这一情况,我们已经开展其他关联方法的研究,力争在未来的工作中,提出一个可以应用于入侵关联环境的测试标准和可以参考的评价基准.

5 结论

本文将模糊积分和模糊认知图应用于警报关联中,提出了一种基于 Choquet 模糊积分的入侵检测警报关联方法,设计并实现了一个警报关联引擎.该方法能够检测已定义的多步攻击,计算量相对较小,且规则易于扩展.如果在分布式入侵检测系统中嵌入本方法,能够实现在线识别多步攻击的入侵检测系统.

在未来的工作中,我们将尝试将本文方法应用到我们研究团队实现的在线入侵警报关联分析系统中,并在真实网络环境下进行算法效能的测试,给出我们方法的实际应用效果参考,为以后的研究者提供入侵警报关联方法在线环境下的实验数据.

参考文献

[1] Corporation Symantec. Symantec Global InternetSecurity Threat Report Trends for 2008[EB/OL]. http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_internet_security_threat_report_xiv_04-2009.en-us.pdf.

[2] Peng Ning, Yun Cui, Douglas S. REEVES, DINGBANG XU. Techniques and tools for analyzing intrusion alerts[J]. ACM Transactions on Information and System Security, 2004, 7(2): 274 – 318.

[3] James M Keller, Jeffrey Osborn. Training the fuzzy integral [J]. International Journal of Approximate Reason, 2002, 15 (1): 1 – 24.

[4] 王熙照.模糊测度和模糊积分及在分类技术中的应用[M].北京:科学出版社,2007.

[5] 哈明虎,吴从▮.模糊测度与模糊积分理论[M].北京:科学出版社,1998.

[6] 杨锋,钟诚,李智.基于概率模糊认知图的 Mstream 攻击检测方法[J].计算机工程,2006,32(10):125 – 127.
YANG Feng, ZHONG Cheng, LI Zhi. Mstream attack detection approach based on probabilistic fuzzy cognitive map[J]. Computer Engineering, 2006, 32(10): 125 – 127. (in Chinese)

[7] Luo Xiang-feng, Gao Jun. Probabilistic fuzzy cognitive map based on belief knowledge database[J]. Journal of Computer Research and Development, 2003, 40(7): 925 – 933.

[8] 钟诚,杨锋.基于概率模糊认知图的混合入侵检测方法[J].小型微型计算机系统,2006,27(5):783 – 787.
ZHONG Cheng, YANG Feng. Hybrid intrusion detection approach based on probabilistic fuzzy cognitive map[J]. Journal of Chinese Computer Systems, 2006, 27(5): 783 – 787. (in Chinese)

[9] Valdes Alfonso, Skinner Keith. Probabilistic alert correlation [A]. Proceedings of the 4th International Symposium on Re-

cent Advances in Intrusion Detection, 2001[C]. London: Springer-Verlag, 2001. 54 – 68.

[10] Cuppens F, Miege A. Alert correlation in a cooperative intrusion detection framework[A]. IEEE Symposium on Security and Privacy, 2002[C]. Oakland, CA: IEEE, 2002. 202 – 215.

[11] Dain O, Cunningham R K. Fusing a heterogeneous alert stream into scenarios[A]. The 2001 ACM Workshop on Data Mining for Security Application[C]. Philadelphia, Pennsylvania: ACM, 2001. 1 – 13.

[12] Li G, Hamilton H J. Basic association rules[A]. Proceedings 2004 SIAM International Conference on Data Mining (SDM'04)[C]. Lake Buena Vista: Soc Ind & Appl Math, 2004. 166 – 177.

[13] Wai Hoau, Chan K C C. Mining changes in association rules: A fuzzy approach[J]. Fuzzy Sets and Systems, 2005, 14(1): 87 – 104.

[14] QIN XZ, LEE W. Discovering novel attack strategies from INFOSEC alerts[A]. ESORICS 2004[C]. Sophia Antipolis: LNCS(3139), 439 – 456.

[15] 郭帆,叶继华.基于 IDMEF 和分类的报警聚合[J].计算机应用,2008,28(1):250 – 253.
GUO Fan, YE Ji-hua. Alert aggregation based on IDMEF and category[J]. Journal of Computer Applications, 2008, 28(1): 250 – 253. (in Chinese)

[16] 郎良.网络攻击分类描述与典型攻击对策研究[D].西安:西安电子科技大学,2004.

[17] 何信振,胡维华.一种基于警报数据关联的入侵检测系统模型[J].计算机工程与科学,2009,31(8):30 – 32.
HE Xin-zhen, HU Wei-hua. An IDS model based on alert correlation[J]. Computer Engineering & Science, 2009, 31(8): 30 – 32. (in Chinese)

[18] 於晓兰. DDos/DRDos 攻击的研究与设计[J].武汉职业技术学院学报,2008,7(3):83 – 86.
YU Xiao-lan. Research on DDoS/DDoS attacks and its design [J]. Journal of Wuhan Institute of Technology, 2008, 7(3): 83 – 86. (in Chinese)

[19] 姜千,胡亮.入侵检测系统评估技术研究[J].吉林大学学报(信息科学版),2009,27(4):383 – 388.
JIANG Qian, HU Liang. Research on security of telecommunication accounting system[J]. Journal of Jilin University(Information Science Edition), 2009, 27(4): 383 – 388. (in Chinese)

作者简介

努尔布力　男,哈萨克族,1981 年生,博士,研究方向为网络安全、入侵检测. E-mail: Nurbol_mail@163.com

柴　胜(通信作者)　男,1976 生,博士,主要研究领域为安全软件工程. E-mail: chaisheng@jlu.edu.cn