昆明学院学报　2017，39(3)：54～57　ISSN 1674－5639　CN 53－1211/G4
Journal of Kunming University

基于 Snort 输入的智能神经网络计算机取证模型研究

贾学明，袁　策
（云南警官学院 信息网络安全学院，云南 昆明 650223）

摘要：传统的计算机取证技术大多是规则检测，通过构造合理的规则库和关键词，提取出有效的数据证据. 针对传统取证技术的不足，设计了一种新的智能神经网络计算机取证模型，结合成熟产品的成功检测结果对神经网络输入进行学习，以专家库预处理方式对规则库输入进行调整，并以 Snort 开源入侵检测的软件输出进行改造以适应神经网络学习训练. 用神经网络对可疑信息提前进行预警，再重点对相关信息进行检测.

关键词：神经网络；计算机取证；Snort；取证模型
中图分类号：TP391　**文献标识码**：A　**文章编号**：1674－5639(2017)03－0054－04
DOI：10.14091/j.cnki.kmxyxb.2017.03.013

Study Computer Forensic Model Based on Intelligent Neural Network with Snort Input

JIA Xueming，Yuan Ce
（Information Security College，Yunnan Police College，Kunming，Yunnan，China　650223）

Abstract：The traditional computer forensic methods are based on the rule detection to achieve the evidence by constructing the rule database and recognizing some keywords. Due to the shortage of the traditional computer forensic technology，a new intelligent neural network computer forensic model is designed，which was combined with the successful detecting results of the well-performed products to study the neutral network input，to adjust the rule database with the expert database preprocessing method，and to transform the soft output with Snort open source intrusion detection to adapt neutral network study training，to give early-warning of the doubtful information and to focus on detecting the information concerned.

Key words：neural network；computer forensic；snort；forensic model

随着信息技术和信息产业的迅猛发展，计算机犯罪形势越来越严峻. 特别对于一些存在争议的网络犯罪，采用计算机取证则是一种相对直接的解决方式.

由于信息数据爆炸式增长，以及机犯罪手段和技术不断翻新，给计算机取证提出了新的挑战，因此各种计算机取证自动化技术和工具相继出现. 但是，目前计算机取证人才仍极为稀缺. 同时要求取证人才不断更新知识来适应不同的取证需求[1]. 如能实现取证过程的自动化经验共享，将极大地提高取证效率.

1　神经网络在计算机取证系统中的应用

1.1　计算机取证的相关技术问题

计算机取证是指对计算机犯罪行为，利用计算机软硬件技术，依法对电子存储的数据进行提取、存档、分析和获取数字证据的过程. 它是一个新的研究领域，且是交叉学科，同时涉及的问题十分广泛. 现有的计算机取证模型，主要依靠取证人员的经验来完成. 而对复杂的实际案例，特别是面对 TB 级数据时，基于经验的取证方法注定陷入困难. 但事实上，一次成功的取证案例是难以用规则知识的手段进行描述、存储、再现的. 因此，即使是非常相近的案例，仍要求有同样经验的专业人员进行重复分析. 这就意味着，将基于经验的方法和基于取证模型的方法结合，才是计算机取证的关键所在[2].

1.2　规则知识与非规则知识

对泛在的知识结构来说，按照其描述方式可分

收稿日期：2017－04－19
基金项目：公安部科技创新资助项目(2013YYCXYNST078)；云南省刑事科学重点实验室研究资助项目(YJXK16005).
作者简介：贾学明(1975—)，男，云南昆明人，副教授，研究生，主要从事神经网络、计算机取证研究.

为两类:一类是采用显式语法进行描述的,这类知识称为规则知识. 对规则知识而言,一般可采用构造规则库的方法,让计算机进行高效扫描. 目前的大部分计算机取证软件都使用了规则库进行可疑数据选取. 而另一类知识,难以用显式语法进行描述,称为非规则知识. 对这类知识进行取证,一般凭专家经验来获得相应结果. 同时,非规则知识较难以构造规则库. 由于计算机取证中存在大量非规则知识,因此这也是难以进行自动化取证的原因.

1.3　神经网络在计算机取证中的应用

神经网络具有较强的样本容错特性,能够从海量数据中挖掘信息,同时具有自适应、动态学习更新的能力. 因此,可将神经网络方法用于计算机取证模型构建,从而得到智能化取证系统[3]. 一般采用的方法是:对每一次成功取证的非规则知识进行学习,让智能化的取证系统先对海量数据进行分类,提取可疑数据,且能够迅速减少需要取证人员进行分析的数据[4],大大减轻工作量. 同时,通过对非规则知识的学习,解决了计算机取证人员知识、经验共享的难题.

2　系统模型设计与 Snort 改造

2.1　系统模型

本文将给出基于神经网络构造的计算机取证模型设计. 所设计的系统由两部分组成,第 1 部分是对现有的成熟计算机取证系统进行敏感数据提取、改进以适应神经网络学习的需要,主要是对目前国内外流行的 Snort 入侵检测结果进行提取转化[5];第 2 部分是用神经网络工具对每一次取证人员的取证过程(有效取证与失败取证)进行记录学习,主要是利用 Matlab 中的神经网络工具箱进行设计[6],对一次成功的取证,记录其进行过的分析步骤和相关数据以方便系统自适应学习. 系统模型流程图如图 1 所示.

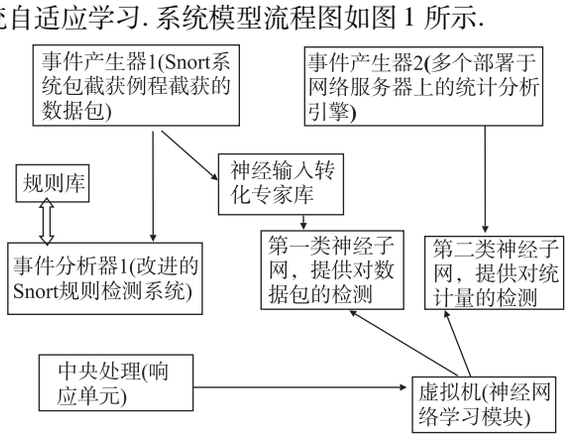

图1　系统流程图

1)正常工作时. 对 Snort 入侵检测分析筛选犯罪数据进行提取学习,学习单元自主通过学习规则库以提高分类能力,达到基本要求:专业取证软件能分析筛选出可疑数据,神经网络取证系统也能正常、正确分类出.

2)当取证人员进行一次有效的取证过程时. 记录取证人员的每一个操作和得到的值,当形成一个有效取证时,神经网络进行强化学习,强化分类能力,使无法描述的非规则知识自动加入到神经网络.

对每一次成功取证的非规则知识进行学习. 让智能化的取证系统先对海量数据进行分类,提取可疑数据,能迅速减少需要取证人员分析的数据,从而减轻工作量. 同时通过对非规则知识的学习,则可解决计算机取证人员知识、经验共享的难题.

2.2　基于神经网络取证的 Snort 改造

Snort 常用于网络入侵检测系统. 它是由全球无数程序员共同维护和升级的,能够在 IP 网络上进行实时的流量分析和数据包记录[7]. 与其他入侵检测系统相比,其具有系统小、易安装、便于配置、功能强大、使用灵活等特点. Snort 的源代码是完全公开和免费的,这为对系统进行深入分析提供了方便. 本文对其源代码进行了分析,并提出相应的改进方案以适应数据传送及神经网络学习.

对于某一类特定的入侵发生时,必伴随某一特定的审计记录变化,而这在规则设计时一般就能有明确的确定. 其实,审计记录是多种多样的,神经网络对每一审计记录都进行学习是不现实的. 为了神经网络的有效及高效学习,必须对入侵进行分类,每一类入侵进行一种小网络学习. 学习应在规则设计中进行明确. 为双系统的数据融合,应进行如下的规则系统改进.

1) 规则头中的规则动作改进.

(a)原有的 Snort 系统规则动作有 3 个关键字确定可能的动作.

Alert:使用选定的报警方式产生报警信号,然后启记录该数据包.

Log:记录该数据包.

Pass:忽略该数据包.

(b)应增加对神经网络数据传送的两个关键字.

Nni:让神经网络进行进一步入侵检测确认.

Nns:入侵明确,让神经网络进行学习,同时调

用 Alert 报警.

2)规则选项中的入侵种类与审计分类设计. 原有的 Snort 系统规则选项使用了 21 种选项关键字. 之间用分号“;”分隔,而关键字与参数之间使用冒号“:”分隔. 21 种选项关键字不需要列出. 为了对神经网络的数据传递,额外增加以下关键字.

Ii:入侵分类.

Es:审记种类.

NnID:子神经网络 ID.

以上 3 个关键字只是简单说明,可能的取值将在系统实现部分进一步说明. 此 3 个关键字一般设计在 1 种入侵发生时规则的最后端.

3)数据结构改进.

RuleTreeNode 数据结构,int type 规则类型中增加:

RULE_NNI:神经网络检测;

RULE_NNS:神经网络学习.

OptTreeNode 数据结构,int type 规则类型中增加:

入侵分类;

审记种类;

增加子神经网络指针.

RuleFpList 规则头函数列表中加入 2 个与神经网络有关的规则头函数指针列表.

RuleOptList 规则选项函数列表中加入 3 个与神经网络有关的规则选项函数指针列表.

3 神经网络计算机取证系统设计

3.1 系统组成

SA_NN IDS 采用更为复杂的检测技术,构建包括中央控制台、通信系统在内的完整入侵检测系统. 其基本结构包括:

1)探测器. 用于收集可疑数据包,重组、分解包头,作为规则系统检测的输入端.

2)检测引擎. 按照大小子网的形式来接收不同数据类型的数据包,从而实现针对不同类型的入侵检测进行侦探.

3)通信控制单元. 包括虚拟机统计系统、神经网络通信系统等控制单元.

4)神经网络学习系统. 通过设置运行参数,从虚拟机、通信单元中提取统计数据以用于神经网络系统进行自适应学习. 同时决策学习,分析收敛情况,进行检测更新.

5)存储系统. 主要由一系列数据库构成. 包括报警事件数据库、专家库、检测规则库、神经子网模型等组成.

6)中央控制台. 即用户交互界面. 用于操作整个系统. 同时,控制台还控制整个通信子系统. 本文只提出基于技术的内部总控制细节,未对用户交互界面进行设计.

3.2 组网学习模型

组网学习模型的子网化分图如下图 2 所示.

由图 2 可知,在进行数据包检测时,按照子网协议划分 4 个子网. 每个协议子网再进行字段划分,从而得到更细致的子网划分结果. 例如,对 UDP 协议划分子网时,可得到源端口子网、目标端口子网、报文长度子网、检验子网等.

此外,统计检测子网是也很重要的子网之一,包括日志子网、监控子网. 日志子网用于将日志数据进行分类,然后形成特定事件的不同子网. 这样的划分,可以对网络流量、CPU 情况进行子网划分,也可以更好地监视文件操作、注册表修改等事件.

需要说明的是,本系统以 BP 神经网络组网.

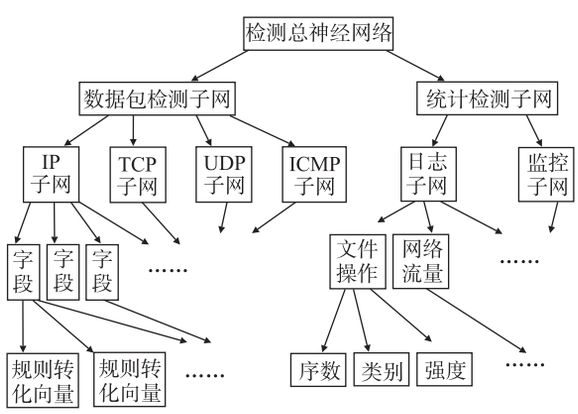

图2 子网划分模型示意图

3.3 基于虚拟机运行的分布探测与神经检测通信

为保证安全,含有入侵代码的数据包是不能在实际系统上运行的,但为了神经网络学习需要,又必须运行入侵代码. 这样,所有基于统计的学习都必须在虚拟机上运行.

虚拟机与宿主机是运行在同一计算机上的两个系统,虚拟机操作系统由宿主机创建和维护,虚拟机是基于文件型的,当发生系统崩溃等事故时(由于运行入侵代码,可能经常发生),只需进行简单的文

件拷贝操作就能恢复.

虽然两个系统运行于同一台计算机,统计数据提取却不能直接获得(这是由于现有的虚拟机技术所决定的).这样,我们在虚拟机和宿主机上设计了一条专供通信的虚拟网络(几乎所有虚拟机都提供虚拟网络及虚拟网卡服务),仍然可通过 Socket 方式进行通信.虚拟机上的统计模块不对虚拟网络进行审计,以保证统计不受通信影响.

4　实验结果与讨论

为提高检测系统效率,在实验中,针对不同类型的攻击采集了相应数据,并将这些数据用于系统前期的初始化学习.实验结果表明,针对不同类型的攻击数据,本系统的学习收敛周期较短.在针对 5 000 个攻击数据包的 10 000 迭代学习后,学习收敛程度可达到 0.89.同时,针对多种计算机入侵与攻击的数据,我们的系统在 100 G 的干扰数据源中,有效识别率可达到 92%.这些实验结果说明,本系统具有较强的自适应能力和自主学习能力.当然,在实际应用中,面临的情况可能更为复杂.但是自适应学习机制能够保证本系统适应大多数取证应用.

另一方面,相比防火墙,本文的计算机取证系统的自动化尚存在一些问题.目前在检测系统研究中已经引入了多方面的技术,单纯的引入某一种技术并不能解决存在的所有问题.因此作为一个神经网络自动化取证的系统模型,仍存在一些基本问题需进一步改进.

此外,本系统运行结果与有经验专家手工取证比较,自动化取证系统能从海量数据中有效分类出大部分的可疑数据,极大地降低了工作人员工作强度,但比较下来,存在以下问题:

1)误报与漏报.误报与漏报问题,是目前所有自动化取证系统面临的一个主要问题.但误报率与漏报率的高低,关系到该系统是否能够应用于实际领域.

2)神经网络学习权值漂移.神经网络在学习过程中,权值的漂移是一个不能忽略的问题.权值漂移是这样一种现象:当学习样本中存在大量的入侵数据,而正常数据较小时,学习后的神经元权值将向入侵出口方向漂移,此时对正常数据检测时,有可能检测出入侵,造成误报;当学习样本中存在大量正常数据,而入侵数据较小时,学习后的神经元权值将向正常出口方向漂移,此时对入侵数据检测时,有可能检测出正常,造成漏报.

当然,作为一个理论的模型,实际设计中还存在一些前面提到的问题.在今后的工作中,还需进一步研究计算机安全与神经网络的结合,并探讨模型存在的问题,形成更为完善的模型和设计方案.

[参考文献]

[1] BIENKOWSKI M, FENG M, MEANS B. Enhancing teaching and learning through educational data mining and learning analytics: an issue brief [J]. Education Department of America, 2012,51:336 - 339.

[2] BRYNJOLASSON E. A revolution in decision-making improves productivity [EB/OL]. [2016 - 12 - 06]. http://mitsloanmit. edu/erik-brynjolfsson.

[3] LEE Y H, HSIEH Y A, CN HSU C N. Adding innovation diffusion theory to the technology acceptance model: porting employees' intentions to use E-learning systems[J]. Educational Technology & Society, 2011, 14 (4): 124 - 137.

[4] BAPLER P, MURDOCH C J. Academic analytics and data mining in higher education[J]. Intemational Jounlal for the Scholarship of Teaching and Leanling, 2010, 4(2): 67 - 28.

[5] ZHU Y. The innovation of the basic computer instruction under the computational thinking[J]. Journal of Shenyang Agricultural University, 2014, 16(3): 338 - 341.

[6] CHEN G L. The innovation of the university to calculate the revitalization of education science and engineering research [EB/OL]. [2016 - 12 - 22]. http://blog. sciencenet. cn/blog - 512355 - 520901. html.

[7] FORCHHAMME S, WU X. Context quantization by minimum adaptive code length[J]. Proceedings of IEEE International Symposium on Information Theory, 2007, 35(6): 246 - 250.