维普资讯 http://www.cqvip.com

计算机科学 2004Vol. 31№. 2

入侵检测系统评估技术述评*)

董晓梅 肖 珂 于 戈
(东北大学信息科学与工程学院 沈阳 110004)

摘 要 随着对入侵检测技术的深入研究和入侵检测产品的广泛应用,对入侵检测系统的评估成为一个重要的研究领域。本文说明了对入侵检测系统进行评估时的主要评价指标,包括入侵检测系统的检测率和误报、检测范围、检测延迟与负荷能力等方面。介绍了入侵检测评估的相关工作。分析了在1999年DARPA的离线评估中,所使用的测试网络环境和评估模型,并对其进行了评价。讨论了在对检测系统进行评估中的网络流量和主机使用模拟、攻击模拟以及评估报告的生成等关键技术。

关键词 网络安全,入侵检测,评估,仿真,攻击

Survey on Evaluation of Intrusion Detection Systems

DONG Xiao-Mei XIAO Ke YU Ge
(School of Information Science and Engineering, Northeastern University, Shenyang 110004)

Abstract With the deeply research on intrusion detection techniques and the widely use of intrusion detection products, the evaluation of intrusion detection systems becomes an important research field. The primary aspects of intrusion detection systems in an evaluation are expressed, including detection rate and false positive, detection range, detection delay, loading ability and so on. Relative works of the evaluation of intrusion detection systems is introduced. The test network environment and evaluation model in 1999 DARPA off-line evaluation are analyzed and criticized. The critical techniques when evaluating intrusion detection systems, such as emulation of network traffic and host usage, emulation of attacks and evaluation report generation, are discussed.

Keywords Network security, Intrusion detection, Evaluation, Emulation, Attack

随着计算机技术的飞速发展,信息网络已经成为社会发展的重要保证。信息网络涉及到国家的政府、军事、文教等诸多领域,政府宏观调控决策、商业经济信息、银行资金转帐、股票证券、能源资源数据、科研数据等许多重要信息在网络中进行存贮、传输和处理。其中有很多是敏感信息,甚至是国家机密。因此网络安全对于网络的应用至关重要。

防范网络攻击,保证系统安全的传统方法是使用防火墙。防火墙为内部网络提供安全的边界,实现了访问控制,在保护系统安全中起到了一定的作用。但是,仅仅使用防火墙来保障网络安全是远远不够的。为了弥补防火墙的不足,入侵检测系统(Intrusion Detection System, IDS)应运而生[1]。入侵检测系统被认为是防火墙之后的第二道闸门,在不影响网络性能的情况下能对网络进行监测,从而提供对内部攻击和误操作的实时保护[2~4]。IDS提高了系统管理员的管理能力,有助于提高信息安全基础结构的完整性。

目前有越来越多的机构开始进行入侵检测技术的深入研究,每年都会有新的入侵检测系统问世。投资方迫切需要知道他们投资巨大的研发系统的实际功效及运行情况。同时,随着入侵检测技术的不断发展,很多计算机用户已经开始把IDS作为他们计算机安全机制不可缺少的一部分,用户在选择使用IDS时应该知道IDS的各项性能指标,以便选择更加适合自身系统的IDS产品,确定可以在多大程度上依靠IDS。对于IDS的研制和开发人员来说,对IDS的经常性评估可以了解技术发展的现状和存在的不足,这样可以及时将研究重点放在关键的技术问题上,有利于IDS的开发,使IDS的开发取向适应Internet的发展。因此,评估IDS的技术也就变得与入侵检测技术同等重要了。

但是,评估IDS是一项很困难的工作。首先,在任意一台主机上获取囊括所有可能出现攻击的数据集是不现实的,因此评估过程中使用的数据集是不完整的;不可能得到所有曾经被其他IDS检测到的攻击信息;不断会有新的系统漏洞被发现,攻击者会利用这些漏洞发起新的攻击,这使得评估的结果难以持久。在评估IDS过程中的另一困难是计算机网络环境的不同条件会影响IDS的检测结果。这就是说一种在正常情况下IDS能够检测得到的攻击,出现在系统繁忙时很可能会成功躲避IDS的检测[5]。所有这些都使评估IDS变得十分复杂。

1 评估的内容

对一个入侵检测系统进行评估,主要是评价入侵检测系统的检测率和误报、检测范围、检测延迟与负荷能力等方面[6]。

(1)检测率和误报。这是IDS最重要的指标。令I表示入侵行为,$\neg$I表示正常行为,A表示IDS发出了报警,$\neg$A表示IDS未发出报警。检测率(Detection Rate)可定义为P(A|I),即发生入侵检测时,IDS发出报警的概率。误报率(False Positive)定义为P(A|$\neg$I),即没有发生入侵行为时,IDS发出报警的概率。为了方便起见,一般用在一定时间内IDS发

*)基金项目:国家863计划CIMS主题(2003AA414210);国家自然科学基金(60173051);教育部优秀青年教师科研教学奖励计划。董晓梅 讲师,博士研究生;于 戈 教授,博士生导师。

维普资讯 http://www.cqvip.com

出误报的次数来衡量 IDS 的误报率。

(2)检测范围。通常情况下,一个 IDS 能检测到的攻击是有一定范围的。检测范围的考察,就是在一定的攻击分类标准下,考察 IDS 对不同类型攻击的检测能力。

(3)检测延迟。检测延迟指的是在攻击发生至被 IDS 成功检测之间的延迟时间。

(4)负荷能力。IDS 有其设计的负荷能力,在超出负荷能力的情况下,性能会出现不同程度的下降。考察 IDS 的负荷能力,就是观察不同大小的网络流量,不同强度的 CPU、内存等系统资源的使用,对 IDS 的关键指标如检测率、误报的影响。

(5)其它指标。包括每个警报的附加信息量,IDS 安装、配置、使用的方便程度,IDS 对 CPU、内存等系统资源的占用情况以及 IDS 本身的鲁棒性。

这些指标对于 IDS 的开发者和使用者都很重要。开发者应该测试 IDS 的各项性能指标,尽可能保证 IDS 的检测率高、检测范围宽而误报率低,同时其他指标不受影响;用户在选择 IDS 产品时也可以根据不同 IDS 的性能指标来衡量 IDS 的优劣,选择更合适的 IDS 产品。

2 相关工作

在评估 IDS 的领域中,相关的工作不多。1993 年加州大学的 Puketza 等人第一次对 IDS 进行了评估,并详细记载了评估的过程。1996 年 Puketza 的论文描述了他们评估 IDS 的方法和使用的软件平台[5,7]。他通过使用脚本模拟网络正常流量并生成攻击数据包的方法来提供多重的动态数据流,这种方法在一定程度上提供了能够实时检测 IDS 的可重复环境。但评估过程中仅有唯一的 IDS NSM(Network Security Monitor)参评。评估考虑到不同的运行环境会影响 IDS 的检测结果,因此,评估是在单独运行 NSM 和有负荷的前提下运行 NSM 两种情况下进行的,在这里,负荷仅以主机并行任务的数量为标准,没有考虑其他的负荷情况。这次评估只是简单地记录了 NSM 可以检测到几种简单入侵的能力,不能算作是正式的评估;1998 年 IBM 的一个小组在苏黎世进行了另一次评估 IDS 的试验,在论文中着重论述了如何模拟理想的网络流量等一系列相关问题;2000 年在苏黎世还有人在进行另一种评估方法的研究,该方法通过分析设计原理而不是基于 IDS 实际运行的情况进行评估,但这种研究还只是处于初级阶段[8];迄今为止最权威的 IDS 评估是 DARPA 在 1998 和 1999 年进行的两次 IDS 离线评估[8~12]。

2.1 DARPA 的 IDS 离线评估

Lippmann 等人精心设计的测试网络环境可以模拟正常的网络流量和大量的攻击实例。图 1 为 1999 年 DARPA 测评 IDS 的网络环境。

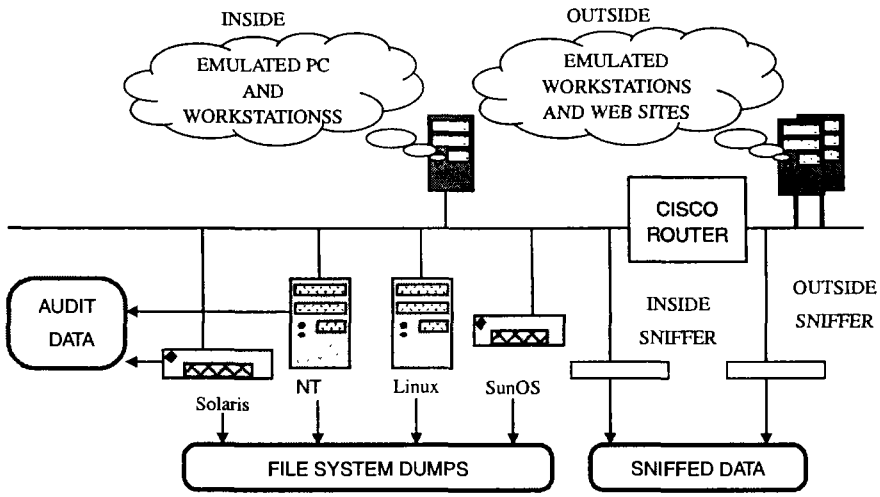

图 1 1999 年评估 IDS 的测试网络环境

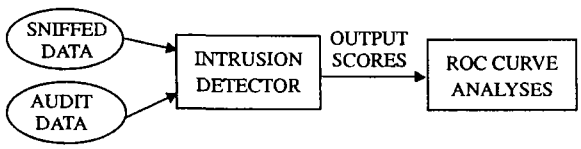

图 2 Lippmann 的评估模型

如图 1 所示,测试网络用来仿真被路由器分隔开的内外两部分网络。外部(OUTSIDE)网络包括两台工作站,它们用来模拟与之相连的 Internet。内部(INSIDE)网络包括很多类型的主机(Linux,Solaris,SunOS,NT),攻击分别由内外网向这些主机发起。还有一台用以模拟路由器内部其他工作站的主机。在路由器两侧都使用了"sniffer"这样的工具,这样就可以收集到所有在网络中流通的数据,将它们连同审计数据和文件信息一起交由参评的 IDS 进行离线分析,根据测试结果得出最后的评估报告。图 2 所示为 Lippmann 的评估模型。

2.2 评价

同 1998 年相比,1999 年的评估进行了很大改善。比如在测试网络中加入了运行 Windows NT 和 Windows 98 的主机,并增加了很多新的攻击方法和来自内部的攻击方式[9~11]。总体来说,1999 年 DARPA 对 IDS 进行的离线评估是成功的。系统根据参评 IDS 的检测结果对 IDS 打分,并详细分析了漏报和误报的原因,攻击数据集中增加了很多新的攻击类型(58 种攻击类型的 200 种实例),包括针对 Windows 系统的攻击和来自内部的攻击方式[12]。然而这次评估仍然存在不足。

Lippmann 先对正常的网络流量进行统计分析,然后在评估过程中用模拟的数据流代替真实的网络流量。虽然 Lippmann 认为此模拟流量与实际流量十分接近,但分析真实网络流量所采用的统计方法并没有公开。至于模拟流量是否会对评估结果造成影响,也没有相关的完整性论证。因此测试环境中使用的模拟流量会在多大程度上影响测试结果仍属未知。

模拟流量中除了正常数据流外还包括攻击数据。虽然评

维普资讯 http://www.cqvip.com

估过程中模拟了大量的攻击方法，但只是简单地把攻击数据加入模拟网络流量中，并没有逻辑上的次序。而实际的攻击者都会很仔细地制订攻击策略(例如在展开正式攻击前，通常都会扫描目标主机或网络，而在攻击结束后会进行系统清理，清除入侵痕迹)，这势必影响到评估的结果。

Lippmann 在对 IDS 进行离线评估时只是简单地把获取的网络流量、系统日志和文件信息作为输入交给参评的 IDS，并没有区分参评 IDS 的种类。不同类型的 IDS 获取检测所需数据源的途径并不相同。如果把包含攻击数据的网络流量作为输入让基于主机的 IDS 来检测，一定会降低其检测率。类似地，系统日志和文件信息并不是基于网络的 IDS 所需的数据，这些数据对于此类 IDS 来说就是干扰流量，会在一定程度上影响检测结果。因此，这种不区分 IDS 类型使用相同数据源进行评估的结果很难令人信服。

DARPA 的这次评估目标是希望提供一个可以通用的数据集，大多数 IDS 可以把此数据集作为输入用来测试自身的各项性能指标，因此 Lippmann 采用了离线评估的方式。但是不同 IDS 运行于不同的网络环境中，需要处理的数据量和数据类型都各不相同，获取可以通用的数据集是不可能的。而且，离线评估的方法忽略了对 IDS 鲁棒性的测试。因为很多攻击者能够发现 IDS 的位置，他们首先攻击 IDS，使其瘫痪，丧失检测入侵的能力。因此鲁棒性是 IDS 的重要性能指标，应该在评估范围内。

评估报告中使用 ROC(Receiver Operating Characteristic)曲线来反应 IDS 检测率和误报之间的变化关系，这种最初应用于检测信号中的技术能否准确地反应检测率和误报之间的变化关系是人们的另一个疑问[8,13,14]。另外，针对 Windows 系统的攻击大都由手动发起，这种做法费时且繁琐。因此，如何自动发起攻击和模拟网络流量是未来评估 IDS 的重要研究方向之一[15]。

3 关键技术

3.1 流量控制

评估 IDS 时应该在两种不同的环境下分别进行，一是运行 IDS 的主机单独运行 IDS，没有其它需要处理的任务，同时网络负载不大；二是在系统繁忙时运行 IDS。系统繁忙指运行 IDS 的系统上并行任务的数量、干扰流量的多少(background noise，指非入侵的网络流量)、会话的大小(攻击性的会话中执行的命令数量)、同时出现的会话数量等。这样做的目的是为了测试系统负载对参评 IDS 的影响。因为我们不能假想运行 IDS 的主机系统不会出现系统繁忙的情况，同时 IDS 保护的网络环境也可能出现过载。如果这些情况对 IDS 的检测结果影响很大，那么即使它在无负载情况下运行良好，也没有实际应用价值。流量调度模块通过模拟流量来控制测试网络环境中的流量及系统负载等，测试外界环境对参评 IDS 的影响。需要模拟的流量包括两部分：

(1)网络流量的模拟。只有在有网络背景流量的情况下，IDS 的评估结果才是客观和全面的。入侵行为在背景流量的掩护下，被 IDS 发现的几率会大大降低。而 IDS 也可能将正常的流量误判为攻击，产生误报。模拟的网络流量可以是一次网络联接(基于联接的流量，TCP)，也可以是一个网络数据包(非面向联接的数据流量，UDP)。

(2)主机使用情况的模拟。对主机使用情况的模拟主要是为了测试基于主机的 IDS。让攻击行为混杂在正常使用中，可以得到更为真实准确的评估结果。对主机使用情况的模拟可以是网络服务的每次使用，也可以是一条用户执行的指令。

3.2 攻击模拟

攻击的模拟是评估环境的核心，也是对 IDS 进行测试的关键。模拟攻击的第一步是要尽可能多地收集各种攻击方法。Internet 上的各类安全站点、论坛和邮递列表中有大量的相关信息。一些国家和公司已建立了相当完备的系统漏洞和攻击数据库，Lippmann 两次离线评估所使用的数据集也公布在网上，可以下载。这些都是模拟攻击所需的攻击方法来源。由于各种攻击的数量过于庞大，所以不可能模拟所有的攻击方式。对现有攻击方式进行分类，然后在每类中取一个有代表的子集进行模拟是比较实际的做法。

Lippmann 根据攻击造成的后果对攻击进行分类，以此方法把攻击分为四类：U2R(User to Root)指普通用户利用某种攻击方法非法拥有超级用户的权限；DoS(Denial-of-Service)指攻击者加大系统的负载，使系统无法提供原有的正常服务；R2L(Remote-to-Local)指远端用户非法访问本地主机这样的攻击方式；Data 指破坏文件数据完整性的攻击[16,17]。在实际的攻击过程中，黑客会仔细地制定攻击程序，攻击通常都是分阶段进行的。例如攻击的起始阶段攻击者会使用工具对目标网络或主机进行扫描，以获取目标机更多的信息(如使用的操作系统版本，系统漏洞是否被补丁等信息)。为了得到更接近真实的评估结果，可以根据入侵的阶段对攻击方式进行分类。

3.3 评估报告

在进行评估的过程中，应该记录测试环境中的网络流量，主机日志等数据。以此为依据，通过比较被测 IDS 的检测结果和数据记录模块所记录的数据，计算 IDS 的各项性能指标，可以得出评估报告，包括检测率和误报率等。

目前，对于 IDS 的整体性能的综合评估还没有一套成熟的评价模型。Lippmann 的评估报告中使用 ROC 曲线来反应 IDS 检测率和误报之间的变化关系。然而，也有人对这种评价方法进行过批评。如何能真实反应 IDS 的检测结果，包括检测率和误报之间的关系及 IDS 的自动响应和恢复情况等，仍然是一个需要进一步研究的课题。

结语 没有哪个系统管理员会允许在他的网络上进行评估测试，所以大多评估都需要在仿真的网络环境中进行，测试环境的设计至关重要。评估时应仔细选取测试环境中的流量，使其最大可能地接近真实的流量。测试的数据集中应包含尽可能多的攻击方式，对攻击的模拟还应该包括并发的方式。在评估 IDS 的过程中，要考虑运行环境对 IDS 检测性能的影响。对 IDS 自身安全性的测试也应该在评估范围内。最后，评估报告要能真实反应 IDS 的检测结果，包括检测率和误报之间的关系及 IDS 的自动响应和恢复情况等。

从 20 世纪 90 年代 IDS 受到重视以来，评估 IDS 的脚步就没有停止过。到现在评估 IDS 的研究已经得到了长足的发展，也涌现出很多的评估方法，有些方法取得了很好的效果。IDS 及其评估技术正成为计算机网络安全的核心研究问题。从本文的分析中可以看出，IDS 的评估技术仍然是个年轻的研究领域。随着入侵检测技术和 Internet 的发展，评估 IDS 的技术从理论研究到实际应用中还存在很多的问题有待研究、探索和发展。

参考文献

1 蒋建春，马恒太，任党恩，等. 网络安全入侵检测：研究综述. 软

维普资讯 http://www.cqvip.com

件学报,2000,11(11):1460～1466
2 Debar H,Dacier M,Wespi A. Towards a taxonomy of intrusion-detection systems. Computer Networks,1999,31(8):805～822
3 Paxson V. Bro:A system for detecting network intruders in real-time. Computer Networks,1999,31(23):2435～2463
4 Manganaris S,Christensen M,Zerkle D,et al. Data mining analysis of RTID alarms. Computer Networks,2000,34(4):571～577
5 Puketza N,Zhang K,Chung M,et al. A methodology for testing intrusion detection systems. IEEE Transactions on Software Engineering,1996,22(10):719～729
6 蔡忠闽,孙国基,卫军胡,等. 入侵检测系统评估环境的设计与实现. 系统仿真学报,2002,14(3):377～380
7 Puketza N,Chung M,Olsson R A,et al. A Software Platform for Testing Intrusion Detection Systems. IEEE Software,1997,14(5):43～51
8 Mchugh J. Testing Intrusion Detection Systems:A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory. ACM Transactions on Information and System Security,2000,3(4):262～294
9 Lippmann R,Fried D,Graf I,et al. Evaluating Intrusion Detection Systems:the 1998 DARPA Off-Line Intrusion Detection Evaluation. In:Proc. of the 2000 DARPA Information Survivability Conf. and Exposition (DISCEX),Hilton Head:IEEE. 2000,2:1012～1035
10 Lippmann R,Haines J,Fried D,et al. The 1999 DARPA Off-Line Intrusion Detection Evaluation. Computer Networks,2000,34(4):579～595
11 Haines J,Rossey L,Lippmann R,et al. Extending the DARPA Off-Line Intrusion Detection Evaluations. http://www.cs.rpi.edu/～brancj/publications/discex01_paper.pdf
12 Lippmann R,Haines J,Fried D,et al. Analysis and Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation. http://www.cs.fit.edu/～pkc/id/related/lippmann-raid00.pdf.
13 Durst R,Champion T,Witten B,et al. Testing And Evaluating Computer Intrusion Detection Systems. Communications Of The ACM,1999,42(7):53～61
14 Pickering K. Evaluating The Viability of Intrusion Detection System Benchmarking. http://www.cs.virginia.edu/～evans/theses/pickering.pdf.
15 Korba J. Windows NT Attacks for the Evaluation of Intrusion Detection Systems. http://www.ll.mit.edu/IST/ideval/pubs/2000/jkorba_thesis.pdf
16 Kendall K. A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems,http://www.ll.mit.edu/IST/ideval/pubs/1998/kkendall_thesis.pdf.
17 Das K. Attack Development for Intrusion Detection Evaluation. http://www.cc.gatech.edu/～wenke/ids-readings/attack_development_thesis.pdf

(上接第10页)

根据定理6的第二个结论,可知Pos({a,b},{d})={1}。DT=({1,2,3,4,5},{a,b},{d}),根据定理6的第三个结论,可知b在DT的条件属性简化中不能去掉。根据相关的定义,可知$\tilde{F}$(DT)={{a}},({a,b},$\tilde{F}$(DT))所有的Ⅰ型极小子集为{b},因此根据定理6的第四个结论,可知决策表({1,2,3,4,5},{a,b}{d})所有的条件属性简化为{b}。

定理7 S=(U,A,V)是一个知识表达系统,|A|>1,d∈A,有下面的结论成立:(1)∅≠C⊆A－{d},决策表(U,C,{d})协调,当且仅当C⊆$\tilde{M}$(x,y)蕴含d∈$\tilde{M}$(x,y),x,y是U中任意的元素。(2)决策表(U,A－{d},{d})协调,当且仅当A－{d}不是$\tilde{M}$(S)的元素。(3)a^*是A－{d}中任意一个元素,a^*在(U,A－{d},{d})的条件属性简化中不可去,当且仅当A－{a^*,d}是$\tilde{M}$(S)中的元素。(4)定义$\tilde{F}$(DT)={A^*|∅≠A^*⊂A－{d},A^*是$\tilde{M}$(S)的元素},其中DT=(U,A－{d},{d}),如果DT协调,则对A－{d}的任意非空子集C而言,C是决策表DT条件属性的简化,当且仅当C是(A－{d},$\tilde{F}$(DT))的Ⅰ型极小子集。

证明:(1)(U,C,{d})协调,等价于∀x∈U,x∈Pos(C,{d}),根据定理6的第一个结论,可知前面的命题等价于∀x,y∈U,C⊆$\tilde{M}$(x,y)蕴含d∈$\tilde{M}$(x,y)。(2)(U,A－{d},{d})协调,等价于∀x∈U,x∈Pos(A－{d},{d}),根据定理6的第二个结论,可知前面的命题等价于∀x∈U,不存在y∈U,使得$\tilde{M}$(x,y)=A－{d},等价于A－{d}不是$\tilde{M}$(S)的元素。(3)a^*是A－{d}中任意的一个元素,根据定理6的第三个结论,可知a^*在(U,A－{d},{d})的条件属性简化中不可去,等价于∃x,y∈U,x∈Pos(A－{d},{d})并且$\tilde{M}$(x,y)=A－{a^*,d},因为(U,A－{d},{d})协调,所以前面的命题等价于A－{a^*,d}是$\tilde{M}$(S)的元素。(4)定义F'(DT)={A^*|∅≠A^*⊂A－{d},并且存在x,y∈U,满足x∈Pos(A－{d},{d}),$\tilde{M}$(x,y)=A^*},因为(U,A－{d},{d})协调,所以F'(DT)={A^*|∅≠A^*⊂A－{d},A^*是$\tilde{M}$(S)的元素},即F'(DT)等于这里定义的$\tilde{F}$(DT),根据定理6的第四个结论,可知C是(U,A－{d},{d})条件属性的简化,等价于C是(A－{d},F'(DT))的Ⅰ型极小子集,进而等价于C是(A－{d},$\tilde{F}$(DT))的Ⅰ型极小子集。证毕。

表3

U	a	b	c	d
1	1	0	2	0
2	2	1	0	1
3	2	1	2	2
4	1	2	2	0
5	1	2	0	2

表3中的知识表达系统S=({1,2,3,4,5},{a,b,c,d},{0,1,2})的不可分辨矩阵

$$\tilde{M}(S)=\begin{bmatrix} & 1 & 2 & 3 & 4 \\ 2 & \varnothing & & & \\ 3 & c & ab & & \\ 4 & acd & \varnothing & c & \\ 5 & a & c & d & ab \end{bmatrix}$$

因为{a,b,c}不是$\tilde{M}$(S)的元素,所以根据定理7的第二个结论,决策表DT=({1,2,3,4,5},{a,b,c},{d})协调。因为{a,b}是$\tilde{M}$(S)的元素,所以根据定理7的第三个结论可知c在DT条件属性的简化中是不可去的。根据定义,$\tilde{F}$(DT)={{c}{a}{a,b}},({a,b,c},$\tilde{F}$(DT))所有的Ⅰ型极小子集为{c,a}{c,b},因此,根据定理7的第四个结论,可知决策表({1,2,3,4,5},{a,b,c},{d})所有的条件属性简化为{c,a}{c,b}。

至此,论文使用不可分辨矩阵将知识表达系统的三大简化问题转化为生成集族的Ⅰ型极小子集问题,可以使用类似的方法导出用Skowron[3]定义的分辨矩阵将知识表达系统三大简化问题转化为生成集族的Ⅰ型极小子集问题。

参考文献

1 Pawlak Z. Rough Sets-Theoretical Aspects of Reasoning about Data. Dordrecht:Kluwer Academic,1991
2 曾黄麟. 粗集理论及其应用—关于数据推理的新方法(修订版). 重庆大学出版社,1998
3 Skowron A,Rauszer C. The discernibility matrices and functions in information systems. In:Slowinski R,ed. Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht,Kluwer:Academic Publishers,1992. 331～362
4 李小霞,陈绵云. 决策表条件属性的简化. 华中科技大学学报(已录用)