第 42 卷 第 6A 期
2015 年 6 月

计 算 机 科 学
Computer Science

Vol. 42 No. 6A
June 2015

基于粗糙集的实时网络安全态势评估的研究

吴朝雄 王晓程 王红艳 石 波

(中国航天科工集团二院 706 所 北京 100854)

摘 要 针对网络安全态势评估中评估精度以及实时性不足的问题,提出了基于粗糙集的实时网络安全态势评估方法。通过粗糙集理论从多样本数据中发现高质量的攻击规则集,生成多级规则树,将规则与实时攻击感知引擎结合,实现对动态数据流的在线分析检测,最后将实时检测的结果作为态势评估的依据,并根据层次化态势评估模型实时计算整体网络的安全态势值。经测试证明该方法有效地提高了态势评估的客观性、实时性、准确性。

关键词 网络安全,态势评估,粗糙集,动态,实时,规则树,攻击感知

中图法分类号 TP393.08 文献标识号 A

Research on Real-time Network Security Situation Assessment Based on Rough Set

WU Chao-xiong WANG Xiao-cheng WANG Hong-yan SHI Bo

(Institute 706, Second Academy of China Aerospace Science and Industry Corporation, Beijing 100854, China)

Abstract Aiming at the problem of the accuracy and real-time of situation assessment, a real-time network security situation assessment based on rough set method was proposed. It acquires high-quality rule sets from multi-sample through rough set theory, and generates multi-level rule trees, then integrates rule into real-time attack awareness engine to achieve online analysis and detection of dynamic data stream at the same time. And the result as the evidence of situation assessment is used to compute the value of situation in whole network according to the model of situation assessment at last. The method improves the assessment on accurate, real-time, objective sufficiently by experiments.

Keywords Network security, Situation assessment, Rough set, Dynamic, Real-time, Rule trees, Attack awareness

1 引言

网络安全态势感知旨在能够将事后处理转换为事前预防,尽量减少因网络攻击带来的灾害和影响。网络安全态势评估作为其重要的一环,贯穿于整个态势感知中,是整体网络安全状态最直接的体现,也是系统管理员采取措施和制定决策的依据。在网络态势评估研究方面,龚正虎等人[1]总结了态势评估的 3 大类方法。陈秀珍等人[2]综合考虑流量、主机、服务等方面的因素建立了层次化态势评估模型。王娟等人[3]从网络的脆弱性、容灾性、威胁性、稳定性 4 个方面归纳总结了态势指标体系。卓莹等人[4]分析了基于粗集的评估分析模型。赖积保等人[5]结合安全威胁与主机防御配置角度构建评估流程图。石波等人[6]构建了从攻击态势评估、防御态势评估到安全态势评估的评估层。但是网络安全态势评估方面的研究仍然存在以下一些问题:

1)态势评估性能指标没有统一的标准。

2)态势评估缺乏一定的实时性和客观性。传统的态势评估属于静态离线评估,通过累积 Δt 时间段的数据,对数据经过处理后再评估,评估的实时性和客观性主要依赖于 Δt 的取值。

3)评估准确度不高。首先评估采用的理论和方法对评估的准确度产生了很大的影响。如基于贝叶斯的评估[7]在构建贝叶斯因果关系图、条件概率表时依赖于专家的经验和水平;基于 D-S 证据理论的评估[8]方法对证据体概率的分配,同样也依赖于专家经验。其次,攻击感知的识别度不高也对评估的结果产生很大的影响,目前对于攻击识别主要还是依赖于个体的如 IDS 等设备,但网络环境复杂多变,单个设备对于攻击的识别难以准确把握,导致报警质量粗糙且数量庞大。

本文针对 2)、3)两点,提出了基于粗糙集的实时态势评估方法。该方法在一定程度上消除了很多人为因素的干扰,为后面攻击感知和态势评估提供了可靠的依据,并且通过在线评估,能够获得较为准确和客观的态势评估值。

2 基于粗糙集的实时网络安全态势评估分析

2.1 粗糙集理论

粗糙集是由波兰科学家 Z. Pawlak 在 1982 年首先提出来的[9]。粗糙集将客观世界抽象为一个信息系统。信息系统由四元组 S 表示,$S=\langle U, A, V, f\rangle$。$U$ 是对象或者事例的有限集合,称作论域,记为 $U=\{x_1, x_2, \cdots, x_n\}$,$A$ 是属性的有限集合,记为 $A=\{A_1, A_2, \cdots, A_n\}$,属性集 A 又常常分为两个集合 C 和 D,即 $A=C\cup D, C\cap D=\emptyset$,$C$ 表示条件属性集,D 表示决策属性集,将带有条件属性集和决策属性集的系统作为决策系

吴朝雄(1988—),男,硕士生,主要研究方向为计算机网络安全,E-mail:yangguangxiaoxiong@126.com;王晓程(1973—),男,硕士,研究员,主要研究方向为可信计算与网络安全;王红艳(1978—),女,硕士,高级工程师,主要研究方向为数据挖掘、网络安全;石 波(1988—),男,硕士,工程师,主要研究方向为信息安全。

统，记为 $S=\langle U,C\cup D,V,f\rangle$，$V$ 是属性值的值域，记为 $V=\{V_1,V_2,\cdots,V_n\}$，f 是信息函数，即 $f:U\times A\rightarrow V$，$f(x_i,A_j)\in V_j$。对于 A 中任意一个属性 A_i，如果有记录 x_i,x_j 对于属性 A 的取值相同，则称 x_i,x_j 基于属性 A_i 等价。$EQ=\{EQ_1,EQ_2,\cdots,EQ_n\}$ 称为基于 A 属性集的等价划分。

经过 20 多年的发展，粗糙集理论已被成功应用于知识发现、智能控制、神经专家系统、决策分析、股票数据分析等领域。其中粗糙集在知识发现中的应用主要体现在使用不可分辨（等价）关系对数据进行聚类形成等价集合，对属性、对象进行约简计算，生成决策规则[10]。在知识发现中，粗糙集与概率论、模糊集理论等最显著的差别在于它不需要任何的先验知识[11]，消除了人为的主观性，保证了分析结果的真实性。

2.2 基于粗糙集的实时态势评估思想

本文基于粗糙集的实时态势评估的基本思想，首先对获取到的态势样本数据集进行分类统计，并基于粗糙集对统计后的数据进行态势决策，其中态势决策包括构建态势决策表、约简安全态势属性、从约简后的决策表中提取高质量的安全规则集，然后对安全规则集进行归类合并，形成安全规则树，将其加载到实时攻击感知分析引擎，再对流经引擎的安全数据流实时动态分析，最后将检测到的攻击作为态势评估的输入，利用层次化评估模型计算态势评估值。基于粗糙集的实时态势评估流程如图 1 所示。

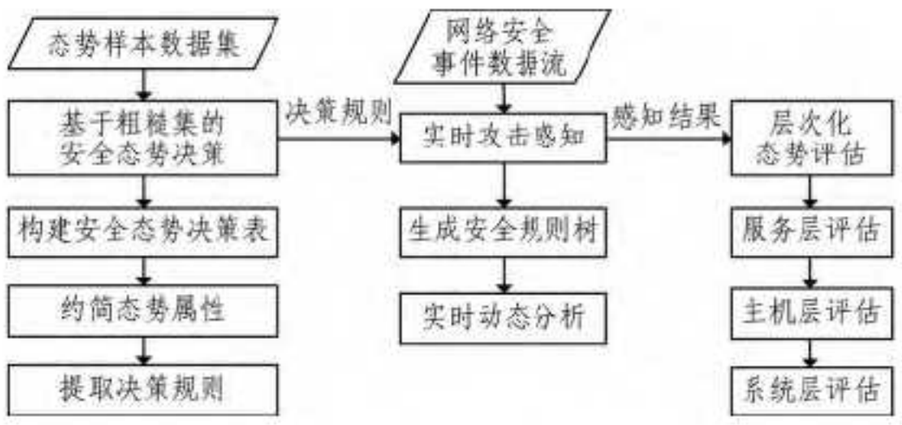

图 1 基于粗糙集的态势评估

2.3 基于粗糙集的安全态势决策

2.3.1 构建安全态势决策表

在网络安全态势分析中，采集的安全事件对应论域 U，即单个攻击相互组合后所形成的复杂攻击，各单个攻击为条件属性集 C，复杂攻击所造成的威胁程度集对应决策属性集 D。由于样本数据集为做了多次实验而得到的数据，可能存在重复性，因此为了避免构建的决策表过于庞大而不利于计算，在构建决策表之前先对实验数据集中的数据按照条件属性集和决策属性集进行统计分类得到带有事件发生次数的决策表。经过处理后，建立如表 1 所列的安全态势信息决策表。

表 1 安全态势信息决策表

U	NUM	条件属性 C			……		D
		C_1	C_2	C_3		C_n	
1	N_1	$C_{1,1}$	$C_{2,1}$	$C_{3,1}$		$C_{n,1}$	D_1
2	N_2	$C_{1,2}$	$C_{2,2}$	$C_{3,2}$		$C_{n,2}$	D_2
…	…						
n	N_n	$C_{1,n}$	$C_{2,n}$	$C_{3,n}$		$C_{n,n}$	D_n

表 1 中，NUM 为样本数据中对象次数，为了更好地理解和分析决策表，做如下的一些定义。

定义 1 对于任意 $P\in A,x_i,x_j\in U$，称 $U/IND(P)$ 为对属性 P 的不可区分关系。

$$U/IND(P)=\{(x_i,x_j\in U\times U)\mid\in P,p(x_i)=p(x_j)\}$$

定义 2 对所有的 $p\in P,f(x,p)=f(y,p)$，称为 P 对 U 的等价关系，记为 U/P。

定义 3 粗糙集是以上近似和下近似来近似定义粗糙集，其中对于集合 $X\subset U,X$ 的下近似表示为 $R\text{-}(X)$。

$$R\text{-}(X)=\bigcup\{y_i\subset U/IND(R):y_i\subset X\}$$

定义 4 等价关系 R 的子集 C 和 D，定义 D 的 C 正域为 $POS_C(D)$：

$$POS_C(D)=\bigcup C\text{-}(X)$$

定义 5 对任意的 $C_i\in C$，如果删除属性 C_i 使得

$$POS_{C-C_i}(D)=POS_C(D)$$

则称 C_i 属性为无效态势因子。

定义 6 对于提取的态势决策规则 $C\rightarrow D$，规则的可信度由 k 表示：

$$k=\frac{|pos_C(D)|}{|U|}=\frac{\sum_1^m N_i}{\sum_1^n N_i} \tag{1}$$

选取样本数据集中的一个样本集进行分析，其中选取的样本数据中，$C=\{C_1,C_2,C_3,C_4,C_5,C_6,C_7\}=\{$ping，SNMP，whois，SYN Flood，Finger，缓冲区溢出，特洛伊木马$\}$，$C_i=1$ 表示对象中存在该攻击，$C_i=0$ 表示不存在该攻击。$D=\{1,2,3,4,5\}$，D 值越大说明攻击产生的影响越大，建立如表 2 所列的安全态势决策表。

表 2 安全态势决策表

U	NUM	条件属性 C							D
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	
1	539	0	1	0	0	1	1	1	5
2	354	1	1	0	0	1	1	1	5
3	423	1	0	0	1	1	0	0	3
4	104	0	1	0	1	0	0	1	4
5	234	0	0	1	0	1	1	0	2
6	378	1	1	1	0	1	0	1	4
7	200	1	1	0	1	1	1	0	5
8	493	0	0	0	1	1	0	0	3
9	256	0	1	0	1	1	1	0	3
10	237	0	1	1	0	1	1	1	5
11	106	0	1	0	0	1	1	0	2

2.3.2 约简态势属性

对表 2 的数据按照定义 5 分别计算各个态势属性的 $POS_{C-C_i}(D)$，并与 $POS_C(D)$ 进行比较，消除无效的态势因子，简化决策表。约简如下：

$U/D=\{(1,2,7,10),(4,6),(3,8,9),(5,11)\}$，$U/C=\{(1),(2),(3),(4),(5),(6),(7),(8),(9),(10),(11)\}$，$POS_C(D)=\{(1,2,3,4,5,6,7,8,9,10,11)\}$，计算各属性的 $POS_{C-C_i}(D)$，如表 3 所列。

表 3 属性必要性

	$POS_{C-C_i}(D)$	无效态势因子
C_1	(1,2,3,4,5,6,8,10,11)	是
C_2	(1,2,3,4,5,6,7,8,9,10,11)	否
C_3	(1,2,3,4,5,6,7,8,9,10,11)	否
C_4	(1,2,3,4,6,7,8,10)	是
C_5	(1,2,3,4,5,6,7,8,9,10,11)	否
C_6	(4,5,8,9,11)	是
C_7	(2,3,4,5,6,7,8,9)	是

从表中计算结果可知必要属性为 C_1、C_4、C_6、C_7，对表 2 的属性进行约简合并后得到的决策表如表 4 所列。

表 4　安全态势约简决策表

U	NUM	条件属性 C				D
		C_1	C_4	C_6	C_7	
1	776	0	0	1	1	5
2	354	1	0	1	1	5
3	423	1	1	0	0	3
4	104	0	1	0	1	4
5	340	0	0	1	0	2
6	378	1	0	0	1	4
7	200	1	1	1	0	5
8	493	0	1	0	0	3
9	256	0	1	1	0	3

2.3.3　提取决策规则

对经过约简后的决策表进行规则提取，通过组合态势属性，发现决策表中高可信、高质量的决策规则，取 $C=\{C_6, C_7\}$，可得 $U/C=\{(1,2),(3,8),(4,6),(5,7,9)\}$，$U/D=\{(1,2,7),(4,6),(3,8,9),(5)\}$，$C-\{\{1,2,7\})=\{1,2\}$，$C-\{\{3,8,9\})=\{3,8\}$，$C-\{\{5\})=\varnothing$，$C-\{\{4,6\})=\{4,6\}$，$POS_C(D)=\{1,2,3,4,6,8\}$，计算可信度 $k=0.761$，说明 $C_6C_7\rightarrow D$ 规则的可信度为 0.761，且该规则符合实际中的组合攻击情况，说明通过粗糙集提取的态势决策规则是可信的、符合客观事实的。同理，可以依次组合其他态势属性提取出态势决策规则，并计算决策规则的可信度，从而得到决策表中所有可能的决策规则，其中一部分规则如表 5 所列。

表 5　部分态势决策规则

态势决策规则	$POS_C(D)$	可信度 k
$C_1\ C_6\rightarrow D$	{2,7}	0.167
$C_6C_7\rightarrow D$	{1,2,3,4,6,8}	0.761
$C_1C_6C_7\rightarrow D$	{2,3,6,7,9}	0.408
$C_4C_6\rightarrow D$	{1,2,5,6}	0.556
…	…	…

提取出的规则将被用来动态对攻击的感知，同时也能提高攻击感知的准确性。

2.4　实时攻击感知

2.4.1　生成安全规则树

实时攻击感知根据提取的态势决策规则集动态分析识别数据流中可能的攻击，同时结合网络的具体环境计算组合攻击规则的威胁值。它实际上是在线对动态数据流进行分析检测和处理。实时攻击感知首先对由粗糙集理论和方法提取到的规则按照攻击事件类别进行合并和归类，形成一棵有序的多级安全规则树，如图 2 所示。

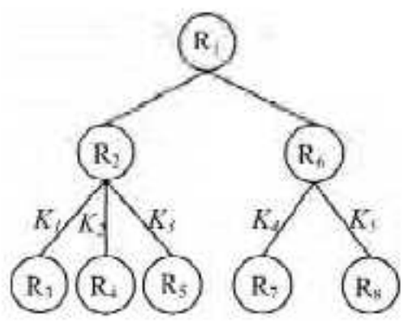

图 2　多级安全规则树

图 2 中，攻击规则归并的原则是攻击规则的起点相同，如组合规则 C_6C_7 与 $C_1C_6C_7$ 不能进行合并，因为这两个规则头不一样。通过对获取到的规则合并成多颗规则树能够清楚地看到不同攻击之间的相互组合情况。规则树中与叶子节点相连的连线表示该组合攻击的可信度 k，k 由式(1)计算得到。多级安全规则树中的节点攻击规则用八元组 $AttackRule$ 表示，记为 $AttackRule=\langle rulename, src\_ip, dst\_ip, dst\_port, timestamp, timeout, success, importance\rangle$，$rulename$ 表示规则名称，src_ip 表示源 IP，dst_ip 表示目的 IP，dst_port 表示目的端口，$timestamp$ 表示时间戳，$timeout$ 表示允许的后续攻击发生的超时时间，$success$ 表示攻击成功度，$success$ 值越大也说明攻击成功的可能性越高，取值范围为 1－5，$importance$ 代表该攻击在所有对系统攻击中的重要性，取值范围为 1－5。

2.4.2　实时动态分析

实时攻击感知引擎由过滤器、触发器、关联器、计算 $threat$ 组成，其结构图如图 3 所示。

图 3　攻击感知结构图

过滤器负责除杂和分流的任务，触发器根据上级分析结果触发下级的分析执行，关联器按照规则对数据流进行关联分析，$threat$ 模块根据当前的环境计算当前攻击的威胁值 $threat$。以组合规则 $C_6C_7\rightarrow D$ 为例来说明攻击感知的过程。分析引擎首先检测安全数据流中是否符合缓冲区溢出攻击的安全事件，如果没有，则说明此时没有缓冲区溢出攻击发生，分析引擎在给定的时间窗口内会根据规则一直检测缓冲区溢出这一安全事件。如果发现有缓冲区溢出安全事件发生，则触发分析引擎在 $timeout$ 时间窗口内关联特洛伊木马事件，并且此时将提升组合攻击的成功度即 $success$ 的值，以此来表明该组合攻击成功的可能性越大。如果在时间窗口内检测到特洛伊木马事件，则结束当前的关联分析，认为此次的组合攻击结束。如果超出时间窗口后仍未发现特洛伊木马事件，则认为当前组合攻击只成功了一半，同时也结束本次的组合攻击关联分析。

2.5　层次化态势评估

本文中采取分层感知方法，从服务、网内主机、网络系统 3 个层次对网络安全态势进行评估。

定义 7　复杂攻击链路 i 对本服务造成的侵害程度为服务期望威胁指数，用 T_{ser} 表示。

$$T_{ser}=k_i * count_i * 10^{threat_i} \tag{2}$$

$count_i$ 为发生的次数，$threat_i$ 为检测到的包括当前攻击的所有威胁值的最大值。

$$threat_i=\max\{threat_{i,1}, threat_{i,2}, \cdots, threat_{i,j}\} \tag{3}$$

定义 8　多个服务受到攻击后对单台主机系统的影响程度指数，用 T_h 表示。

$$T_h=\sum_1^n S_p * T_{ser_i} \tag{4}$$

S_p 表示该服务在所有主机开通的服务中所占的比重。S_p 的值一般根据主观的因素来确定。

定义 9　网内所有主机受攻击后对整个网络系统的影响程度指数，用 T_{sys} 表示。

$$T_{sys}=\sum H_p * T_h \tag{5}$$

H_p 表示各主机在网络系统中所占重要度的权重。

3　实验测试与结果

3.1　实验环境

为了测试和验证系统态势感知的性能，本实验搭建了如

（下转第 458 页）

[16] Yuan X T, Zhang T. Truncated power method for sparse eigenvalue problems[J]. The Journal of Machine Learning Research, 2013, 14(1): 899-925
[17] Saad Y. Numerical methods for large eigenvalue problems[M]. Manchester: Manchester University Press, 1992
[18] Mackey L W. Deflation methods for sparse pca[C]// Advances in Neural Information Processing Systems. 2009: 1017-1024
[19] Cadima J, Jolliffe I T. Loading and correlations in the interpretation of principle compenents[J]. Journal of Applied Statistics, 1995, 22(2): 203-214
[20] Vines S K. Simple principal components[J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 2000, 49(4): 441-451
[21] Jolliffe I T, Trendafilov N T, Uddin M. A modified principal component technique based on the LASSO[J]. Journal of Computational and Graphical Statistics, 2003, 12(3): 531-547

(上接第 437 页)

图 4 所示的环境进行测试和验证。

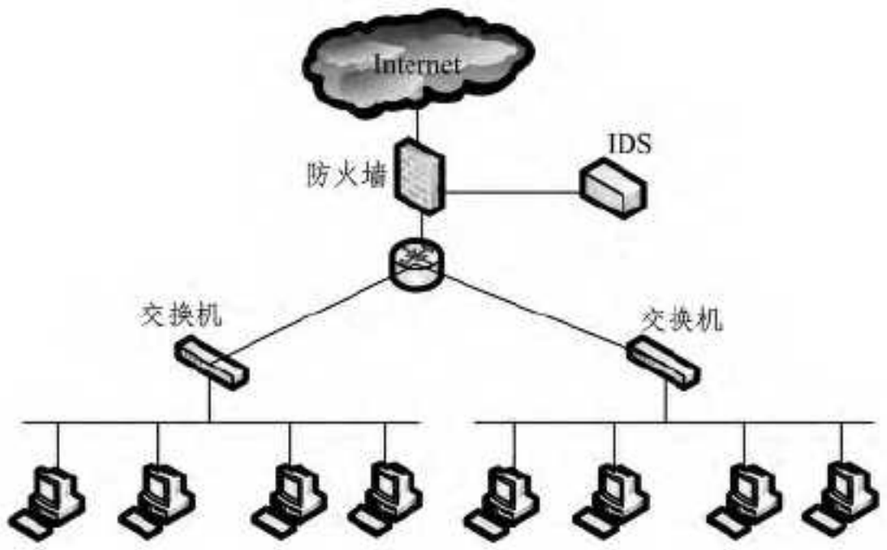

图 4 实验环境图

实验中放置了 10 台配置为 window XP/i3/3G PC 主机连接在互联网中，防火墙采用天融信千兆防火墙，IDS 为启明星辰千兆 IDS。

3.2 实验结果

首先对所有样本集中数据生成态势决策表，提取出大量高质量高可信的规则，并将规则转换成机器能识别的语言。其次，对文中涉及的指标进行量化。实验测试环境中开通了 FTP、Telnet、Http、DNS、SNMP 服务，各服务所占的比 $S_p = \{0.1, 0.25, 0.35, 0.15, 0.15\}$，按照每台主机上所开服务、拥有资源以及漏洞情况，将环境中 10 台主机的权重量化为 $H_p = \{0.103, 0.138, 0.172, 0.069, 0.069, 0.138, 0.172, 0.034, 0.172, 0.034, 0.069, 0.034\}$。根据实时攻击感知到的攻击结果，结合式(2)—式(5)以及各变量的量化值计算每时每刻的态势值。态势值越大，说明网络越处于不安全状态。取 2014 年 9 月 27 日 18:00—20:30 检测到的数据，计算得到各个时刻的态势值。实验中每 5 分钟评估一次态势值。得到如表 6 所列的态势数据。

表 6 实验结果

时间	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
态势值	845	785	2034	560	1467	365	895	768	803	654	358	295	103	206	489
时间	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
态势值	579	156	267	810	2327	754	796	1450	976	895	1320	1020	846	591	768

根据表 6 的数据绘制得到图 5 所示的态势曲线图。

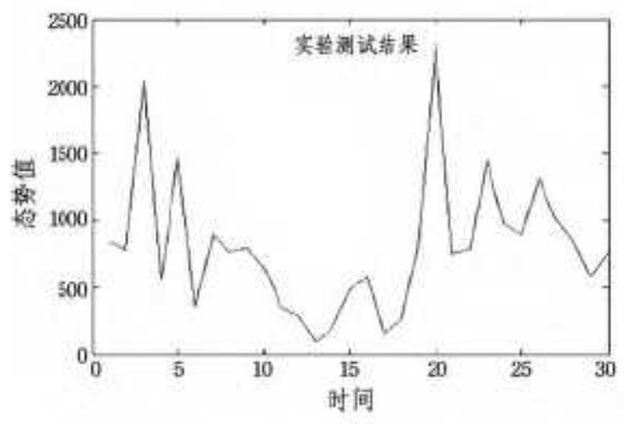

图 5 实验结果

从实验结果可以看出有几次的安全态势值的峰值较高，网络处于不安全状态，需要管理员采取一些措施改善网络状况。

结束语 本文提出了基于粗糙集的实时态势评估，首先用粗糙集中属性约简确定对安全态势产生影响的有效因子；其次，采用决策表进行规则的提取和发现。这一过程没有人为因素的干扰，保证了提取规则集的可靠性、准确性。同时由于需要及时发现复杂攻击，引入实时攻击检测引擎加载提取到的规则集，采用实时流计算思想对流经的安全事件数据流在线检测和分析，并将分析检测的结果作为实时态势评估的依据。实时攻击检测的结果在一定程度上保证了网络安全态势评估的准确性、实时性和客观性。最后经实验验证，文中态势评估的结果具有较高的实时性和准确性。由于攻击检测引擎不仅要将规则树加载到内存，同时还需要具有一定的存储能力，因此当规则集数量多而复杂时，可能需要消耗较多的内存资源，因此，下一步需要对方法进行分布式化，进行资源的合理分配。

参 考 文 献

[1] 龚正虎，卓莹. 网络态势感知研究[J]. 软件学报，2010，21(7)：1605-1609
[2] 陈秀真，郑庆华，管晓宏，等. 层次化网络安全威胁态势量化评估方法[J]. 软件学报，2006，17(4)：885-897
[3] 王娟，张凤荔，傅翀，等. 网络态势感知中的指标体系研究[J]. 计算机应用，2007，27(8)：1907-1909
[4] 卓莹，何明，龚正虎. 网络态势评估的粗集分析模型[J]. 计算机工程与科学，2012，34(3)：1-5
[5] 赖积保，王颖，王慧强，等. 基于多源异构传感器的网络安全态势感知系统结构研究[J]. 计算机科学，2011，38(3)：144-149
[6] 石波，谢小权. 基于 D-S 证据理论的网络安全态势预测方法研究[J]. 计算机工程与设计，2013，34(3)：821-825
[7] 康长青，郭立红，罗艳春，等. 基于模糊贝叶斯网络的态势威胁评估模型[J]. 光电工程，2008，35(5)：1-5
[8] 王琳，寇英信. Dempster-Shafer 证据理论在空战态势评估方面的应用[J]. 电光与控制，2007，14(6)：155-157
[9] Pawlak Z. Rough Sets[J]. International Journal of Information and Computer Science, 1982, 11(5): 341-356
[10] Pawlak Z, Gzymala Busse J, Slowinski R. Rough sets[J]. Communications of the ACM, 1995, 38(11): 88-95
[11] 王国胤，姚一豫，于一洪. 粗糙集理论与应用研究综述[J]. 计算机学报，2009，32(7)：1229-1246