一种基于粗糙集理论的网络安全态势感知方法*)

梁 颖 王慧强 赖积保

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)

摘 要 网络安全态势感知是目前网络安全领域研究的一个热点问题。本文提出利用粗糙集理论进行网络安全态势感知,该方法把网络攻击行为作为网络安全要素,定量分析了各安全要素或安全要素组合对网络安全的威胁程度,最终建立了具有攻击行为、网络服务和安全态势3个层次的网络安全态势感知模型,并通过仿真实验生成了明确的网络安全态势图。采用粗糙集理论使得系统能够高效处理海量网络安全状态数据,生成易理解的攻击检测规则,清楚直观地反映攻击行为对网络整体安全状况的影响。

关键词 网络安全,态势感知,粗糙集理论,网络安全要素

A Method of Network Security Situation Awareness Based on Rough Set Theory

LIANG Ying WANG Hui-Qiang LAI Ji-Bao

(Computer Science and Technology College, Harbin Engineering University, Haerbin 150001)

Abstract Network security situation awareness is a hot research field in the network security domain. A method of network security situation awareness based on rough set theory is first used, in which network attack is regarded as the network security factor, and the threat degrees of each security factor or combination of them on network security are analyzed quantitatively, finally network security situation awareness model with three layers—attacks, network services and security situation is established and definite network security situation graph is created by the simulation experiment. The use of RST makes the network security situation awareness system be able to deal with large volumes of network data effectively, generate explainable attack-detection rules, then the influence of network attack on the whole network security can be reflected clearly and intuitively.

Keywords Network security, Situation awareness, Rough set theory, Network security factor

1 引言

随着网络的普及,其面临的威胁也越来越大,计算机病毒、木马程序、DoS/DDoS攻击日益猖獗。为保证网络安全运行,人们采用了入侵检测、防火墙、病毒检测等技术。然而这些技术属于被动防御手段,只能对攻击行为的某一局部进行检测,对一些分布式、规模化、复杂化的攻击则束手无策,更无法反映某些攻击行为的组合对网络安全的危害。因此,网络安全专家Bass[1]提出了网络安全态势感知(Network Security Situation Awareness, NSSA)研究,旨在解决上述种种问题。然而,目前该领域研究尚处于起步阶段,网络安全态势感知还没有形成一个明确的定义,也未能提出一个一致认可的解决方法。Bass[1]只提出并建立了网络空间态势感知框架,并没有实现具体的原型系统;Stephen G. Batsell[2]等开发了一个集成现有网络安全系统的安全框架,以提供对大规模网络的实时态势感知,该方法对网络结构部署要求有一定的局限;Jason Shifflet[3]采用"纵深防御"的思想,集成当前网络攻击检测技术,搭建了一个模块化技术无关框架结构,然而该方法只能对有限攻击进行检测,无法实现真正的网络安全态势感知。此处网络安全态势[3]是指某时刻由各种网络设备运行情况、网络服务状况及用户行为等因素构成的整个网络所处的安全状况。态势是一种状态,一个趋势,是一个整体、全局的概念。网络安全态势感知则指在大规模网络环境中,对能够引起网络安全态势发生变化的安全要素进行提取、理解、显示并预测未来发展趋势。这就需要融合海量网络安全状态数据。而粗糙集理论(Rough Set Theory, RST)借助信息系统表达和处理知识,具有能从海量数据中发现有用规律,并将其转化为逻辑规则的优势,已经在网络安全领域得到初步应用[4~6],并显示出了其他方法无法比拟的优势。

基于此,本文提出将粗糙集理论用于网络安全态势感知,把网络攻击行为作为安全要素,并对其进行冗余要素约简和要素重要性度量,量化网络攻击行为对网络安全的影响,进而实现网络安全态势感知,帮助网络管理人员更好地了解网络运行状况。

2 模块化系统体系结构模型

粗糙集理论是由波兰数学家Z. Pawlak[7,8]于1982年提出的。该理论建立在分类机制的基础上,将分类理解为在特定空间上的等价关系,这也就构成了对该空间的划分。笔者根据网络数据的特点,引入了粗糙集理论的思想,将基于RST的网络安全态势感知分为在线感知和离线感知两部分,分别实现网络安全态势的实时感知和非实时感知。该系统主

*)高等学校博士学科点专项科研基金项目(20050217007)、国防预研重点资助项目(413150702)、武备预研基金资助项目(51416060104CB0101)。梁 颖 博士研究生,研究方向计算机网络及应用、数据融合;王慧强 教授,博士生导师,研究方向为可靠性理论、计算机网络;赖积保 博士研究生,研究方向计算机网络、信息安全。

要由网络安全状态数据源集成平台、逻辑推理、异常发现、态势分析和安全态势可视化等模块组成，如图 1 所示。网络安全状态数据源集成平台实现多源异构网络安全状态数据的集成处理，为上层模块提供数据支持。逻辑推理模块利用异常行为库中的异常行为模式对当前网络安全事件进行在线推理，得到网络安全态势感知结果。异常发现模块采用匹配技术，根据异常行为库来检测网络中可能存在的各类攻击行为，并对异常行为库进行实时更新。态势分析模块将攻击行为作为安全要素，采用粗糙集理论分析各安全要素重要性，并对照态势知识库中的态势信息，最终生成网络整体态势，同时负责态势知识库的实时更新。保证准确实现网络安全态势感知。安全态势可视化模块负责根据下层模块生成的态势信息，生成直观的网络安全态势图，并将其呈现给网络管理员，为做出合理准确的决策提供依据。

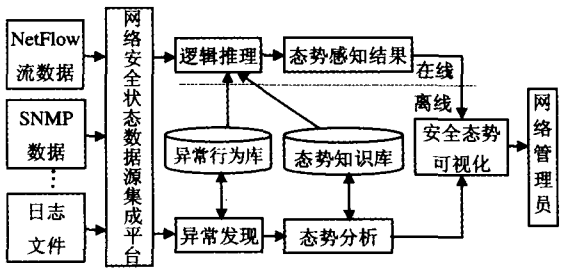

图 1　基于 RST 的 NSSA 系统结构模型

3　基于 RST 的态势感知方法

大规模网络环境下的安全态势感知，其处理的网络安全状态数据是海量的。而 RST[9] 借助于信息系统 $S=\langle U, R=C\cup D, V, f: U\times R\to V\rangle$ 表达和处理知识，是一种刻画不完整性和不确定性的数学工具，能有效分析和处理不精确、不一致各种不完备信息，并能从中发现隐含知识，揭示潜在规律的特点。采用 RST 保证了系统能从海量网络数据中提取各种网络安全要素，定量分析安全要素在网络安全中的重要性，并能进一步挖掘出安全要素组合对网络的威胁，最终建立大规模网络安全态势感知模型。感知当前网络安全状况并预测网络安全态势走向。

3.1　问题描述

结合粗糙集理论的思想，建立层次化 NSSA 模型，需做如下说明：各网络安全事件记录对应对象集 U，安全要素集 F 对应条件属性集 C，各攻击行为对网络的威胁程度集作为决策属性 r。本文把对网络的各种攻击行为作为系统的安全要素，参考 Snort 用户手册[10] 给出的攻击类别及威胁程度，规定威胁程度集 $t=\{$高、中、低$\}$ 作为系统的决策属性，选定条件属性集 $C=\{V_1, V_2, \cdots, V_k\}$ 对网络安全态势进行感知，建立态势感知信息决策表，如表 1 所示。

表 1　态势感知信息决策表

U	条件属性 C				r
	V_1	V_2	…	V_k	
1	V_{11}	V_{12}	…	V_{1k}	r_1
2	V_{21}	V_{22}	…	V_{2k}	r_2
	…	…	…	…	…
n	V_{n1}	V_{n2}	…	V_{nk}	r_n

结合上述分析，为了进一步说明网络安全态势的生成，我们给出如下定义。

定义 1　按安全要素分类对象集。不分明关系 F 对对象集 U 的划分，表示为 $U|N(F)$。

$$U|N(F)=\{(x,y)|(x,y)\in U^2\cap f\in F(f(x)=f(y))\} \tag{1}$$

定义 2　按威胁程度分类安全要素。设 R 是由威胁程度集 t 导出的分类，则 R 的 F 正域可表示为 $P_F(R)$：

$$P_F(R)=\bigcup E(R) \tag{2}$$

式中，$E(R)$ 表示所有通过用分类 $U|F$ 表达的知识，能确定划入 $U|R$ 类的对象的集合。

定义 3　属性重要性。各安全要素属性在 F 中的重要性，表示为 $I(V_i)$。

$$I(V_i)=\frac{|P_F(R)|-|P_{F-\{V_i\}}(R)|}{|U|} \tag{3}$$

结合 RST 思想，根据各安全要素属性的重要性约简态势感知信息决策，删掉对网络安全威胁性为 0 的安全要素，最终使得每一个记录代表一类具有相同威胁特性的样本，就得到了与态势感知规则相对应的结果，也就是所需的态势感知模型，其形式为“$A\to B$”，表示“若 A，则推出 B”。

3.2　层次化网络安全态势感知模型

网络安全态势感知的研究正处于起步阶段，尚需借鉴态势感知在其他领域运用的成熟理论和技术。Endsley[11] 于 1995 年提出了 Endsley 态势感知模型，该模型强调对与态势感知理解和量度有关的因素的研究，将实现态势感知分为态势提取、态势理解和态势预测三个层次。大规模网络的安全态势取决于网络提供的各项服务，同时系统所遭受的攻击行为又影响着各项网络服务的运行。为了对大规模网络做出整体的安全状况判定，分析网络所遭受攻击行为及攻击行为的组合对网络安全的影响，从构成整体网络安全态势的攻击行为层、网络服务层和网络安全态势层进行分析，结合 Endsley 态势感知模型，我们制定了层次化网络安全态势感知（NSSA）模型，如图 2 所示。

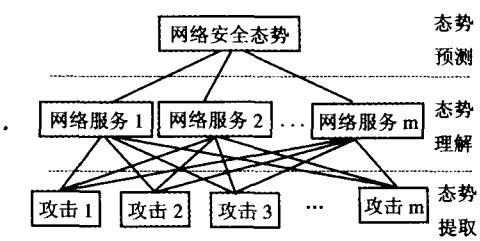

图 2　层次化 NSSA 模型

网络安全态势 H_T 定义为大规模网络环境中各项网络服务所占的比重向量 β 与服务层安全指数向量 H_S 的点积，即

$$H_T=\beta\cdot H_S \tag{4}$$

其中 $\beta=\{\beta_1, \beta_2, \cdots, \beta_m\}$，$m$ 为大规模网络中运行的服务数，元素 β_i 为各项网络服务所占的比重，可由管理员按各个服务在网络中的重要性来确定。$H_S=\{H_{S1}, H_{S2}, \cdots, H_{Sm}\}$ 是各项网络服务的安全风险值，是各安全要素在网络安全中所占的重要性比重向量 W_i 与安全风险等级向量 V_i 的点积，即

$$H_S=W_i\cdot V_i \tag{5}$$

其中 $W_i=\{w_{i1}, w_{i2}, \cdots, w_{in}\}$ 来自收集的网络数据样本集。根据粗糙集理论中属性重要性度量思想，考虑各安全要素在不同网络服务中所占的比重是不同的，故定义 w_{ij} 为影响网络服务 S_i 的第 j 个安全要素在网络安全中所占的比重，即

$$w_{ij}=\frac{I(V_{ij})}{\sum_{j=1}^{n}I(V_{ij})} \tag{6}$$

而 $V_i=\{V_{i1},V_{i2},\cdots,V_{in}\}$ 则代表安全要素在网络安全中的威胁等级。网络安全态势即通过量化网络攻击行为对网络服务运行的影响，最终得到整体网络安全状况的变化趋势 H_T。

4 仿真实验

4.1 实验环境搭建

为测试系统态势感知能力，搭建如图 3 所示的实验环境，在实验室局域网内进行测试。实验室采用高性能千兆交换机、Firewall、天阗 6.0IDS 系统及模块化多业务路由器。所选设备均支持选定的 NetFlow、SNMP、日志文件等多源异构数据源。另外有不少于 50 台 Windows XP/512M/160G 的 PC 机、千兆局域网。同时在局域网外部采用各种网络软件对局域网实施网络攻击，以获得更多、更全面的网络攻击数据。测试时间为 2006 年 7 月中旬～2006 年 8 月中旬。

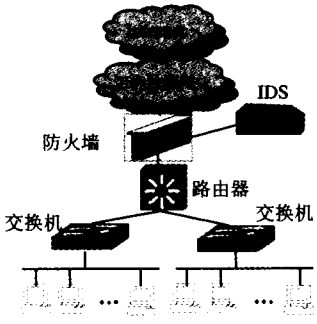

图 3 实验环境

4.2 网络安全态势建模

利用网络软件在外网对本实验室局域网进行网络攻击实验，选取其中典型的网络攻击行为作为安全要素。规定安全要素集 $F=\{V_1,V_2,V_3,V_4,V_5,V_6,V_7,V_8\}$，$V_i$ 代表实验局域网所遭受的第 i 个攻击。网络安全要素集 F 代表一组网络攻击行为，具体表示为 $F=\{V_1$(网络蠕虫)，V_2(smurf 攻击)，V_3(端口扫描)，V_4(病毒)，V_5(SYN Flood)，V_6(缓冲区溢出)，V_7(木马)，V_8(会话劫持)$\}$。F 中各项均为二进制取值，0 代表该攻击行为不存在，1 则代表存在。就某一时刻选用对象集 $U=\{1,2,3,\cdots,14\}$，根据专家经验对各安全要素及安全要素组合对整体网络安全状况的威胁程度进行赋值，同时建立态势感知决策表，如表 2 所示。

根据态势感知决策表，得到

$U|R=\{(1,4,6,9,10),(2,3,5,8,11,13,14),(7,12)\}$

又 $P_{F-\{V_1\}}(R)=\{(1),(2),(3),(4),(5,8),(6,10),(7),(9,14),(12),(13)\}$，从而 $|P_{F-\{V_1\}}(R)|=12$，同时 $|U|=14$。将上述三个结果代入属性重要性计算公式(5)，得到

$$I(V_1)=\frac{|P_F(R)-P_{F-V_1}(R)|}{|U|}=\frac{12-12}{14}=0$$

$P_F(R)=\{(1),(2),(3),(4),(5),(6,10),(7),(8),(9,14),(11),(12),(13)\}$，从而 $|P_F(R)|=12$。

其他属性的重要性计算与此类似。得到各安全要素的重要性分别为{0, 1/14, 0, 0, 1/7, 1/7, 0, 1/14}。从中删掉重要性为 0 的安全要素，最终得到约简的态势感知决策表，如表 3 所示。

表 2 态势感知决策表

U	条件属性								r
	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	
1	0	0	1	1	1	0	0	1	中
2	0	0	0	1	0	1	0	1	高
3	1	1	1	0	1	0	1	0	高
4	0	1	0	1	0	0	1	0	高
5	0	1	1	1	1	0	0	0	高
6	0	0	1	1	0	0	1	0	中
7	0	0	0	1	0	0	0	1	低
8	1	1	1	1	1	0	0	0	高
9	0	0	0	1	0	1	0	0	高
10	0	0	1	1	0	0	1	0	高
11	0	0	1	0	1	0	1	0	高
12	0	0	1	1	0	0	0	1	低
13	0	1	0	1	0	1	0	0	高
14	0	0	0	1	0	1	0	0	高

表 3 态势感知决策表约简结果

U	条件属性				r
	V_2	V_5	V_6	V_8	
1	0	1	0	1	中
2	0	0	1	1	高
3	1	1	0	0	高
4	1	0	0	0	中
5	0	0	0	0	中
6	0	0	0	1	低
7	0	0	1	0	中/高
8	0	1	0	0	高
9	1	0	1	0	高

4.3 实验结果分析

在所建立的层次化 NSSA 模型的基础上，为了更直观地表示网络安全态势，我们对各项数值进行量化处理，最终生成网络安全态势图。

各数值量化处理：

(1)在实验进行时间内，主要的网络服务数为 8 项，同时根据网络服务的重要性确定各项服务所占的比重向量为 $\beta=\{4/21, 1/21, 2/21, 7/21, 1/21, 3/21, 1/21, 2/21\}$；

(2)各安全要素在网络安全中所占威胁性比重可根据各属性重要性近似进行量化赋值，结果为{1, 5, 2, 1, 5, 4, 1, 4}；

(3)根据式(6)可计算得到各安全要素在大规模网络系统安全状况中所占的比重为 $W=\{0, 1/6, 0, 0, 1/3, 1/3, 0, 1/6\}$；

(4)对网络安全状况量化赋值以表明网络目前所处安全程度。规定网络安全风险值∈[0,1]区间。当网络安全风险值趋于 0 时，说明网络当前正处于极不安全状况；而当网络安全风险值趋于 1 时，则表示网络安全状况良好。

表 4 网络安全风险值

日期	3	6	9	12	15	18	21	24	27	30
风险值	0.42	0.13	0.6	0.75	0.5	0.92	0.4	0.8	0.5	0.67

由此，根据式(4)得到整个局域网在某个时刻的网络安全风险值。整个实验进行时间内，我们每两天计算一次网络安全风险值，得到各时刻的风险值，如表 4 所示。按照各时刻风

(下转第 147 页)

3.1 实验 1

使用 PSO 变异算子，子种群数目为 5，采用均匀分割，目的是测试 PSO 变异算子与种群分割策略的性能。实验结果如表 1 所示。

表 1 两种算法的进化代数对比

函数	收敛次数		平均收敛代数	
	PSOGA	SGA	PSOGA	PSOGA
六峰值驼背函数	49	41	35.2	242.1
DeJong 函数 2	50	36	51.4	276.5
RoseBrock 函数	50	43	41.0	182

实验结果表明，PSOGA 的收敛性能和搜索能力比 SGA 有很大提高，加速比最大可到 9.61，最小也有 4.4。由于采用了平均收敛代数指标来衡量算法的性能，表中的收敛代数比首次搜索到最优解时的代数要大出许多。在 50 次重复实验中，有几例甚至不超过 10 代就能首次搜索到最优解。

3.2 实验 2

在实验 1 的基础上，让 SGA 也均匀分割，单独测试 PSO 变异算子，实验结果见表 2。

表 2 PSO 变异算子的性能测试

函数	收敛次数		平均收敛代数	
	PSOGA	SGA	PSOGA	PSOGA
六峰值驼背函数	50	44	34.0	180.0
DeJong 函数 2	50	40	48.8	203.5
RoseBrock 函数	49	43	42.3	137

从表中数据可以看出，PSO 变异算子比传统变异算子性能要高出许多，三个函数几乎都能取到最优解。

对表 2 和表 1 进行比较，发现 SGA 由于使用了本文的种群分割策略，其性能得到了明显提升，平均收敛代数显著下降；PSOGA 则趋于稳定，两张表得出的数据基本一致。

两组实验结果表明，本文所提算法在改善收敛速度、克服早熟收敛方面有较好的表现，并且在增强局部搜索能力方面也有明显的改进。

结束语 本文将 PSO 算法的思想引入 SGA 算法，提出了一种改进的遗传算法新方案 PSOGA。新方案的核心是采用有覆盖的种群分割和 PSO 变异算子，前者是从维护种群多样性的角度来避免早熟收敛，后者则是从避免传统变异算子的随机性和盲目性的角度来克服早熟收敛，提高收敛效率。初步实验结果表明，PSOGA 算法达到了预期效果。

未来的工作，仍需进一步分析 PSO 变异算子的性能和数学性质，对于覆盖因子 α 和子种群数目对算法性能的影响也需要作进一步的实验和分析。

参 考 文 献

1 Eshelma I. J. The CHC adaptive search algorithm: How to have safe search when engaging in non-traditional genetic recombination. In: Founfation of Genetic Algorithm. Morgan Kaufmann Publishers, 1991. 265～283
2 Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutations in Gas. IEEE Trans. On SMC, 1994, 24(4): 656～667
3 马钧水，刘贵忠，贾玉兰. 改进遗传算法搜索性能的大变异操作. 控制理论与应用，1998，15(3)：404～407
4 苏小红，杨博，王亚东. 基于进化稳定策略的遗传算法. 软件学报，2003，14(11)：1863～1868
5 蒙祖强，蔡自行. 一种基于超群体的并行遗传算法. 计算机工程与应用，2001，21：28～30
6 林焰，郝聚民，纪卓尚，戴寅生. 隔离小生境遗传算法研究. 系统工程学报，2000，15(1)：86～91
7 丁永生. 计算智能-理论、技术与应用. 北京：科学出版社，2004
8 Skinner B T, Nguyen H T, Liu D K. Performance study of a multi-deme parallel genetic algorithm with adaptive mutation. In: 2nd international conference on autonomous rebots and agents, 2004
9 周明，孙树栋. 遗传算法原理及应用. 北京：国防工业出版社，1999

（上接第 97 页）

险值列表，最终得到实验进行期间网络安全态势演化图，如图 4 所示。图中结点表示网络在所处时刻的安全风险程度值。从整个网络安全态势图中，可以看出整个局域网在一个月时间内的安全状况。在第 12、18、24 天，网络安全风险值较大，表示攻击行为对网络安全威胁程度大；而在第 6 天，网络安全风险值较小，网络遭受攻击少，安全状况良好。

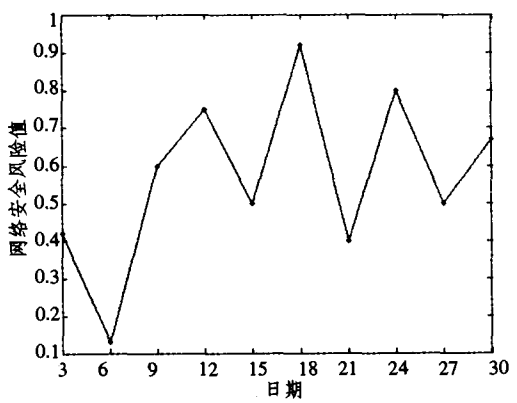

图 4 网络安全态势演化图

结论 采用粗糙集理论进行网络安全态势感知的方法具有能够度量各网络安全要素重要性、量化单个网络攻击行为或某些网络攻击行为组合对网络安全状况影响的特点，有利于将网络安全态势感知结果清楚直观地表示出来，方便网络管理员及时掌握网络安全状况并做出合理决策。但目前该领域的研究刚刚起步，利用粗糙集理论进行网络安全态势感知方法还存在一定的局限性。诸如粗糙集理论按等价关系进行分类，其准确性尚需完善，更高效的粗糙集属性约简算法也有待进一步深入研究。

参 考 文 献

1 Bass T. Intrusion Detection System and Multisensor Data Fusion [J]. Communications of the ACM, 2000, 43(4): 99～105
2 Batsell S B, Rao N S, Shankar M. Distributed Intrusion Detection and Attack Containment for Organizational Cyber Security [EB/OL]. http://www.ioc.ornl.gov/projects/documents/containment.pdf
3 ShifflетJ. A Technique Independent Fusion Model for Network Intrusion Detection. In: Proceedings of the Midstates Conference on Undergraduate Research in Computer Science and Mathematics, 1(3): 13～19
4 蔡忠闽，管晓宏，邵萍，等. 基于粗糙集理论的入侵检测新方法[J]. 计算机学报，2003，26(3)：361～366
5 陈秀真，郑庆华，管晓宏，等. 基于粗糙集理论的主机安全评估方法[J]. 西安交通大学学报，2004，38(12)：1228～1231
6 王旭仁，许榕生，王彦丽. 基于 Rough Set 理论的安全审计日志分析[J]. 计算机科学，2004，31(10)：109～111
7 Pawlak Z. Rough Sets[J]. International Journal of Computer and Information Science, 1982, 11: 341～356
8 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版社，2001，5：23～55
9 张文修，吴伟志. 粗糙集理论介绍和研究综述[J]. 模糊系统与数学，2000，14(4)：1～11
10 Roesch M, Green C. Snort Users Manual [M]. http://www.snort.org/docs/snortman-ja.pdf. 2006
11 Endsley M R. Toward a Theory of Situation Awareness in Dynamic Systems[J]. Human Factors, 1995, 37(1): 32～64