摘要

课题介绍

## 课题背景

## 研究意义

## 研究内容

# 相关基础知识

随着网络的发展，数据传输的速度也越来越快。密码学的重要性也越来越重要，消息认证是密码学的一个重要内容。本章详细的对密码学级密码学中消息认证的基础知识做出详细的介绍。具体包括：困难问题、椭圆曲线、认证、Hash函数、以及典型的Hash算法SHA和MD5

## 困难问题

困难问题是密码学的基础，所有的加密算法和认证算法必须依附于困难问题，即没有困难问题就没有密码学。想构造一个密码学算法，必须找到一个与之相对的困难问题。

大整数因数分解问题

给定两个素数p，q，计算比较容易；

给定大整数n，求n的素因数p，q使非常困难。

虽然说大整数因数分解是一个困难问题，但随着计算机的硬件发展，这个困难问题也越来越容易破解。依据这个问题构造的RSA算法现在512bits和1024bits相继与1999年和2009年被破解。现在要保证RSA算法的安全n大约要去2048bits。

离散对数问题

已知有限循环群及其生成元g和阶。

给定整数a，计算元素很容易；

给定元素h，计算整数x，0≤x≤n,使得非常困难。

目前，有限域上的离散对数问题仍没有找到一种方法能有效破解。它的难易程度和生成元的取值无关，只和群中的元素数量有关。一般来说当k取值很小的时候，我们可以利用计算机穷举得到结果，现在的k一般取为1024bits。

椭圆曲线

椭圆曲线是在射影平面上满足方程

（其中a、b、c、d、e为常数）

上所有点的集合。

在诸多困难问题中，椭圆曲线上的离散对数问题要比其他问题难解的多，据研究，160位ECC加密安全性相当于1024位RSA加密，210位ECC加密安全性相当于2048位RSA加密。

椭圆曲线密码体制是比较安全且密钥长度较短的公钥密码体制

## 认证

认=检查、确认。证=身份、证明。认证其实就是为了确认身份、检查权限。

认证包括身份认证和消息认证

身份认证是指验证信息的发送者是否为真，包括信源、信宿等。

消息认证是指验证信息的完整信，验证数据在存储或传输过程中未被篡改、重放等

常见的攻击方式：

1. 重放攻击：偷听者记录当前的消息，在适当的时间再次发送。

对策：增加时间戳

1. 字典攻击：利用常用的口令（例如123456、password等）对消息进行破解。

对策：在口令里面增加额外信息

### Kerberos认证系统

认证系统是防止消息被篡改、重放等攻击的一种有效方式，确保接受者确认消息的真伪。

Kerberos是身份验证协议、是Windows 2000的默认身份验证服务，是基于tcp/ip网络协议设计的可信协议。

步骤如下:

1. client(客户端)发送请求到验证服务器（AS），验证服务器创建一个“session”（根据你的请求），并产生一个代表服务器的随机值。
2. client将这个”session”发送给授权服务器（TGS），TGS会确认”session”，并返回一个消息，并打上时间戳。
3. client向消息服务器发送TGS返回的消息，选择接受或拒绝，并执行相应的服务。

## Hash函数

Hash函数，即散列函数，音译为哈希函数，目的是将任意长的消息压缩为一个固定长度的摘要。

Hash的特点：

1. 输入域中的元素是无穷的
2. 输出域中的元素是有限的
3. 当输入发生微小变化时，输出会发生巨大变化
4. 对于任意输入x计算H(x)是比较容易的，知道H(x)，寻找x是比较困难的

Hash构造方法

1. 直接寻址法

2. 数字分析法

3. 平方取中法

4. 折叠法

5. 随机数法

6. 除留余数法

MD5

MD5 (message-digest algorithm 5)，经由MD2、MD3和MD4发展而来。主要用途：1防止信息被篡改、2防止直接看到明文、3防止抵赖。

MD5原理

**1、数据填充**

对消息进行数据填充，设消息长度为a，a满足(a+x) mod 512=448。可以算出需要填充长度x。填充方法：在消息后面进行填充，填充第一位为1，其余为0。

再把x的长度转换为二进制共64位，如果长度大于那么只取其后64位二进制数。并把这64位填充到消息上，现在消息长度为512的整数倍。

**2、数据初始化**

准备需要用到的数据：

* 4个常数： A = 0x67452301, B = 0x0EFCDAB89, C = 0x98BADCFE, D = 0x10325476;
* 4个函数：F(X,Y,Z)=(X & Y) | ((~X) & Z); G(X,Y,Z)=(X & Z) | (Y & (~Z));  H(X,Y,Z)=X ^ Y ^ Z; I(X,Y,Z)=Y ^ (X | (~Z));

**3、数据分组**

四种操作（Mj表示第j快的32位数据，ti是floor(abs(sin(i)) × 2^32)，<<<s表示循环左移s位）

FF(a,b,c,d,Mj,s,ti)表示a=b+((a+F(b,c,d)+Mj+ti)<<<s)

GG(a,b,c,d,Mj,s,ti)表示a=b+((a+G(b,c,d)+Mj+ti)<<<s)

HH(a,b,c,d,Mj,s,ti)表示a=b+((a+H(b,c,d)+Mj+ti)<<<s)

II(a,b,c,d,Mj,s,ti)表示a=b+((a+I(b,c,d)+Mj+ti)<<<s)

把最终消息以512位为一组，分成若干组，以第一组为例：把这512位分成16快32位

每种操作进行16次，得到最后一组a，b，c，d。

每组操作完成之后A=a+A，B=b+B，C=c+C，D=d+D，进行下组计算，最后将ABCD级联

输出的结果就是MD5码。

SHA-1

SHA-1是由美国标准技术局（NIST）颁布的国家标准，这个算法也是有MD4发展而来。它和MD5的原理基本一样，这里我就只说MD5和SHA的区别。

SHA-1 与 MD5 的比较

l 对强行攻击的安全性：SHA-1得到的结果比MD5码长32 位。使用强行技术， MD5是

量级的操作，而对SHA-1则是量级的操作。这样，SHA-1能保证更大的安全强度。

2 速度：在相同的硬件上，因为SHA-1的循环次数多且缓存区大（160bits：128bits），所以SHA-1 的运行速度比 MD5 慢

三 云存储的完整性验证

云存储作为一种线上的存储方式，我们肯定会怀疑存储的数据是否安全，是否被别人修改，数据认证成为了云存储中的一个重点研究对象。本章简述了云存储中数据认证的重要的数据结构Merkle Tree，以及典型的数据认证算法。

数据完整性方案模型：

在现有的数据完整性认证模型中：PDP和POR是最典型两种模型。这两种模型的出现为以后的数据认证模型奠定了基础。

PDP模型高效，因为它不需要对数据进行加密。

POR可以恢复以损坏的文件。

PDP模型

Ateniese等人在2007年提出了PDP模型，它是一个基于概率统计的一个模型，是采用抽样的形式，通过一个或几个子文件来判断整个文件的完整性。客户端在做预处理时，利用少量的元数据去向服务器发起挑战。

图片

PDP包括以下算法

KeyGen() ->(pk,sk):客户端生成公钥pk=（N,g）和私钥sk=（e，d，v）其中e是事先说明的大数e>且d>，v，输出pk，sk。

TagBlock(pk,sk,m,i) ->:客户端利用pk和sk生成，。输出，称为元数据的标签。

GenProof(pk,F,chal,∑) –>V:服务端利用输入的pk和chal，遍历所有的元数据，得到V=(T,)

其中T=，H(g mod N),输出V；

CheckProof(pk,sk,chal,V)->{success,failure}:客户端利用pk和sk和chal生成，比较V和是否相等，若相等则输出success，否则failure。

我们生成PDP协议和规范分为两步：setup和challenge。

setup：客户端运行KeyGen()生成公钥pk=（N,g）和私钥sk=（e，d，v）并在客户端上存储。将原始文件分块，再利用TagBlock(pk,sk,m,i)生成，并将原始文件和集合{}一同发送个服务器。

Challenge：客户端以n为单位按周期向服务器发起挑战，即在原始文件中随机抽取c个样本组成chal集合，并chal请求发送个服务器。服务器利用GenProof(pk,F,chal,∑)生成证明V，并返回客户端。收到V，CheckProof(pk,sk,chal,V) 来确定原始文件是否完整。

POR模型

Juels等人于同年提出了POR模型的概念。POR和PDP一样都可以保证数据的完整性，但是POR模型设置了客户端发送挑战的次数的最大值，导致了POR模型只能支持静态数据验证和恢复。

图片

POR包括以下算法

KeyGen()->K

Encode(F;K,) ()->()

Extract(;K,) ()->F

Challenge(;K,) ()->c

Respond(c,)->r

Verify((r,); K,)->b

POR算法和PDP算法执行基本一致，只有

Encode(F;K,) ()->()算法中对文件进行了加密，并在文件中加入”哨兵”数据和原始数据混合生成新的数据。

Extract(;K,) ()->F算法是通过正确的验证来返回原始数据，由于”哨兵”的产生是随机的有限的，每次数据恢复就会返回一个”哨兵”，从而导致只能做有限次验证。

具体的POR算法请参考文献。

公开验证

在本地存储的文件，我们一般通过访问全部数据来验证完整性，如果在云存储中也采用同样的方式进行数据完整性的校验，会耗费计算机大量的资源。即使不考虑计算机的性能问题，我们也不能每时每刻的让客户端保持连接状态。因此，我们把数据完整性校验委托给第三方来代替我们进行验证工作。

在公开验证中，引入了一个第三审计员，在客户端与服务端建立可信机制，所以又把这种验证方式称为数据审计。

公开验证的关键：

在PDP和POR协议中，只有客户端拥有原始数据的私钥sk，其他人无法进行验证。同时这两种算法的客户端计算复杂度、存储开销、通信开销都比较高。如果提高验证效率和扩展算法得以支持公开验证成了学者关注的主题。现在的方案大多都是基于同态验证标签（HVTs）来实现的。

同态验证标签具有两个基本性质：

1. 无数据块验证：只需要提供原始数据由标签构造的证据，就能进行数据审计
2. 同态性：已知数据块m，n它们分别对应的同态标签，。数据块f=m||n。

可以通过，来表示数据块f的同态标签。

动态数据更新

四 跳表

跳表是一种比较重要的数据结构，在数据认证中起着很大的作用。本章详细介绍了跳表这种数据结构，包括跳表的增删改查算法及其实现，同时还介绍了跳表的复杂度，更能直观的反应出跳表的优缺点。最后介绍了跳表的实际应用场景。

### 为什么选择跳表？

在云存储中我们经常要存储大量的数据文件。如果我们直接用二叉树或者链表用来记录文件节点，会影响我们的查询效率。

然而 在目前经常使用的平衡二叉树结构中，它是一种严格的平衡，虽然它的最差的查询时间一定是O(logn)，但它在插入方面需要旋转和overhead等操作。

跳跃表（Skip List）于1987年诞生，其查询 、插入和删除操作的预期时间复杂度为O（logn）。虽然跳表的最差查询时间是不稳定的，但跳表对插入更快，对并发的处理更友好。

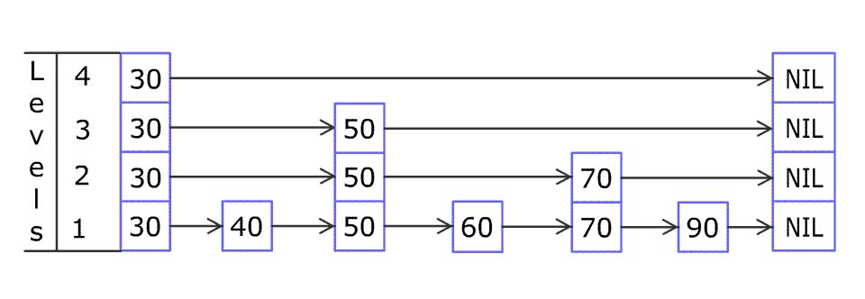
## 跳表的基本原理

跳跃表由多条链构成（，， ……，），满足一下条件：

最底层链表包含所有的元素，链表中的元素升序

每个链表有两个特殊节点 -∞(起始节点)和+∞(终止节点)

S0包含所有元素，对任意的i>0， +1是的子集，且中的每一个节点，有一个下行指针，指向中与之有相同索引值的节点。

 （图片换）

跳表的特点

1. 节点重排，以及顺序性的访问方式
2. 灵活的增删节点

跳表的相关操作：

查询节点：（如果当前节点存在，则返回当前节点。如果不存在则返回离此节点最近的）

1. 从最上层的链（Sh）的开头开始
2. 假设当前位置为p，它向右指向的节点为q（p与q不一定相邻），且q的值为y。将y与x作比较

(1) x>y 从p向右移动到q的位置

(2) x=<y 从p向下移动一格

3. 当p在底链（S0）中且x=<y 输出当前p节点

删除节点

1. 通过查询 判断节点是否存在

2. 存在 则将该节点所在整列从表中删除

3. 将多余的“空链”删除

增加节点

高度的生成方法

（1）level = 0，

（2）产生一个0到1的随机数r，如果r小于一个常数p（），

（3）则执行方案level=level+1且重新执行（2）（3）

（4）否则，输出level。

1. 通过查询 判断节点是否存在
2. 在当前节点p后增加高度为level的节点

跳表的性能

空间复杂度： O(n) （期望）

相关操作的时间复杂度：

查找： O(logn) （期望）

插入： O(logn) （期望）

删除： O(logn) （期望）

空间复杂度分析：

令由高度的生成方法，每个元素插入到第i层（Si）的概率为，则在第i层插入的期望元素个数为n，跳跃表的元素期望个数为 ，当的数时，次数总和小于2n。

S = 

当

S = 2n

空间复杂度为O(n)

实际测试： 共执行了次 maxLevel=3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P | 平均插入时间 | 平均高度 | 总结点数 | 平均查找时间 |
|  | 0.53976 | 1.7503845 | 17503845 | 0.28027 |
|  | 0.46651 | 1.5754064 | 15754064 | 0.25341 |
|  | 0.60256 | 1.3146518 | 13146518 | 0.31451 |
|  | 0.78464 | 1.1734221 | 11734221 | 0.35154 |

一般我们把p取，此时的平均查询时间最短，如果有要求空间的话，我们也可以退而求其次取p=。这样节点就不会占据大量空间。

跳表的实际应用：

应用场景：

不考虑空间，但要求查询时间快，和节点增加更新不是很频繁。

1. Redis

Redis能存储五种数据类型string、list、set、zset、Hash。其中zset结构使用跳跃表提高查询效率(以score为顺序的跳跃表)

五 认证跳表

认证跳表是基于跳表实现对数据认证的一种算法。

基于有向哈希树的认证跳表