9.词典

（xc）MD5

以小明大，见一叶落而知岁之将暮，
睹瓶中之冰，而知天下之寒

邓俊辉

deng@tsinghua.edu.cn

MD5

- 网络下载：除了数据包本身，你是否还注意过对应的MD5？

  MD5？有何作用？

- UNIX：/etc/passwd

```
#LOGNAME:PASSWORD:UID:GID:USERINFO:HOME:SHELL
root:zPDeHbougaPpA:0:0:Super User:/:/bin/sh
ftp:xyDfccTrt180xMy8:3:3:FTP User:/usr/spool/ftp:
smart:xmotTVoyumjls:6:4:THUDSA Student:/:/bin/csh
#......
```

- PASSWORD域，是如何计算出来的？如何验证的？反过来，会暴露原文吗？

Note: The i32 Intelligent Updater package cannot be used to update Symantec AntiVirus Corporate Edition servers, but can be used to update Corporate Edition clients.

File Name	Creation Date	Release Date	File Size	MD5 | all
20101113-003-v5i32.exe | FTP	11/13/2010	11/13/2010	80.1 MB	6DEBC39595EFC409C308AA0CA79667D0

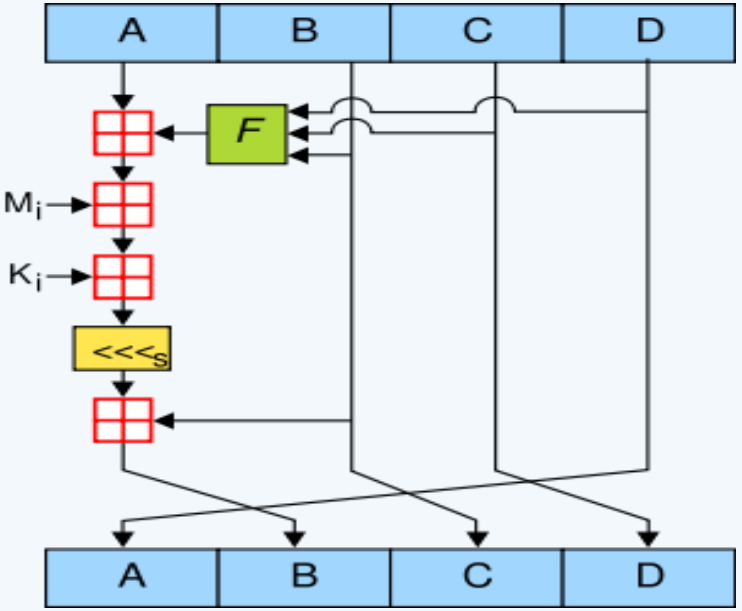
傅位于四子

以小见大——MD5

数字签名

16^32 = 2^128 = 10^38

正向易，逆向难



MD5

- 在很多场合，MD5都可大显身手...
- 数字签名：授权文档一经篡改随即失效 //电子形式签署的法律文书
- 身份验证：OS或应用软件，根据口令做账号授权 //充分而不必要
- 隐私通知：“你我知道是指谁，别人不知道是指谁” //手机尾号
- 寻渊探源：文档之间的（局部、逻辑）复制关系 //学术诚信鉴别
- 信道校验：带宽有限时，快速确认大数据的一致性 //分布式存储、云存储
- 甄别赝伪：系统文件是否被病毒篡改？被何种病毒篡改？ //特征
- ...

散列指纹

❖ 方法：在文档中选取若干序列，通过散列生成指纹

❖ 简单的累加？极易冲突！

`I am Lord Voldemort` = `Tom Marvolo Riddle` = `He's Harry Potter`

❖ 问题：序列越多越好吗？

技巧：在足以刻画文档的前提下，如何选取更少的序列？

❖ 单向函数：$f()$可快速计算，但$f^{-1}()$的计算却十分...十分地耗时

❖ 整数乘法：$f(x, y) = x \cdot y$ //O(1)时间——RAM模型

$f^{-1}(x \cdot y) = \langle x, y \rangle$ //O(??)——质因数分解

MD5算法

❖ Message-Digest Algorithm version.5, 1991

MIT Lab for Computer Science & RSA Data Security Inc.

将信息统一视作比特流，每512 = 16×32 bit作为一个处理分组

经过四轮位运算变换，最终得到一个128 bit的大整数，即为MD5指纹

❖ 初始化： 如有必要，追加一个1及多个0，使总长形如512*k - 64

将原信息的长度附加到末尾，使总长形如512*k

❖ 128 bit，足够复杂？ //$2^{128} = 3.4 \times 10^{38}$

要重构（破译）符合指定指纹的原始信息，似乎不那么简单...

课后

- 了解MD5算法的更多细节 //google("MD5 algorithm")
- 学习MD5相关软件的使用 //google("MD5 tool")
- 验证：某个文件的MD5，可能与另一个文件的MD5雷同
- 思考：如果有人掌握了MD5冲突的原理，则意味着什么？
- 了解MD5安全性的最新结论 //google("MD5 'WANG Xiaoyun'")
- 提交作业包时，尝试同时提交MD5指纹
- 了解分布式散列表的

  原理、方法与 //Overlay Network, Distributed Hash Table

  实现 //Napster, Gnutella; Chord, CAN, Tapestry, Pastry, ...