

简要分析雅虎泄露的密码

© 2012 Passcape Software
Passcape Software

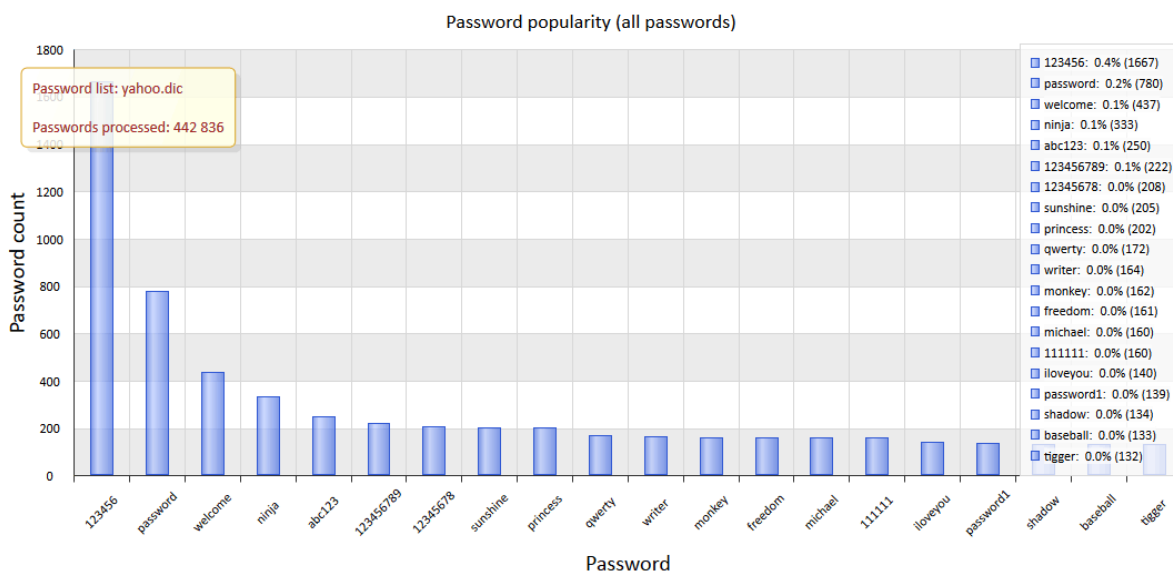
1.	简要分析雅虎泄露的密码	3
1.1	摘要	3
1.2	最受欢迎的20个密码	3
1.3	密码的独特性	4
1.4	密码长度分布图	5
1.5	字符集的多样性	6
1.6	字符集独特性	7
1.7	字符集的顺序	8
1.8	密码格式	8
1.9	总结	9

1 简要分析雅虎泄露的密码

1.1 摘要

2012年7月初, 黑客组织D33D公司在其网站上公布了一个包含45万雅虎用户的电子邮件地址和密码的文本文件。最有可能的是, 这种大规模的盗窃是由于SQL注入, 因为该文件包含MySQL表中的变量、列和行的名称。这篇文章是对被泄露的雅虎密码的概述, 可以与之前的Rockyou的分析进行比较。为这篇文章收集的统计数据是用 [Windows Password Recovery](#)这个密码审计工具整理出来的。

1.2 最受欢迎的20个密码



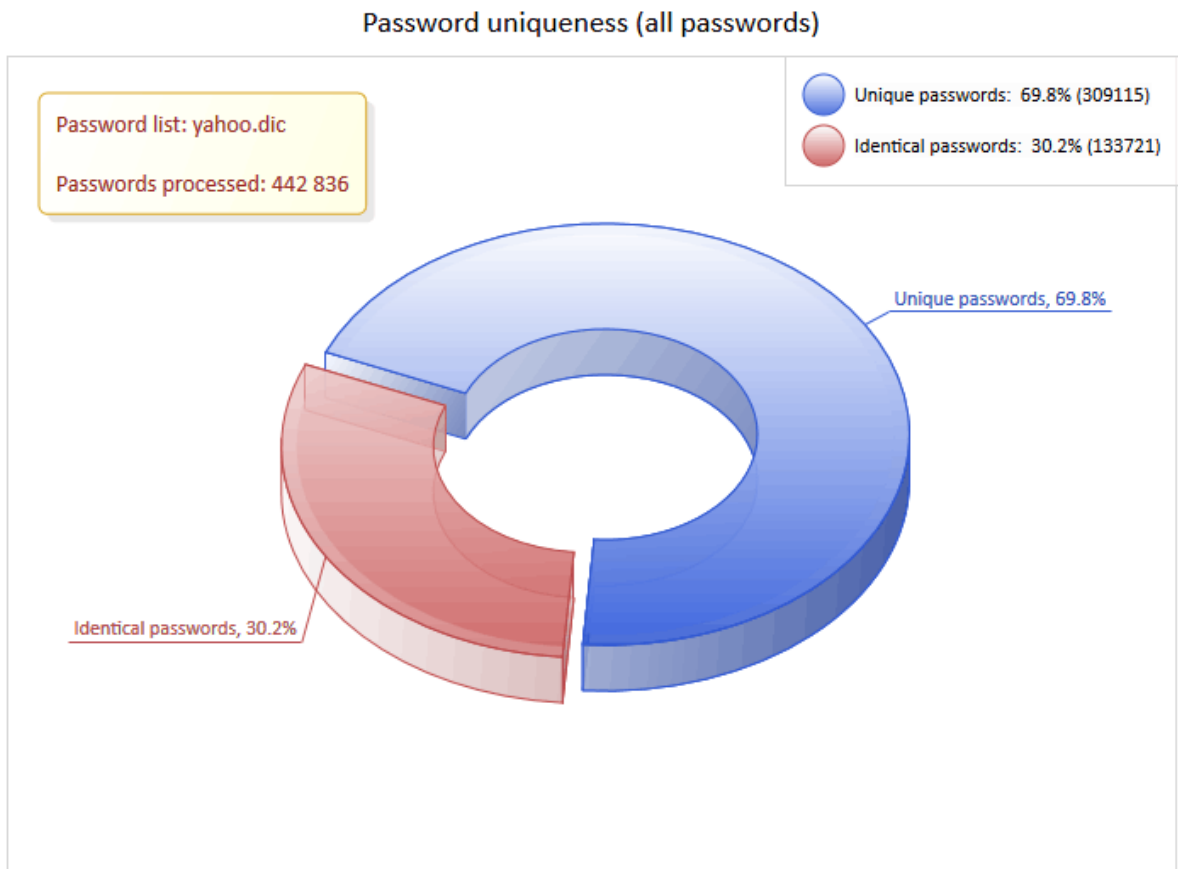
臭名昭著的123456开始了20个最受欢迎的密码名单。毫无疑问, 这个密码应该被称为 "有史以来最流行的密码"。除此以外, 前20名中还有四个数字密码。据统计, 每100个雅虎用户中就有1人的密码来自这个前20名排行榜。

- 123456
- password
- welcome
- ninja
- abc123
- 123456789
- 12345678
- sunshine
- princess
- qwerty
- writer
- monkey

freedom
michael
111111
iloveyou
password1
shadow
baseball
tiger

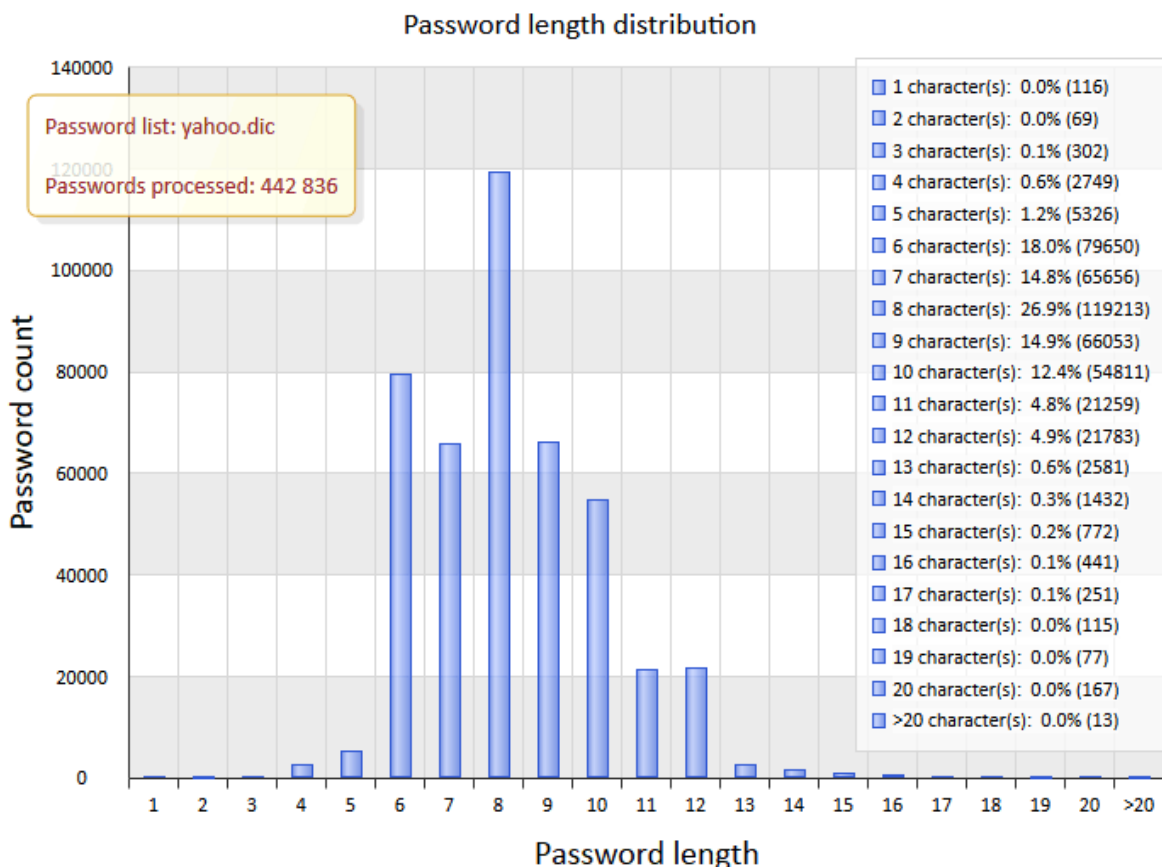
We had previously [proposed](#) that if 10 most popular passwords were used in over 1% of user accounts, such a database became insecure. In the database we're analyzing, this parameter is close to 1%.

1.3 密码的独特性



尽管少量的处理过的密码不能提供系统的全貌，但总体趋势可能跨越整个数据库，这告诉我们，超过30%的密码不是唯一的，也就是说，系统中每三个用户的密码都有重复的。

1.4 密码长度分布图

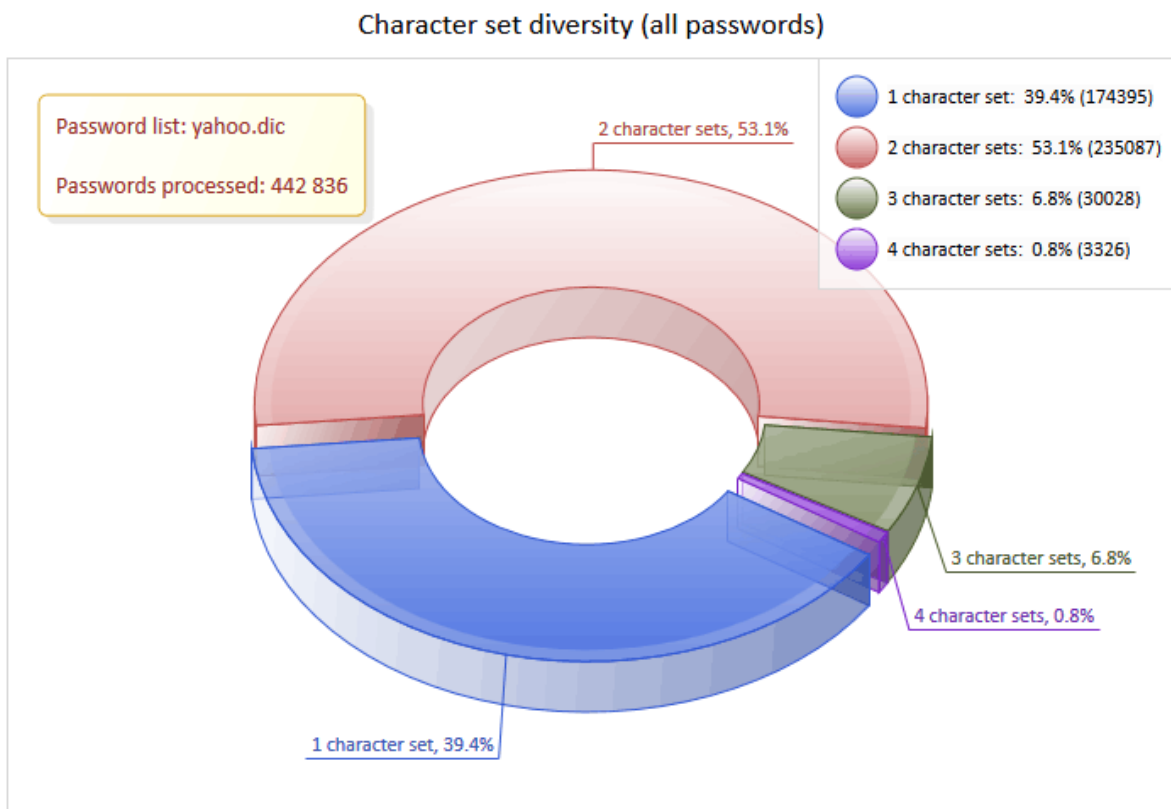


根据长度表, 大多数密码由6-10个字符组成。它们约占所有密码的87%, 这与Rockyou数据库 (86,5%) 的情况相同。最受欢迎的密码由8个字符组成, 几乎占总数的27%。

这里有一些外来的东西:

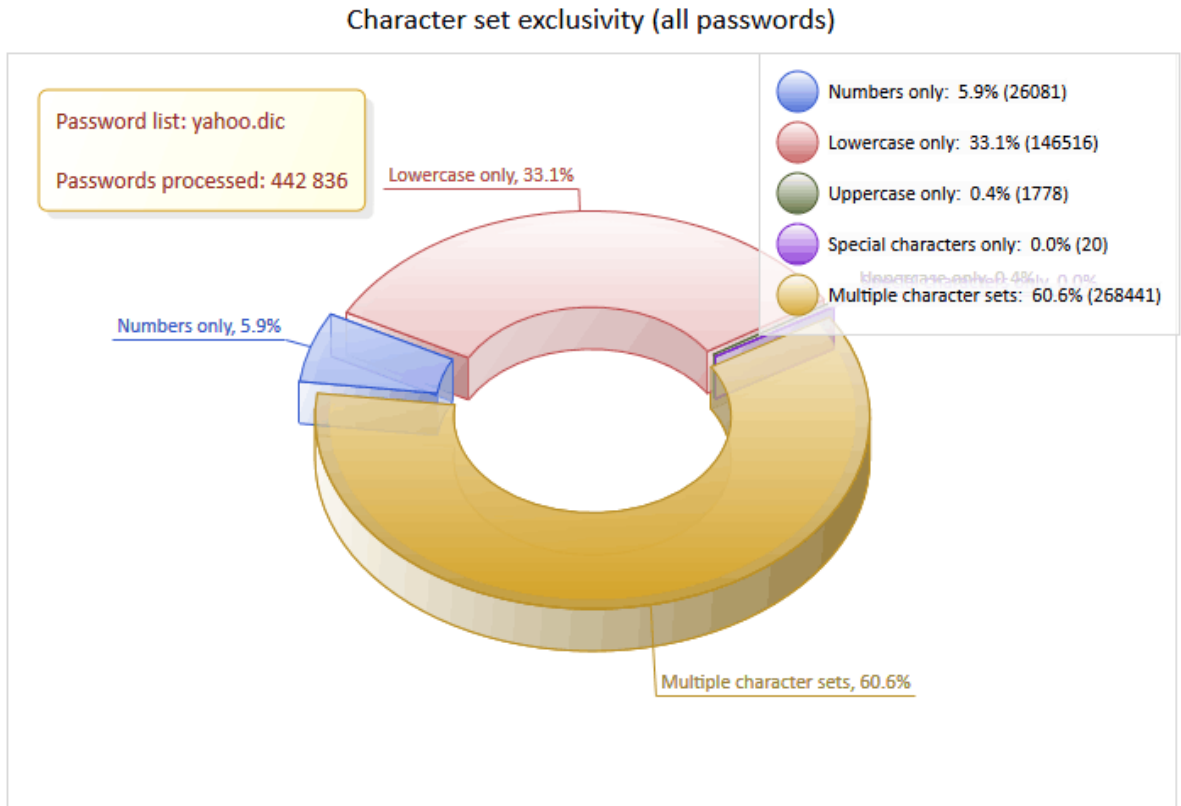
- pritamranashamarana** - prita-what ?
- j sucks donkey dicks** - parental advisory required
- imnotafraidofdragons** - me too, if I don't smoke too much
- secretlymysterious** - oh, not at all
- ihateroaches24/7!** - me too!
- tonylovesbiscuits** - who doesn't like biscuits?
- pleasedietoday123** - please don't
- platinumwindow861** - owner of a new version of Windows?
- fromheretothere6** - hither and yon, that's right
- 1my2artical3pass** - easy as 1-2-3

1.5 字符集的多样性

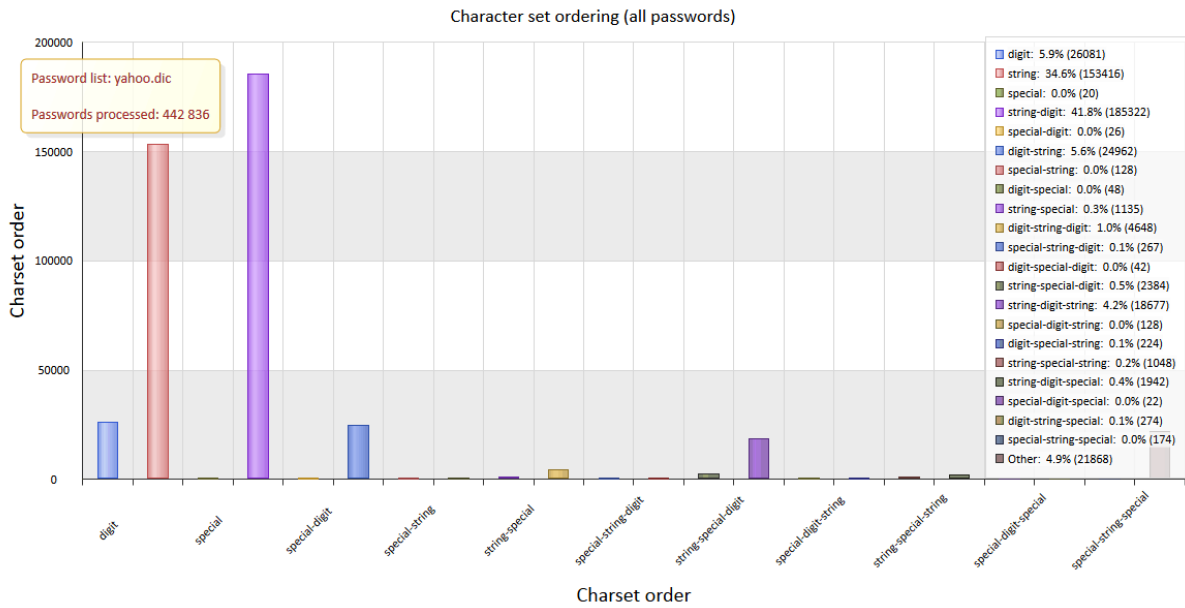


密码的长度是决定其可靠性的关键因素之一。然而，其元素的多样性是一个更重要的特征。尽管有传统的说法，但这种多样性通常是以使用的标准字符集的数量来衡量的（例如，一组小写拉丁字符或数字）。统计显示，39.4%的用户选择由单一标准字符集组成的简单密码。此外，有5.9%的人使用纯粹的数字密码。双套密码占密码总数的53.1%。剩下的7.6%是使用3个或更多字符集的复杂密码。

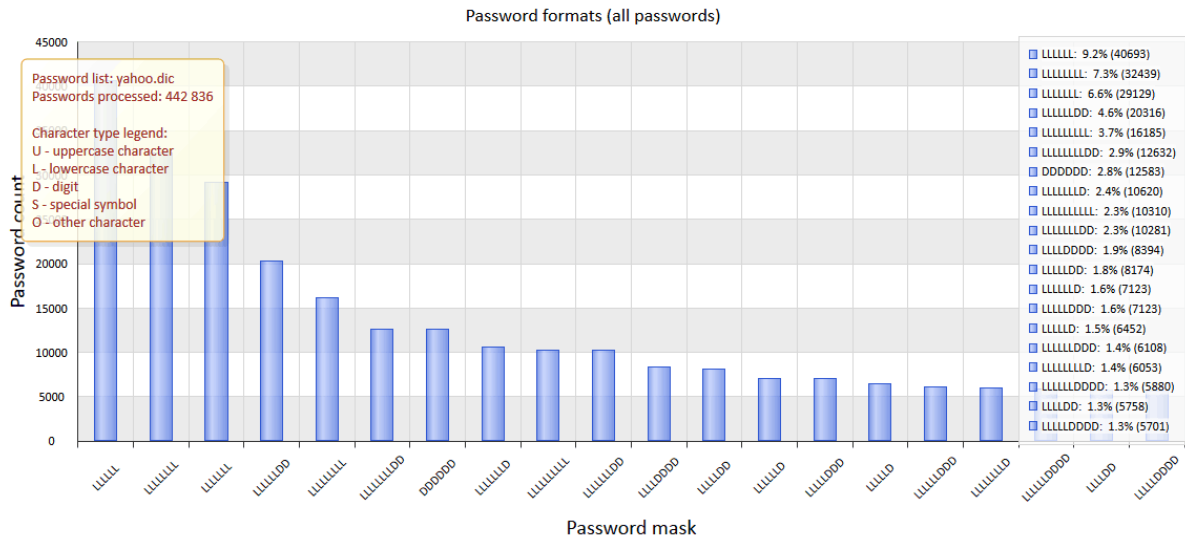
1.6 字符集独特性



1.7 字符集的顺序



1.8 密码格式



如果用掩码表示每个密码字符，下表将告诉您所使用的密码格式。下面是密码掩码图例：

- U** – uppercase character
- L** – lowercase character
- D** – digit

S – special character

O – other

下面是20种流行掩码的列表：

LLLLLL
LLLLLLLLL
LLLLLLL
LLLLLLDD
LLLLLLLLL
LLLLLLLLLDD
DDDDDD
LLLLLLLLD
LLLLLLLLLLL
LLLLLLLLLDD
LLLLDDDD
LLLLLDD
LLLLLLD
LLLLLDDDD
LLLLLDD
LLLLLLLLDDDD
LLLLLLLLLDD
LLLLLLLLDDDD
LLLLDD
LLLLLDDDD

这些掩码所涵盖的密码数量为261954个，占总数的59.2%。从图中可以看出，许多密码都是由小写拉丁字母组成的。然而，大多数用户仍然在密码的字母部分的末尾添加了一两个数字。

1.9 总结

作为结论，这里有几个图表显示了不同角色的出现统计。

