

Brief analysis of Yahoo leaked passwords

© 2012 Passcape Software
Passcape Software

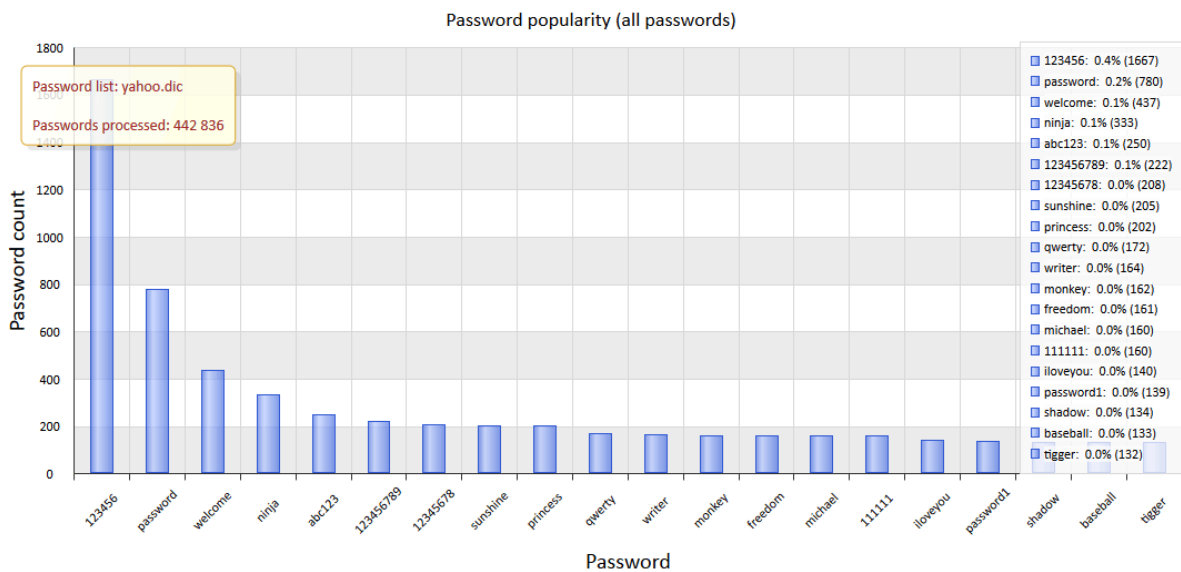
1.	Brief analysis of Yahoo leaked passwords	3
1.1	Abstract	3
1.2	Top 20 popular passwords	3
1.3	Password uniqueness	4
1.4	Password length distribution	5
1.5	Charset diversity	6
1.6	Charset exclusivity	7
1.7	Character set ordering	8
1.8	Password formats	8
1.9	Conclusion	9

1 Brief analysis of Yahoo leaked passwords

1.1 Abstract

In early July 2012, the hacker group D33Ds Company published on its website a text file with email addresses and passwords of over 450 thousand Yahoo users. Most probably, this massive theft became possible thanks to an SQL injection, since the file contains the names of variables, columns and rows from MySQL tables. This article is an overview of compromised Yahoo passwords that can be compared with our previous Rockyu's analysis. The statistics collected for the article were put together using [Windows Password Recovery, a password audit tool](#).

1.2 Top 20 popular passwords



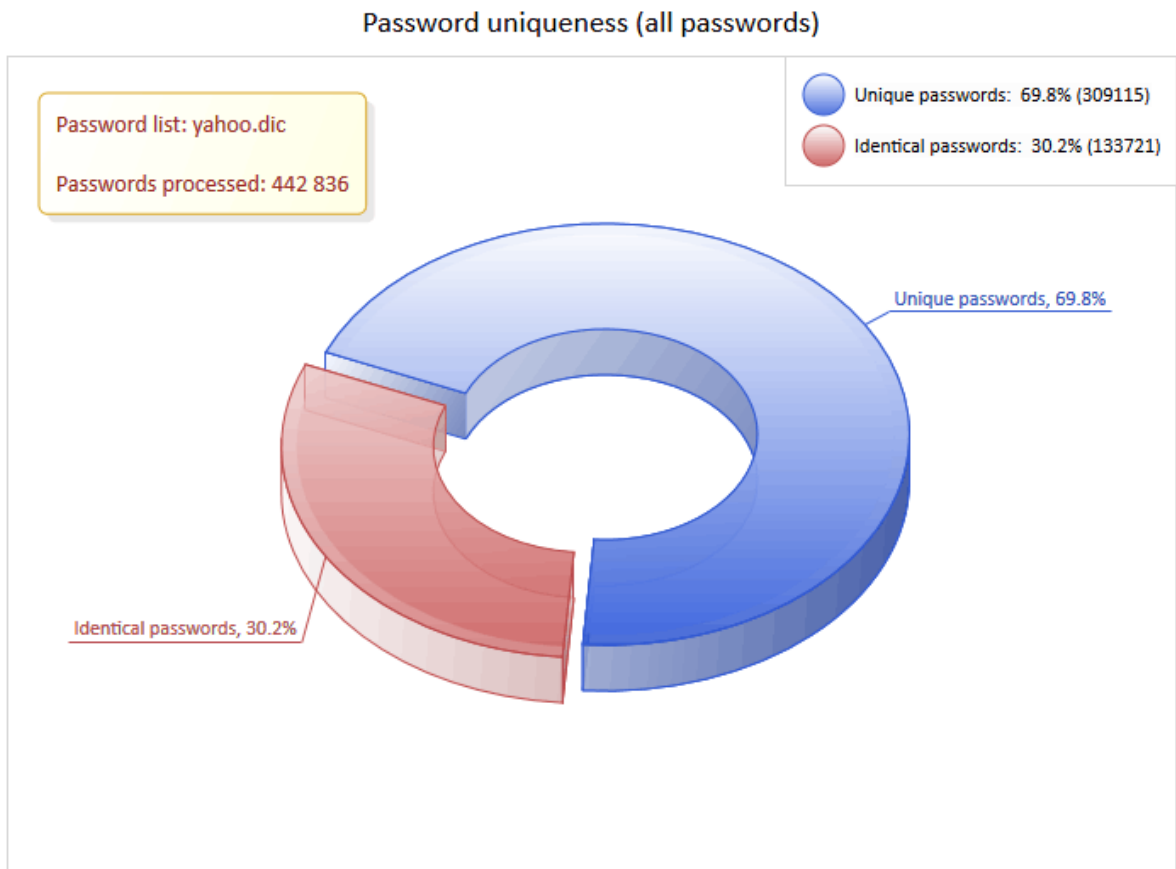
The notorious **123456** starts the list of the 20 most popular passwords. Without doubt, this password should be dubbed the "most popular password of all times". Apart from this one, the Top 20 has four more numerical passwords. According to statistics, **1 out of 100 Yahoo users has a password from this Top 20 chart**:

- 123456
- password
- welcome
- ninja
- abc123
- 123456789
- 12345678
- sunshine
- princess

qwerty
writer
monkey
freedom
michael
111111
iloveyou
password1
shadow
baseball
tigger

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%.

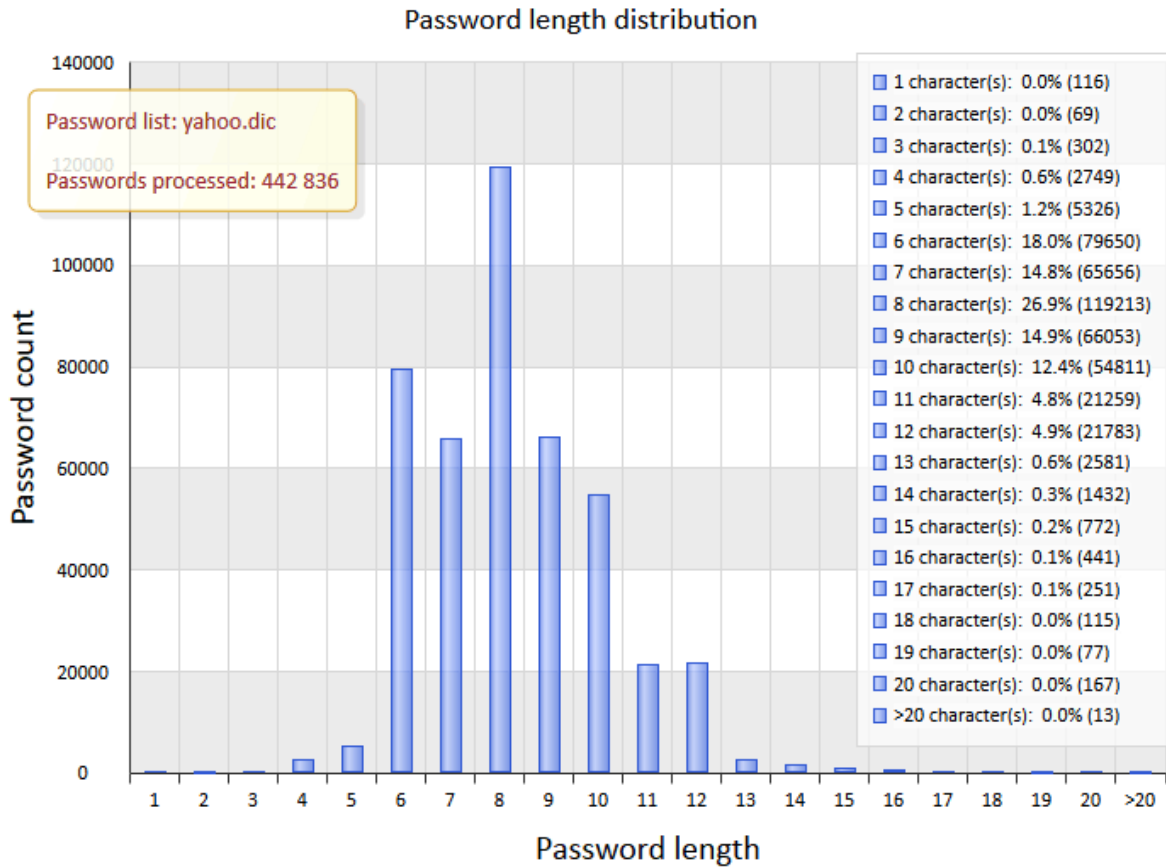
1.3 Password uniqueness



Although the small number of processed passwords do not provide the full picture of the system, the overall trend is likely to span across the entire database, which tells us that over

30% of all passwords are not unique - that is, **the password of each third user of the system has a duplicate.**

1.4 Password length distribution



According to the length chart, most passwords consist of **6-10 characters**. They account for around 87% of all passwords, which is in the same ballpark with the Rockyou database (86,5%). **The most popular passwords consist of 8 characters**, which is almost 27% of the total number.

Here are some exotic stuff:

pritamranashamarana - prita-what ?

j sucks donkey dicks - parental advisory required

imnotafraidofdragons - me too, if I don't smoke too much

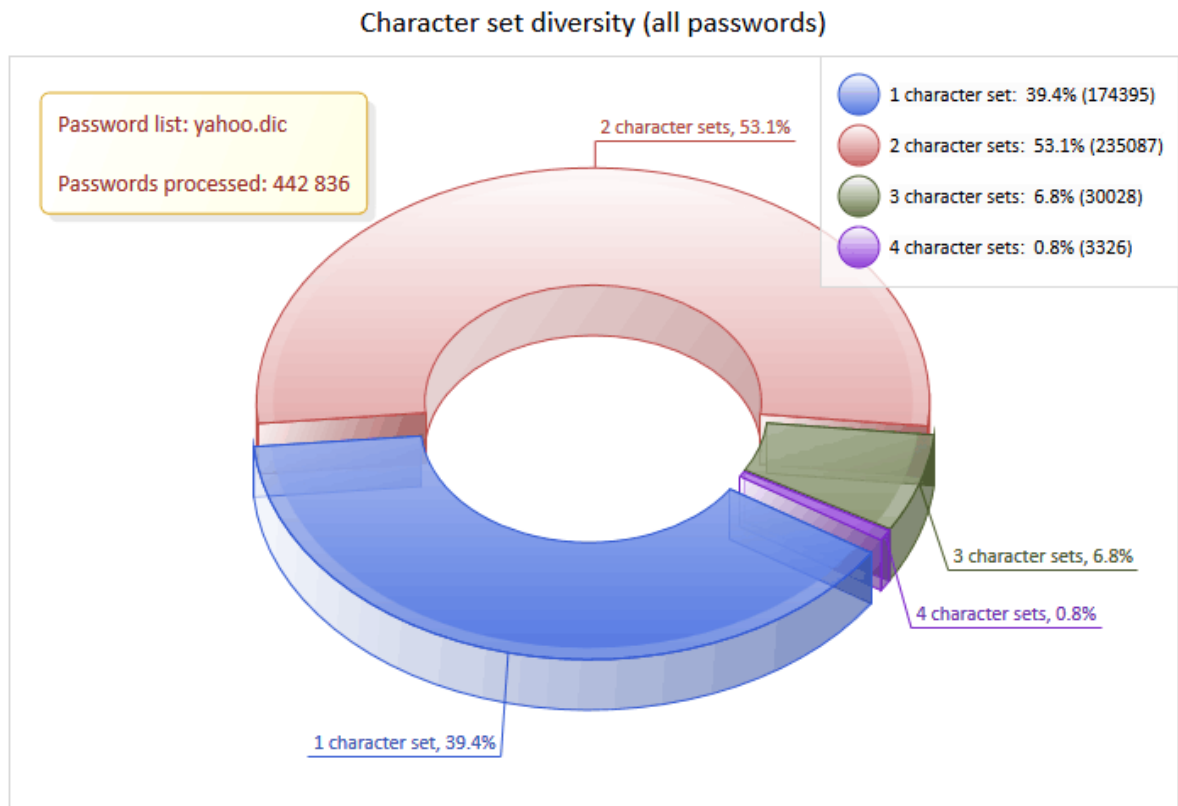
secretlymysterious - oh, not at all

ihateroaches24/7! - me too!

tonylovesbiscuits - who doesn't like biscuits?

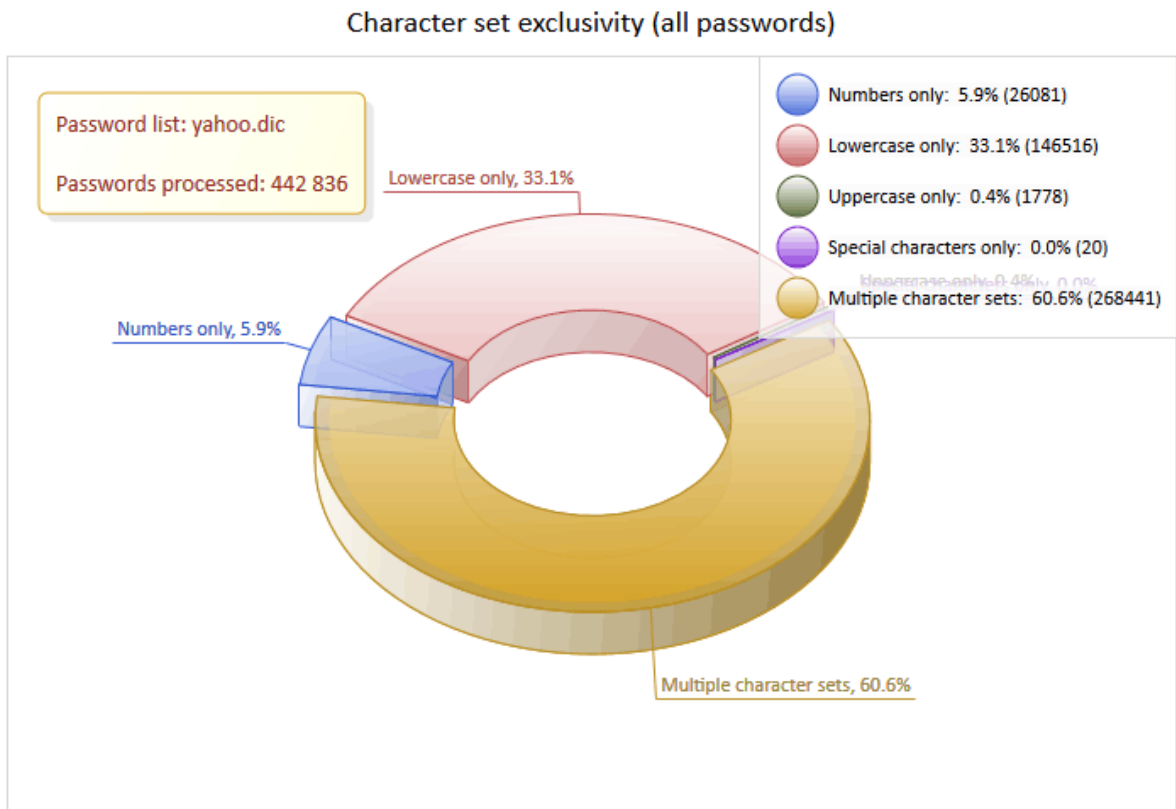
pleasedietoday123 - please don't
platinumwindow861 - owner of a new version of Windows?
fromheretothere6 - hither and yon, that's right
1my2artical3pass - easy as 1-2-3

1.5 Charset diversity

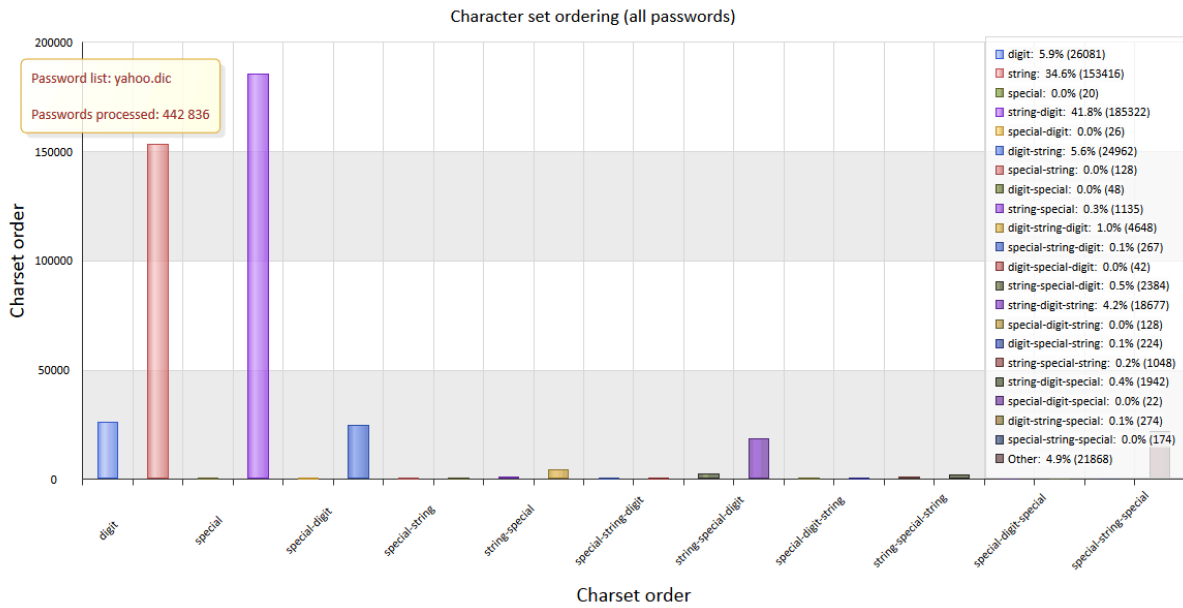


The length of a password is one of the key determinants of its reliability. However, the diversity of its elements is a lot more important characteristic. Despite the conventionalism, this diversity is usually measured in the number of standard character sets used (for instance, a set of lower-case Latin characters or digits). Statistics show us that **39.4% of all users pick simple passwords** that consist of a single standard character set. Besides, **5.9% of them use purely numerical passwords**. Two-set passwords account for **53.1%** of the total number of passwords. The remaining **7.6% are complex passwords** using 3 or more character sets.

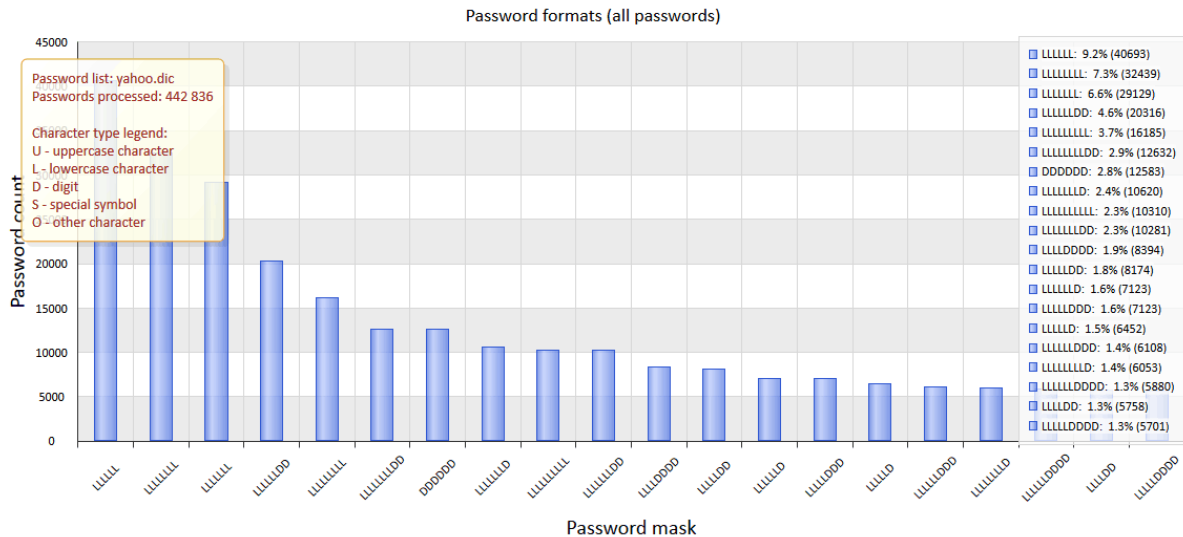
1.6 Charset exclusivity



1.7 Character set ordering



1.8 Password formats



If you represent each password character with a mask, the following chart will give you an idea about the format of passwords used. Here is the password mask legend:

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

- D** – digit
- S** – special character
- O** – other

Below is a list of 20 popular masks:

- LLLLLL
- LLLLLLLL
- LLLLLLL
- LLLLLLDD
- LLLLLLLLL
- LLLLLLLLLDD
- DDDDDD
- LLLLLLLLD
- LLLLLLLLLLL
- LLLLLLLLLDD
- LLLLDDDD
- LLLLLDD
- LLLLLLD
- LLLLLLDDD
- LLLLLLD
- LLLLLLDDD
- LLLLLLD
- LLLLLLDDDD
- LLLLDD
- LLLLLDDDD

The number of password covered by these masks is 261954 or **59.2%** of the total number. As you can see in the chart, many passwords consist of lowercase Latin characters. However, **most users still add a digit or two to the end** of the letter part of their passwords.

1.9 Conclusion

As a conclusion, here are a couple of charts showing the occurrence statistics for different characters.

