基于 SNORT 体系的实时入侵检测研究

赵鑫玺，郑春厚，王春芳
（曲阜师范大学 信息技术与传播学院，山东 日照 276826）

摘 要：在充分发挥 SNORT 开源和以插件形式进行功能扩展的优势基础上，将 BP 神经网络优化算法运用到系统的规则训练模块和检测模块，构建了 SNORT 实时入侵检测系统。结合 SNORT 系统以规则匹配进行异常检测的特点，把从传输层捕获的数据包分为 TCP、UDP、ICMP 三类并分别编码，把编码之后的数据输入到神经网络中训练、检测。最后，通过实验验证了该方法的可行性。

关键词：SNORT；入侵检测；神经网络；网络安全

A Real-time Intrusion Detection Study Based on SNORT

ZHAO Xin-xi, ZHENG Chun-hou, WANG Chun-fang
(Qufu normal university,Dept of Information Technology and communication College,Rizhao,Shandong 276826,China)

Abstract:Based on the characteristics of open source and plug-ins of SNORT, this paper designed a real-time intrusion detection of SNORT system by using BP neural network in the training module and the testing module. The system first uses the Snort system's anomaly detection characteristic to capture data from the transport layer which contains TCP, UDP, ICMP and encoding them respectively, and then puts the data into neural network for training, testing. Experiments prove the feasibility of the method.

Key words:SNORT; Intrusion detection; neural network; network security

1 引言

基于特征匹配的入侵检测系统（IDS，Intrusion Detection System），因其成熟性和实用性已经被广泛应用在各类网络环境中。IDS 捕获网络数据流以匹配用户定义的一些规则，并根据检测结果做出响应。SNORT 是一个轻便的开源 IDS，可以完成实时流量分析和对网络数据包的安全性进行测试等功能，能完成协议分析、内容查找或匹配，能用来探测多种攻击和嗅探。SNORT 将所有已知的攻击以规则的形式存放在规则库中，是一个经典的、使用规则语言表示入侵特征的 IDS[1]。

SNORT 的规则检测算法使用了 BM(Boyer—Moore)、Wu-Man-ber 和 Aho-Corasick 等算法，它们都是字符串匹配算法，特点就是快速高效。以 BM 算法为例，其主要优势就是在文本匹配的过程中，匹配字符串越长，字符串匹配所需要的相对时间越短[2]。基于这样优秀的算法和同样优秀的系统结构，SNORT 能够在百兆网络满负荷的情况下对抓取到的每一个数据包进行检测。此外，由于其开源的特性，SNORT 具有很好的可扩展性，能够很容易地对系统功能进行添加和修改，一度成为 IDS 研究者所热衷的轻量级入侵检测工具。但是归根到底，SNORT 是一个基于模式匹配的检测系统，检测效率主要取决于入侵特征规则库的描述能力。如果规则能够将入侵特征准确地描述出来，SNORT 会拥有极高的检测效率[11]。但在实际情况下，SNORT 规则语言有限的描述能力在面对日益多样化的网络攻击方式和未知特征的攻击行为时，往往显得力不从心，更加难以追求实时的规则更新，因此，若引进一种具有自动学习和规则匹配的检测方法，将有望大大增强 SNORT 的检测效率。成熟而实用的 BP 神经网络技术正是一种合适的方法，而 SNORT 灵活的插件机制为这种思路提供了途径。

由于 SNORT 引入了插件机制，所以能够很容易地对 SNORT 进行功能上的修改、增加。SNORT 包含三种插件：预处理插件、检测插件和输出插件。本文所研究的 BP 神经网络技术的应用即处于预处理插件的位置，它工作在包解码之后，规则匹配之前，灵活性最强。

2 将 BP 神经网络应用于 SNORT 系统的理论分析

2.1 基于 BP 网络的 SNORT 入侵检测系统的优点

(1)BP 神经网络实质上实现了一个从输入到输出的映射功能，而且已有数学理论证明了它具有实现任何复杂非线性映射的能力。它能够通过学习带正确解的实例集合自动提取“合理的”求解规则，即具有自学习的能力，这使得它特别适合于求解内部机制复杂的问题。这一特点对 SNORT 系统实现自学习和动态规则更新有重要意义。

(2)BP 神经网络具备高度的自适应能力。通过对输入正常样本和异常样本的不断训练学习，神经网络不仅能够以很高的准确率识别出训练样本中已知的入侵行为特征，而且能够以一定的概率识别出新的入侵行为特征和已知入侵行为的变种形式。这种通过学习能够识别全新入侵行为特征的能力，可以克服基于模式匹配检测技术的局限性。

(3)BP 神经网络具有对输入信息概括和抽象的能力，表现为以一定程度的容错能力处理不完整的输入信息。在

网络环境中，常常会出现信息丢失、不完整或者变形失真的情况。在这种情况下，SNORT 通常会产生漏报和误报，而神经网络的非线性处理和概括抽象的特性对于处理此类情况是非常合适的[3]。

当然，事有利弊，神经网络在 SNORT 系统中的应用也不例外。在对 SNORT 起到巨大促进作用的同时，BP 神经网络面临着如何解决自身算法缺陷的问题。

2.2 标准 BP 算法的缺陷

标准的 BP 神经网络的缺点主要存在于以下两点：

(1) 由于 BP 算法本质上为梯度下降法，而它所要优化的目标函数又比较复杂，因此，容易出现“锯齿形现象”，使算法低效；在神经元输出接近 0 或 1 的情况下，会出现一些平坦区，在这些区域内，权值误差改变很小，使训练过程几乎停顿。

(2) 从数学角度看，BP 算法为一种局部搜索的优化方法，但它要解决的问题为求解复杂非线性函数的全局极值，因此，算法很有可能陷入局部极值，使训练失败[6]。

2.3 算法改进方案

从 BP 神经网络本身看，缺陷存在的根本原因在于算法。因此，本文拟采用 Levenberg-Marquarat 算法（以下简称 LM 算法）和动量法分别作为神经网络的训练函数和学习函数将其算法做适度改进，以增强算法的有效性。最后通过实验验证思路的可行性。

(1) 改进学习规则

标准 BP 算法在调整权值时，只按 T 时刻误差的梯度降方向调整。而没有考虑 T 时刻以前的梯度方向。从而常使训练过程发生振荡，收敛缓慢。而基于数值优化的 LM 算法则是在利用了目标函数一阶导数信息的基础上，进一步获取其二阶导数信息，从而缩短学习时间[4]。LM 算法的迭代公式为：

$$x_{k+1}=x_k-[J^T(x_k)J(x_k)+\mu]^{-1}J^T(x_k)V(x_k) \quad (1)$$

与牛顿法类似，LM 算法也是为了使训练快速收敛而设计的，而与牛顿法的不同之处就在于该算法避免了计算海森矩阵（Hessian matrix），取而代之的是使用海森矩阵的近似矩阵。计算权值调整率的公式为：

$$\Delta W=[J^TJ+\mu]^{-1}J^Te \quad (2)$$

式中，e 是误差向量；J 是网络误差对权值导数的雅可比（Jacobian）矩阵；μ是标量，当μ很大时上式接近于梯度法，当μ很小时上式变成了 Gauss-Newton 法，但使用的是近似的海森矩阵。牛顿法能够更快更准确地逼近一个最小误差，所以，应尽快地将上式向牛顿法转换，于是在每一步转换成功后，μ值都会减小，只有当发现下一步输出变坏时才增加μ值。依照这种方法，算法的每一步运行都会使评估函数向理想的方向发展，从而达到算法优化的目的[5]。

(2) 用动量法改进学习函数

动量法权值调整算法的具体做法是：将上一次权值调整量的一部分迭加到按本次误差计算所得的权值调整量上，作为本次的实际权值调整量，即：

$$\Delta W(n)=-\eta\nabla E(n)+\alpha\Delta W(n-1) \quad (3)$$

其中：α为动量系数，通常 $0<\alpha<0.9$；η是学习率，范围在 0.001～10 之间。这种方法所加的动量因子实际上相当于阻尼项，它减小了学习过程中的振荡趋势，从而改善了收敛性。动量法降低了网络对于误差曲面局部细节的敏感性，有效抑制了网络陷入局部极小[7]。

3 基于 BP 神经网络的 SNORT 检测体系

通常 SNORT 利用 Libpcap/Winpcap 从网络上获取数据包，经过对数据包的协议解码、预处理，把数据送入检测引擎，检测引擎根据规则库中的规则对每一个数据包进行内容查找及匹配，从而判断出是否发生了入侵行为[3]。而基于神经网络优化的 SNORT 系统的工作流程如图 1 所示。

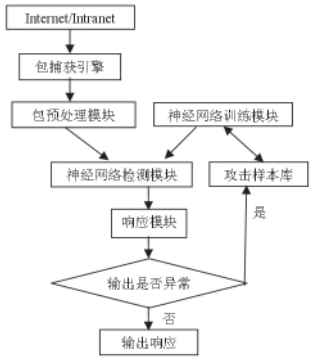

图 1 基于 BP 网络优化的 SNORT 系统结

首先，利用包捕获引擎从检测网络中捕获数据包，对收集到的网络数据包进行预处理，构建待检测数据矢量，即网络所传输的数据负载内容，该数据矢量就成为检测引擎的检测对象。与数据包预处理同时，将选定的攻击样本进行相应归一化等预处理后，输入神经网络的训练模块训练，一方面为检测引擎提供检测规则，以备其在检测时进行数据匹配；另一方面构建攻击样本库，为新一轮的训练和实时规则更新提供支持。

接下来，将第一步构建的待检数据矢量进行同样的归一化处理后，送入神经网络检测模块进行判断，如果发现有可疑攻击，则此时的数据包交给响应模块；响应模块负责对可疑攻击的处理，可以是手工的，也可以是机器自动的。

最后，响应为预设值的结果将作为输出直接呈现给用户，而响应为异常的数据矢量将被送入神经网络训练模块作为下一轮的异常数据样本库进行训练。经常的训练就保证了 SNORT 检测规则库的实时更新，只要异常响应阈值设定合理，就可以大大提高规则库的自动构建和匹配效率。

可见，数据矢量贯穿检测过程的始终，而各数据包所构成的数据矢量表现出很大的差异性，因此对数据进行归一化处理显得尤为重要。基本思想是完成将各检测值数据到[0，1]区间的映射，为简化运算，本文将采用较为简单的线性转换方式，运算公式为：

$y=(x-MinValue)/|MaxValue-MinValue|$，其中 x、y 分别为转换前后的值，MaxValue、MinValue 分别为样本的最

大值和最小值。

4 实验与分析

4.1 选取神经网络训练样本

本文采用 MIT 林肯实验室发布的入侵检测系统离线测试评估数据集 DARPA2000[10]，选取的数据是 LLDOS 2.0.2-Scenario Two，利用 Winpcap 工具对 dump 形式的文件进行解析，按照 UDP，TCP，ICMP 三种网络数据类型从中提取特定的数据包特征，选取带有总共 27 类特征的 TCP 包、UDP 包、ICMP 包各 500 个作为样本集。从样本集中随机选取 800 个作为训练集参与到三个不同神经网络进行训练，将训练样本分别存放在文件 Ttcp.txt，Tudp.txt 和 Ticmp.txt 中，输出则放在 Rtcp.txt，Rudp.txt 和 Ricmp.txt 文件中。剩余样本则作为验证集和测试集均分。

由于复杂多变的网络环境中所产生的数据差异可能是几十倍甚至上百倍，这对于神经网络的训练十分不利，故在神经网络训练之前需将原始数据进行归一化处理，以方便在程序运行时加快收敛速度。对数据包特征所提取的数据项进行归一化处理的主要目的是使数据间的变化差异不要太大，最好保证各数据项均在 [0，1] 区间内取值。本文就采用前文所提及的归一化方法，待处理的数据包信息包括 SIP、DIP、Sport、Dport、TTL、Flags、Protocol 和 Length 等。以下是部分 TCP 数据样本归一化处理后的格式。

表 1　部分 TC P 训练数据样本

SIP	SPORT	LENGTH	DPORT	TTL	ACK
0.000000003705	0.009073	0.149708	0.002592	0.000648	0.2891009703
0.714131073462	0.009073	0.149708	0.002592	0.001296	0.3239385527
0.902455244148	0.010369	0.154893	0.003240	0.001296	0.0000505701
0.909090907566	0.039533	0.348671	0.008425	0.001296	0.3239385527
0.909090907621	0.040181	0.370058	0.009073	0.000648	0.3658161638

4.2 神经网络训练过程

本研究中所需神经网络拟采用包含输入层、隐含层和输出层的经典三层反向前馈结构。以下对该网络在本研究中的具体应用方法做详细论述。

(1) 确定输入输出

实验借助 SNORT 系统误用入侵检测的优势，使用攻击数据样本进行训练。输入层的节点数取决于输入样本数据的特征向量，即一个节点对应一个输入变量，所以输入层节点为 27，输出层节点数为 3，每个节点为 0 和 1 两种输出。在进行实际网络数据检测时可根据需要增加。

(2) 确定隐层

BP 神经网络的应用中，隐层的节点数目设计是一个十分复杂的问题。虽然根据 Kolmogorov 定理 [8]，可以大致确定三层神经网络中隐含层节点的数目，但由于 TCP、UDP、ICMP 三种包各自的特性，他们在神经网络中训练时，隐层的节点数目不可能是相等的，具体数目需要在试验中反复实验才能得到。通过 mat lab 进行仿真对比，最终确定出隐层的节点数分别取以下值时，可得到较好结果：TCP 数据包 26-47 个，UDP 数据包 35-50 个，ICMP 包 30-55 个。

(3) 传递函数

BP 网络的传递函数有多种。Log-sigmoid 型函数的输入值可取任意值，输出值在 0 和 1 之间；tan-sigmod 型传递函数 tansig 的输入值可取任意值，输出值在 -1 到 +1 之间；线性传递函数 purelin 的输入与输出值可取任意值 [9]。在本研究中，输入层与中间层间的传递函数采用正切函数特性的 tan-sigmod 函数，中间层与输出层间的传递函数使用对数特性的 Log-sigmoid 函数，如此，输出将限定在 0 和 1 之间，但输出不可能是绝对的 0 或 1，故规定两个阈值。若输出单元的值低于 0.3，则认为是 0 输出，而高于 0.7，则认为是 1 输出。若输出处于 0.5 左右，则该包将被记录作为一种异常在下一轮训练中作为样本。

(4) 其他相关参数

本研究在实验中发现，在对三种网络数据包进行训练时，学习率所取值越大，所需训练次数越小，也就是函数收敛速度越快，因此将三种数据包在训练时的学习率统一设定为 0.9。单样本的误差精度 value=0.001，单样本的训练的最大次数 count=1000，正确率 M=98%，而隐层节点数有所不同，根据前述较好效果时的节点数范围，TCP 神经网络取 32；UDP 神经网络取 40；ICMP 神经网络取 37。所有神经网络参数设置完成后，其网络结构如图 2（以 TCP 网络为例）所示。

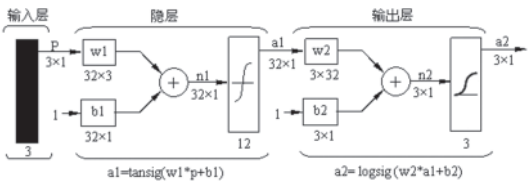

图 2　TCP 的三层 BP 神经网络结构

(5) 训练结果呈现

图 3 呈现了 SNORT 系统中各数据包在 BP 神经网络训练模块中的训练情况。从图 3 中可以看出三种数据包的训练误差皆呈逐渐下降趋势，在接近 900 次时都得到了收敛，训练效果理想。

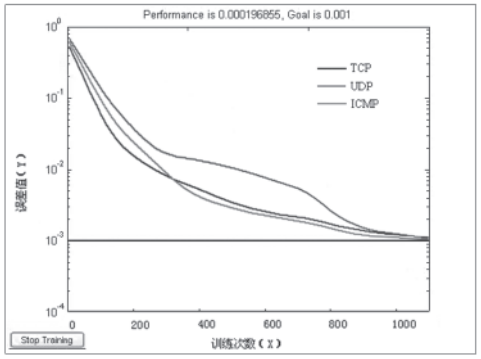

图 3　BP 神经网络训练曲线

4.3 实验结果与分析

(1) 实验环境

实验环境采用具有多台 PC 工作站的局域网，其中一

台作为网络服务器，安装 SNORT 入侵检测系统和 SQL 数据库，负责抓取所有网络数据包并进行检测；另外选取一台 PC 工作站作为攻击主机发送攻击数据流；其余则作为正常工作站产生正常的网络数据流。局域网采用 10M/100M 以太网卡和 10M/100M 以太网交换机搭建，各工作站配置皆为 Pentium（R）4CPU2.8Hz，512M 内存和 WindowsXP Sp2 操作系统。

(2) 实验实施方案

为了呈现出经 BP 神经网络优化过的 SNORT 系统的优越性能，本研究采用了对比实验的方法，即先由传统 SNORT 和优化过的 SNORT 分别对测试样本进行检测，之后对比其检测结果，评估优劣。另外在测试样本中加入若干新类型攻击包，以测试优化后的 SNORT 系统自动生成规则和实时检测的能力。新类型攻击包是指原本不包含在攻击样本库且在规则训练模块未曾训练的攻击包，在实验中的具体分布为 TCP 包 5 类，UDP 包 5 类，ICMP 包 3 类。因此，实验中所输入的攻击包数也按原类型与新类型的比例分配。

(3) 实验结果

攻击检测结果如下，表 2 为传统 SNORT 系统人工定义规则集检测结果，表 3 为基于 BP 神经网络优化的 SNORT 检测结果。表中所描述的检测率、虚警率和漏检率是衡量入侵检测系统检测效果的参数。他们的计算公式为：检测率 = 被正确标识的异常样本数目 / 异常样本总数，虚警率 = 被错误标识为异常的正常样本 / 正常样本数，漏检率 = 被漏检的异常样本数 / 异常样本总数。

表 2　传统 SNORT 系统人工定义规则集检测结果

数据包类型	检测流量	攻击包		被正确标志为攻击的包		被错误标志为攻击的正常包		检测率	漏检率	虚警率
		原类型	新类型	原	新	原	新			
TCP	1600	101	48	92	0	0	0	61.547%	38.453%	0.000%
UDP	952	83	40	76	0	1	0	61.869%	38.131%	0.121%
ICMP	768	66	32	60	0	0	0	61.425%	38.575%	0.000%

表 3　基于 BP 神经网络优化的 SNORT 检测结果

数据包类型	检测流量	攻击包		被正确标志为攻击的包		被错误标志为攻击的正常包		检测率	漏检率	虚警率
		原类型	新类型	原	新	原	新			
TCP	1600	101	48	98	42	3	0	93.658%	6.342%	0.207%
UDP	952	83	40	81	36	4	1	95.246%	4.754%	0.603%
ICMP	768	66	32	65	28	1	0	95.209%	4.791%	0.149%

通过检测结果可以看出，通过 BP 神经网络优化的 SNORT 检测引擎对 TCP，UDP，ICMP 三种攻击数据包的检测，在以低于 0.6% 的虚警率的情况下换取了高达 93% 以上的检测率，且漏检率稳定在 7% 以下。相比之下，由于传统的 SNORT 没有额外定义新类型攻击包的检测规则，因此不能识别新类型的攻击包，而基于 BP 神经网络优化的 SNORT 系统，具有自动学习自动构建新规则的能力，除漏检率有小范围提高之外，其他两项都能得到比较好的检测结果。

4 小结

文章在对传统 SNORT 的检测原理分析的基础上指出了其不足之处，提出以算法研究和实际应用都比较成熟的 BP 神经网络技术对其进行优化，在针对检测模块进行相应的编程之后以插件的形式应用于 SNORT 系统中，最后以实验的形式验证了这种思路的可行性。

本文创新点体现在两个方面，一是将 BP 神经网络自学习的特点与 SNORT 系统强大的插件功能结合，从而实现了 SNORT 规则库的自动更新，使 SNORT 实时检测成为现实。二是针对 BP 神经网络本身算法上的缺点提出了从学习函数和训练函数进行优化的方法，实验证明文中所用的优化方法使神经网络在预设的条件下达到了收敛，取得了理想的结果。

参考文献：

[1] 张亚玲，谢少春，汤来锋．基于活跃规则集的 Snort 高效规则匹配方法 [J]. 计算机工程与应用 .2008,44(24).

[2] 李洋，王康，谢萍 .BM 模式匹配改进算法 [J]. 计算机应用研究 .2004(4):58-59.

[3] 傅德胜，高建，柳亚婷 .Snort 平台下基于 BP 网络的预处理插件 [J]. 计算机工程与设计 .2008.5.29 卷 10 期 .

[4] 胡志刚 .Levenberg-Marquarat 算法及其在测量模型参数估计中的应用 [J]. 测绘工程 .2008.8.17 卷 4 期 .

[5] 李佳，周铁军．BP 神经网络优化算法在入侵检测中的应用研究 [J]. 计算机与信息技术 .2009.1.

[6] J.Cannnady, "Next generation intrusion detection:Autonomous reinforcement learning of network attacks," inProc.23rd Nat. Information Systems Security Conf, Oct.2006.

[7] Karl Levitt.Intrusion Detection: Current Capabilities and Future Directions. In Proceeding of the 18th Annual Computer Security Applications Conference.2007.

[8] Kristopher Kendall. A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems, Department of Electrical Engineering and Computer Science, May 21, 1999.

[9] Arboleda A F, Snort development diagrams [EB/OL], [2005-04-14], http://afrodita.unicauca.edu.co/~cbedon/snort/snortdevdiagrams.pdf.

[10] Sinha S,Jahanian F,Patel J M.WIND:workload-aware intrusion de-tection[C]//Proceedings of Recent Advances in Intrusion Detection(RAID),Hamberg,Germany,2006.

[11] Khan A,Revett K.Data mining the PIMA dataset using rough settheory with a special emphasis on rule reduction[C]//INMIC2004,IEEE,2004:334-339.

作者简介：**赵鑫玺**（1984—），男，在读研究生，主要研究领域：信息安全；**郑春厚**（1971—），男，副教授，主要研究领域：人工智能；**王春芳**（1983—），女，在读研究生，主要研究领域：数字水印技术。

收稿日期：2009-06-05