• 信息安全技术 •

入侵检测数据集 KDD CUP99 研究

张新有， 曾华燊， 贾 磊

（西南交通大学 信息科学与技术学院，四川 成都 610031）

摘 要：为了更好评价各种入侵检测算法的性能，指出了入侵检测数据集应当具备的特点。分析了两种有影响的入侵检测数据集：MIT LL 入侵检测数据集和由此整理形成的 KDD CUP99 入侵检测数据集的特点及构成，重点分析了 KDD CUP99 训练数据集和测试数据集的各攻击类型及详细分布、数据集中每条连接的特征分类及其各个特征的含义，并对数据集的使用进行了说明。最后，对 KDD CUP 数据集存在的问题及相应改进措施给出了建议。

关键词：入侵检测算法； 训练数据集； 测试数据集； 攻击类型； 特征

中图法分类号：TP393 文献标识码：A 文章编号：1000-7024 (2010) 22-4809-04

Research of intrusion detection system dataset-KDD CUP99

ZHANG Xin-you， ZENG Hua-shen， JIA Lei

（School of Information Science and Technology，Southwest Jiaotong University，Chengdu 610031，China）

Abstract：To evaluate the performance of the various intrusion detection algorithms effectively， the necessary features of intrusion detection dataset are presented. The structures and features of two important intrusion detection datasets are analyzed including MIT Lincoln Labs intrusion detection dataset and KDD CUP99 intrusion detection dataset formed based on the former. All kinds of attack types and their detailed distribution， feature classes of each connection， the meaning of each feature， which are included in KDD CUP99 training dataset and testing dataset， are detailedly analyzed， and the usage of the KDD CUP99 dataset is also explained. Finally， the existing problem and improving direction for KDD CUP99 dataset are pointed out.

Key words：intrusion detection algorithms； training dataset； testing dataset； attack types； features

0 引 言

入侵检测系统(IDS)通过获取系统的安全审计数据，分析安全审计数据，从中提取系统的各种行为模式及行为特征，进而检测出系统中存在的某些入侵行为。目前入侵检测系统大多数采用误用检测和异常检测相结合的方法来分析检测系统中的入侵行为。

误用检测是根据已知的入侵或攻击方式的行为特征，建立攻击模式库或入侵特征库，通过分析被检测系统中的数据，采用某种模式匹配算法判断被检测系统中是否存在攻击或入侵行为。误用检测可以检测出已知模式或特征的入侵行为，但随着攻击手段不断变化，无法在一个系统中对所有的攻击行为进行特征编码。因此误用检测很难准确识别新的攻击模式。

异常检测方法是目前入侵检测系统的主要研究方向，其特点是通过对系统异常行为的检测，可以发现未知的攻击模式。该方法首先定义一组系统“正常”情况的数值，如网络数据流、CPU 利用率、内存利用率、文件校验和、系统日志等，然后将系统运行时的各种特征与所定义的“正常”情况特征比较，从而识别出网络内是否存在攻击及何种攻击。

入侵检测是一个涉及很多方面的多变量、动态、高复杂度的问题。近年来，许多学者从各种角度对入侵检测技术进行了研究，提出了很多新的思想和算法，其中基于数据挖掘的入侵检测算法已成为一个研究热点。数据挖掘也称数据库中知识发现(knowledge discovery in database，KDD)，通过对海量的安全审计数据进行智能化的处理，从中提取系统安全相关的行为特征，从而识别出入侵行为。此外还有基于模型推理的检测算法，基于专家系统的检测算法，基于神经网络的检测算法，基于免疫原理的检测算法等。

上述入侵检测算法都基于入侵检测模型。建立入侵检测模型过程包括：数据集选择；数据预处理；确定知识发现算法；数据挖掘；知识评价。数据集选择是研究和评价各种入侵检测算法的第一步。数据集的好坏直接决定各种入侵检测算法的评价结果。因此需要有一个公认的、性能优良的评测数据集，否则各种算法以及改进就没有比较的基础和平台。目前

收稿日期：2009-11-09；修订日期：2010-01-13。

基金项目：国家自然科学基金项目（60773102）。

作者简介：张新有（1971－），男，四川成都人，硕士，副教授，研究方向为对等网络、网络安全、网络体系结构； 曾华燊（1945－），男，四川成都人，博士，教授，博士生导师，研究方向为网络体系结构、网络计算、网络安全； 贾磊（1984－），男，河南开封人，硕士研究生，研究方向为网络安全、对等网络、计算机网络、P2P 技术。E-mail：xyzhang@swjtu.edu.cn

大家公认的入侵检测数据集是基于MIT LL采集[1]、由哥伦比亚大学IDS实验室整理形成的安全审计数据集KDD CUP99[2,3]。许多论文及研究成果都基于该数据集为基础，但对该数据集的分析和使用上存在差异[4-6]。本文对KDD CUP99入侵检测数据集的结构、攻击分布和特征选择等方面进行分析研究，并指出其不足之处以及改进方向。

1 入侵检测数据集及特点

数据集的分布和结构集中反映了受保护系统运行状态和行为，为入侵分析程序提供原始的安全审计数据。这些安全审计数据是入侵检测算法的处理对象，算法的部分任务就是从这些原始审计数据集中提取特定的入侵模式。

入侵检测算法的实现需要有足够的先验知识。为了验证算法的性能，将通用的数据集分为训练数据集和测试数据集。作为入侵检测算法进行学习和处理的对象，要求所选择的数据集应具备如下特点：①需要有足够的、全面而真实的反映系统安全状态的先验数据；②安全审计数据集在正常情况下是非常稳定的且全面的；③先验数据集中要能够反映各类攻击事件的特征分布；以便各种入侵检测算法学习到入侵模式；④攻击行为总是使安全审计数据的某些特征变量明显地偏离正常值。

1.1 MIT LL入侵检测数据集

1998年，DARPA入侵检测评估项目由MIT林肯实验室承担[7]；其主要目的是测量和评估入侵检测系统。项目的成果之一是建立了一个模拟军事网络中各种入侵的安全审计数据集。该数据集采用压缩二进制的tcpdump格式文件存储。其中训练数据集包括7周的网络流量，共500万条连接记录；测试训练集包含2周网络流量，共两百万条连接记录。该研究共模拟了5大类网络攻击：① Denial-Of-Service(DOS)：非法企图中断或干扰主机或网络的正常运行；② Remote to Local(R2L)：远程非授权用户非法获得本地主机的用户特权；③ User to Root (U2R)：本地非授权用户非法获取本地超级用户或管理员的特权；④ Surveillance or probe(Probe)：非法扫描主机或网络，寻找漏洞、搜索系统配置或网络拓扑；⑤ Date Compromise(data)：非法访问或修改本地或远程主机的数据。

除了网络流量外，MIT LL的入侵检测数据集还包含了模拟网络系统的其它信息，如各种被攻击主机系统(Solaris, SunOS, Linux)的审计数据、日志信息和相关的文件系统配置等；其中训练数据集中还包含相应的攻击类型标记。

考虑到数据集的公开性和定量的精确评测，MIT LL采用合成的背景流量生成方法[8]，该方案既能较好的解决数据流量中的隐私问题，又能保留真实的网络特性。同时LL也充分考了攻击流量的生成方法：对攻击行为分类，每一类攻击中选取了几种有代表性的攻击，采用自动化调度的方法形成攻击，部分攻击用脚本实现，一些比较复杂的攻击用手工实现。

1.2 KDD CUP99入侵检测数据集

MIT LL发布的入侵检测数据集，不仅是IDS综合测试系统的典范，也是目前学术界最有影响力和公信力的入侵检测数据集。然而在采用该数据集进行入侵检测分析时有如下缺点：

(1) 数据集庞大。不仅包含网络流量，还包含各种主机的审计数据和日志信息，其中主机审计数据、日志信息与系统运行环境及其配置有关，不利于各类入侵检测算法的公平性评比。

(2)tcpdump二进制格式数据处理不便。

(3)每条记录的特征信息表示复杂，不同协议有不同的数据格式，不利于特征选取(MIT LL采用tcpdump捕获网络数据，命令为 tcpdump -s 66000 -F option -w <datafile>)。

基于此，Stolfo教授领导的哥伦比亚大学IDS实验室等公布的安全审计数据集KDD CUP99是从1998年MIT LL的IDS数据集整理来的[2]，其中仅包含了网络流量数据，是目前大家公认的、实用的网络安全审计数据集，许多论文及研究成果都基于该数据集为研究基础。该数据集包括：

(1) 全部数据集：训练数据集 kddcup.data.gz, 18M；测试数据集 kddcup.testdata.unlabeled.gz, 11.2M；

(2) 10%数据集：训练数据集 kddcup.data_10_percent.gz(含攻击标记), 2.1M；测试数据集 kddcup.newtestdata.unlabeled_10_percent.gz(不含攻击标记), 1.4M；

(3) corrected.gz 为含有攻击标记的测试数据集，研究人员可用此数据集对自己算法的检测结果进行对照与分析。

迄今为止网络安全审计研究人员使用较多的是10%数据集。本文以10%数据集为分析对象。

2 KDD CUP99数据集攻击类型及分布

在KDD CUP99提供的10%数据集(包括训练数据集和测试数据集)中，包括了MIT LL入侵检测数据集中特征比较明显的4大类网络攻击类型[9-10]：

(1)Denial-Of-Service(DOS)；

(2)Surveillance or probe(Probe)；

(3)User to Root(U2R)；

(4)Remote to Local(R2L)。

在两种10%数据集中，上述4大类攻击各含攻击行为数量不同。训练数据集包含23种攻击行为；测试数据集包含38中攻击行为，具体分布情况如表1所示。表1中将正常数据包(normal)也算一种攻击类型。需要说明以下几点：

(1)部分攻击行为只包含在测试数据集中而没有包含在训练数据集中。这样设计是因为优秀的入侵检测算法通过对训练数据集的学习，可以识别一些没有遇到过的、新的攻击行为。

(2)训练数据集和测试数据集正常连接的数据记录(Normal)分布基本一致，但攻击数据分布不一致。有的差别较大，如表2所示。

(3)训练数据集中包含的U2R和R2L攻击很少(0.238%)，测试数据集二者所占的比例较多(5.278%)，综合比较现有的各种入侵检测算法分类结果，这两类攻击的检错率较高。实际上通过网络流量检测并区分U2R和R2L攻击十分困难，因为二者都只有少量的格式化网络连接信息。因此一些入侵检测算法中将二者作为一种攻击来考虑[3,9]。

(4)在对入侵检测算法进行评估测试时使用未标注的测试数据集；带有标注的数据集是用来验证算法的性能的。

3 特征及选取

KDD CUP99两种数据集都采用文本格式存储，并使用相同的记录格式（未标注的测试数据集没有最后一项：攻击类

表 1　KDD CUP99 10%数据集中的攻击行为及分布

攻击类型 (category)	10%训练数据集 Training Data		10%测试数据集 Test Data(labeled)	
	攻击行为(23)	数量	攻击行为(38)	数量
Normal(0)	normal	97278	normal	60593
Probe(1)	ip sweep	1247	ip sweep	306
	nmap	231	nmap	84
	portsweep	1040	portsweep	354
	satan	1589	satan	1633
			saint	736
			mscan	1053
DOS(2)	back	2203	back	1098
	land	21	land	9
	neptune	107201	neptune	58001
	pod	264	pod	87
	smurf	280790	smurf	164091
	teardrop	979	teardrop	12
			apache2	794
			mailbomb	5000
			udpstorm	2
			processtable	759
U2R(3)	perl	3	perl	2
	rootkit	10	rootkit	13
	loadmodule	9	loadmodule	2
	buf_overflow	30	buf_overflow	22
			httptunnel	158
			ps	16
			sqlttack	2
			xterm	13
R2L(4)	ftp-write	8	ftp-write	3
	guess_passwd	53	guess_passwd	4367
	multihop	7	multihop	18
	phf	4	phf	2
	imap	12	imap	1
	spy	2	spy	
	warezclient	1020	warezclient	
	warezmaster	20	warezmaster	1602
			named	17
			xsnoop	4
			xlock	9
			sendmail	17
			worm	2
			snmpgetattack	7741
			snmpguess	2406
总数	494021		311029	

型)；每行表示一个记录，每条记录包含从一条连接中提取的包括 41 个特征(未包括最后标注的攻击类型)。如下显示了 2 条连接记录(每条连接记录表示相同源主机/端口和相同目的主机/端口的一次完整会话过程)的数据格式，记录的每个特征用逗号分隔。

0, udp, private, SF, 105, 146, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 255, 254, 1.00, 0.01, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, normal.

0, tcp, private, REJ, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 212, 20, 0.00, 0.00, 1.00, 1.00, 0.09, 0.07, 0.00, 255, 20, 0.08, 0.07, 0.00, 0.00, 0.00, 0.00, 1.00, 1.00, neptune.

表 2　攻击数据分布

攻击类型	训练数据集		测试数据集	
	数量	百分比	数量	百分比
Normal(0)	97278	19.691066	60593	19.481463
Probe(1)	4107	0.831341	4166	1.339425
DOS(2)	391458	79.239142	229853	73.900826
U2R(3)	52	0.010526	228	0.073305
R2L(4)	1126	0.227926	16189	5.204981

3.1　特征的分类

上述每条连接记录特征可以分为如下 3 类：

3.1.1　网络连接的基本特征

每条网络连接的基本特征如表 3 所示。

表 3　网络连接的基本特征

序号	特征	描述	类型
1	duration	连接的长度（秒）	连续
2	protocol_type	协议类型，如 TCP，UDP	离散
3	service	目的站网络服务，如 HTTP, Telnet 等	离散
4	flag	连接的状态（正常或错误）	离散
5	src_bytes	源到目的站的数据字节数	连续
6	dst_bytes	目的站到源站的数据字节数	连续
7	land	如果连接从/到相同的主机/端口，则为 1；否则为 0	离散
8	wrong_fragment	错误段（fragment）的数量	连续
9	urgent	urgent 包的数量	连续

3.1.2　网络连接的基于内容的特征

不同于 DOS 攻击和扫描探测攻击在短时间内向同一目的主机发起大量连接，U2R 和 R2L 攻击在连接记录中没有明显的频繁连续攻击模式，这两种攻击行为嵌入在报文的数据部分，通常只涉及单个连接。

当前对于报文的非结构化的数据部分的有效挖掘算法仍是一个值得研究的问题。为查找数据部分的可疑行为，Stolfo 等为每条记录增加了一些特征，如登录失败的次数、超级用户登录的次数等。这些特征称“基于内容的特征”，如表 4 所示。

3.1.3　2 秒时间单元的流量特征

“相同主机”特征用于检查在过去 2 秒内和当前连接有相同目的主机的所有连接，并计算其相应协议的行为服务等统计数据。“相同服务”特征用于检查在过去 2 秒内和当前连接有相同服务的所有连接。“相同主机”特征和“相同服务”特征一起被称为连接记录的“基于时间的流量特征集”，如表 5 所示。

其中 23~31 特征用来说明某条连接的源主机统计数据(注意序号不是顺序排列的)，32~41 用来说明某条连接的目的主机的统计数据；二者特征基本含义相同。目的主机多了一个

表 4 连接的内容特征

序号	特征	描述	类型
10	hot	“hot”指示器的数量	连续
11	num_failed_logins	登录失败的次数	连续
12	logged_in	成功登录为 1，否则为 0	离散
13	num_compromised	满足被攻击条件的数量	连续
14	root_shell	获得超级用户 shell 为 1，否则为 0	离散
15	su_attempted	若企图执行“su root”为 1，否则为 0	离散
16	num_root	root 访问的数量	连续
17	num_file_creations	文件创建操作的数量	连续
18	num_shells	Shell 提示符的数量	连续
19	num_access_files	访问控制文件操作的数量	连续
20	num_outbound_cmds	ftp 会话中带外命令的数量	连续
21	is_hot_login	login 属于“hot”列表，为 1，否则为 0	离散
22	is _guest_login	login 用户为 guest，为 1，否则为 0	离散

表 5 连接的流量特征（2 秒时间单元）

序号	特征	描述	类型
基于某 TCP 连接的源主机特征			
23	count	2 秒内和当前连接有相同目的主机的连接数	连续
	以下 4 个特征指的是这些相同主机的连接		
25	serror_rate	SYN 错误的连接数%	连续
27	rerror_rate	REJ 错误的连接数%	连续
29	same_srv_rate	相同服务的连接数%	连续
30	diff_srv_rate	不同服务的连接数%	连续
24	srv_count	2 秒内和当前连接有相同服务的连接数	连续
	以下 3 个特征指的是这些相同服务的连接		
26	srv_serror_rate	SYN 错误的连接数%	连续
28	srv_rerror_rate	REJ 错误的连接数%	连续
31	srv_diff_host_rate	不同主机的连接数%	连续
基于某 TCP 连接的目的主机特征			
32	dst_host_count		连续
33	dst_host_srv_count		连续
34	dst_host_same_srv_rate		连续
35	dst_host_diff_srv_rate		连续
36	dst_host_same_src_port_rate	相同源端口的连接数%	连续
37	dst_host_srv_diff_host_rate		连续
38	dst_host_serror_rate		连续
39	dst_host_srv_serror_rate		连续
40	dst_host_rerror_rate		连续
41	dst_host_srv_rerror_rate		连续

dst_host_same_src_port_rate 特征，指的是目的主机上有相同源端口的连接数百分比。

3.2 数据集的预处理

各种入侵检测算法对数据集进行操作的第一个重要步骤就是预处理。数据预处理就是对所提供的数据集进行再加工，检查数据的完整性和一致性，对干扰数据进行过滤以及丢失的数据进行填补等。

特征选择也是预处理的一个重要环节。KDD CUP99 数据集中包含 3 类 41 种特征，有些特征对入侵检测算法没有影响，只会是算法运行效率降低，甚至可能导致算法的结果出现偏差。因此一般采用特征选择技术进行特征简化，挑选认为能够提高算法分类准确性和运行效率的特征集。许多研究结果表明，针对同一算法，由于选取的特征集不同，入侵检测算法的分类性能和效率有较大差异。

4 KDD CUP99 数据集存在的问题与改进

MIT LL 入侵检测数据集是目前公认的最优秀、较有影响力的网络安全审计数据集。KDD CUP99 对 MIT LL 数据集的网络流量进行了格式化处理，改进并规范了特征的表示，提供了便于适用于各种入侵检测算法进行测试和评估的数据集，自公布以来，出现了许多基于此数据集的研究成果和研究论文。

KDD CUP9 数据集仍有一些缺点，或者说该数据集待改进与完善的方面：

(1)特征选择缺陷。一些 Probe 攻击扫描主机或端口的时间间隔大于 2 秒(如每分钟扫描一次)的攻击不易识别。改进方法：连接记录可以按目的主机分类，如特征的构建使用 100 个有相同目的主机的连接作为检查单元而不用时间单元，亦即采用“基于主机的流量特性集”。

(2)数据集缺少主机数据(主机日志)信息，且由于网络数据报文段的非结构化信息中的特征不易提取，因此U2L和U2R类攻击识别效率比较低。

(3) MIT LL 的背景流量生成主要依据专家经验和统计模型，所采用的用户行为模型非常简单而且固定，模拟时缺乏灵活性，而灵活的配置是新一代数据集应具备的特点和功能。

(4)随着时间发展和新应用的出现，新的攻击以及变体不断涌现，原先数据集中的攻击样本显得陈旧。

(5)MIT LL数据集的网络应用环境是 10M 以太网，而根据目前网络的实际应用环境，100M/1000M 以太网环境才能符合当前应用需求。

(6)不能反映网络节点被入侵后的丢包行为，不适用于如 Ad Hoc 网络的入侵检测系统，需要改进，填加相应特征。

5 结束语

入侵检测系统已经成为确保网络安全的不可缺少的设施，而目前层出不穷的所谓的入侵检测“新技术”不仅不能明显提高网络安全等级，反而使人们对技术的改进和革新产生怀疑。入侵检测系统的评测成为人们关注的焦点。公认的优秀的入侵检测数据集是 IDS 系统以及安全审计算法性能和运行效率评测的基础。本文首先介绍了两种有影响的入侵检测数据集：MIT LL 入侵检测数据集和由此整理形成的 KDD CUP99 入侵检测数据集的特点以及构成，重点分析了 KDD CUP99 训练数据集和测试数据集的各自攻击类型以及详细分布，数据集中每条连接的特征分类及其各个特征的含义，并对数据集的使用进行了说明，最后对KDD CUP数据集存在的问题与改进措施提出了建议。

(下转第 4816 页)

Anonce和消息1中的Anonce进行比较,若不同就丢弃消息3,若相同再根据MIC值进行数据完整性校验,若成功就可把消息中在Key DATA字段中存放的RSN IE和初始关联时候AP发送的RSN IE进行比较,若不同,说明AP为假冒,断开关联,若相同就根据Key ID的前两个字节中放入的GTK长度获取Key DATA字段中的GTK值,随之就可加载PTK和GTK,并构造消息4,消息4在EAPOL-Key帧的key nonce字段中放入Snonce的值,置消息回应GTK接收有效位,在Key DATA字段中存放的RSN IE,生成MIC。

AP接受消息4后检查GTK是否装载有效,再根据Key RSC进行重放攻击检查,然后把Snonce和AP中已经接收的Snonce比较,相同的情况下进行MIC数据完整性校验,若成功再将消息4所带的RSN IE和AP RSN IE比较,不同就断开连接,相同就可加载PTK。

2.2 方案性能分析

改进后的方案采用EAPOL-Key来传输各种消息,没有对其帧结构进行任何调整和修改,保持了对原802.11i协议的兼容,同时改进方案由四次握手加组密钥握手完成PTK和GTK分发变成只需四次握手就同时实现PTK和GTK分发,缩短了密钥分发的过程,减少了分发环节,从而使得漫游环境下的密钥分发效率大大提高,减少了用户接入WLAN的延迟。具体性能体现在:

(1)方案在对原协议兼容的基础上保持了原有协议的安全性,因为PMK是基于802.x认证成功后获得的,是可信和安全的,PTK是由PRF函数根据随机数生成,也是可信的和安全的。

(2)方案减少了信息交换的次数,在保持消息1和消息2不改变的基础上,利用消息3和4捎带了GTK分发的基本信息,而GTK又可利用消息1、2生成的PTK进行加密传送。所以方案是可行的。

(3)在改进方案上利用了EAPOL-Key的保留字段Key ID来GTK的长度及其有效性标志,不涉及对EAPOL-Key的修改,保持了对802.11i协议的兼容性。

3 结束语

新无线网安全协议802.11i旨在从加密和认证、密钥管理3个方面加强无线局域网的安全性能,其中密钥管理是无线网络安全的很重要的一环[8]。但由于802.11i没有充分考虑现实可用性问题,使得密钥计算和分发效率低下。为了使移动环境下的STA能够高效迅速的获得密钥,本文在分析无线网络802.11i密钥分发机制的基础上,提出了新的替代方案,新方案通过利用了EAPOL-Key的保留字段Key ID传递GTK分发的基本信息,减少了AP和STA之间密钥分发握手的消息数量,大大降低了计算负荷和信道通信量,提高了无线网络漫游环境下的密钥分发效率。

参考文献:

[1] Arbaugh W A,Shanker N,Wan Y C J,et al,Your 802.11 wireless network has no clothes[J].IEEE Wireless Communication,2005(12):44-51.

[2] He C,Mitchell J C.Analysis of the 802.11i 4-way handshake[C].New York:Association for Computing Machinery,2004:43-50.

[3] 文远保,刘涛.RSN密钥分发机制的研究及实现[J].华中科技大学学报(自然科学版),2004(1):41-43.

[4] 韩平,朱艳琴,罗召喜.基于逻辑密钥树的IEEE802.11i GTK更新方案[J].计算机工程与应用,2009,45(10):116-118.

[5] 曹秀英,耿嘉,沈平.无线局域网安全系统[M].北京:电子工业出版社,2004.

[6] Jon Edney,William A Arbaugh.无线局域网安全实务——WPA与802.11i[M].北京:人民邮电出版社,2006.

[7] Hayriye Altunbasak,Henry Owen.Alternative pair-wise key exchange protocols for robust security networks (802.11i) in wireless LANs[C].Proceedings of IEEE Southeast Con,2004:3-9.

[8] 曹利,杨凌凤,顾翔,等.基于802.11i的EAP-TLS认证机制的安全分析[J].计算机工程与设计,2010,31(4):756-759.

(上接第4812页)

参考文献:

[1] Lippmann R P,Fried D J,Graf I,et al.Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation[C].Los Alamitos,CA:Proc of the 2000 DARPA Information Survivability Conference and Exposition (DISCEX),2000:12-26.

[2] Hettich S, Bay S D. KDD cup 1999 data [EB/OL]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html,1999.

[3] Ke W,Salvatore J S.Anomalous payload based network intrusion detection[C].Proc of the 7th International Symposium on Recent Advanced in Intrusion Detection(RAID),2004:201-222.

[4] 刘自伟,蔡勇,陈波.KDD在入侵检测中的应用[J].微型机与应用,2003,12:55-58.

[5] 祖宝明,詹永照,卿林.一种针对MANET入侵检测Agent分布的分簇方法[J].微计算机信息,2007,5(3):41-43.

[6] 刘密霞,张秋余.入侵检测报警相关性及评测数据集研究[J].计算机应用研究,2008,25(10):13-11.

[7] McHugh J.Testing intrusion detection systems:A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory[J].ACM Transactions on Information and System Security,2000,3(4):262-294.

[8] MIT Lincoln Lab. 2000 DARPA intrusion detection scenario-specific datasets [EB/OL]. http://www.ll.mit.edu/IST/ideval/data/2000/2000_data_index.html,2000.

[9] Stolfo S J,Wenke Lee,Chan P K,et al.Dataming-based intrusion detectors: An overview of the columbia IDS project [J]. ACM SIGMOD Record,2001,30(4):5-14.

[10] 史美林,钱俊,许超.入侵检测系统数据集评测研究[J].计算机科学,2006,33(8):1-8.