彩虹表原理详解及工具介绍

（https://blog.csdn.net/gscaiyucheng/article/details/17113073）

彩虹表（Rainbow Table）是一种破解哈希算法的技术，是一款跨平台密码破解器，主要可以破解 MD5、HASH 等多种密码。它的性能非常让人震惊，在一台普通 PC 上辅以 NVidia CUDA 技术，对于 NTLM 算法可以达到最高每秒 103,820,000,000 次明文尝试（超过一千亿次），对于广泛使用的 MD5 也接近一千亿次。更神奇的是，彩虹表技术并非针对某种哈希算法的漏洞进行攻击，而是类似暴力破解，对于任何哈希算法都有效。

一、彩虹表原理

先讲述一些基本概念：

Tables

可以说长期进行密码学研究的人很少有不知道这个的。在很多年前，国外的黑客们就发现单纯地通过导入字典，采用和目标同等算法破解，其速度其实是非常缓慢的，就效率而言根本不能满足实战需要。之后通过大量的尝试和总结，黑客们发现如果能够实现直接建立出一个数据文件，里面事先记录了采用和目标采用同样算法计算后生成的 Hash 散列数值，在需要破解的时候直接调用这样的文件进行比对，破解效率就可以大幅度地，甚至成百近千近万倍地提高，这样事先构造的 Hash 散列数据文件在安全界被称之为 Table 表(文件)。

Rainbow Tables
最出名的 Tables 是 Rainbow Tables，即安全界中常提及的彩虹表，它是以 Windows 的用户帐户 LM/NTLM 散列为主要破解对象的。简单说明一下，在 Windows2000/XP/2003 系统下，账户密码并不是明文保存的，而是通过微软所定义的算法，保存为一种无法直接识别的文件，即通常所说的 SAM 文件，这个文件在系统工作时因为被调用所以不能够被直接破解。但我们可以将其以 Hash 即散列的方式提取，以方便导入到专业工具破解，提取出来的密码散列类似于下面：

```
Administrator:500:96e95ed6bad37454aad3b435b51404ee:64e2d1e9b06cb8c8b05e42f0e6605c74:::
Guest:501:aad3b435b51404eeaad3b435b51404ee:31d6cfe0d16ae931b73c59d7e0c089c0:::
user1:1001:732b2c9a2934e481cd0a8808b19097ef:778620d5d5de064154e689fa4790129f:::
user2:1002:a042f67a99758fd727b99b2375d829f9:6127ee12a83da34fc19953e538e4d580:::
```

若是以传统破解方式而言，无论是本地，还是内网在线破解，效率都不是很高。据

实际测试，单机环境下，破解一个 14 位长包含大小写字母以及数字的无规律密码，一般是需要 3~~9 小时的，这个时间值会随着密码的复杂度及计算机性能差异提升到几天甚至数月不等。虽然说大部分人都不会使用这样复杂的密码，但对于目前很多密码足够复杂并且长度超过 10 位的密码比如“Y1a9n7g9z0h7e”，还是会令黑客们头痛不已。

2003 年 7 月瑞士洛桑联邦技术学院 Philippe Oechslin 公布了一些实验结果，他及其所属的安全及密码学实验室(LASEC)采用了时间内存替换的方法，使得密码破解的效率大大提高。作为一个例子，他们将一个常用操作系统的密码破解速度由 1 分 41 秒，提升到 13.6 秒。这一方法使用了大型查找表对加密的密码和由人输入的文本进行匹配，从而加速了解密所需要的计算。这种被称作“内存-时间平衡”的方法意味着使用大量内存的黑客能够减少破解密码所需要的时间。

于是，一些受到启发的黑客们事先制作出包含几乎所有可能密码的字典，然后再将其全部转换成 NTLM Hash 文件，这样，在实际破解的时候，就不需要再进行密码与 Hash 之间的转换，直接就可以通过文件中的 Hash 散列比对来破解 Windows 帐户密码，节省了大量的系统资源，使得效率能够大幅度提升。当然，这只是简单的表述，采用的这个方法在国际上就被称为 Time-Memory Trade-Off ，即刚才所说的“内存-时间平衡”法，有的地方也会翻译成“时间—内存交替运算法”。其原理可以理解为以内存换取时间，可由下图表示：

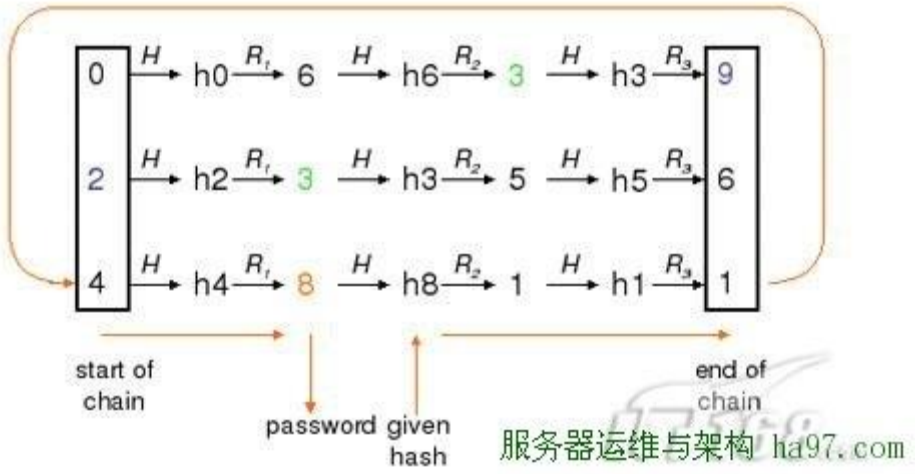

图 著名的“内存-时间平衡”法运算原理图

具体算法方面的内容本文就不再涉及，对于想进行更高深层次探究的读者，可以仔细参考 2003 年的这篇详细文档《Making a Faster Crytanalytical Time-Memory Trade-Off》以及 2005 年的文档《Time-Memory Trade-Offs: False Alarm Detection Using Checkpoints》。

正是由于 Rainbow Tables 的存在，使得普通电脑在 5 分钟内破解 14 位长足够复杂的 Windows 帐户密码成为可能。

ID	USERNAME/LMHASH	LMpasswd1	LMpasswd2	NTpasswd
1001	user1	MZXWSWS	WWS	mzxwswswws
1010	user10	CJCHNWS	CQWSBSY	cjchnwscqwsbsy
1011	user11	HAHA123	.....	Not found
1012	user12	CJCHNWS	.....	Not found
1013	user13	1984031	2	19840312
1014	user14	/EMPTY/		Not found
1015	user15	/EMPTY/		Not found
1016	user16	/EMPTY/		Not found
1017	user17	/EMPTY/		Not found
1018	user18	LIURUXI	NSHIZHU	liuruxinshizhu
1019	user19	ILOVETH	ISDAY	ilovethisday
1002	user2	CJCHNWS	7843	cjchnws7843
1020	user20	.....	.....	Not found
1003	user3	C78J33C	6HNWS	c78j33c6hnws
1004	user4	CKLASDF	HJCUKLH	cklasdfhjcuklh
1005	user5	2178724	3040236	21787243040236
1006	user6	3891177	0	38911770
1007	user7	TESTYEA	RBOY	testyearboy
1008	user8	HAPPYPI	G123	happypig
1009	user9	YEMAWAN	GLUO178	yemawangluo178

图 对 Windows 账户进行 Rainbow Tables 破解

如上图可以看到，类似于 c78j33c6hnws、yemawangluo178、38911770 这样的 Windows 帐户密码几乎全部在 180 秒即 3 分钟内破出，最短只用了 5 秒，个别稍长的密码破解开也没有超过 3 分钟。

这几乎是令人难以置信的，Roger 迫不及待的去看了 http://www.project-rainbowcrack.com 所介绍的原理。这其实已经不是新的技术了，但是很遗憾的是，搜索“彩虹表原理”出来的文章对彩虹表原理的介绍都有不太正确，Roger 就在这里简单的介绍一下，主要参考的是 Wiki 上的这篇 http://en.wikipedia.org/wiki/Rainbow_tables，英文好的可以去看这篇论文 http://lasecwww.epfl.ch/pub/lasec/doc/Oech03.pdf。

我们先来做点科普，哈希（Hash）算法就是单向散列算法，它把某个较大的集合 P 映射到另一个较小的集合 Q 中，假如这个算法叫 H，那么就有 Q = H（P）。对于 P 中任何一个值 p 都有唯一确定的 q 与之对应，但是一个 q 可以对应多个 p。作为一个有用的 Hash 算法，H 还应该满足：H(p)速度比较快； 给出一个 q，很难算出一个 p 满足 q = H(p)；给出一个 p1，很难算出一个不等于 p1 的 p2 使得 H(p1)=H(p2)。正因为有这样的特性，Hash 算法经常被用来保存密码————这样

不会泄露密码明文，又可以校验输入的密码是否正确。常用的 Hash 算法有 MD5、SHA1 等。

破解 Hash 的任务就是，对于给出的一个 q，反算出一个 p 来满足 q = H(p)。通常我们能想到的两种办法，一种就是暴力破解法，把 P 中的每一个 p 都算一下 H(p)，直到结果等于 q；另一种办法是查表法，搞一个很大的数据 库，把每个 p 和对应的 q 都记录下来，按 q 做一下索引，到时候查一下就知道了。这两种办法理论上都是可以的，但是前一种可能需要海量的时间，后一种需要海量 的存储空间，以至于以目前的人类资源无法实现。

我们可以简单的算一下，对于 14 位的大小写加数字（先不算特殊字符了）组成的密码的集合有多大？自然就是（26*2+10）^14 = 62^14 = 1.24 * 10^25，这个就约等于 12 亿亿亿，即使我们每纳秒可以校验一个 p(一秒钟 10 亿次，目前 PC 做不到)，暴力破解法也大概需要 4 亿年；如果我们采用查表 法，假定 Hash 的结果是 128Bit 即 16 字节的，光存放 Hash（不存放明文 P）就需要 10^26 字节的存储空间。什么？现在硬盘很便宜？没错现在 1GB 硬盘大概是五毛钱，那么按这个来算光存储这个 Hash 大概需要 5 亿亿人民币来买硬盘。所以有些文章说彩虹表就是依赖查一个巨大的表来破解 Hash， 简直是个无知的玩笑。

也正因为如此，我们一直都认为 Hash 是足够安全的，十几位的密码也是强度足够的，直到彩虹表的出现。现在我们来看看彩虹表是怎么干的。

彩虹表的根本原理就是组合了暴力法和查表法，并在这两者之中取得一个折中，用我们可以承受的时间和存储空间进行破解。它的做法是，对于一个 Q = H（P），建立另一个算法 R 使得 P = R(Q)，然后对于一个 p，这样进行计算：

p0 -H-> q1 -R->p1 -H-> q2 -R->p2 -H-> q3 -R->p3 ... -H-> q(n-1) -R->p(n-1) -H-> qn -R->pn

简单的说，就是把 q 用 H、R 依次迭代运算，最后得到 pn，n 可能比较大。最后我们把 p0 和 pn 都存储下来，把其他的结果都丢弃。然后用不同的 p0 代入计算，得到多个这样的 p 的对子。

我们在做破解的时候，给出了一个 q，我们来寻找 p。我们先把 q 做一次 R 运算得到一个值例如叫 c1，然后把 c1 和每一个 p 对的最后一个做比较，假如和某一个 pn 相等，那么有可能这个 pn 所对应的 p(n-1)就是我们在追寻的 q，为了验证我们把 pn 对应的 p0 再做一次链式计算，比对 qn 是否就是给出的 q，如果 是，很明显 p(n-1)就是我们在追寻的 p，因为 p(n-1) -H-> qn。如果不是就继续寻找直到遍历所有的 q0qn 对。

事情还刚刚开始，我们再算 q -R-> c1 -H-> -R-> c2，再比对 c2 是否是 qn，如果是，那么 p(n-2)就可能是 p；再算 c3、c4 直到 c（n-1），不知道这样说你明白了吗？

总的来说，就是用一个 popn 对来存储了一个链子的数据，如果 n 很大，就可以大大减小了存储的空间。这样带来的问题是必须做 n 次比对，时间更长，但是我们不需要瞬间破解，等待几秒乃至几天破解一个密码都是可以接受的。

当然这里只是讲述了最粗浅的原理，仔细想一下还有很多的问题，例如 R 的选择，Hash 冲突的处理，如何选择 po 来实现足够的覆盖，如何在有限资源下生成彩虹表等等。对这些感兴趣的可以去看看 RainbowCrack 的源码 http://www.project-rainbowcrack.com

二、获得彩虹表

彩虹表可以使用 RainbowCrack 或 Cain 来生成。表分割得越细，成功率就越高，生成的表体积也越大，所需时间也越长。但下载比生成快得多，有人做过测试，4 核 4GB 内存的机器，生成 2GB 彩虹表，需要花费 7 天时间，而 7 天按 2MB 的带宽（256K/S 左右）几乎可以下载 48GB 左右，效率明显要超过生成。当然，你要是有超级计算机群生成的话，也不妨自己生成。对于广大网络安全爱好者来说，还是直接下载来得靠谱！

Ophcrack 彩虹表 官方下载地址：
http://ophcrack.sourceforge.net/
120G 彩虹表 BT 下载（这是种子文件，迅雷上有资源，如果是会员使用迅雷下载还是很快的，我 8M 带宽，下了 3 天左右下完了。）：
http://www.ha97.com/code/tables.rar

三、彩虹表的使用

彩虹表工具很多，常用到的彩虹表工具有 Ophcrack、rcracki_mt、Cain、RainbowCrack 等，主流的彩虹表有以下四种。

Cain: http://www.onlinedown.net/soft/53494.htm
Free Rainbow Tables
官方网址：http://www.freerainbowtables.com/en/tables/
镜像下载：http://tbhost.eu/rt.php
提供了多种类型的彩虹表下载，LM、NTLM、MD5、SHA1 等。千万别把人家法语字符的表也下了，对国人来说，几乎没什么用，不过如果你有特殊需要，那就下

吧……这里提供的都是.rti 格式的，有别于传统的.ri 格式，.rti 比.rt 的多了一个目录.index 文件，据说遍列速度比.rt 的更快（未曾对比过，无法确定是否属实）。比较新的，用的索引和压缩，所以速度更快，体积更小，而且支持分布式破解。
支持 HASH 类型：LM，MD5，NTLM，SHA1，HALFLMCHALL
网上有已经生成好的表可供下载，真是造福于民。
扩展名：rti

Ophcrack
官网网址：http://ophcrack.sourceforge.net/tables.php
最常用的，界面友好，与众不同，压缩储存，有自己独特的彩虹表结构，还有 Live CD。
支持的 HASH 类型：LM，NTLM
扩展名：乱七八糟的。
高级的表要花钱买，免费的表有（推荐只下 2 和 5，要求高的可以下载 3 和 5）：
1.XP free（LM 表：包含大小写+数字）380MB（官网免费下载）
2.XP free fast（和前一个一样，但是速度更快）703MB（官网免费下载）
3.XP special（LM 表：大小写+数字+所有符号包括空格）7.5G
4.Vista free （NTLM 表：包含常用密码）461MB（官网免费下载）
5.Vista special（NTLM 表：包含 6 位的全部可打印字符，7 位的大小写字母数字，8 位的小写和数字）8G

RainbowCrack
官网网址：http://project-rainbowcrack.com/table.htm
通用的，一般的破解软件如 saminside 都支持，命令行界面，黑客的最爱，支持 CUDA。
可以自己生成表，不要钱，传说中的 120G 就来自于此。
支持 HASH 类型：LM, NTLM, MD5, SHA1, MYSQLSHA1, HALFLMCHALL, NTLMCHALL.
扩展名：rt

最小彩虹表是最基本的字母数字表，就这样它的大小就有 388MB，这是 Ophcrack 启动盘默认的表，该表可以在 11 分钟内破解所有可能 14 位数字字母密码组合中的 99.9%。国内有比较流行的传说中的 120G 的彩虹表，国外还有几 T 的海量彩虹表。win2003 及以前的 windows 操作系统的密码采用的 LM 算法加密，而 Vista、Win7、Win2008/R2 采用的是 NTLM，NTLM 比 LM 安全得多。

LM 和 NTLM 详解：

1、话说在远古时期，DES 当道。微软在考虑 9X 系统口令加密的时候就自然地采用了国家标准 DES 一次加密 8 字节，留一位校检，还剩 7 字节（下文有解释），也就是 LM（Lan Manage）的核心。

2、那有人要问了，万一我的口令是 8 位怎么办呢？不用怕，微软的程序员很“聪明”：先把前 7 位加密，后一位补 6 个 0，再当 7 位一起加密不就可以了吗，结果就真的这么做了。

3、这就导致破解 LM 密码只需 7 位一分割，然后再逐块破解，这大大减低了破解的难度。因为最后一块往往不够 7 位，一般瞬间即可得出结果。也就是 7 位和 13 位的密码，在破解者眼里几乎是一样的，因为 13 位的后 6 位很快就能破解出来，而且可以根据后 6 位猜测出前 7 位的密码，这就是为什么我们破解 XP 和 2003 密码很快的原因，因为他们都使用了 LM 加密方式。

4、由于 LM 的种种不安全性，微软在设计 NT 系列操作系统时采用了新的口令存储手段，即 NTLM 技术（New Technology Lan Manage），采用 MD4+RSA 存储，立马安全性要高很多。但是为了保证兼容性，直到 2003 微软仍然保持着 LM 的加密方式，也就是在 2000、2003 和 XP 中，我们的口令同时保存了两份，一份 LM 一份 NTLM，我们仍然可以通过 LM 破解 2003 的口令。

5、在 Vista 和 2008、Win7 中，微软终于下定决心对 LM 斩草除根，只留下 NTLM，破解难度增大。

6、回到彩虹表，由于 LM 最多只有 7 位，所以它的彩虹表很小。而 NTLM 用了散列函数，所以理论上口令是可以无限长的，所以 NTLM 的彩虹表往往很大，120G 肯定是不够完成所有可打印字符的，最大的彩虹表已经是 T 量级了。

LM 和 NTLM 验证机制：
http://www.nsfocus.net/index.php?act=magazine&do=view&mid=1665

某人用彩虹表测试破解 MD5 的小结：

ophcrack 的表不支持破解 md5，具体讲.rt .rtc .rti 格式的，只需对比一组数据就可以。同样是破解 12 位的纯数字密码：

.rt 的需要 20GB .rtc 的需要 8.75GB .rti 的需要 1.67+1.67+1.68+1.71=6.72GB

明显是.rti 的小，但是我试过，我下了上面.rti 格式破解 12 位的 6.72GB 的表中的 1.67GB，其破解效果很让我惊讶，我本以为纯数字的破解出来的可能性是四分之一，因为我只下了 4 个表中的一个，我只下了那 1.67GB，但我试着破解了几个 12 位数字加密的 32 位 md5，结果大多数都能跑出来，很少有跑不出的，汗。但当我惊喜时发现他并不支持破解 16 位的 md5，然后去那国外的官方论坛去逛了逛，才发现这并不支持破解 16 位的 md5。原来老外不来 16 位这一套，但我们国内的网

站用 16 位的 md5 占绝大多数，所以入侵时大部分得到的是 16 位的 MD5 密码，而老外的就不来 16 位的，郁闷。

Ophcrack 文档描述了它所能使用的彩虹表之间的差异：
字母数字表 10k 388MB 包含所有字母数字混合密码中 99.9%的 LanManager 表。这些都是用大小写字母和数字组成的密码（大约 800 亿组合）。
由于 LanManager 哈希表将密码截成每份 7 个字符的两份，我们就可以用该表破解长度在 1 到 14 之间的密码。由于 LanManager 哈希表也是不区分大小写的，该表中的 800 亿的组合就相当于 12*10 的 11 次方（或者 2 的 83 次方）个密码。

字母数字表 5k 720MB 包含所有字母数字组合的密码中 99.9%的 LanManager 表。但是，由于表变成 2 倍大，如果你的计算机有 1GB 以上的 RAM 空间的话，它的破解速度是前一个的 4 倍。

扩展表 7.5GB --**xp special** 包含最长 14 个大小写字母(密码大于 14 LM-HASH 会全显示 0 或以 AA3D 开头，详见另一篇文章 Windows LM/NTLM HASH 加密)、数字以及下列 33 个特殊字符（!”#$%&’()*+,-./:;?@[]^_`{|} ~）组成的密码中 96%的 LanManager 表。该表中大约有 7 兆的组合，5*10 的 12 次方（或者 2 的 92 次方）密码。

NT 8.5 GB--**vista special** 我们可以使用该表来破解计算机上的 NT 哈希表，这是 LanManager 哈希表所做不到的。该表包含了用如下字符组成的可能密码组合的 90%：
·最高 6 位字符由大小写字母、数字以及 33 个特殊字符（同上面列举的一样）
·7 大小写字母及数字
·8 小写字母及数字
该表包含 7 兆种组合，对应 7 兆的密码（NT 哈希表不存在 LanManager 哈希表的弱点）。

注意：所有这些彩虹表都有其特定适用的密码长度和字母组合。太长的密码（如数十位），或者包含表中没有的字符，那么用彩虹表就无法破解。