基于规则的 IDS 中的 CBR 研究*)

李玲娟[1,3]　王汝传[1,2,3]
(南京邮电大学计算机科学与技术系　南京 210003)[1]
(中国科学院研究生院信息安全国家重点实验室　北京 100039)[2]
(苏州大学计算机科学与技术学院　苏州 215006)[3]

摘　要　本文在入侵检测系统(IDS)中引入基于案例的推理(CBR)来降低基于规则的精确匹配所造成的漏报率,有效地检测由已知攻击变异成的攻击。描述了实现 CBR 的步骤;给出了由规则设计和构造案例库的启发式方法;分析了实现 CBR 的有关算法;最后给出在入侵检测系统 Snort 上扩充 CBR 功能的实验结果。
关键词　入侵检测,规则,基于案例的推理,Snort

Research on CBR in Rule-Based IDS

LI Ling-Juan[1,3]　WANG Ru-Chuan[1,2,3]
(Dept. of Computer Science and Technology, NJUPT, Nanjing 210003)[1]
(National Key Laboratory of Information Security Graduate School, Academia Sinica, Beijing 100039)[2]
(Computer Science & Technology School, Soochow University, Suzhou 215006)[3]

Abstract　In this paper the case-based reasoning (CBR) is introduced to reduce the false negative rate caused by rule-based precise matching in intrusion detection system and to detect the variation of known attack. The steps of implementing CBR are described, several illuminative methods for designing and constructing case-base from rules are proposed, and algorithms for implementing CBR are analyzed. Finally, the experiment results of applying CBR to Snort are shown.
Keywords　Intrusion detection, Rule, Case-based reasoning, Snort

1　引言

作为健壮网络的第二道安全防线,入侵检测系统(IDS: intrusion detection system)检测能力的强弱直接决定了网络系统的安全性能。入侵检测可分为误用检测和异常检测两类。目前 IDS 中的误用检测多通过基于规则的精确匹配来实现,所使用的规则是基于某些特征对已知攻击模式的基本描述。尽管这种机制能实现快速检测,但也存在一些局限,比如当攻击模式很常见时会产生误报,而当攻击模式过于特殊时又会产生漏报,同时也无法识别出新的攻击。造成这些缺陷的原因之一就是规则的表达能力有限。目前已出现了一些解决办法,但结果不够理想。

基于案例的推理(CBR: case-based reasoning)是一种重要的人工智能方法,在数据挖掘中用于分类。它是对新案例在案例库(Case Base)中检索出旧案例,进行修改,给新案例提供一种解的推理模式[1]。它在新、旧案例之间采用类比,而不强调精确匹配;其"案例"是复杂的符号描述,能够最大限度地描述对象的特征;而规则和案例只是知识的不同表示,从已知规则转化出案例极具可行性和高效性。

为此,本文尝试在入侵检测中用 CBR 方法来代替基于规则的精确匹配,以期降低漏报率,并在 Snort 这个基于规则的开放源代码的网络入侵检测系统上进行了实验。值得指出的是,将 CBR 用于 IDS 与用于商业相比,在实现技术上有其特殊性。本文将介绍在 IDS 中实现 CBR 的步骤和由规则构造案例的启发式方法;分析实现 CBR 的有关算法;给出在 Snort 上扩充 CBR 功能的实验结果。

2　CBR 的推理过程

2.1　CBR 的一般推理过程

如图 1 所示,CBR 的推理过程包括:

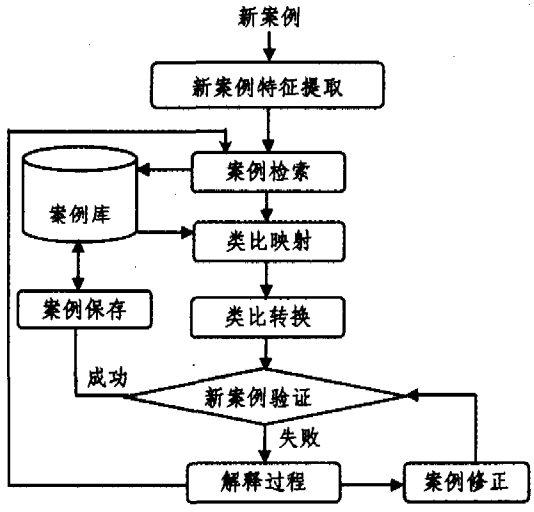


图 1　CBR 的推理过程

(1)新案例特征提取。在基于案例的推理中,首先对新案例(或目标案例)进行分析和特征提取,以产生检索条件。

*)本课题得到国家自然科学基金(60173037 和 70271050)、江苏省自然科学基金(BK2005146)、江苏省自然科学基金预研项目(BK2004218)、江苏省高技术研究计划(BG2004004、BG2005037、BG2005038)、国家高科技项目八六三(2005AA775050)、江苏省计算机信息处理技术重点实验室基金(kjs050001)资助。李玲娟　副教授,博士生,主要研究方向为数据挖掘、信息安全以及移动代理等;王汝传　教授,博士生导师,研究方向为计算机软件理论、计算机网络及信息安全、移动代理技术等。

(2)案例检索。利用新案例的以上特征信息从案例库中检索并选择潜在、可用的源案例,检索过程将参照一个合理的相似度评价标准来判断新、旧案例之间的可类比性,并将获得的源案例集按照与新案例的可类比程度进行优先级排序。

(3)类比映射。从案例库中选择一个最优的源案例,建立它与新案例之间的一致的对应关系。

(4)类比转换。对源案例中与新案例相关的信息进行转换,以便将其应用于新案例的求解过程中,获得新案例的完整的(或部分的)求解方案。

(5)新案例验证。验证新案例与源案例类比关系的有效性。

(6)解释过程。对把转换过的源案例应用到新案例时所出现的成功与失败做出解释,得出成功的经验或者给出失败的因果分析报告。如果得到的仅是新案例的部分解答,则把解答的结果加到新案例的初始描述中,重新开始对案例库的搜索;如果得到的新案例的求解方案未能给新案例以正确的解答,则需要解释方案失败的原因,且调用修补过程来修改所获得的方案。CBR 系统应记录失败的原因,以避免以后再出现同样的错误。

(7)案例修补。当复用阶段产生的结果不好时,需要对其进行修补。这一过程类似于类比转换。修补的第一步是对复用结果进行评估,如果成功,则不必进行修改;否则需要针对错误采用修补措施。第二步是对结果进行评估,可以依据实际运行结果的反馈,也可以通过向专家询问来完成。

(8)案例保存。新问题得到解决后,新案例就可以作为可利用资源加入到案例库中去,这既是学习也是知识获取。案例保存时要考虑选取哪些信息予以保存、如何把新案例有机地集成到案例库中,以及对案例库的修改和精化,其中包括泛化和抽象等过程。

需要保留的案例信息一般包括:与问题有关的特征描述、问题求解的结果、成功或失败的原因及解释。

2.2 在 IDS 中的附加工作

在基于规则的 IDS 中应用 CBR,已有的检测规则仍是判断入侵的重要依据,而多数规则都已经为用于检测的判别特征定义了阈值。当发现用户行为使得判别特征值大于等于该阈值时,就判定为攻击。如果对每个新案例都采用图 1 所示的推理过程,那么对与检测规则完全匹配的攻击行为的检测必然因不必要的推理而多浪费系统开销。

为此,本文认为,在基于规则的 IDS 中应用 CBR 时,推理之前需附加一级预处理和一级过滤。预处理是按规则的定义将时间相关的判别特征按规则定义的时间段重新计算其值。过滤是识别出那些在判别特征上的值大于等于相应阈值的新的案例,对这些案例直接判定为攻击,不传递给 CBR 推理引擎。

3 实现 CBR 的步骤

3.1 一般步骤

实现 CBR 至少包括以下两大步骤:

(1)设计和构造案例库。

这是实现 CBR 的第一步和基础。案例库的设计应遵循以下目标:能提供检测所需的重要信息,以提高检测准确率;尽量精小,以利于加快检索速度。案例库的构造可借助数据库产品提供的功能来完成。

(2) 编写用于实现推理的代码。

用于实现推理的代码包括检索算法代码、用于案例类比的相似度算法代码等;如果案例库基数很大,还需对案例施加聚类算法,以缩小检索范围,提高检索速度;如果案例的修正涉及到特征权值的产生和修正,则需要编写权值估算代码。

3.2 在 IDS 中的实现步骤

CBR 在 IDS 中的应用与其在商业中的应用的重要区别之一就是在相似比较过程中参照的特征的权值不能简单地主观决定,而要依据历史数据用合理的方法加以推导。并且,各种攻击的案例中的判别特征不同;即使是相同的判别特征在不同的案例中的权值也不尽相同。因此在 IDS 中,设计案例时应该针对具体案例,对不同特征赋予合适的权值,以提高推理的准确性。

4 案例库的构造方法

4.1 一般方法

CBR 所使用的案例是复杂的符号描述。案例表示、案例索引、案例库的组织是案例设计的重要内容。CBR 系统的性能和效率在很大程度上依赖于案例的表示和组织,其目标是使案例既能最大限度地描述对象的特征,又易于使用和管理。

目前,较为常用的知识表示方法有一阶谓词逻辑表示法、产生式表示法、语义网络表示法、框架表示法、过程表示法、面向对象表示法等 10 余种[2]。其中,框架表示法将知识用一种类似框架的结构来存储,当遇到新事物时,从中找出一个合适的框架,并根据新的情况对其细节加以修改、补充,从而形成对这个新事物的认识,其基本思想与 CBR 的相一致。因此,本文采用框架来进行案例表示。

典型的框架表示结构如下:

〈框架〉::=〈框架名〉〈槽〉{〈槽〉}

〈槽〉::=〈槽名〉〈侧面〉{〈侧面〉}

〈侧面〉::=〈侧面名〉〈数据〉{〈数据值〉}

〈数据值〉::=〈数值常数〉|〈字符串〉|〈逻辑常数〉|〈其他常数〉

在框架的一般表示中,有一些约束条件,主要是指出什么样的值才能填入到槽或侧面中去。

案例库的组织形式和案例的表示形式是相关联的,目前较为常用的是以数据库形式来进行组织和存储,以方便使用和管理。

以框架为案例表示形式时,主框架和槽对应数据库中的表名和数据项。当槽作为子框架展开时,子框架和子特征就对应了数据库中的其他表,这些表和主框架对应的表之间构成了数据库中表的关系。案例以表的结构形式存储于案例库中,而案例的主要特征以数据项的形式来体现。

案例的知识一般表示为三元组:〈问题描述、解描述、效果描述〉。其中,

(1)问题描述:对求解的问题及周围世界或环境的所有特征的描述。

(2)解描述:对问题的求解方案的描述。

(3)效果描述:描述解决方案后的结果情况,是失败还是成功。

案例库的使用和维护包括:案例的录入、检索、删除和修改。其中,

案例的录入大致分为 3 种情况:

(1)在案例求解的过程中,通过 CBR 的自学习功能,将变换调整得到的案例加入到案例库中。

(2)根据用户的需求录入新案例。

(3)通过对系统的情况分析,或者在案例求解时认为有必要或者必须通过录入某新案例才能完成求解过程时,由专家来进行新案例的录入。

案例的删除通常是在系统维护时认为某案例过于特殊,失去其应用价值,或者是出现了比该案例更具有代表意义的新案例时,由系统维护人员对其进行。

案例的修改通常是在检索后对相似案例的评价和修改。通常情况下,通过检索得到的相似案例,一般会与当前问题存在较多差异,这时应对相似案例进行修改,以检查其对待求解案例约束条件的满足程度。

4.2 IDS中的特殊方法

由于CBR应用于入侵检测还是一项新思路,没有成熟的案例集可供参考。从无到有地构造案例库将会是一项繁重而且不可能全面概括的工作。为此,本文主张移植或借鉴基于规则的IDS中的较为完善的规则库(或文件),直接完成4.1节中案例录入的(2)和(3),然后通过CBR的自学习功能完成案例录入工作中的(1)。基于这种思想的初始案例库的构造过程见图2。

以Snort为例,图3描述的一个Snort规则将转化为图4所示的一个案例描述。

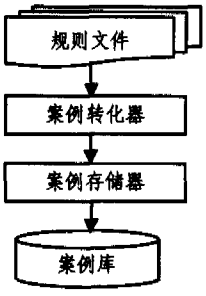


图2 由规则构造案例库

```
alert tcp !192.168.1.0/24 any ->
    192.168.1.0/24 111 (content: "|0001 86 a5|";
      msg: "external mountd access";)
```

图3 Snort规则示例

特征:
协议:tcp ; 源IP:非192.168.1.0/24; 源端口:任意; 目标IP: 192.168.1.0/24
目标端口:大于 111; 包内容:000186a5
行动:
输出报警: "external mountd access"

图4 Snort规则转化为案例示例

5 实现CBR的有关算法

5.1 用于类比的相似度算法

CBR利用相似度知识和特征索引从案例库中找出与当前问题相关的一个或多个最佳案例。检索要达到两个目标:①检索出来的案例应尽可能地少;②检索出来的案例应尽可能与当前案例相关或相似。相关或相似度评价算法的选择对判断近似案例中谁是最适合的案例来说至关重要。

最邻近(NN:Nearest Neighbor)算法[3]是一种应用广泛的非参数分类算法,是一种典型的懒散分类算法,即用一个或多个与待预测的新样本相似的训练样本来显示分类结果,直到有新的样本需要进行分类时才进行分类处理,这种思想符合CBR的推理需求。本文使用NN算法评价相似度,但从实现的角度对之做了改进。

NN算法的基本目标是在多维空间的k个点中寻找与目标样本点最邻近的点。因为最邻近算法易受噪声数据的影响,所以测试点与样本的距离由平均值决定。同时,由于一个对象受其各类特征的影响是不同的,因此引入距离加权的概念。

在IDS中对于一个目标样本O有r个特征,O_i表示目标样本的第i个特征;案例库中一共有n个案例,C^k表示第k个案例;C_i^k表示第k个案例的第i个特征;W_{ki}表示第i个特征在第k个案例中的权值。

数值型特征i的最大值、最小值和范围如下:

$$MAX[i]=\max(C_i^1,C_i^2,\cdots,C_i^k,\cdots,C_i^{n-1},C_i^n)$$

$$MIN[i]=\min(C_i^1,C_i^2,\cdots,C_i^k,\cdots,C_i^{n-1},C_i^n)$$

$$RANGE[i]=MAX[i]-MIN[i]$$

案例及目标样本的归一化特征值为:

$$Cnv[C_i^k]=\frac{C_i^k-MIN[i]}{RANGE[i]}$$

$$Onv[O_i]=\frac{O_i-MIN[i]}{RANGE[i]}$$

C_i^k与目标样本O的距离:

$$\delta_i^k=Onv[O_i]-Cnv[C_i^k]$$

C^k与目标样本O的加权平均距离:

$$D^k=\sqrt{\sum_{i=1}^{r}(W_{ki}*\delta_i^k)^2}$$

设 $$maxD=\sqrt{\sum_{i=1}^{r}W_{ki}^2}$$

C^k和目标样本O间考虑权重后的综合相似度:

$$Sim(O,C^k)=1-\frac{D^k}{\max D}$$

5.2 特征权值的估算法

对于商业应用中的CBR,案例中特征权值可以由用户根据自己的兴趣度来确定。但是在IDS中,案例中特征的权值必须尽可能客观地反映各特征对判断入侵的重要程度。文[4]介绍并比较了多种权值估算法,但我们的研究和实验结果表明,这些特征权值估算法对IDS的适用度较低。本文提出借助特征可视化这种特征选择方法来确定判别特征的权值。

特征可视化是特征选择的重要手段之一,具有直观、准确的特点。本文用林肯实验室1998年发布的用于评估网络IDS的TCPDUMP数据[5]进行特征可视化实验,使用了Ethereal、SPSS、数据库、多维展示算法,得到的部分结果示于图5和图6。

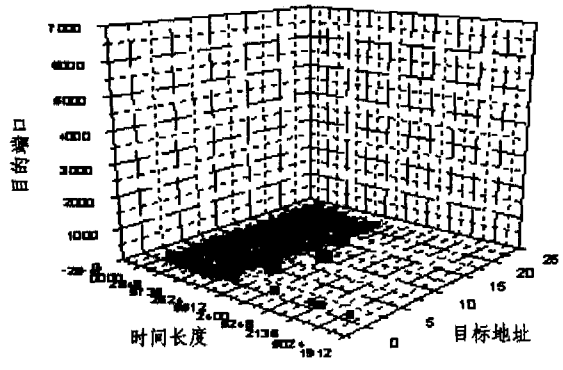


图5 选择的DARPA第一周星期四的数据(不含有攻击)

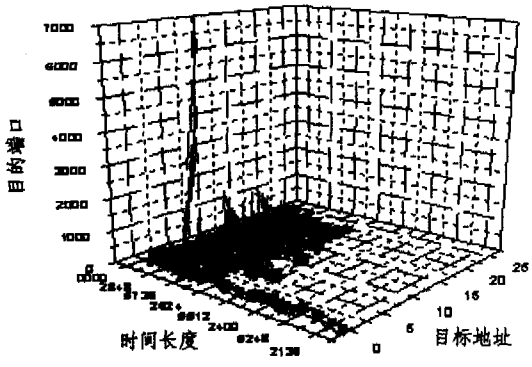


图 6　选择的 DARPA 第三周星期四的数据(含有攻击)

图 6 与图 5 相比,在目的端口特征轴上出现两条短垂直线和一条长垂直线。出现这些现象的原因是第三周的星期四安排了端口扫描攻击和 Neptune 这种拒绝服务攻击。以上是三维分析的结果,通过对开源软件 GGobi 编程还可以提供多维数据展示能力。然而,这个简单的例子已经说明,通过特征的可视化可以分析出判断攻击时可用的重要特征。

无论对基于规则的 IDS 中规则的建立,还是对 CBR 案例的构造而言,以找出与入侵最紧密相关的特征为其目的的特征选择都是增强 IDS 检测效率的必要步骤。因此,可以在借助特征可视化手段离线进行特征选择的同时,进行初始案例库的构造和判别特征权值的确定,将那些伴随着攻击的出现而表现异常的特征选择出来,并根据异常程度的高低赋予相应的权值。实验证明,该方法比文[4]的方法对基于规则的 IDS 更具客观性和有效性。

6　在 Snort 上扩充 CBR 的实验结果

为了测试 CBR 应用于 IDS 的可行性和有效性,本文以 Snort 为平台做了实验,开发了一个在线攻击平台、一个 Snort 在线数据可视化模块、一个案例库(图 7 给出一例)和一个用于入侵检测的 CBR 引擎。

协议名	W1	源主机数	W2	目的主机数	W3	目的端口数	W4	时间间隔	W5	同一字符数	W6	相连字符	W7	HTTP 请求的字符	W8	缓存占有量	W9	攻击名
TCP	3	1	0	5	4	20	10	60	4	0	0	0	0	0	0	0	0	Port Scan Attack

图 7　案例库中的案例示例

实验的设计及结果示例如下:

(1)利用在线攻击平台发出符合 Snort 检测规则的端口扫描攻击。

Snort 中对端口扫描定义了如下检测规则[6]: target_limit 5 port_limit 20 timeout 60。

本规则定义:在 60 秒内,如果一台主机向 5 台主机的 20 个端口建立连接,就认为是端口扫描,预处理器就报警。

当通过攻击平台对目的 IP 地址 10. 10. 138. 19,扫描目的端口 1～200 时,可视化模块显示的检测结果见图 8,前 6 项都是检测到的端口扫描。

(2)利用在线攻击平台发出不完全符合 Snort 检测规则的端口扫描攻击。

为了逃避 Snort 的检测,利用攻击平台向 10. 10. 138. 19 的 18 个端口发出攻击。检测结果见图 9。可见,基于规则精确匹配的 Snort 没能检测出这个意图明显的端口扫描攻击,产生漏报。

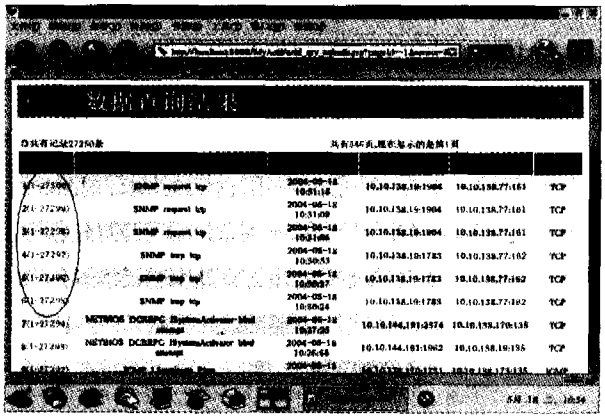


图 8　Snort 对符合规则的攻击的检查结果

扩充了 CBR 功能之后的检测结果见图 10。可见,采用 CBR 机制后,只要确定合理的相似度阈值,就能检测出由已知攻击变异而来的攻击形式,漏报率可大大降低。检测结果可以用于构造存入案例库的新案例,为以后的推理和检测提供依据。

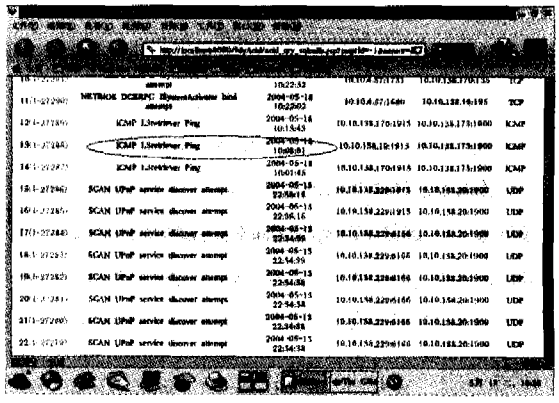
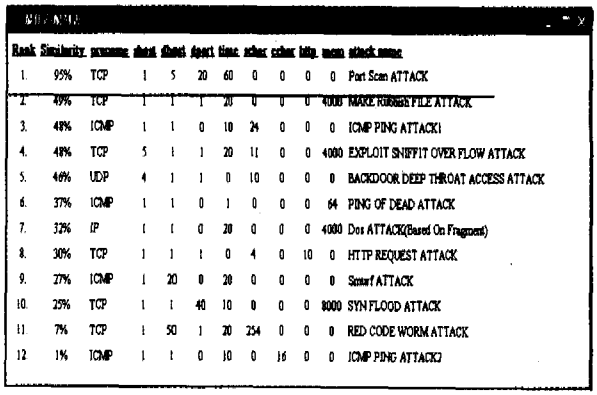


图 9　Snort 对不完全符合规则的攻击的检测结果(漏报)



Rank	Similarity	protocol	dest	dport	sport	time	other	other	http	num	attack name
1.	95%	TCP	1	5	20	60	0	0	0	0	Port Scan ATTACK
2.	49%	TCP	1	1	1	20	0	0	0	4000	MAKE Ribbon FILE ATTACK
3.	48%	ICMP	1	1	0	10	24	0	0	0	ICMP PING ATTACK1
4.	48%	TCP	5	1	1	20	11	0	0	4000	EXPLOIT SNIFFIT OVER FLOW ATTACK
5.	46%	UDP	4	1	1	0	10	0	0	0	BACKDOOR DEEP THROAT ACCESS ATTACK
6.	37%	ICMP	1	1	0	1	0	0	0	64	PING OF DEAD ATTACK
7.	32%	IP	1	1	0	20	0	0	0	4000	Dos ATTACK(Based On Fragment)
8.	30%	TCP	1	1	1	0	4	0	10	0	HTTP REQUEST ATTACK
9.	27%	ICMP	1	20	0	20	0	0	0	0	Smurf ATTACK
10.	25%	TCP	1	1	40	10	0	0	0	8000	SYN FLOOD ATTACK
11.	7%	TCP	1	50	1	20	254	0	0	0	RED CODE WORM ATTACK
12.	1%	ICMP	1	1	0	10	0	16	0	0	ICMP PING ATTACK2

图 10　CBR 对不完全符合 Snort 规则的攻击的检测结果

(下转第 127 页)

果当然比4基算法要好，但同时整个程序的分支也变得比较多，程序也显得比较复杂。但对于大素数的情形，这还是值得的。

6 算法的应用

在RSA公钥算法中大量用到的模幂运算，可以用Montgomery乘和指数加减链的方法来实现[3]。要用加减链，就要求底数的Montgomery逆。如果加减链的单位是多比特的话，所需要算的Montgomery逆就更多了，因此这个算法可以用来加速以加减链为基础的模幂运算。虽然在讨论中我们假定 p 是素数，其实我们只要求输入参数互素即可。这在RSA实际应用是可以保证的，因此算法是可以运用在RSA中的。

另外，在现在广为应用的椭圆曲线密码体制中，倍点运算一直是运算的瓶颈。考虑素域 F_p 上的椭圆曲线（p 是奇素数）$F(x,y)$，对于上面的点 p,Q，要计算 $p+Q,2p$，都将要计算 $\lambda=\frac{y_2-y_1}{x_2-x_1}$ 和 $\lambda=\frac{3x_1^2+a}{2y_1}$。一般，我们是将平面坐标转换成射影坐标，然后将所有参数都转化到对应的Montgomery数，用Montgomery模乘来完成计算，这是以乘法来换取求逆运算。现在我们可以直接利用Montgomery逆来进行计算，因为整个Montgomery逆算法要比普通的求逆算法要快得多。

同时，正如我们前面提到的，本文提出的算法也可以运用于求基本的模逆。

结论 本文通过对Montgomery逆算法核心部分进行两次改进，得到了两种分别以4为基和以8为基的快速算法。其中算法D是以4为基的算法，在基本没有增加算法实现复杂度的基础上，将迭代次数的平均上限从 $2n$ 降到了 $\frac{7}{6}n$，平均迭代次数也从 $\frac{3n}{2}$ 降到了 $\frac{7}{8}n$。而以8为基的算法，其复杂度会稍有增高，主要是内部分支增多，但体现在程序上也只是增加了一些判断和跳转的运算量，其迭代次数的平均上限则降到了 $\frac{25}{24}n$，平均迭代次数也降到了 $\frac{25}{32}n$。另外，由于新算法只要求 u,v 中有一个变为1就可以结束操作，因此实际的迭代次数可能比上述的还要少。同时，这两种算法都可以运用于求基本的模逆。两种算法比较而言，一般情况下建议用4基算法，它效率较高，算法实现也是比较轻量级的。但对于特别大的素数的应用场合，在8基算法的多分支比较跳转所造成的开销相对可以接受的情况下，还是建议用8基算法，它的效率更高。本文所提出的算法在RSA公钥体制以及ECC公钥体制实现中有广泛的应用。

参考文献

1 Kaliski B S Jr. The Montgomery Inverse and its Applications [J]. IEEE Trans on. Computers, 1995, 44(8): 1064～1065

2 Savas E, Koc C K. The Montgomery Modular Inverse-Revisited [J]. IEEE Trans on Computers, 2000, 49(7): 763～766

3 Gutub A. A scalable hardware for montgomery modular inverse computation: [Tech Report]. Corvallis, Oregon 97331: Information Security Laboratory, Electrical and Computer Engineering Department, Oregon State University, 2001

4 Gutub A, Tenca A F, Koc C K. Scalable VLSI architecture for GF(p) Montgomery modular inverse computation [C]. In: Proceedings of IEEE Computer Society Annual Symposium on VLSI, 2002. 53～58

5 L'orencz R'. New Algorithm for Classical Modular Inverse [C]. In: 4th International Workshop on Cryptographic Hardware and Embedded Systems, 2002. 57～70

6 Gutub A A A, Tenca A F. Efficient scalable hardware architecture for montgomery inverse computation in GF(p). In: Signal Processing Systems, 2003, SIPS 27～29

7 McIvor C, McLoone M, McCanny J V. Improved Montgomery modular inverse algorithm [J]. IEEE Electronics Letters, 2004, 40(18)

8 Menezes A, van Oorschot P, Vanstone S. Handbook of Applied Cryptography. CRC Press, 1996

9 Knuth D E. The Art of computer Programming, vol 2. Addison-Wesley, 1981

（上接第120页）

结束语 本文提出了将CBR用于IDS中减少漏报率的思想，也提出了在基于规则的IDS中应用CBR时借助特征可视化手段确定案例中各特征的权值的思路；介绍了实现CBR的步骤、CBR的推理过程和所采用的算法；给出了在Snort上的实验结果。实验证明：CBR方法中对于度量特征明晰的案例的推理效果较之基于规则的方法有了长足的进步，可以有效地检测出Snort无法检测的由已知攻击变异出的攻击，效果良好。

需要指出的是，在案例不多的情况下，CBR的案例并不一定用数据库存储。另外，如果在CBR的推理过程中，案例修改或产生的新案例涉及到特征权值的重确定，那就需要在线估算特征权值，这是本文要进一步完善的工作。

参考文献

1 Dutta S, Wierenga B, Dalebout A. Case-based reasoning systems: from automation to decision-aiding and stimulation. IEEE Transactions on Knowledge and Data Engineering, 1997, 9(6): 911～922

2 王万森. 人工智能原理及其应用[M]. 北京：电子工业出版社，2000

3 Han Jiawei, Kamber M. Data mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers, Inc, 2001

4 Wettschereck D, Aha D W, Mohri T. A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms. Artificial Intelligence Review, 1997, 11: 273～314

5 http://www.ll.mit.edu/SST/ideral/1998/1998data/index.html