**消息摘要算法第五版MD5**

1. 算法特点

消息摘要算法第五版(Message Digest Algorithm 5th, MD5)是一种常见的hash函数，该标准在RFC1321中被确立。和其他的hash算法一样，MD5被用于验证信息的**完整性**，防止信息被篡改。MD5算法具有以下特点：

①有损压缩。对于任意长度的数据得到的MD5长度都是固定的。

②容易计算。从原始数据计算出MD5很容易。

③**抗修改**性和**抗碰撞**性。这两点提供了MD5的安全性保证。对原始数据进行的微量修改都会导致最终计算出的MD5产生很大差别；即使已知原始数据和MD5值，想要找到一个具有相同的MD5的数据是非常困难的。

2. 算法应用

MD5主要有两个方面的应用：

第一个是完整性验证。例如从网站上下载软件，下载地址旁一般会附上其MD5或SHA-1, 这样用户在将软件下载到本机以后，计算出软件的MD5和SHA-1就可以和网站提供的数据进行比对，如果符合则说明数据是完好的，没有被篡改过。

第二个是密码保存。对于服务器而言，以明文保存密码是非常危险的举动（一旦服务器被攻破则所有用户的明文密码都会泄露，不仅本站沦陷，还可以造成更严重的撞库攻击），因此一般的策略是保存密码的hash值（如MD5值），在用户输入密码时，将密码的MD5和已经保存的MD5进行比对，如果符合则说明密码是正确的。

然而即使保存hash也并不是那么保险。注意到密码无非是数字、大小写字母和有限字符的组合，且字串长度一般都不会太长(<20), 由于MD5正向计算是简单的，因此我们有从MD5反推密码的**彩虹表**攻击技术：即迭代生成所有可能的密码并计算它们的MD5值（这一系列的键值对一般称为彩虹表），当我们需要寻找某个MD5所对应的明文密码时，只需要进入彩虹表查询即可。

3. 理论破解及其利用

MD5的安全性只是算法设计者的美好幻想. 2004年山东大学王小云加速了MD5碰撞（并且在2005年进一步搞定了SHA-1）使得寻找到具有相同的MD5的字节串在短时间（几个小时）内成为可能。

自此以后，国内外对MD5碰撞就开始了广泛的研究和利用。360注意到：**新型的碰撞样本**在2014年首次出现，当时还处于测试阶段，所以只有少数样本在传播。到了2015年初的时候样本大规模爆发，360认为这两波的样本是出于同一作者之手（并称之为“碰撞作者”）。由于安全软件的介入，2015年9月开始，碰撞作者开始结合**数字签名利用技术**与安全软件进行对抗，到了11月开始利用**双签名**对抗查杀。流程表示如下：

