基于误用和异常技术相结合的入侵检测系统的设计与研究

田俊峰 张 喆 赵卫东
(河北大学数学与计算机学院 保定 071002)

摘 要 目前，入侵检测系统(IDS) 的漏报率和误报率高一直是困扰 IDS 用户的主要问题，而入侵检测系统主要有误用型和异常型两种检测技术，根据这两种检测技术各自的优点，以及它们的互补性，将两种检测技术结合起来的方案越来越多地应用于 IDS 中。该文提出了基于统计的异常检测技术和基于模式匹配的误用检测技术相结合的 IDS 模型，减少了单纯使用某种入侵检测技术时的漏报率和误报率，从而提高系统的安全性。

关键词 入侵检测系统，异常检测，误用检测，模式匹配，统计分析
中图分类号：TP393.08 **文献标识码**：A **文章编号**: 1009-5896(2006)11-2162-05

The Design and Research of Intrusion Detection System Based on Misuse and Anomaly

Tian Jun-feng Zhang Zhe Zhao Wei-dong
(*College of Computer and Mathematics*, *Hebei University*, *Baoding* 071002, *China*)

Abstract Currently, the false positive and the false negative of Intrusion Detection System are very high. It was always the main problem that bothered the user of IDS. But there are tow main technologies applied in IDS. To this problem, because both the technologies have its own advantages and they can supply for each other. So IDS combined with the tow technologies was used more and more widely. This paper presented a model of IDS based on combination of misuse detection and anomaly detection. In this model, misuse detection is based on pattern matching and Anomaly Detection is based on statistical analysis. It combined the tow technologies to reduce the false positive rate and the false negative rate in only one detection technology, and then to improve security of IDS.

Key words Intrusion Detection System (IDS), Anomaly detection, Misuse detection, Pattern matching, Statistical analysis

1 引言

入侵检测系统(Intrusion Detection System，IDS)通过从计算机网络或计算机系统中的若干关键点收集信息并对其进行分析，以发现网络或系统中是否有违反安全策略的行为和遭到袭击的迹象[1]。入侵检测系统主要通过以下几种活动来完成任务：监视、分析用户及系统活动；对系统配置和弱点进行审计；识别与已知的攻击模式匹配的活动；对异常活动模式进行统计分析；评估重要系统和数据文件的完整性；对操作系统进行审计跟踪管理，并识别用户违反安全策略的行为。

目前入侵检测系统所采用的技术主要是异常检测与误用检测(又称特征检测)两种[2]。但它们都不同程度地存在误报和漏报的情况，且特点各异，优势互补。

最近几年间，异常检测型IDS又出现了许多新的方法，如Fumio Mizoguchi提出的基于机器学习的可视化异常型IDS[3]，基于网络状态的IDS模型[4]。但是异常检测型的IDS依赖于异常模型的建立，由于入侵性活动并不总是与异常活动相符合，不同的异常模型，可能检测出不同的入侵行为，从而产生一定的误报和漏报信息。

误用检测型IDS的研究也在不断地发展，Garvey和Lunt[5]首先提出的基于模型的误用检测方法，目前仍在广泛使用。误用检测型IDS比较常用的有基于特征匹配的IDS[6]；基于状态转换分析的IDS，如STAT和USTAT[7, 8]。但是误用型IDS特别是基于模式的IDS只能检测出那些已知的攻击，因此，需要不断地更新它们使用的攻击特征库来检测新的攻击，对攻击特征库以外的攻击却无能为力，因此会产生很高的漏报信息。

由于异常检测和误用检测这两种方法各有所长，被其中一种方法忽视掉的入侵，很可能会被另外一种所识别，目前这两种技术结合使用的IDS通常是将异常检测探测器和误用检测探测器并联使用，然后通过数据融合技术来产生警报，如ISA-IDS[9]，但它的误报现象还很严重。基于上述原因，本文提出并设计了一种基于模式匹配和统计分析入侵检测系统模型，称之为MAIDS(Misuse and Anomaly-based IDS)。

2 基于模式匹配和统计分析入侵检测系统(MAIDS)的设计

模式匹配方法的 IDS 将所有入侵行为和手段及其变种表达为一种模式或特征，检测主要判别网络中搜集到的数据特征是否在入侵模式库中出现，若出现，则视为入侵行为，否

2005-03-08 收到, 2005-09-26 改回
河北省自然科学基金 (F2004000133)资助项目

则视为正常。但是某个入侵行为的数据特征并未出现在入侵模式库中，则产生漏报现象。统计分析方法通过流量统计分析建立系统正常行为的轨迹，那么理论上就可以把所有与正常轨迹不同的系统状态视为异常活动，从而只要检测到所有的异常活动，则可检测出所有的入侵性活动。但是这种活动存在 4 种可能性：(1)入侵性而非异常；(2)非入侵性且异常；(3)非入侵性且非异常；(4)入侵且异常。显然，异常入侵检测技术的局限性在于并非所有的异常活动都是入侵性活动，即第(2)种情况时，将产生严重的误报。

基于统计分析的异常检测技术可以使系统检测新的和未知的攻击或其它情况。模式匹配检测技术可以防止那些耐心的攻击者逐步改变行为模式，使得异常检测器将攻击行为误认为是合法行为，从而保证异常检测的完整性。由于两种检测技术各有优缺点，而且可以互补，所以将这两种方法结合起来，可能会获得更好的性能，有助于提供一个功能较完整的入侵监测系统。本文提出的 MADIS 就是基于误用检测和异常检测技术相结合的入侵检测系统，它主要采用模式匹配技术和统计分析技术来实现对入侵行为的检测。

2.1　MAIDS 的系统结构

MAIDS 的结构如图 1 所示，主要有以下几个部分组成：探测器、中心控制服务器、异常检测服务器、特征检测服务器。

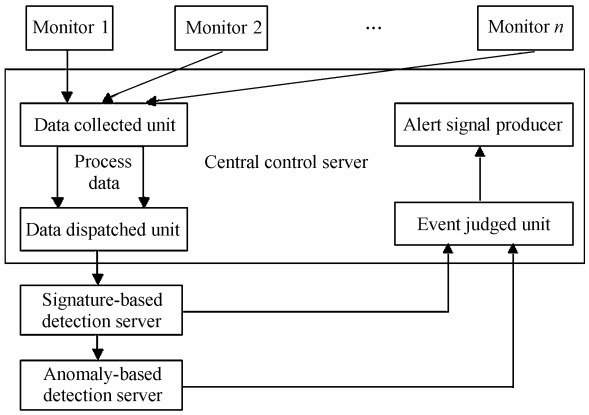

图 1　MAIDS 的系统结构

Fig.1 The architecture of MAIDS

(1)探测器　位于一台计算机系统中，主要担任收集审计数据的作用。它负责按照一定的要求或者规则收集用户、服务、系统或者网络数据流信息，把它们组织成适当格式的审计数据并送交中心控制服务器。

(2)中心控制服务器　是整个系统的控制台，由它可以配置系统，控制系统行为，接收由探测器传来的审计数据，经过一定的处理后将审计数据送交特征检测服务器，并且接收由异常检测服务器和特征检测服务器检测的结果，做出对审计数据的判断，并对入侵行为产生报警。

(3)异常检测服务器　由轮廓引擎和异常检测器组成。轮廓引擎提供了非异常情况下，即正常情况下的原始轮廓，由异常检测器通过对审计数据进行统计分析，判断该行为是否位于这个轮廓中，如果偏离轮廓，则视为异常行为，否则视为正常行为。

(4)特征检测服务器　由模式匹配器、入侵模式库和合法模式库组成。模式匹配器是将审计数据与两个模式库中的模板进行匹配运算的部件，它通过模式匹配算法判断审计数据符合哪一个模式库；入侵模式库中存储的是已知的入侵行为模板，凡是符合这个模式库的行为都视为入侵行为；合法模式库中存储的是已知的合法行为模板，其中主要是合法但产生异常的行为，凡是符合这个模式库的行为都视为合法行为。

2.2　MAIDS 的工作环境及其并行性检测机制

(1) MAIDS 的工作环境　MAIDS 是基于网络的入侵检测系统，探测器分布于局域网中的各个主机上，探测器将收集到的审计数据包传送给中心控制服务器。中心控制服务器位于局域网中的一台主机上，作为整个系统的控制中心。中心控制服务器将这些数据通过处理，由数据调度管理器发送审计数据到特征检测服务器。特征检测服务器由位于局域网中两台主机组成，一台用于审计数据与入侵模式库的特征检测，另一台用于审计数据与合法模式库的特征检测，这样可以实现对审计数据进行模式匹配的并行处理，加快系统速度。特征检测服务器对数据进行特征检测，符合入侵模式库的审计数据做出入侵标记，符合合法模式库的审计数据做出合法标记，发送到事件判断单元；不符合任何一个模式库的数据发送到异常检测服务器。异常服务检测器也位于局域网中的一台主机上，对审计数据作异常检测。异常检测服务器将收到的数据进行异常检测，超出轮廓的审计数据视为异常行为，否则作为正常行为，分别做出标记，发送到事件判断单元。事件判断单元对收到的数据进行判断，带有入侵标记的数据发送到报警产生器。报警产生器，对收到的数据发出报警信号。

(2) MAIDS 中模式匹配的并行检测机制　当审计数据包的数量很少时，无论先与哪个模式库进行匹配，系统速度不会受到影响；但是当审计数据包的数量很大时，如果串联与两个模式库进行匹配，将大大降低系统检测速度，所以 MAIDS 采用与两个模式库进行匹配并行处理的方法。这种方法是：首先将审计数据包集分成两组，一组先与入侵模式库进行模式匹配，同时另一组先与合法模式库进行模式匹配，当匹配完成后，再将两组互换，将未匹配成功的审计数据与另外的模式库进行模式匹配，这样始终保持两个模式匹配器同时工作，从而加快了系统速度。

2.3　MAIDS 的检测流程

在漏报方面，统计分析方法只要建立当系统运行时比较合适的阈值，则对检测未知的入侵行为有很好的效果，从而对模式匹配方法未能检测到的入侵予以很好的弥补，也就减少了漏报率；而有些非异常的攻击，很可能在模式匹配中已经被检测出来，而这些攻击正是异常检测系统无法检测的，

从而也减少了漏报率。

在误报方面，MAIDS 解决这种误报的方法是增加一个合法模式库，它含有一些已知的异常且非入侵性活动的模式，将收集到的数据特征与之匹配，匹配成功则视为合法，不成功再对其采用异常检测。这样一些非入侵性且异常活动可以避免报警，从而减少了误报率。

MAIDS 的审计数据检测流程如图 2 所示。图 2 中的审计数据源是从网卡接收的数据包，首先采用特征检测技术，模式匹配器根据入侵模式库中的入侵模板和合法模式库中的合法模板过滤审计数据源中的数据包，如有与入侵模板相匹配的包或包集合，则视为入侵行为，产生报警；如有与合法模板相匹配的包或包集合，则视为合法行为；两个模式匹配器并发执行，如果审计数据第 1 次匹配成功，直接确定其为入侵行为或是合法行为，而不再进行第 2 次匹配及以下的检测；否则对审计数据进行第 2 次匹配，第 2 次匹配成功，也直接确定其为入侵行为或是合法行为，也不再进行异常检测。如果都未匹配成功，则向下采用异常检测技术，异常检测器统计审计数据包中的各项属性值，并分析产生其异常值，若数据包的异常值过高，超出系统预先设定阈值，则视为攻击行为，产生报警。

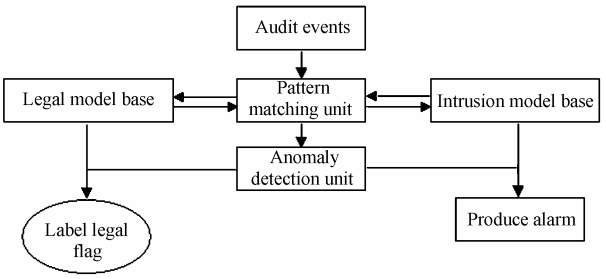

图 2 MAIDS 的数据流程图

Fig.2 The data flow of MAIDS

2.4 MAIDS 的检测技术

MAIDS 主要应用了两个入侵检测技术，一个是误用检测系统(本文称之为 Matching)中的基于模式匹配检测技术；另一个异常检测系统(本文称之为 Statistician)中的基于统计分析检测技术。

Matching 是通过建立证据模型，并且根据证据模型与审计数据的匹配程度，做出判断结论。Matching 中的证据模型分为两部分，一部分是攻击剧本的模型数据库，其中每个剧本表示一个入侵行为，而其中一些入侵行为是不会产生异常的；另一部分是合法剧本的模型数据库，其中记录的都是合法行为，而其中一些合法行为会产生异常。Matching 是将当前的活动行为与存储在模型数据库中的特征模型进行比较和匹配，再根据匹配的结果来决定当前的活动行为是入侵行为、合法行为或是未知行为。入侵行为将产生报警，未知行为将再对其进行异常检测，合法行为则不产生报警也不再进行异常检测。

Statistician是基于平均值和方差模型的异常入侵检测技术，根据异常检测器观察主体的活动，然后刻画出这些主体的行为轮廓，每一个轮廓记录都表示主体的当前行为，然后将当前的行为轮廓与已储存的正常轮廓进行比较，来判断其异常行为。Statistician中设f为正常轮廓的阈值，f是人为设定的一个界限值，它是对大多数合法行为进行统计分析而得来的，当审计数据的统计值大于f时，被视为异常行为。$G_1, G_2, G_3, \cdots, G_n$($n$为特征变量的个数)为轮廓的特征变量，这些特征变量分别是CPU的使用、I/O的使用、使用地点、使用时间、邮件使用，编辑器使用、访问或改变的目录及文件、文件访问数量、网络会话时间等。用$S_1, S_2, S_3, \cdots, S_n$分别描述特征变量$G_1, G_2, G_3, \cdots, G_n$的异常程度值，$S_i$($i$=1,2,$\cdots$,$n$)越大，则表示$M_i$异常程度越大。用这些异常程度值的加权平方和来表示轮廓异常值G：

$$G = a_1 S_1^2 + a_2 S_2^2 + \cdots + a_n S_n^2 \quad a_i > 0, \quad i = 1,2,\cdots,n \tag{1}$$

式(1)中a_i表示每一特征的权值，这是由于$M_1, M_2, M_3, \cdots, M_n$之间不是相互独立的。再选用标准方差作为判别准则，则标准方差：

$$d = \sqrt{\frac{G}{n} - \left(\frac{S_1 + S_2 + \cdots + S_n}{n}\right)^2} \tag{2}$$

如果 d 的值超出了正常轮廓的阈值 f，即当 $d>f$ 时，就认为出现异常，并且产生报警。

3 MADIS 与其它类型 IDS 的比较分析

下面以一组审计数据的集合为例，对各种方法的漏报率和误报率做一下分析比较。为分析方便，以下误用型IDS均采用模式匹配技术，异常型IDS均采用统计分析技术。审计数据集合的关系如图 3 所示。作如下定义：设这组审计数据所有数据包为全集U，其数据包的个数记作u(u为自然数)。其中有真正的合法行为集合J，其数据包的个数记作j(0<j<u)；真正的入侵行为的集合I，其数据包的个数记作i(0<i<u)。所以有如下关系：$U=J\cup I$；$u=i+j$。可以被模式匹配技术检测到的入侵行为的结合M，其数据包个数记作m(0<m<u)；可以被模式匹配技术检测到的合法行为的集合N，其数据包的个数记作n(0<n<u)；被统计分析检测到并被视为入侵行为的集合A，其数据包的个数记作a(0<a<u)，余下的集合B，其数据包的个数记作b(b=u – a)。以下所讨论的情况，假设误用型IDS均采用模式匹配技术，异常型IDS均采用统计分析技术，数据包的集合均在全集U的范围内，漏报率和误报率分别用Pu_k和Pw_k表示，其中k=1,2,3,4，表示第k种方法。下面就几种常规的情况进行讨论：

(1) 仅用误用检测技术的 IDS，如图 4 所示，这里称之为 MIDS(Misuse Intrusion Detection System)，它采用模式匹配方法。假设这种方法能检测出的入侵行为的集合为 M，因为模式数据库中所存储的模式剧本都是已知的入侵行为，所以 M 是 I 的一个子集。

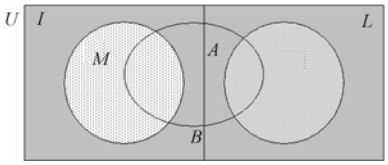

图 3　审计数据中各个集合的关系

Fig.3 The relation of each set in audit events

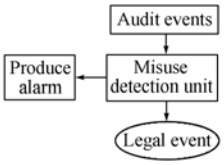

图 4　MIDS 的检测模型

Fig.4 The detection model of MIDS

令 $M'=I-M$，表示未检测到的入侵行为的集合，其个数记作 m'，$(m'=i-m)$。一般来说，$M\neq I$，也就是说 M'不为空，即存在漏报现象，当 M'在 U 中占很大比例时，就有很高的漏报率。

所以有漏报率：

$$Pu_1 = m'/u \tag{3}$$

MIDS 对未知的入侵行为会有很高的漏报现象，所以它并不能很好满足系统的要求。

(2) 仅用异常检测技术的 IDS，这里称之为 AIDS (Anomaly Intrusion Detection System)，如图 5 所示，它采用统计分析方法。设这种方法能检测出的入侵行为的集合为 A；余下的行为的集合记作 B，其个数记作 b。

令$A_i=A\cap I$，表示被AIDS检测到的真正的入侵行为集合(入侵性且异常)，其个数记作a_i。

令$A_j=A\cap L$，表示被AIDS认为是入侵的合法行为集合(非入侵且异常)，其个数记作a_j。

令$B_i=B\cap I$，表示被AIDS未检测到的真正入侵行为集合(入侵且非异常)，其个数记作b_i；

令$B_j=B\cap L$，表示被AIDS认为是合法的行为集合(非入侵且异常)，其个数记作b_j。

所以B_i为漏报的入侵行为集合，A_j为误报的入侵行为集合。

所以有漏报率：

$$\mathrm{Pu}_2 = b_i/u \tag{4}$$

误报率：

$$\mathrm{Pw}_2 = a_j/u \tag{5}$$

AIDS 虽然能够检测出部分未知的入侵行为，减少了一定的漏报率，但是对一些合法的假入侵也认为是入侵行为，从而增加了误报率，同时对那些非异常的入侵行为也会有漏报现象，所以它也不能很好的满足系统的要求。

(3) 两种检测技术并列使用的IDS，如图 6 所示。这种方法是同时采用两种检测方法对审计数据进行检测，记为PIDS。PIDS无论是那种方法检测到“入侵”，都被视为入侵行为。设PIDS检测到的入侵行为的集合记作I_b，其个数为i_b；余下的集合记作J_b，其个数为j_b。假设这里的误用检测和异常检测与前两个情况是完全一样的，那么$I_b=M\cup A$。

令$I_{\mathrm{sa}}=M'\cap A_i$，表示被MIDS漏掉，但被AIDS检测到的真正入侵行为的集合，其个数为i_{sa}。

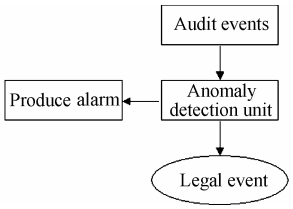

图 5　AIDS 的检测模型

Fig.5 The detection model of AIDS

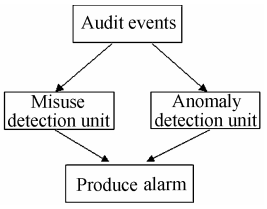

图 6　PIDS 的检测模型

Fig.6 The detection model of PIDS

令$I_{\mathrm{as}}=M\cap B_i$，表示被AIDS漏掉，但被MIDS检测到的真正入侵行为的集合，其个数为i_{as}。

由以上定义可知：

$$i_b = m' - i_{\mathrm{sa}} = b_i - i_{\mathrm{as}} \tag{6}$$

所以有漏报率：

$$\mathrm{Pu}_3 = i_b/u = \mathrm{Pu}_1 - i_{\mathrm{sa}}/u = \mathrm{Pu}_2 - i_{\mathrm{as}}/u \tag{7}$$

误报率：

$$\mathrm{Pw}_1 = \mathrm{Pw}_2 \tag{8}$$

显然，由式(7)可知 PIDS 的漏报率比 MIDS 和 AIDS 减少了，但误报率的问题没有解决，当误报要求较高时，便不能满足系统要求。而且，每个审计数据都要进行两种方法的检测，系统开销很大。

(4) MAIDS 的性能分析，检测流程如图 7 所示。图 7 中有一个合法模式库，其中存放的是被认为是合法行为的模式剧本，而且主要是一些产生异常的合法行为的模式剧本，它们都是已知的合法行为，这个集合为 N。

假设 MAIDS 中的误用检测系统和异常检测系统与 PIDS 完全相同。首先将审计数据用模式匹配的检测方法进行检测，发现与入侵模式相匹配的，直接被认为是入侵行为，并产生报警；发现与合法模式相匹配的，直接被认为是合法行为；然后，匹配不成功的审计数据继续向下进行异常入侵检测，超出异常值，即偏离合法轮廓的审计数据，被认为是入侵行为，则产生报警，否则，被视为合法行为。

MAIDS 在漏报率方面，与 PIDS 相同。所以有漏报率：

$$\mathrm{Pu}_4 = \mathrm{Pu}_3 \tag{9}$$

在误报率方面，令$N'=N\cap A_j$，表示与合法模式库中的合法模板相匹配，但被AIDS误认为是入侵的合法行为集合，其个数为n'。所以有误报率：

$$\mathrm{Pw}_4 = \left(a_j - n'\right)/u = \mathrm{Pw}_2 - n'/u \tag{10}$$

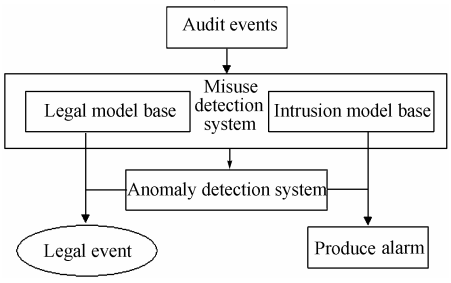

图 7　MAIDS 的检测模型

Fig.7 The detection model of MAIDS

综上所述，有如下关系：

漏报率：

$$Pu_4 = Pu_3 < Pu_2 \tag{11}$$

$$Pu_4 = Pu_3 < Pu_1 \tag{12}$$

误报率：

$$Pu_4 < Pw_3 = Pw_2 \tag{13}$$

显然，由式(11)，式(12)，式(13) 可知 MAIDS 与前 3 种方法相比，漏报率和误报率都有所减少，提高了安全性。而且，MAIDS 相比 PIDS，不用将每个审计数据都进行异常检测，也节省了系统开销，对于已知的行为，由于不必再进行下一步的检测，也节省了一定的时间。

4 结束语

入侵检测技术是目前网络安全领域研究的热点，现有的 IDS 都或多或少地存在着一定的误报与漏报现象，而单纯靠误用检测型或异常检测型技术，很难减少 IDS 误报与漏报的情况，误用型和异常型两种检测技术相结合的方法，弥补了它们各自的缺点，使 IDS 达到减少误报与漏报的目的。文中应用模式匹配的特征检测和统计分析的异常检测相结合的技术，构造了 MAIDS 的模型，通过 Matching 和 Statistician 两个系统有机的合作，在一定程度上减少了漏报率和误报率，实现了对系统的安全管理。但是 MAIDS 在如何提高入侵检测系统的检测速度，以适应网络通信的要求和如何提高入侵检测系统的互动性能，从而提高整个系统的安全性能等问题上还将继续研究。

参 考 文 献

[1] 赵小林, 彭祖林, 王亚彬. 网络安全技术教程. 北京: 国防工业出版社, 2002-1, 245−245.

[2] 蒋建春, 马恒太, 任党恩, 等. 网络安全入侵检测: 研究综述. 软件学报, 2000, 11(11): 1460−1466.

[3] Fumio Mizoguchi. Anomaly Detection Using Visualization and Machine Learning. IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. Gaithersburg, Maryland: March 14-16, 2000: 165−170.

[4] Shan Zheng, Chen Peng, Xu Ke, *et al*.. A Network State Based Intrusion Detection Model. 2001 International Conference on Computer Networks and Mobile Computing. Beijing, CHINA: October 16 -19, 2001: 481−486.

[5] 蒋建春, 冯登国. 网络入侵检测原理与技术. 北京: 国防工业出版社, 2001-7, 39−39.

[6] 李小秋, 孙学涛, 谢余强, 等. 入侵检测系统中的快速多模式匹配算法. 计算机应用与软件, 2004-02, 21(2): 84−86.

[7] Koral Ilgun, Richard A.Kemmerer, Phillip A.Porras. State transition analysis: A rule-based intrusion detection approach. *IEEE Trans. on Software Engineering*, 1995-3, 21(3): 181−199.

[8] Nittida Nuansri, Samar Singh, Tharam S.Dillon. A Process State-Transition Analysis and its Application to Intrusion Detection. 15th Annual Computer Security Applications Conference. Phoenix, Arizona：December 06-10, 1999: 378−387.

[9] Nong Ye, Syed Masum Emran, Xiangyang Li, *et al*.. Statistical Process Control for Computer Intrusion Detection. DARPA Information Survivability Conference & Exposition Anaheim, California: June 12-14, 2001, 1(1): 3−14.

田俊峰： 男，1965 年生，教授，研究领域为信息安全，网络技术.

张 喆： 男，1978 年生，硕士生，研究方向为信息安全.

赵卫东： 男，1979 年生，硕士生，研究方向为信息安全.