简要分析RockYou泄漏的密码

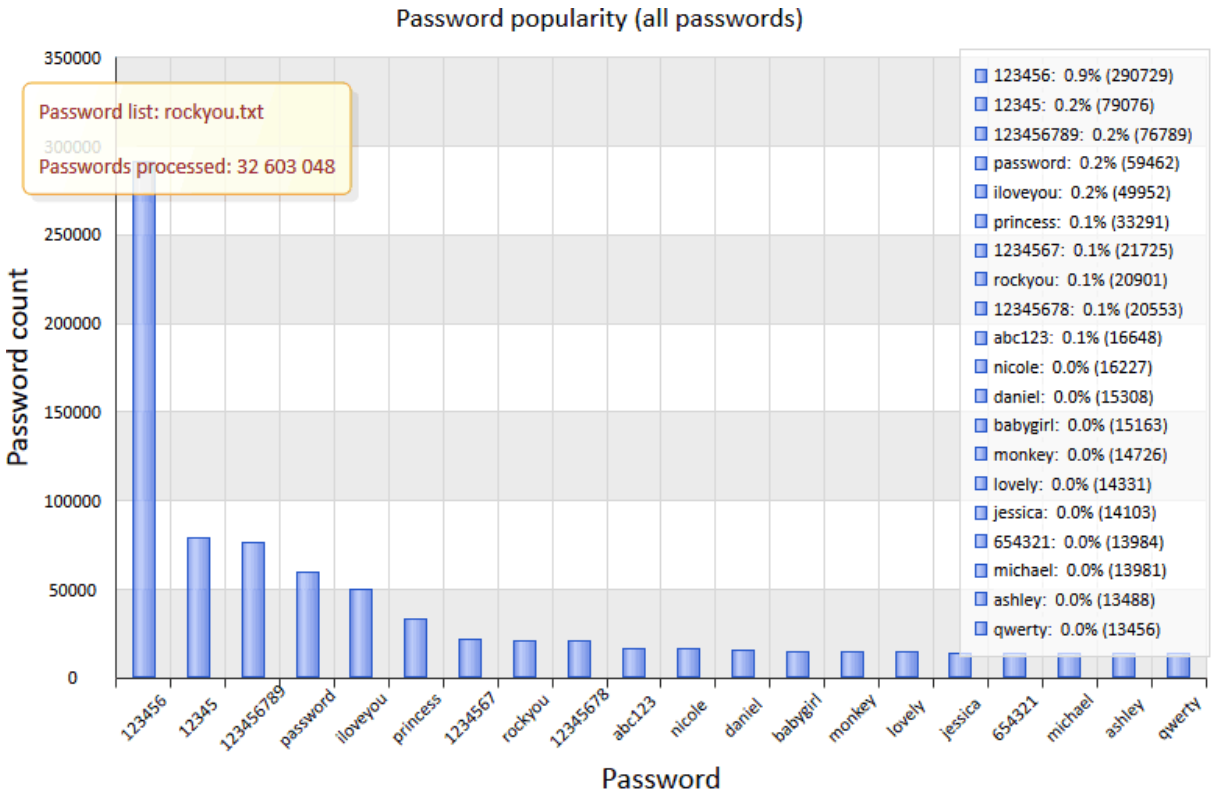
Passcape Software

1 简要分析RockYou泄漏的密码

1.1 摘要

2009年12月，RockYou.com网站遭到了黑客的攻击。这次攻击是成功的，并导致了一个文本数据库的非法披露，其中包含了3260万被盗的网站用户的密码。这一重大机密数据的泄露使我们得以观察人们是如何创建他们的密码的，并从外部观察者的角度做一个非常简单的安全审计。为了收集和处理统计数据，我们使用了 Windows Password Recovery tool.

1.2 前20个流行的密码

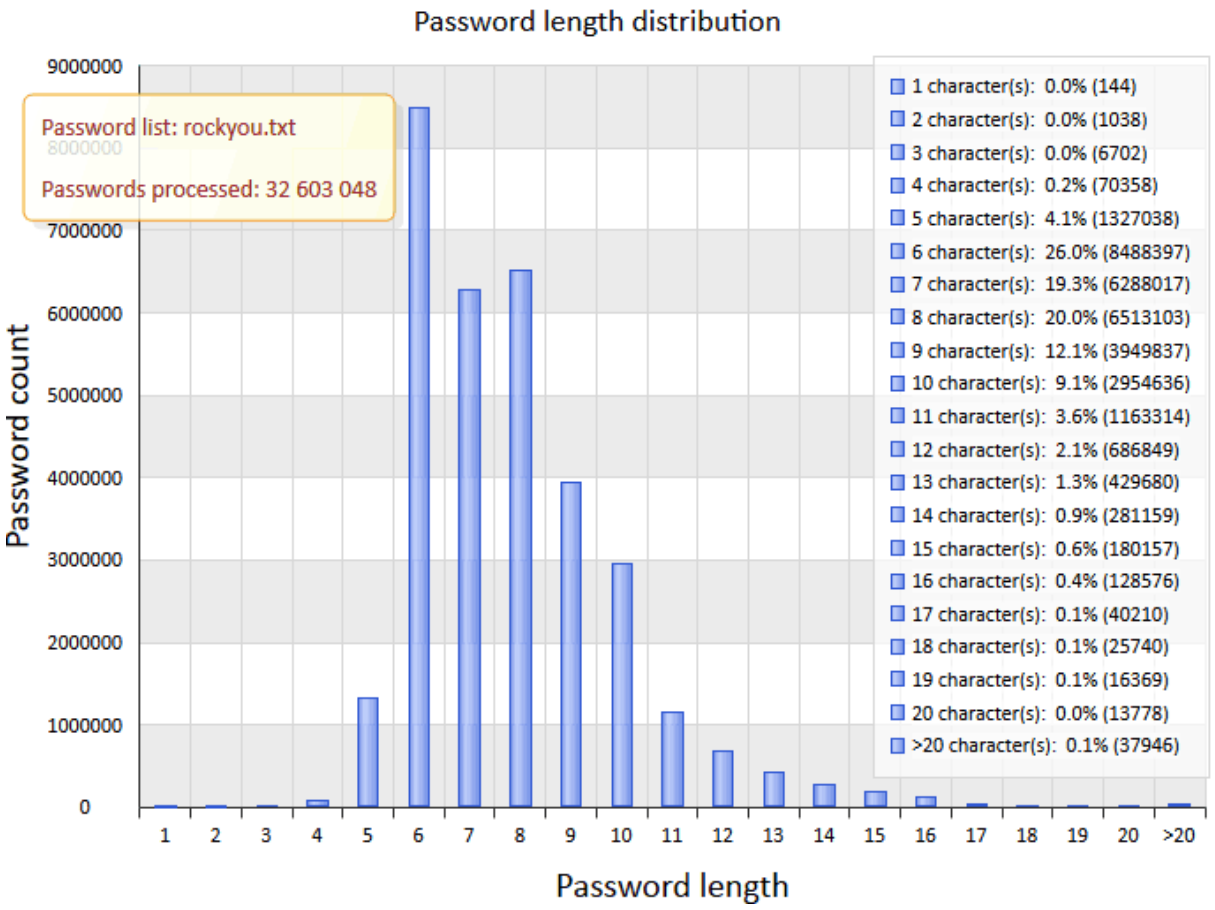


在最受欢迎的密码中很难找到比 "123 "更复杂的密码，所以前20名一点都不奇怪。正如你所看到的，最受欢迎的密码("123456")远远领先于其他密码。它的受欢迎程度甚至在数字上也是惊人的。在3200万条记录中，有29万条。我们的一位同事建议，如果一个数据库中10个最流行的密码占到了1%以上的记录，就应该被认为是不安全的。在RockYou.com的数据库中，这个 "不安全因素 "被超过了2倍。在分析流行的密码时，我们把它们按共同属性分为几组：

- 字典密码，也就是像 "密码"、"猴子 "之类的词，是最稳定的一组。
- 基于容易记忆的数字组合、电话号码、文件号码、出生日期等的数字密码构成了另一个群体，它和前一个群体一样稳定，甚至可能更受欢迎。
- 基于名字及其衍生物的密码。例如，一个用户可以使用他自己的名字，他的宠物，一些城市，一些地方，等等。
- 基于键盘组合的密码，如 "abc123"，"qwerty "等。

- 情感密码，如 "iloveyou"、"hatu"、"lovely"、"ihatemyboss "或 "ILoveJohn"。

1.3 密码长度分布



一个密码的长度是其抵抗攻击的重要因素。你可以说密码越长，就越难破解。但是别忘了，如果用户在一个热闹的聚会上喝了很多酒之后，"成功　"地忘记了长密码，那么这种不稳定的平衡就很容易失去。有些用户把写有密码的便条贴在显示器上，但大多数人更不讲究，他们更喜欢最多只有7或8个字符的容易记住的密码。

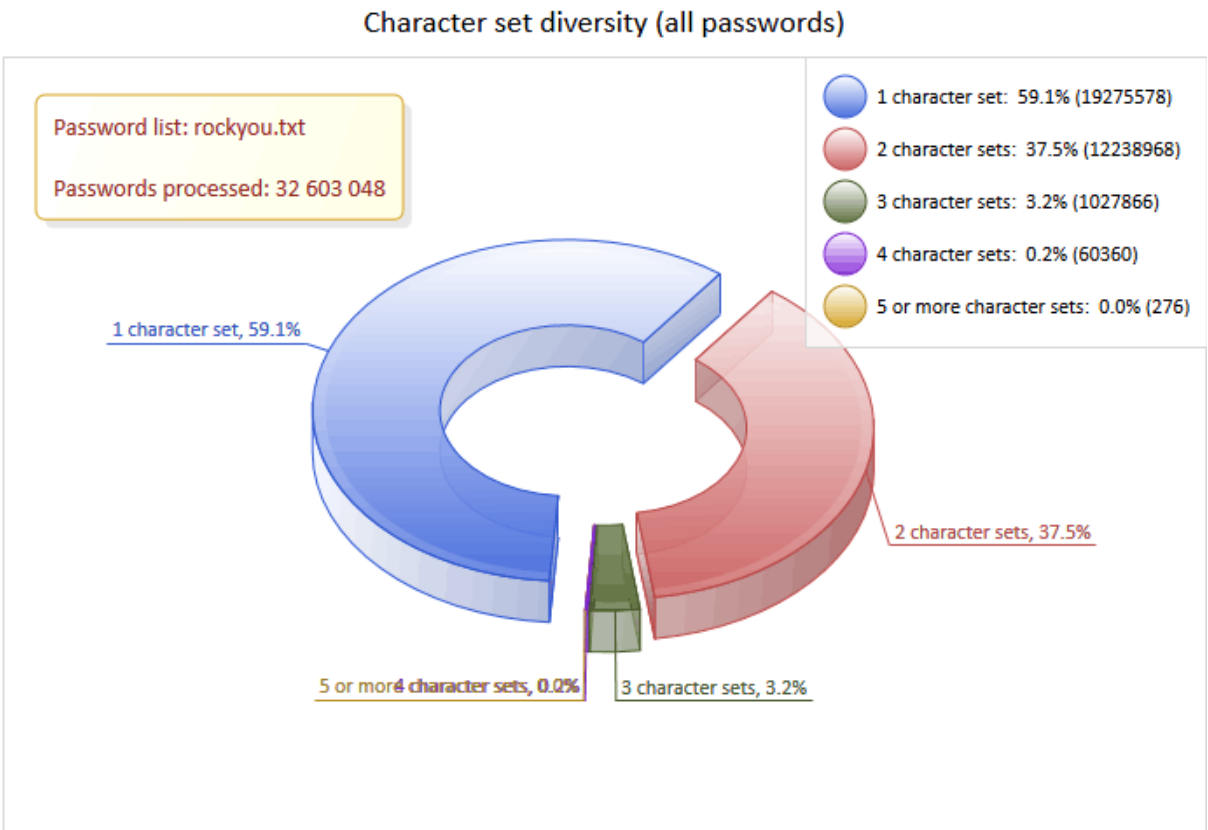
正如你在图表上看到的，最受欢迎的密码是6、8或7个字符的长度。这占所有密码的65%以上。这意味着，3个密码中有2个可以轻易被破解。

奇怪的是，有些用户选择了由20多个字符、单词组合或短语组成的淫秽的长密码。以下是其中一些奇妙的密码（你真的以为没有人知道你在输入什么吗？）：

Hahaithinkilovejessebutthenagainmaybenotcuzheisadiknob
Lets you come back for your Countdown Timer
me plus food equals more sleep each night
tommmmmmmmmmmmmmmmmmmmmmmmmmm

truongcaodangcongdonghaiphong
icantbelievethisshit.12345
banditbanditbandit1bandit1bandit1banditbandit
11111111111111111111111111111111111111111111
aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
FUCKSCHOOLANDALLTHETEACHERSINIT
ilovepalmermyfuturehusband
1delightyourselfinthelord!
Imaprincessbecausemyfatheristheking
iluvanjabisset4evashesmawebaexxx
stuartandchrisrmybestmatesforeva
thisismypasswordyoullnevergetit

1.4 字符集的多样性

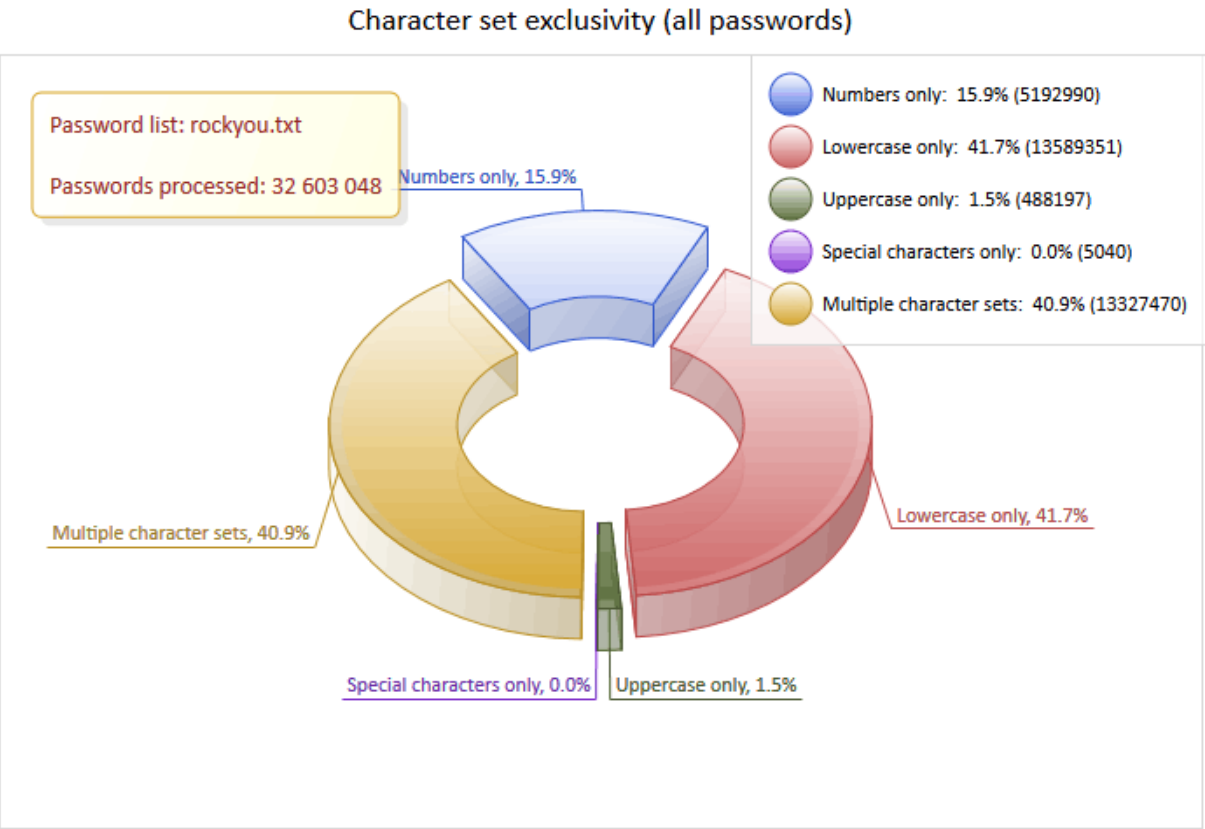


影响密码强度的另一个重要因素是字符的多样性。例如，7个字符的密码 "aB1~cde "的强度是9个字符的密码 "abcdefghi "的10倍以上。

超过一半的用户在他们的密码中只使用一种字符集。例如，他们只使用数字或只使用小写字母。超过96.5%的用户只有一个或两个字符集，只有不到3.5%的用户使用三个或更多的字符集。密码长度和

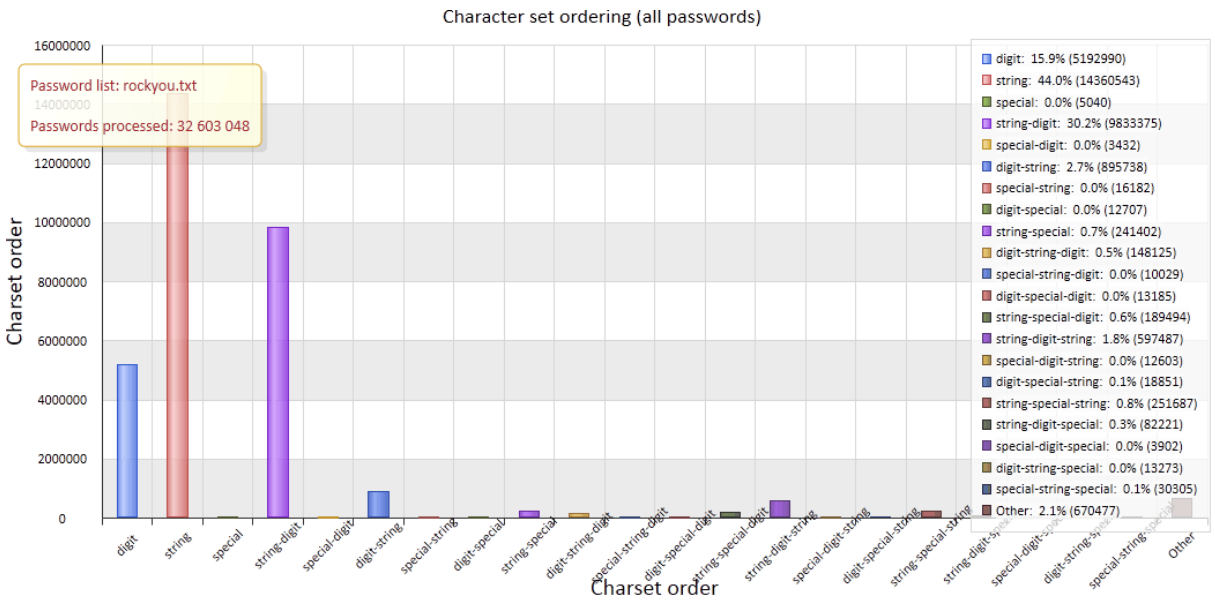
多样性的图表清楚地表明，用户在选择密码时的偏好在过去20年里几乎没有变化。我们的结论。我们成功地揭穿了 "你可以强迫用户使用强密码 "的神话（至少，在这种情况下）。

1.5 选择字符集时的偏好



一般的统计数据显示，在根据一个字符集选择密码时，用户大多喜欢只由小写字母组成的单词（占所有密码的41%以上），只由数字组成的单词（近16%），或只由大写字母组成的单词（1.5%）。仅由特殊字符组成的密码则很少使用。

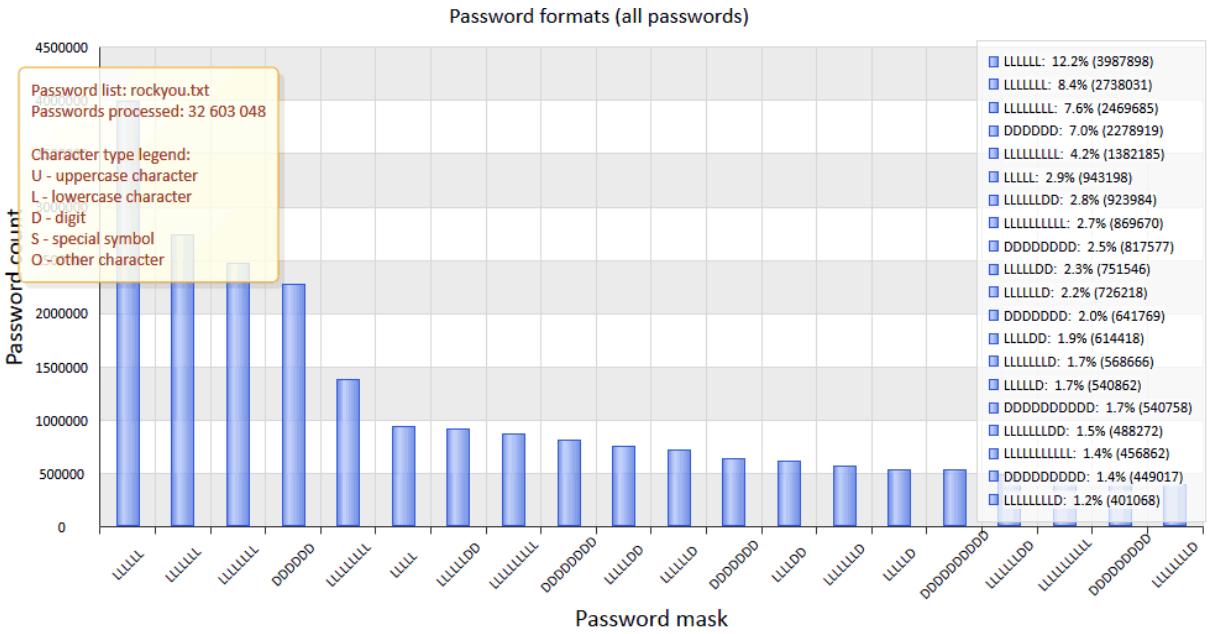
1.6 密码中的字符集顺序



如果你看一下这个图表，你可能会对以下事实感到惊讶：绝大多数由两个或多个字符集组成的密码都是字符串-数字或数字-字符串的组合。熟练掌握社会工程的攻击者可以利用这一观察来破解这类密码。

我们的结论是。"iloveyou12345 "的密码和 "iloveyou "的密码一样弱。

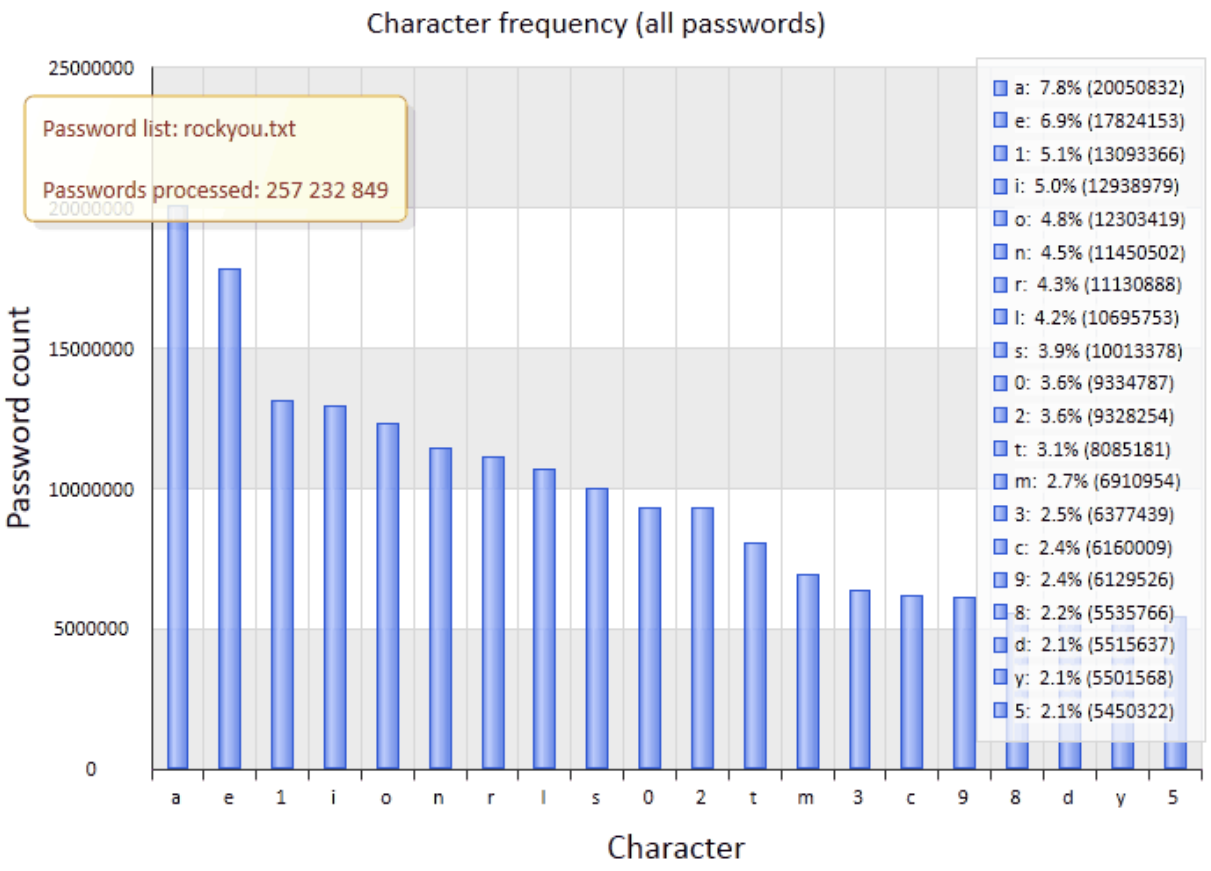
1.7 密码格式



现在让我们来考虑一下密码掩码，逐个字符来考虑。最流行的密码由6、7或8个字符组成，并且只使用一个字符集。可预测的，不是吗？在更复杂的密码中，用户通常在一个词的尾部放一个或多个数字。

我们的结论是。为了加快猜测密码的过程，攻击者可以对密码格式做一个非常简单的分析，并创建特殊的模板（掩码）。

1.8 字母频率



这个字符频率图显示了密码列表中最常使用的字符。此外，这些统计数据可用于数学分析，如构建Markov链。你可能认为这样的统计数据完全没有用处，对吗？唉，事实并非如此。如果你分析rockyou.txt中的所有记录，找出任何由20个最常用字符组成的密码，你会发现数据库中的密码多达4,789,597个，占总数的14.7%。从理论上讲，这样的分析可以用来制作完美的字符集，以便更好更快地进行密码的破解。

1.9 总结

最近，我们一直在观察一个非常令人担忧的趋势。当今计算机的性能按照摩尔定律继续增长，而利用图形处理器(GPU)进行密码破解的机会更加具有威胁性。即使是低端的四核CPU也可以每秒检查100,000,000个以上的Windows NT密码，而廉价的GPU的性能则要高出一个数量级。例如，使用GPU可以让你在短短几秒钟内就能破解一个7个字符的密码！这也是一个很好的例子。看来，软件供应商必须找到替代的数据保护方法，或者保持密码保护系统中的加密算法的更新。

简单的密码破解方法，如暴力破解或字典搜索，正在被基于统计处理、数学分析、模板、社会工程、新型攻击等的更先进工具所取代。

上述情况不应该那么令人惊讶。一位古希腊人说，在我们周围的世界中，没有什么是不变的（除了人的愚蠢）。以下是他这句话的我们的版本。没有什么是不变的，只有人们对自己的安全无所谓的态度。如果你跟着我，一个强大的密码比 "123456 "或 "qwerty "这样的东西要好得多，如果你把它写在你的笔记本上或便利贴上（即使你把它贴在显示器上）。很难让普通用户更关心他们的安全，只是因为他们已经有太多的其他问题要考虑了。这就是为什么软件供应商应该提供更好的安全，而不是让它成为用户的负担，而是在他们的产品中开发和实施更先进的密码保护算法，更好地满足今天的需求。