

文章编号:1001-9081(2016)02-0499-06

DOI:10.11772/j.issn.1001-9081.2016.02.0499

基于时间对抗的网络报警深度信息融合方法

邱辉*, 王坤, 杨豪璞

(信息工程大学, 郑州 450001)

(*通信作者电子邮箱 pioneerqh@126.com)

摘要:针对目前网络报警信息融合方法仅以单时间点为处理单元,无法适应网络攻击逐渐呈现出的隐蔽性强、持续时间长等特点,提出一种基于时间对抗的网络报警深度信息融合方法。面对多源异构报警数据流,首先采集并保存当前一个较长时间窗口内的报警信息,然后利用基于滑动窗口的流聚类算法对报警信息进行聚类,最后引入窗口衰减因子对聚类后的报警进行深度融合。真实数据的实验结果显示,与基本 DS 证据理论(Basic-DS)和指数加权 DS 证据理论(EWDS)融合方法相比,该方法有较高的检测率和较低的误检率,但因为采用了更长的时间窗口,精简率上略低;实际测试与性能分析也表明,该算法的时延较小,能更加有效地检测网络攻击,且能完成实时处理。

关键词:异构数据流;网络报警;深度信息融合;时间对抗;衰减因子

中图分类号: TP393.08 文献标志码:A

Network alerts depth information fusion method based on time confrontation

QIU Hui*, WANG Kun, YANG Haopu

(Information Engineering University, Zhengzhou Henan 450001, China)

Abstract: Due to using a single point in time for the processing unit, current network alerts information fusion methods cannot adapt to the network attacks with high concealment and long duration. Aiming at this problem, a network alerts depth information fusion method based on time confrontation was proposed. In view of multi-source heterogeneous alerts data flow, firstly, the alerts were collected and saved in a long time window. Then the alerts were clustered using a clustering algorithm based on sliding window. Finally, the alerts were fused by introducing window attenuation factor. The experimental results on real data set show that, compared with Basic-DS and EWDS (Exponential Weight DS), the proposed method has higher True Positive Rate (TPR) and False Positive Rate (FPR) as well as lower Data to Information Rate (DIR) because of longer time window. Actual test and theoretical analysis show that the proposed method is more effective on detecting network attacks, and can satisfy real-time processing with less time delay.

Key words: heterogeneous data flow; network alert; depth information fusion; time confrontation; attenuation factor

0 引言

面对网络攻击威胁的日益频繁,网络中的监控与检测设备也随之增多,如入侵检测系统(IIntrusion Detection System, IDS)、防火墙、病毒检测系统(Virus Detection System, VDS)。这些部署于现实网络中的多源传感器,源源不断地产生着海量的异构报警信息,为管理员提供及时的网络安全状况。目前,检测设备产生的安全防护数据普遍存在如下问题:1) 报警信息冗余度大,存在着大量的重复报警。IDS 有时会对同一个安全事件在短时间内产生大量重复的安全报警,从而减弱了报警信息的可读性。2) 漏报误报率偏高。经研究,文献[1]中指出,传统的反病毒软件无法检测 54% 的恶意软件;文献[2]中指出,传统检测技术无法发现 87% 的数据泄露事件;尤其是面对近年来新兴的高级持续性威胁(Advanced Persistent Threat, APT),一种具有隐蔽性强、持续时间长、危害性较大等特点的新型综合性网络攻击手段,传统防护手段检测难度更大。由于存在大量的冗余和误报信息,管理员对真实攻击的发现以及对该攻击作出相应防护的时间被严重拖

延。因此面对日趋严重的网络安全环境,必须融合 IDS、防火墙、VDS 等网络设备的安全防护数据,减小海量报警信息的误报和冗余,以提高网络攻击的实时检测能力。

安全报警信息融合技术最初是为解决 IDS 报警的误报与冗余问题,融合多 IDS 的报警信息,以弥补单一 IDS 设计上的缺陷。最初的融合方法主要分为基于相似概率的融合方法和基于因果关联的融合方法。基于相似概率的融合方法^[3]具有较强实时性,且在某些攻击的报警融合上效果明显,但其相似度衡量标准很难确定;基于因果关联的融合方法^[4-5]则从攻击者的角度来分析并挖掘攻击过程,但其过多地依赖于专家经验,且其扩展性不佳。随后文献[6]提出了多 IDS 环境中基于可信度的警报关联方法,消除各 IDS 报告警报的模糊性和冲突性;文献[7]针对 IDS 报警层次低,且彼此孤立的不足,提出了一种采用攻击策略图的警报综合分析方法,实现了对报警数据的进一步关联分析。但以上信息融合研究都仅局限于 IDS 本身,处理的数据均属于同种类型,不能够克服 IDS 设计上的固有缺陷。随着研究的深入,多源安全报警信息融合逐步开始处理多源异构信息,对 IDS、VDS、防火墙、主机日

收稿日期:2015-06-26;修回日期:2015-09-30。 基金项目:国家自然科学基金资助项目(61309013)。

作者简介:邱辉(1990-),男,河南永城人,硕士研究生,主要研究方向:网络安全、态势感知; 王坤(1975-),男,河南周口人,副教授,博士,主要研究方向:网络安全、数据挖掘; 杨豪璞(1993-),女,河南封丘人,硕士研究生,主要研究方向:网络安全、攻击检测。

志等不同类型的安全信息进行综合处理。文献[8]提出了一种结合 Nessus 和 Snort 的报警关联方法;文献[9]提出了一种基于改进 DS 证据理论的多源异构报警融合方法;文献[10]提出了基于分层融合的安全报警融合环境 (Security Alerts Fusion Environment, SAFE), 将多源安全报警信息融合成综合性安全信息。三者均对多源异构报警信息融合进行了研究, 但仍是基于单时间点的报警融合, 无法适应网络攻击呈现出的新特点。

目前, 网络攻击逐渐呈现隐蔽性强、持续时间长等特点, 安全产品对攻击不能够及时检测, 导致报警信息具有一定的漏报与时延。而现有的安全信息融合方法都是对即时报警信息的单时间点的处理, 因而时延的可疑攻击行为报警不能得到准确的融合结果与检测。因此, 本文对多源异构数据流进行深度的信息融合, 采用“时间对抗”的策略, 保存一段时间的数据流信息, 扩大被检测域, 将基于单个时间点的实时融合转变为基于历史窗的深度融合, 以此来检测发现慢性攻击的蛛丝马迹。该方法采用滑动窗口技术解决处理大量数据的时效问题, 在进行信息融合时引入时间衰减因子以解决融合时间跨度大所带来的准确性问题。

1 报警信息深度融合模型

下面首先对融合模型中出现的术语进行定义。

定义 1 当前窗口。即根据深度融合需要, 保存报警数据的窗口大小, 记作 w 。

定义 2 滑动窗口。更新当前窗口内的报警信息的最小单元, 其大小远远小于当前窗口, 记作 Δw 。

定义 3 窗口衰减因子。为保证最近发生的报警具有更大的权重, 而赋予报警事件一个衰减程度, 越早发生的事件其衰减程度越大, 记作 λ 。

报警信息深度融合模型由四部分组成, 分别为报警信息采集、数据预处理、报警聚类和报警融合, 如图 1 所示。面对多源异构报警信息, 为解决实时性问题, 本文采用分布式处理方法, 将需要消耗大量计算资源的预处理和聚类模块放在各传感节点进行完成。

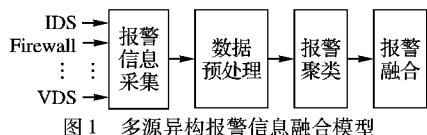


图 1 多源异构报警信息融合模型

步骤 1 报警信息采集。对网络安全传感器中的报警信息进行采集, 保存当前窗口内的报警数据, 并按照滑动窗口不断更新当前窗口内的信息。

步骤 2 数据预处理。首先对数据进行清洗, 通过设置过滤规则, 将不符合规范的数据过滤掉。例如字段缺省, 或参数错误, 超出设定范围, 将此类报警信息视为非法信息, 直接过滤掉。其次为方便后续报警信息聚类和融合的处理, 对多源异构数据统一格式。

步骤 3 报警聚类。面对海量冗余的报警信息, 为减轻后续整体融合的负担, 避免造成网络阻塞, 并增加报警信息的可读性, 便于管理员对网络状况的及时了解, 该模块对报警信息进行聚类, 将重复和相似的多条报警聚合成同一条报警。而本文需要面对保存较长时间的报警数据, 如果直接对当前窗口的大量数据进行聚类, 其无法满足报警融合的实时性要求。因此本文提出了基于滑动窗口的报警流聚类算法, 对一

个时间较短的滑动窗口进行聚类, 再利用其聚类结果不断更新当前窗口的报警聚类信息, 从而大大降低时间开销。其具体算法将于第 2 章详细介绍。

步骤 4 报警融合。报警信息融合是通过数据融合技术, 使多源异构报警相互补充、相互印证, 以融合成更加准确可信的报警信息, 减少单个传感器的漏报与误报, 从而进一步精简安全报警信息的数量, 提高安全报警信息的质量。面对聚类后的多源报警信息, 与传统的融合算法相比, 本文不仅要处理即时的安全报警, 还要面对当前窗口内较大的数据集, 如何快速有效地处理大数据集以及解决融合时间跨度大导致的准确性问题给本文的研究带来了挑战。为解决以上问题, 引入窗口衰减因子的概念, 通过对新近产生的报警赋予更高的权重, 以此来解决融合时间跨度大引起的问题; 利用窗口的不断更新技术, 对数据集进行快速高效的处理; 最后, 选用改进的 D-S 证据理论对当前时间窗口内的数据进行多源融合。其具体算法将于第 3 章详细介绍。

2 基于滑动窗口的报警流聚类算法

在该算法中, 首先通过计算 IP 地址、报警时间、报警类别等属性的差异来确定报警信息之间的相似度; 然后根据报警信息的相似度, 采用基于滑动窗口的流聚类方法对源源不断的报警数据流进行聚类分析。

2.1 报警相似度

假设存在两条具有 m 个属性的报警信息 $Alter_i \{srcIP_i, dstIP_i, srcPort_i, dstPort_i, Time_i, AttackType_i, other_i\}$ 和 $Alter_j \{srcIP_j, dstIP_j, srcPort_j, dstPort_j, Time_j, AttackType_j, other_j\}$, 其属性参数根据各自的特点可以分为数值型变量, 如 $Port$; 时间变量, 如 $Time$; 布尔型变量, 如 IP ; 枚举型变量, 如 $AttackType$ 。在聚类过程中, 不同攻击对不同属性的相似度要求不同, 因此需要分别定义不同属性的相似度函数。各属性的相似度函数定义如下。

1) 数值型变量相似度。

数值型变量相似度函数为:

$$s_n = \begin{cases} 1 - \frac{|N_i - N_j|}{\min(N_i, N_j)}, & \frac{|N_i - N_j|}{\min(N_i, N_j)} < 1 \\ 0, & \text{其他} \end{cases} \quad (1)$$

其中 N_i 和 N_j 为两条报警信息的一组数值型变量参数, 它可以是 $srcPort_i$ 和 $srcPort_j$, $dstPort_i$ 和 $dstPort_j$ 。

2) 时间变量相似度。

文献[11]指出两条报警若在 10 min 之内发生, 其极有可能是对同一事件的重复报警; 若时间差别在 1 h 以上则不可能反映同一攻击行为。因此定义时间变量相似度函数为:

$$s_t = \begin{cases} 1, & T < 10 \\ \frac{60 - T}{50}, & 10 \leqslant T \leqslant 60 \\ 0, & T > 60 \end{cases} \quad (2)$$

其中 $T = |Time_i - Time_j|$ (单位:min)。

3) 布尔型变量相似度。

布尔型变量只有 0 和 1 两个状态, 在判定相似度时采用著名的简单匹配系数法, 其相似度函数定义如下:

$$s_b = x/Sum \quad (3)$$

其中: x 为两个布尔型变量相同位的个数, Sum 为该布尔型变量的总位数。

4)枚举型变量相似度。

枚举型变量有多个取值,在相似度判定时依据其取值是否相同,其相似度函数定义如下:

$$s_e = \begin{cases} 1, & \text{AttackType}_i = \text{AttackType}_j \\ 0, & \text{其他} \end{cases} \quad (4)$$

2.2 报警流聚类算法

假设网络报警信息流形式如 $X_1, X_2, \dots, X_i, \dots$, 其报警产生时间分别为 $T_1, T_2, \dots, T_i, \dots$; 定义当前时间窗口大小为 w , 滑动时间窗口大小为 Δw 。为减轻计算交互负担,当报警数据集时间累积到一个滑动窗口 Δw 时,再将该数据集中的每一条报警信息 X_i 与初始微簇进行匹配,找到与其相似度最大的微簇。若该最大相似度能够满足阈值要求,便将其归入相应的簇;若不满足,便将其作为一个新的微簇加入到初始微簇群中。然后将滑动窗口内的数据保存到内存中,若内存中保存的数据量已超过定义的时间窗口 w ,便将最先到达的一个滑动时间窗口的数据全部清除。最后对当前时间窗口内的报警信息进行再一次处理,统计出当前报警信息中所有的类别。基于滑动窗口的聚类算法具体流程如算法 1 所示。

算法 1 基于滑动窗口的流聚类算法。

```

输入 经过预处理的报警数据流  $X_1, X_2, \dots, X_i, \dots$ ;
输出 当前时间窗口内的报警聚类。
1) BEGIN
2)   Cluster_Database = {Means1, Means2, \dots, Meansk}
      //初始化:设置该类报警常见的 k 个初始微簇
3)   set δ
      // 初始化:设置报警的相似度阈值
4)   S = 0
5)   i = 0
6)   While (i < t)
      //对新到的一个滑动窗口内的报警信息进行聚类
7)     begin
8)       j = 0
9)       while (j < k)
          //计算与窗口内的报警相似度最大的聚类中心
10)      begin
11)        sim = similarity(Xi, Meansj)
12)        if (sim > s)
13)          s = sim
14)        end
15)        if (s > δ)           //判断相似度能否满足阈值
16)          Add (Xi, Meansj)
              //满足则将报警信息加入到相应的微簇
17)        else
18)          AddCluster (Xi)
              //不满足则将其作为一个新的中心加入到微簇群
19)      end
20)      if (Tdata > w)           //更新当前窗口数据
21)        Delete (SWearliest)
22)        Compute_cluster    //统计当前窗口内的聚类信息
23) END

```

3 多源报警信息深度融合算法

在融合过程中,首先通过传感器得到的报警信息与该传感器对相应攻击的检测率 w_h 计算出外部攻击的发生概率 $p(h)$; 然后将各传感器的报警信息经过 D-S 证据理论合成,得到更加精确的攻击发生信息。

3.1 攻击发生概率

在计算攻击发生概率 $p(h)$ 时,其需要面对当前时间窗口内所有的报警信息,报警发生时间跨度大,为保证最近发生的报警具有更大的权重,本文引入窗口衰减因子 λ 的概念,并依据文献[12]中对衰减窗口的定义设计其表达式为:

$$\lambda = (1 - n^{-1})^{n-i} \quad (5)$$

其中: n 为当前时间窗口内滑动窗口的个数; i 为该报警所处的滑动窗口在当前时间窗口内到达的次序。因此结合窗口衰减因子,每一个滑动窗口中的攻击的发生概率 $p_i(h)$ 为:

$$p_i(h) = \begin{cases} w_h \lambda, & \text{attack}_h \text{ 被检测到} \\ 0, & \text{其他} \end{cases} \quad (6)$$

则该攻击不发生的概率 $p_i(\bar{h}) = 1 - p_i(h)$ 。

则融合所有滑动窗口中的攻击报警事件,得到整体的攻击发生概率 $p(h)$ 为:

$$p(h) = 1 - \prod_{i=1}^n p_i(\bar{h}) \quad (7)$$

则该攻击不发生的概率 $p(\bar{h}) = 1 - p(h)$ 。

为快速高效地对数据集进行处理,无需每一次都对所有滑动窗口的攻击发生概率重新进行计算。对于每一个滑动窗口而言,在下一个窗口到来之后,其衰减因子 $\hat{\lambda} = \lambda(1 - n^{-1})$, 则其对应的攻击发生概率 $\hat{p}(h) = (1 - n^{-1})p_i(h)$ 。

3.2 多源数据融合

证据理论是一种不确定性推理方法,最初由 Dempster 在 1967 年提出,后来由 Shafer 推广并形成证据推理。证据推理是建立在一个非空集合 Θ 上, Θ 被称为辨识框架,由一些互斥且穷举的元素组成。对于问题域中任意命题 A , 都应属于幂集 2^Θ 。在 2^Θ 上定义基本概率赋值函数 $m: 2^\Theta \rightarrow [0, 1]$, 且满足条件 $m(\emptyset) = 0$ 和 $\sum_{A \in \Theta} m(A) = 1$ 。

对多源异构数据进行 D-S 证据理论融合,参考文献[7]中对攻击发生支持概率的定义,辨识框架 $\Theta = \{h, \bar{h}\}$, 幂集 $2^\Theta = \{\emptyset, \{h\}, \{\bar{h}\}, H\}$ 。其中: \emptyset 表示“攻击既发生,又没有发生”; H 表示“攻击可能发生,也可能没有发生”。检测设备 i 的 mass 函数表示为: $m_i(\emptyset) = 0; m_i(h) = p_i(h); m_i(\bar{h}) = p_i(\bar{h}); m_i(H) = 0$ 。

但面对网络中的多源报警数据,并不能直接对所有传感器上的报警信息汇集融合。因为部署在网络上的传感器,分为基于网络的和基于主机的,其监控范围与内容具有很大差别。若两个传感器监控的对象完全不一样,在 D-S 融合过程中就会出现高冲突,严重影响融合结果。因此本文通过定义异构报警之间的关联度,将反映范围与内容具有一定关联性的报警进行融合。

定义 4 关联度。指不同传感器产生的报警信息反映同一攻击行为的可能性。

若两个报警信息反映的是同一内容,则其关联度为 1; 否则,则为 0。例如在一个网络中,存在一个网络入侵检测系统 A 、两个主机 B 和 C 。假设在某一时刻, A 产生了一条报警信息 A_i 反映主机 B 正在遭受攻击,若主机 B 防护系统也能检测到该攻击,其产生系统日志 B_i , 则报警 A_i 与 B_i 的关联度为 1, 而报警 A_i 与主机 C 产生的防护日志反映的不是同一内容,其关联度则为 0。主机 B 与主机 C 其防护日志所反映的都是各自主机的安全信息,则其报警关联度为 0。

通过分析,若两个传感器检测的范围有包含关系,那么二者的关联度则为 1, 则关联度 $relevance_{AB} = \begin{cases} 1, & A \subset B \\ 0, & \text{其他} \end{cases}$, 其中

$A \subset B$ 表示传感器 B 的检测范围包含 A 的检测范围。

由于传统 D-S 证据理论无法处理高冲突事件,而报警信息存在大量误报漏报,极易产生冲突,因此本文采用文献 [13] 中改进的 D-S 证据理论进行融合,并将该 D-S 证据理论结合关联度,推导其融合公式为:

$$\begin{cases} m(\emptyset) = 0 \\ m(h) = \prod_{i=1}^n m_i^{relevance_{ai}}(h) + kq(h) \\ m(\bar{h}) = 1 - m(h) \\ m(H) = 0 \end{cases} \quad (8)$$

其中: $k = 1 - \prod_{i=1}^n m_i^{relevance_{ai}}(h) + \prod_{i=1}^n m_i^{relevance_{ai}}(\bar{h})$; $q(h) = \frac{1}{\sum_{i=1}^n relevance_{ai}} \prod_{i=1}^n m_i^{relevance_{ai}}(h)$ 。

4 实验与结果分析

为验证本文模型及算法的可行性与有效性,搭建了一个实验网络,其网络拓扑结构如图 2 所示。网络中包括路由器、交换机、防火墙、入侵检测系统、服务器以及一台用户主机和一台攻击主机。其中 IDS1 和 IDS2 均安装 SNORT 入侵检测系统,Server1 和 Server2 安装 Windows 操作系统,Server3 安装 Solaris 安全操作系统。为采集真实攻击报警数据,在网络运行中,用户正常访问各个服务,而攻击者分别对 Server1 实施了 MSBLAST 蠕虫攻击,对 Server2 实施了 UDP FLOOD 攻击和 Unicode 解码漏洞攻击,对 Server3 实施了 SYN FLOOD 攻击。

在数据采集上选用网络运行中 IDS、防火墙产生的报警数据,以及 Server3 的主机安全审计模块(Base Security Module, BSM)日志。共选用了来自 3 类、4 个网络安全传感器的报警信息作为验证本文提出的融合方法的数据源,因此其能够满足数据的多源异构性。在实际数据采集中,网络运行 4 h,共采集报警数据 41 399 条,数据大小为 121 MB。为训练本文的原始聚类中心,采用林肯实验室的 DARPA1999 数据集。

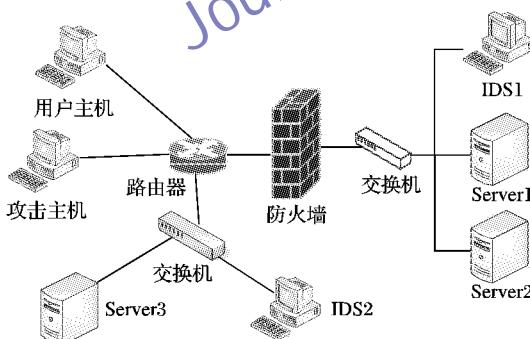


图 2 实验环境网络拓扑图

4.1 数据预处理

为方便后续处理,首先对源数据进行预处理,统一格式。本文采用目前通用的可扩展标记语言(Extensible Markup Language, XML)公共数据模型,对异构报警信息执行格式化。如下所示为主机 Server3 中的 BSM 安全审计日志规范映射成 XML 文件:

```
<IDMEF-Message version = "0.1">
<Alert alertid = "7" impact = "unknown" version = "1">
<Time>
<date>Sun 23 Aug 2015 </date>
```

```
< time > 17:37:31 PM < /time >
< /Time >
< Analyzer ident = "audit_bsm_server3" >
  < name > audit_bsm_server3 < /name >
< /Analyzer >
< Source >
  < Node >
    < name > ppp5-213. iawhk. com < /name >
  < /Node >
< /Source >
< Target >
  < Node >
    < name > server3. eyrie. af. mil < /name >
  < /Node >
< /Target >
< Process >
  < name > sh < /name >
  < pid > 2849 < /pid >
  < path > /usr/bin/sh < /path >
  < Arguments > , - sh < /Arguments >
< /Process >
< /Alert >
< /IDMEF-Message >
```

4.2 报警聚类实验

4.2.1 参数的设定

在初始参数的设定上,设定时间窗口 w 的大小为 2 h,其为已知大部分攻击的一个攻击周期,设定滑动时间窗口 Δw 的大小为 1 min,既能保障检测的时效性,又能使时间窗口保持一定的稳定性,避免因时间窗口更新过于频繁而导致的大量时间开销。因此在一个时间窗口内始终包含 $n = 120$ 个滑动时间窗口的数据。

相似度阈值 δ 的取值参考文献[14]中所给出的确定方法,该方法给出的是基于距离相异度的阈值确定,因此结合本文定义的相似度函数,给出其阈值如表 1 所示。从表 1 中可以看出,不同的攻击类型对不同的属性阈值要求不同,在针对 PortScan、远程未授权访问(Unauthorized Access from A Remote Machine, R2L)、未授权的本地超级用户特权访问(Unauthorized Access to Local Super User Privileges by A Local Unprivileged User, U2R)、拒绝服务(Denial of Service, DoS)攻击报警聚类时要求源地址与目的地址完全一致,而分布式拒绝服务(Distributed Denial of Service, DDoS)攻击则只需目的地址相同;攻击类型对端口号也有相应要求。

表 1 聚类阈值表

报警类型	源地址	目的地址	源端口	目的端口	时间/s
PortScan	1	1	0	0	360
R2L	1	1	1	0	120
U2R	1	1	0	1	120
DoS	1	1	0	0	240
DDoS	0	1	0	0	240
其他攻击类型	0	1	0	0	480

4.2.2 聚类结果及分析

对采集的报警信息进行实时聚类,分别对 4 个传感器产生的报警信息进行聚类分析,聚类结果如表 2 所示。计算聚类的精简率(Data to Information Rate, DIR)来反映聚类算法的有效性。从表 2 中可以看出,该方法能够有效地对多种报警数据进行聚类,大大减小了报警信息的冗余。

表2 聚类结果及分析

检测位置	检测工具	全体报警	聚类报警	DIR/%
IDS1	SNORT	15 341	378	97.54
IDS2	SNORT	8 092	191	97.64
Firewall	Firewall	10 439	364	96.51
Server3	BSM	7 527	52	99.31

4.3 报警融合实验

4.3.1 参数的设定

在得到聚类结果后,对当前时间窗口内的报警信息进行多源融合。首先需要确定各个传感器对不同攻击的检测率,根据所选取传感器系统的性能指标,给出它们对不同攻击的检测率,如表3所示。

表3 传感器对各攻击的检测率

攻击类型	SNORT	BSM	Firewall
Port_Scan	0.9	0.2	0.8
DNS_Attack	0.8	0.8	0.8
ICMP_Attack	0.9	0.4	0.2
Web_Attack	0.8	0.8	0.3
Http_Attack	0.9	0.8	0.6
TCP_Attack	0.9	0.9	0.7
FTP_Overflow	0.8	0.8	0.4

在关联度的确定上,IDS1 和 Firewall 检测位置为同一网络,IDS2 和 Server3 检测位置为同一网络,而两个网络中的传感器检测内容互不交叉。因此,各传感器之间的关联度如表4所示。

表4 传感器之间的关联度

检测位置	IDS1	IDS2	Firewall	Server3
IDS1	/	0	1	0
IDS2	0	/	0	1
Firewall	1	0	/	0
Server3	0	1	0	/

4.3.2 融合结果及分析

对多源报警信息进行 D-S 证据理论深度信息融合后,得到更加准确的报警信息 603 条,与聚类结果相比有了进一步的精简。对最终的融合结果,将从检测率(True Positive Rate, TPR)、误检率(False Positive Rate, FPR)和精简率(DIR)三个指标对算法的有效性进行评价。

TPR 为系统检测出的入侵攻击行为记录在总入侵攻击行为记录中所占的比例;FPR 为系统检测出错误的报警信息占总报警信息的比例;DIR 为原始报警数目与系统检测出的报警数目之差占原始报警数目的比例。

实验中共采集原始报警 41 399 条,攻击者共发送实际攻击 628 条,正确检测到攻击行为 575 条。计算本文方法的检测率、误检率和精简率,并使之与基本 DS 证据理论(Basic-DS)融合方法和文献[14]提出的指数加权 DS 证据理论(Exponential Weight DS, EWDS)融合方法进行对比分析,其结果如图3所示。从图3中可以看出,本文提出的融合方法有较高的检测率和较低的误检率;由于本文融合方法采用了更长的时间窗口,在精简率上略低。

4.4 算法性能分析

4.4.1 时间复杂度

面对海量、高速、动态的报警数据,随着时间的变化,新数

据被不断地读入,传统聚类算法需要对所有数据进行重复扫描,无法完成实时处理。通过上述对本文所提出的流聚类算法的描述与分析,该算法对数据进行一次扫描即可完成挖掘任务。通过分析,本文提出的基于滑动窗口的流聚类算法的时间复杂度为 $O(kt)$ 。其中: t 为一个滑动时间窗口内报警数量, k 为报警微簇中心点的数量。滑动时间窗口是一个比较小的时间窗口,本文采用的是 1 min,因此窗口内的报警数量不会太大,而报警微簇的中心点数量更是有限。因此相比传统聚类算法,该算法能够更好地适应对实时报警数据的挖掘。

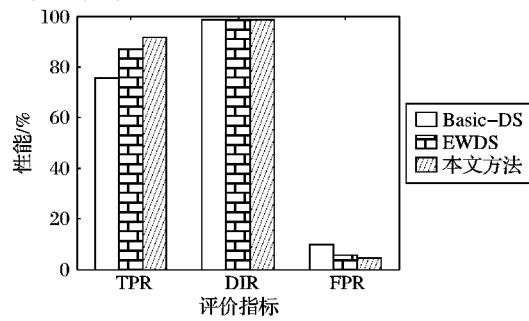


图3 与 Basic-DS 和 EWDS 比较结果

多源报警信息深度融合算法对所有窗口的攻击事件进行处理,并将多个传感器上的报警进行融合,其时间复杂度为 $O(n + mk)$ 。其中: n 为划定的衰减窗口数; m 为安全传感器的数量; k 为聚类后的报警数量,该数值都相对较小且稳定。

通过对上述两个子算法时间复杂度的分析可知,在现有的计算能力下,能够轻松完成对该算法的处理。在实际测试中证实算法时延较小,能够实时完成攻击检测。

4.4.2 网络负载

在网络负载方面,需要将原始报警数据采集到管理端进行数据保存及处理。在实际测试中,网络运行 4 h,共采集报警数据 41 399 条,数据大小为 121 MB。而所采集报警数据会被保存至管理端,无需重复传输,则平均传输大小为 8.6 KB/s。所测试环境其网络带宽为 100 KB/s,因此能够轻松完成对数据的采集,不会对网络造成阻塞。在实际测试中,用户一直能够正常访问各个服务,未出现明显的网络时延现象。

5 结语

面对新型网络攻击隐蔽性强、持续时间长等特点,检测攻击并及时响应的难度越来越大。因此本文研究基于时间对抗的网络报警深度信息融合技术,不再仅仅是对即时产生的报警信息进行融合,而是保存当前一个较长时间窗口的报警数据,采用“时间对抗”的策略,对其进行深度融合。实验表明,该方法能够更加准确地识别出网络中存在的攻击行为,且在性能上能够完成实时处理。下一步将继续研究不同攻击间的内在逻辑关联,以及该方法向并行计算平台的移植问题。

参考文献:

- [1] GROUP N. 2014 Global Threat intelligence report [R/OL]. [2014-03-27]. <https://us.query.ntt.com/en/resources/white-papers/global-threat-intelligence-report.html>.
- [2] Verizon. 2013 Data Breach Investigations Report [R/OL]. [2013-04-23]. http://www.verizonenterprise.com/resources/reports/rp-data-breach-investigations-report-2013_en_xg.pdf.
- [3] 穆成坡,黄厚宽,田盛丰,等.基于模糊综合评判的入侵检测报警信息处理[J].计算机研究与发展,2005,42(10):1679-1685.
(MU C P, HUANG H K, TIAN S F, et al. Intrusion-detection alerts processing based on fuzzy comprehensive evaluation [J]. Journal

- nal of Computer Research and Development, 2005, 42(10): 1679 – 1685.)
- [4] FATEMEH K, BEHZAD A. Automatic learning of attack behavior patterns using Bayesian [C]// IST 2012: Proceedings of the 2012 International Symposium on Telecommunications. Washington, DC: IEEE Computer Society, 2012: 999 – 1004.
- [5] ALIJABAR R, KOUROSH D A, ALI JI. Multi-level fusion to improve threat pattern recognition in cyber defense [J]. Journal of Mathematics and Computer Science, 2014, 8(2014): 398 – 410.
- [6] 梅海彬, 龚俭多. IDS 环境中基于可信度的警报关联方法研究 [J]. 通信学报, 2011, 32(4): 138 – 146. (MEI H B, GONG J D. Research on alert correlation method based on alert confidence in multi-IDS environment [J]. Journal on Communications, 2011, 32(4): 138 – 146.)
- [7] 李龙营. 入侵检测警报综合分析方法的研究与实现[D]. 西安: 西安电子科技大学, 2014: 25 – 30. (LI L Y. Comprehensive analysis approach of intrusion detection alerts and its implementation [D]. Xi'an: Xidian University, 2014: 25 – 30.)
- [8] TIAN Z, FANG B. A vulnerability-driven approach to active alert verification for accurate and efficient intrusion detection [J]. WSEAS Transactions on Communications, 2005, 4(10): 1002 – 1009.
- [9] 韦勇, 连一峰, 冯登国. 基于信息融合的网络安全态势评估模型 [J]. 计算机研究与发展, 2009, 46(3): 353 – 362. (WEI Y, LIAN Y F, FENG D G. A network security situational awareness model based on information fusion [J]. Journal of Computer Research and Development, 2009, 46(3): 353 – 362.)
- [10] 刘靖, 刘建伟, 张铁林, 等. 安全报警融合环境中信息的关联 [J]. 计算机工程与应用, 2011, 47(25): 107 – 111. (LIU J, LIU J W, ZHANG T L, et al. Association of information in security alerts fusion environment [J]. Computer Engineering and Applications, 2011, 47(25): 107 – 111.)
- [11] 何肖慧, 田盛丰, 穆成坡, 等. 分布式入侵检测环境中报警信息整合模型的设计与实现[J]. 计算机科学, 2006, 33(11): 266 – 269. (HE X H, TIAN S F, MU C P, et al. Designing and implementation of distribute intrusion detection system alerts fusion model [J]. Computer Science, 2006, 33(11): 266 – 269.)
- [12] RAJARAMAN A, ULLMAN J D. 互联网大规模数据挖掘与分布式处理[M]. 王斌, 译. 2 版. 北京: 人民邮电出版社, 2014: 110 – 112. (RAJARAMAN A, ULLMAN J D. Mining of massive datasets [M]. WANG B, translated. 2nd ed. Beijing: Posts & Telecomm Press, 2014: 110 – 112.)
- [13] 李弼程, 王波, 魏俊, 等. 一种有效的证据理论合成公式[J]. 数据采集与处理, 2002, 17(1): 34 – 36. (LI B C, WANG B, WEI J, et al. An efficient combination rule of evidence theory [J]. Journal of Data Acquisition & Processing, 2002, 17(1): 34 – 36.)
- [14] 赖积保. 基于异构传感器的网络安全态势感知若干关键技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2009: 54 – 73. (LAI J B. Research on some key technologies for heterogeneous sensors-based network security situation awareness [D]. Harbin: Harbin Engineering University, 2009: 54 – 73.)

Background

This work is partially supported by the National Natural Science Foundation of China (61309013).

QIU Hui, born in 1990, M. S. candidate. His research interests include network security and situation awareness.

WANG Kun, born in 1975, Ph. D., associate professor. His research interests include network security and data mining.

YANG Haopu, born in 1993, M. S. candidate. Her research interests include network security and attack detection.

(上接第 494 页)

- [3] ZHU Z Q, ZHANG Z M, WANG R, et al. Out-of-band ambiguity analysis of nonuniformly sampled SAR signals [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(12): 2027 – 2031.
- [4] PIERROTTET D, AMZAJERDIAN F, PETWAY L, et al. Linear FMCW laser radar for precision range and vector velocity measurements [J/OL]. MRS proceedings, 2008, 1076 [2015-05-02]. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7990284&fileId=s1946427400031225>.
- [5] GAO S, O'SULLIVAN M, HUI R. Complex-optical-field lidar system for range and vector velocity measurement [J]. Optics Express, 2012, 20(23): 25867 – 25875.
- [6] MAO X, INOUE D, MATSUBARA H, et al. Demonstration of in-car Doppler laser radar at 1.55 μm for range and speed measurement [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 599 – 607.
- [7] MAO X S, INOUE D, KATO S, et al. Amplitude-modulated laser radar for range and speed measurement in car applications [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(1): 408 – 413.
- [8] MAYMON S, OPPENHEIM A V. Sinc interpolation of nonuniform samples [J]. IEEE Transactions on Signal Processing, 2011, 59(10): 4745 – 4758.
- [9] SELVA J. FFT interpolation from nonuniform samples lying in a regular grid [J]. IEEE Transactions on Signal Processing, 2015, 63(11): 2826 – 2834.
- [10] XU S Q, CHAI Y, HU Y Q, et al. Reconstruction of digital spectrum from periodic nonuniformly sampled signals in offset linear canonical transform domain [J]. Optics Communications, 2015, 348: 59 – 65.
- [11] LIU W, LIU J, QIAO L. Use of polynomial approximation for reconstruction of periodic nonuniformly sampled signals [C]// ISPEMI '12: Proceedings of the 8th International Symposium on Precision Engineering Measurement and Instrumentation, SPIE 8759. Bellingham, WA: SPIE, 2012: 87590E.
- [12] BABU P, STOICA P. Spectral analysis of nonuniformly sampled data — a review [J]. Digital Signal Processing, 2010, 20(2): 359 – 378.
- [13] LAGUNAS E, NAJAR M. Spectral feature detection with sub-Nyquist sampling for wideband spectrum sensing [J]. IEEE Transactions on Wireless Communications, 2015, 14(7): 3978 – 3990.
- [14] REHFELD K, MARWAN N, HEITZIG J, et al. Comparison of correlation analysis techniques for irregularly sampled time series [J]. Nonlinear Processes in Geophysics, 2011, 18(3): 389 – 404.

Background

This work is partially supported by the Natural Science Foundation of Hubei Province (2014CFB815).

FANG Jianchao, born in 1991, M. S. candidate. His research interests include laser radar signal processing.

MAO Xuesong, born in 1975, Ph. D., professor. His research interests include laser radar signal processing and nonlinear fiber optics.