

椭圆曲线密码体制及其参数生成的研究

于雪燕, 胡金初, 柴春轶

(上海师范大学 数理信息学院 计算机系, 上海 200234)

摘 要:椭圆曲线密码体制因其长度小、破解难度高等特点在公钥密码系统中逐渐得到广泛应用, 目前它已成为公钥密码体制中的研究热点。介绍了椭圆曲线的基本知识以及椭圆曲线上的密码体制, 列举了与其他密码体制相比的优势所在。因为并不是所有椭圆曲线都可应用到公钥密码体制中, 为了保证其安全性, 选取安全椭圆曲线。文中给出了四种寻找安全椭圆曲线的方法。椭圆曲线密码体制在运算速度和存储空间方面具有很大的优势, 促进了公钥密码学的快速发展。

关键词:公钥密码算法; 椭圆曲线; 离散对数

中图分类号:TP309.7

文献标识码:A

文章编号:1673-639X(2006)11-0160-02

Research on Elliptic Curves Cryptosystems and References Generating

YU Xue-yan, HU Jin-chu, CHAI Chun-yi

(Computer Department, Mathematics and Science College, Shanghai Normal University, Shanghai 200234, China)

Abstract: Elliptic curve cryptosystem has many advantages, such as less bytes and difficult to decode, so it is widely used in public key system, and it has been a research hotspot in the area of public key system. Deals with the notion of elliptic curves and elliptic curves based on cryptosystems, and lists the advantages of elliptic curves cryptosystem compared to other public key cryptosystems. Not all of the elliptic curves can be applied to public key system, so must find security elliptic curves. Gives four methods of finding security elliptic curves. Elliptic curve cryptosystem has great advantages in calculating speed and memory space. It promotes the development of public key system.

Key words: public key cryptosystem; elliptic curve; discrete logarithm

0 引言

根据密钥的特点, 可以将密码系统分为私钥和公钥两大密码系统。在私钥密码系统中, 解密密钥和加密密钥相同或者很容易从加密密钥导出, 这一特点致使加密系统变得不安全。1976年 Diffie 和 Hellman 发表了著名的“密码学的新方向”一文^[1], 提出公开密钥密码的思想, 从此开始公钥密码的发展。在公钥密码体制中, 解密密钥和加密密钥不同, 从一个难于推出另一个, 加密和解密是可分离的, 通信双方事先无须交换密钥就可建立起保密通信。目前影响最大的三类公钥密码是 RSA 公钥密码、ElGamal 公钥密码、椭圆曲线公钥密码。其中 RSA 公钥密码的安全性依赖于数学中大整数因子分解问题的难度, 而 ElGamal 公钥密码与椭圆曲线公钥密码分别基于一般有限域离散对数问题(DLP)和椭圆曲线离散对数问题(ECDLP)。在以上三类公钥系统中, 椭圆曲线公钥系统最具有优势, 因为:

(1) 在有限域 F_q 上的椭圆曲线很多, 为用椭圆曲线构造密码系统提供了丰富的资源。

(2) 椭圆曲线公钥密码系统中的主要计算量是计算 $Q = kg$, 且 Q 很容易求出, 而知道 Q, g , 求 k 十分困难。

(3) 要获得同样安全强度, 比 RSA 用的参数规模小得多^[2], 开销较少且速度快。

(4) 椭圆曲线离散对数问题(ECDLP)比有限域离散对数问题(DLP)困难得多。

基于具有无可比拟的优势, 椭圆曲线公钥密码系统被认为是新一代公钥密码系统。无论在数据加密和数字签名上, 椭圆曲线公钥密码系统已成为人们非常感兴趣的研究方向之一, 从而在这方面涌现出了很多有价值的成果。

1 椭圆曲线介绍

定义1: 设 $K = GF(q)$ 为一有限域, K 上椭圆曲线方程 E 为:

$$y^2 = x^3 + ax + b \quad (p \geq 5, a, b \in K, 4a^3 + 27b^2 \neq 0)$$

$$y^2 + xy = x^3 + ax + b \quad (p = 2, a, b \in K)$$

满足椭圆曲线方程 E 的所有点及一个称为无穷远点 O 的点所构成的集合

$E(K) = \{(x, y) \mid (x, y) \in E, \text{且 } x, y \in K\} \cup O$ 为该曲线的 K -有理点集合, 它是一个有限集, 元素个数称为该椭圆曲线 E 的阶, 记 $\#E(K)$ 。在该有限集上定义一个加法运算, 使得这些点对于该加法运算形成一个 Abel 群, 群的单位元为无穷远点 O ^[3]。

定理1(Hasse 不等式): 设 $K = GF(q)$, E/K 为有限域上的椭圆曲线, 有不等式 $|\#E(K) - p^d - 1| \leq$

收稿日期: 2006-02-17

作者简介: 于雪燕(1981-), 女, 吉林吉林人, 硕士研究生, 主要研究网络与多媒体; 胡金初, 教授, 硕士生导师, 主要研究网络与多媒体。

$2(p')^{1/2}$ 成立。

定义 2: 设 E/K 为椭圆曲线, 点 P 为其上的点, 最小的满足条件 $rP = O$, 正整数 r 称为点 P 的阶。根据有限域的知识, 知道这样的 r 总是存在且整除椭圆曲线阶 $\#E(K)$ 。整数 k, l 满足条件 $kP = lP$, 当且仅当 $k = l \pmod{r}$ 。

2 椭圆曲线上的密码体制

椭圆曲线上离散对数问题(ECDLP)定义如下: 给定素数 p 和椭圆曲线 E , 对 $Q = kP$, 在已知 P, Q 的情况下求出小于 p 的正整数 k 。可以证明由 k 和 P 计算 Q 比较容易, 而由 Q 和 P 计算 k 则比较困难。ECDLP 是比整数因子分解问题(IPF)和离散对数问题(DLP)难得多的数学难题。基于该难题, Neal Koblitz^[4]和 Victor Miller^[5]在 1985 年分别利用有限域上椭圆曲线的点构成的群实现了离散对数密码算法, 其中被广泛接受的是椭圆曲线上的 DSA, 称为 ECDSA。此后, 有人在椭圆曲线上实现了类似 ElGamal 的加密算法, 以及可恢复明文的数字签名方案。

3 椭圆曲线签名算法

椭圆曲线密码体制像 RSA, ElGamal 一样, 可以应用于数字签名。DSA 是美国国家标准局制定的数字签名算法, 它是建立在有限域乘法群 F_q^* 上的。对于有限域上的椭圆曲线密码系统, 相应于 DSA, 建议采用椭圆曲线数字签名算法(ECDSA)。下面给出算法。

设椭圆曲线公钥密码系统参数为 (F_q, E, g, n, a, b, h) , 其中 F_q 是有限域, E 是 F_q 上的椭圆曲线, g 是 E 上的一个基点, n 是椭圆曲线 E 的阶, a, b 是椭圆曲线 E 的系数, h 是一个小的素数。

3.1 密钥生成

用户 A 随机选择一个整数 x , 作为私钥, 公钥是 $y = xg$ 。

3.2 签名过程

- (1) 用户 A 随机选取一个整数 k , 其中 $1 < k < n$, 计算 $kg = (x_1, y_1)$, $r_1 = x_1 \pmod{n}$;
- (2) m 为消息, 计算 $e = h(m)$;
- (3) 计算 $s = k^{-1}(e + r_1 x_1) \pmod{n}$;
- (4) m 的签名为 (s, r_1) ;

3.3 签名的验证

- (1) 计算 $e = h(m)$;
- (2) $u = s^{-1}e, v = s^{-1}r_1$;
- (3) $(x_2, y_2) = ug + vx, r_2 = x_2 \pmod{n}$;
- (4) 如果 $r_1 = r_2$, 则接受这个签名。

从上面的签名算法可以知道, 为了计算 s 的值, 必须计算 k^{-1} , 也就是必须进行求逆运算。对一个大整数求逆, 运算是很慢的, 若是采用扩展欧几里得算法来求逆, 平均需完成 $0.8413 \log_2(n) + 1.47$ 次除法, 将耗费一定的运行时间。这就是上述 ECDSA 算法的不足之处^[6]。

4 参数的选取

椭圆曲线上的公钥密码体制的安全性是建立在椭圆曲线离散对数的基础上, 但并不是所有椭圆曲线都可以应用到公钥密码体制中, 为了保证其安全性, 必须选取安全椭圆曲线, 即阶为大素数或含大素数因子的椭圆曲线为安全椭圆曲线。一般来说有 4 种^[2]寻找安全椭圆曲线的方法:

1) 有限域 $GF(q)$ 上随机生成一椭圆曲线, 直接计算其阶, 判断阶是否为大素数或含大素数因子, 若是即确定, 否则继续选取曲线, 直至符合条件。

2) 取具有一定特殊性椭圆曲线的系数, 计算该椭圆曲线的阶, 对该阶进行判断, 直至找到所需要的安全曲线。

3) 如果 $q = 2^m$, 其中 m 能被一个比较小的整数 d 整除, 首先在有限域 $GF(q_1)$ ($q_1 = 2^d$) 上选择一椭圆曲线 E' 并计算其阶, 根据此值, 利用 Weil 定理^[2]计算该曲线在其扩域 $GF(q)$ 上的阶, 若此阶符合安全标准, 再找曲线 E' 在域 $GF(q)$ 上的嵌入 E , 则 E 即为所需的安全椭圆曲线。

4) 首先给出具有安全条件的曲线阶, 然后构造一具有此阶的椭圆曲线。

目前国内外比较流行的计算椭圆曲线阶的算法有 complex multiplication 算法、SEA 算法、Satoh 算法。应用广泛的椭圆曲线公钥密码体制(ECC)中大多是基于特征 2 的有限域上。

5 结束语

分析了将安全椭圆曲线引进公钥密码体制的优点。与目前应用较普遍的 RSA 算法相比, 在同等安全的情况下, 其所需的密钥长度远比 RSA 低, 因而 ECC 的特性更适合当今电子商务需要快速反应的发展潮流, 在快速加密、密钥交换、身份认证、数字签名、移动通信、智能卡的安全保密等领域, 具有广阔的市场前景。

椭圆曲线公钥密码体制实现的关键是安全曲线的构造问题, 首先计算小基域上一椭圆曲线阶, 通过 Weil 定理可知该曲线在其有限扩域上的阶, 若该阶符合椭圆曲线公钥密码体制的要求, 则将曲线嵌入到其扩域上即可求得一条安全曲线。

参考文献:

- [1] Diffie W, Hellman M E. NEW Directions in Cryptography[J]. IEEE Trans Informat Theory, 1976, IT-22: 644-654.
- [2] Silberman J H. The Arithmetic of Elliptic Curves[M]. New York: Springer-Verlag, 1986: 46-61, 130-136.
- [3] Stallings W. 密码编码学与网络安全——原理与实践[M]. 第 3 版. 刘玉珍, 王丽娜, 傅建明, 等译. 北京: 电子工业出版社, 2004. 219-224.
- [4] Miller V. Use of Elliptic Curves in Cryptography[C]// Odlyzko A M. Advances in Cryptology - Proceedings of CRYPTO 1986, volume 263 of Lecture Notes in Computer

(下转第 198 页)

有名词特征词,如果专有名词不在分词的碎片中,那么默认地名特征词的前两个词是专有名词的上界,地名特征词是专有名词下界。

4) 如果专有名词存在于“专有名词特征词……连词……专有名词特征词”框架对模式中,对于是“人名特征词……连词……人名特征词”框架对模式,“人名特征词……连词”间的专有名词的上界是人名特征词,下界是连词的前序词,“连词……人名特征词”间的专有名词的上界是人名特征词,下界是人名后续词。对于是“地名特征词……连词……地名特征词”框架对模式或“机构特征词……连词……机构特征词”框架对模式,“地名特征词……连词”或“机构特征词……连词”间的专有名词的上界是分词的碎片中的起始位置或者是地名特征词、机构名的前序词;“连词……地名特征词”或“连词……机构特征词”框架间的专有名词的上界是连词后续词,下界是地名特征词或机构特征词。

3.5 专有名词归类

如果专有名词的边界确定之后,专有名词的归类就相对容易多了,专有名词的归类是指识别出来的专有名词是属于人名、地名还是机构名类型。专有名词的归类只要看识别专有名词的模式,如果模式是“人名上文……下文”,“人名上文-姓氏特征词”,“姓氏特征词……人名下文”,那所识别专有名词就属于人名类型;如果模式是“地名上文……下文”,“地名上文……地名特征词”,“地名特征词-地名下文”,那所识别专有名词就属于地名类型;如果模式是“机构名上文……下文”,“机构名上文……机构特征词”,“机构特征词-机构名下文”,那么所识别专有名词就属于机构名类型。

4 试验结果与分析

下面给出在实验过程中采用的语料和指标,然后给出试验的一个初步结果及相应的分析。

4.1 试验用语料和评测标准

试验使用了北京大学计算语言学研究所的标注语料库(1998年1月)。在此语料的基础上根据自定义的属性重新进行标注后,作为试验用语料。

针对专有名词的识别,采用了两个评测指标,即准确率(P)、召回率(R)。其定义如下:

准确率 = 系统识别的正确词数 / 系统识别的总词数 $\times 100\%$

召回率 = 系统识别的正确词数 / 总的正确词数 $\times 100\%$

4.2 试验结果

根据模式提取中的阈值设定的不同,在封闭测试中的试验比较如表 1 所示。

表 1 试验结果

阈值(f)的取值	准确率(%)	召回率(%)	自学习到的规则数
1	95.3%	92.5%	6759
2	87.2%	83.7%	8432

4.3 试验结果分析及后续工作

封闭测试中,阈值的选取直接决定了规则的提取和准确率与召回率的结果。阈值的取值越小,造成采用的模板增多,在基于转换的错误驱动进行规则学习的时候学习到的规则就越多。但是,在封闭测试试验中,准确率和召回率反而增高。但是规则的减少,使模板应用的局限性加大了,这样不利于其在开放测试中的应用。所以,在更大的语料里面进行阈值确定和规则的大量提取是以下将要进行的工作。

5 结论

首先分析现阶段专有名词识别存在的问题和局限性,从人自身在阅读时候区别专有名词和普通用词的特点,提出了基于属性标记的专有名词的识别。此方法从专有名词自身特点(姓氏用词等)和上下文环境特点出发,重新标注语料,然后采用基于转换错误驱动和基于实例相结合的学习方法,找出了一系列专有名词出现的上下文环境和规则。在此基础上进行了小规模语料的封闭测试识别,取得了相当好的效果。目前实验表明基于属性标记的专有名词识别方法是行之有效的。但此方法的有效运用,需建立在拥有大量的熟语料库的基础上,而且正确的模板阈值(f)的确定,也关系着系统的准确度。这些存在的问题也正是需要进一步完善的地方。

参考文献:

- [1] 张华平,刘群.基于角色标注的中国人名自动识别研究[J].计算机学报,2004,27(1):85-91.
- [2] 孙宏林,俞士汶.浅层句法分析方法概述[J].当代语言学,2000(2):74-83.
- [3] Brill E. Transformation-based error-drive learning and natural language processing; a case study in part of speech tagging[J]. Computational Linguistic, 1995, 21(4): 543-565.
- [4] Brill E. A Simple Rule-based part of speech tagger[C]//In: Proc 3rd Conference on Applied Natural Language Processing. Trento: ACI, 1992.
- [5] 陈文亮,朱靖波,吕学强,等.词性标注规则的获取和优化[J].术语标准与信息技术,2004(2):23-26.
- [6] 万建成,杨春花.书面汉语的全切分分词算法模型[J].小型微型计算机系统,2003,24(7):1247-1251.

(上接第 161 页)

Science. New York: Springer, 1986, 417-426.

- [5] Koblitz N. Elliptic Curve Cryptosystems[J]. Mathematics of Computation, 1987, 48: 203-309.

- [6] 李道丰,揭金良.基于椭圆曲线的数字有序多签名方案[J].通讯和计算机,2005,2(2):36-38.