MD5 的彩虹表攻击

1. 实验概述

彩虹表是一种典型的时间空间折中算法，主要用于对杂凑值的搜索，他由 Martin E. Hellman 提出。在彩虹表出现前，已经出现了称为”预计算的哈希链集“的杂凑值破解方法。彩虹表并不是破解杂凑函数的单向性，而是使用约减函数构造杂凑值到消息的映射，最终创建一个逆向映射表格，用来查找所需的明文。比如，一个系统采用用户名、密码的方式进行登录，密码由 10 位数字组成。密码数据库中存放用户名和直接调用杂凑算法进行压缩密码对应的杂凑值。该系统的密码杂凑值空间为 10^{10}（大约 2^{34}），可以使用现代计算机进行暴力破解。彩虹表可以更快速的破解该系统，对消息空间很大的密码系统威胁很大。目前，主流的彩虹表都大小都在 100G 以上，保存的消息空间非常大。彩虹表常用于恢复由有限集字符组成的固定长度的纯文本密码。彩虹表对加盐的杂凑函数攻击效果不理想，原因在于盐值使相同密码的杂凑值不同。

实验目的：通过实验掌握彩虹表的构建流程。能够使用 C 语言实现基于 MD5 的彩虹表，并能用实现的彩虹表恢复 10 位纯数字密码。

2. MD5 的彩虹表攻击

计算 H = MD5(m)对应的消息 m，朴素的方法是穷举全部消息，搜索符合要求的 m。时间成本太高。还可以采用字典攻击，存储大量的(H, m)对，空间成本又很高。为了平衡时间和空间成本，出现了彩虹表攻击。构建彩虹表的重点是设计有效的约减函数 R，约减函数 R 是从杂凑值到消息的映射，对于单个 R 函数的彩虹表结构如图 1 所示。

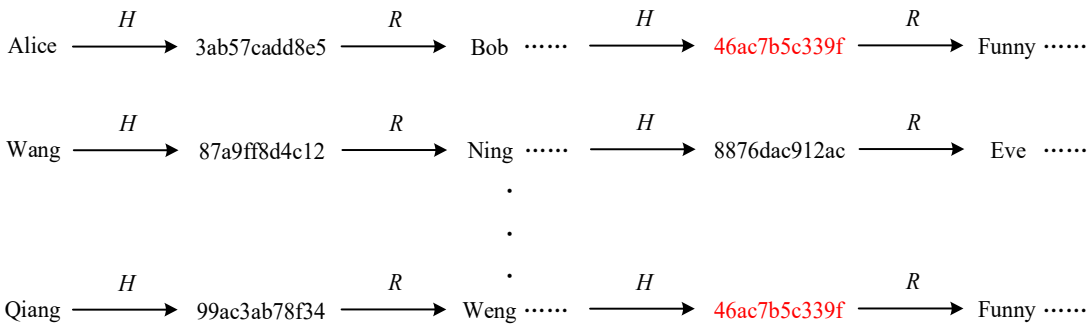

图 1 单个 R 函数的彩虹表

如果两个链出现相同的节点，那么后续全部节点都相同。为了提高彩虹表的存储效率，常采用多个 R 函数，减少相同节点对不同链的影响。多个 R 函数的彩虹表结构如图 2 所示。

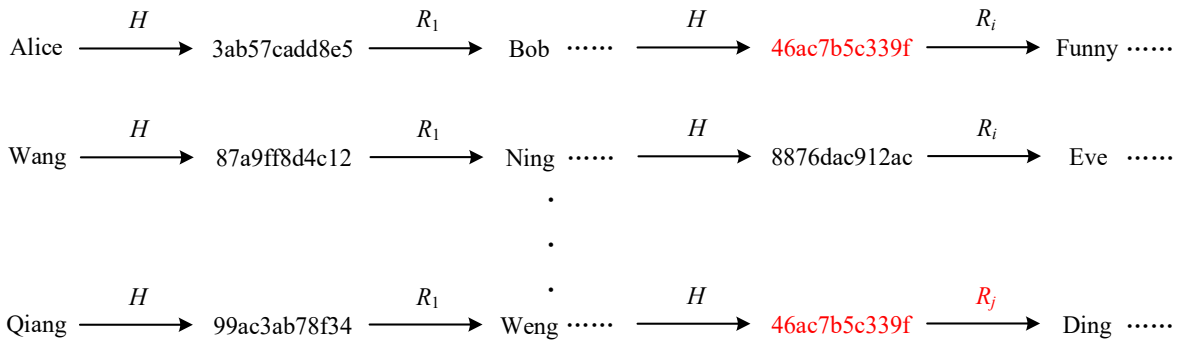

图 2 多个 R 函数的彩虹表

不难从图 2 得出如下结论：$R_i \neq R_j$ 的情况下，即便杂凑值 46ac7b5c339f 相同。这样，即使彩虹表中有相同的杂凑值出现，只要后续约减函数 $R_i \neq R_j$，即节点所处的列不同，后续结果也不同，最大限度降低了重复节点的出现。

彩虹表的每一条链只存储首节点和终节点。假定图 2 的第一条链的最后一个 R 函数执行后的消息是 Wangming，图 2 的第一条链的存储内容为(Alice, Wangming)。假定彩虹表共有 x 条链，每条链执行 y 个不同的 R 函数，此时，彩虹表中共有 $x \times y$ 个节点，只序存储 $2x$ 个表项。彩虹表的构造方法可以描述为：

1. 随机选取消息 m_1，依次计算 $H_1^1 = \mathrm{MD5}(m_1)$，再计算 $m_2 = R_1(H_1^1)$，依次计算直到 m_y，将(m_1, m_y)存到表中。
2. 依次随机选取消息 $m_1^i (1 \leq i \leq x)$，并使 m_1^i 与前边节点均不相同，使用步骤 1 计算并存储(m_1^i, m_y^i)。

彩虹表的查找方法可以描述为：

1. 对给定的 H，首先调用约减函数 R_{y-1} 得到 m_y，搜索 m_y 是否出现在表中。
2. 如果 m_y 第 i 项匹配，则密码存在于第 i 条链中。用 m_1^i 依次计算找到相等的杂凑值，此时前 1 个消息即为符合要求的密钥；

3. 否则，依次调用约减函数 R_{y-2}、R_{y-1} 得到 m_y，搜索彩虹表，查找是否有表项与匹配 m_y；
4. 依次向前进行，直到找到表项与之匹配，输出对应的密码；
5. 否则，搜索失败。

3. 实验步骤

1、登录到虚拟机，密码为 osr；

2、进入桌面 RainbowAttack 文件夹，双击 RainbowAttack.c 文件可查看彩虹表攻击的部分代码。MD5 算法及其他必要算法已给出，可以使用相应的文件调用该算法。

3、请根据彩虹表攻击原理和参考文献 1，补充完整彩虹表的构建函数 void CreateRainbow(unsigned char ** rainbow)和搜索函数 void SearchRainbow(unsigned char * pwd, unsigned char ** rainbow)。可以添加其他必要的代码。**注释一：可以使用单个 *R* 函数构造彩虹表，在实验分析部分增加多个 *R* 函数的对比结果。本实验使用的彩虹表共有 1000 条链(编号从 1 到 1000)，每条链的初始密码为链编号对应的字符串，如第 990 条链的初始密码为字符串“0000000990”。每条链包含 10000 个节点(编号从 1 到 10000)，不包含该条链的初始密码，*R* 函数已经给出。**

4、根据实现的彩虹表恢复 bb0dd49137b4087b8efce1372de635d1 对应的密码。

5、在终端输入编译命令“gcc -w RainbowAttack.c -o RainbowAttack”，将源文件 RainbowAttack.c 编译成一个可执行文件 RainbowAttack，完成实验，并根据要求提交实验报告。

4. 思考及参考文献

4.1 思考

请设计不同的 *R* 函数对攻击速度进行对比和分析?编写程序判断本实验构造的彩虹表中是否存在相同的节点，如果存在请给出你认为较好的改进方案。

4.2 参考文献

[1] Martin E. Hellman. A cryptanalytic time-memory trade-off. IEEE Trans. Inf. Theory, 26(4): 401-406 (1980).

5. 实验结果分析及提交

5.1 实验结果分析

请根据彩虹表的存储大小，分析彩虹表大小对攻击成功率和速度的关系。

5.2 实验结果提交

已知杂凑值为 bb0dd49137b4087b8efce1372de635d1，请基于本实验的示例代码编写程序，将恢复的 10 位数字密码后连接上该密码对应的初始值作为本实验的 Flag（加入密码为 0000000000，初始值为 9999999999，则 flag 为 00000000009999999999）填写到实训教学系统中。同时，学员须将整个实验过程和全部实现代码以报告的形式记录，并上传提交到实训教学系统中。实验报告还可以包含思考和实验结果分析等。