Brief analysis of RockYou leaked passwords

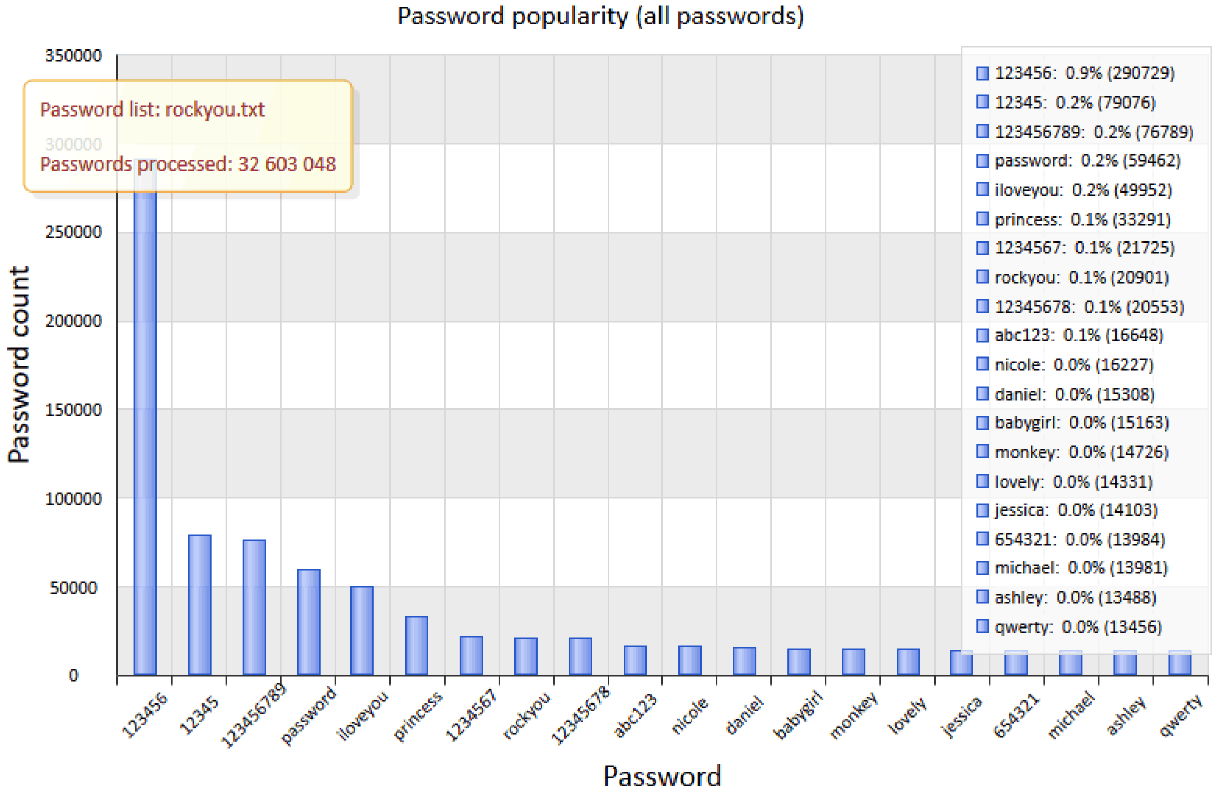
Passcape Software

1 Brief analysis of RockYou leaked passwords

1.1 Abstract

In December 2009, the RockYou.com website was attacked by hackers. The attack was successful and resulted in the illegal disclosure of a text database containing 32.6 million stolen passwords of the website's users. This major leak of confidential data has allowed us to take a look at how people create their passwords and to do a very simple security audit from the viewpoint of an external observer. To collect and process statistical data, we used Windows Password Recovery tool.

1.2 Top 20 popular passwords

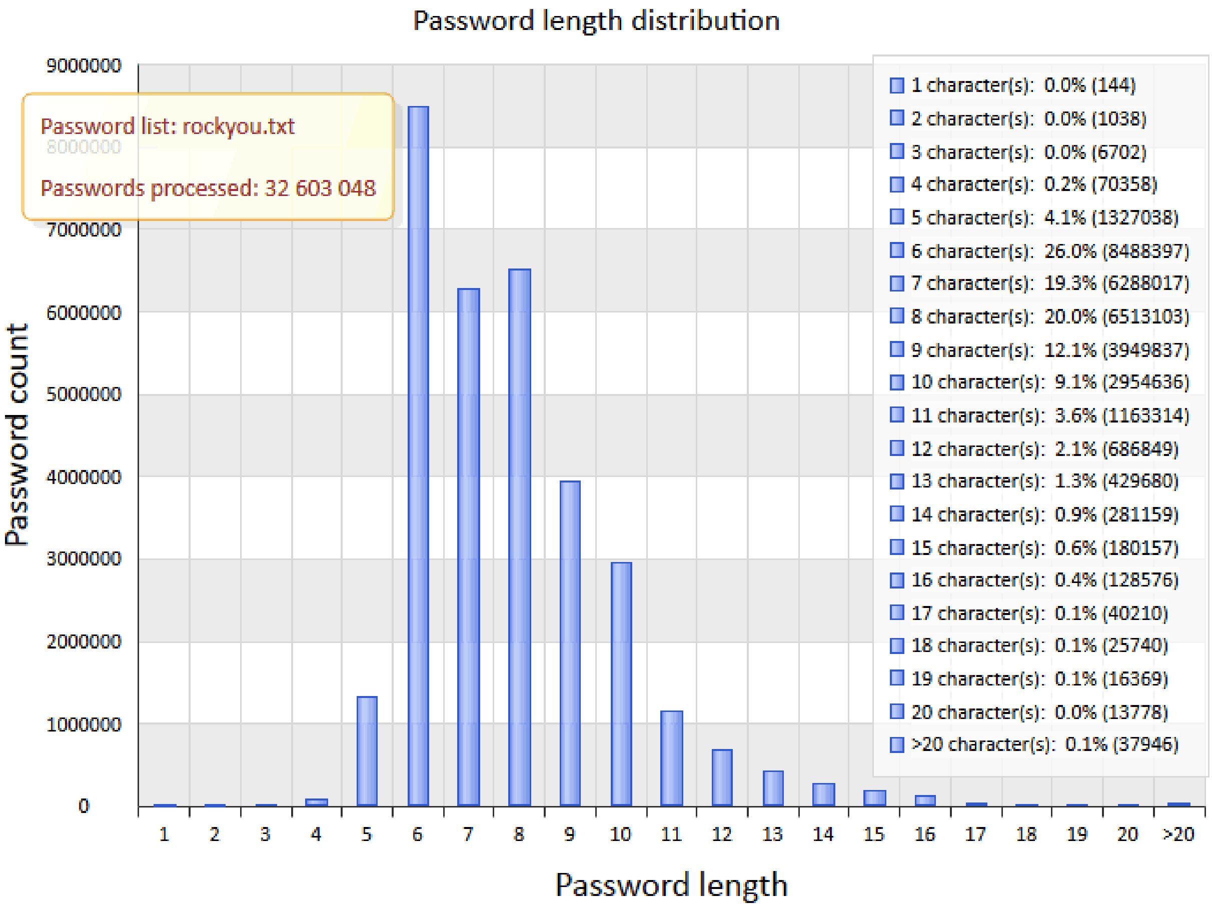


It's hard to find something more complex than "**123**" among the most popular passwords, so the top 20 is not surprising at all. As you can see, the most popular password ("**123456**") is far ahead of the pack. Its popularity is amazing even in numbers: 290,000 out of 32 million records. One of our associates suggested that a database where the 10 most popular passwords count more than 1 percent of records should be considered as insecure. In the RockYou.com database, this "insecurity factor" is exceeded 2 times. When analyzing popular passwords, we categorized them by common attributes into several groups:

- **Dictionary passwords**, that is, words like "*password*," "*monkey*" and so on, make one of the most stable groups.

- **Digital passwords** based on easily memorized numeric combinations, phone numbers, document numbers, birth dates, and more make another group, which is as stable as the previous one, and maybe even more popular.
- **Passwords based on names** and their derivatives. For example, a user may use the name of himself, his pet, some city, some place, and so on.
- Passwords based on **keyboard combinations**, such as "*abc123*," "*qwerty*" etc.
- **Emotional passwords**, such as "*iloveyou*", "*hateu*" "*lovely*", "*ihatemyboss*" or "*ILoveJohn*".

1.3 Password length distribution



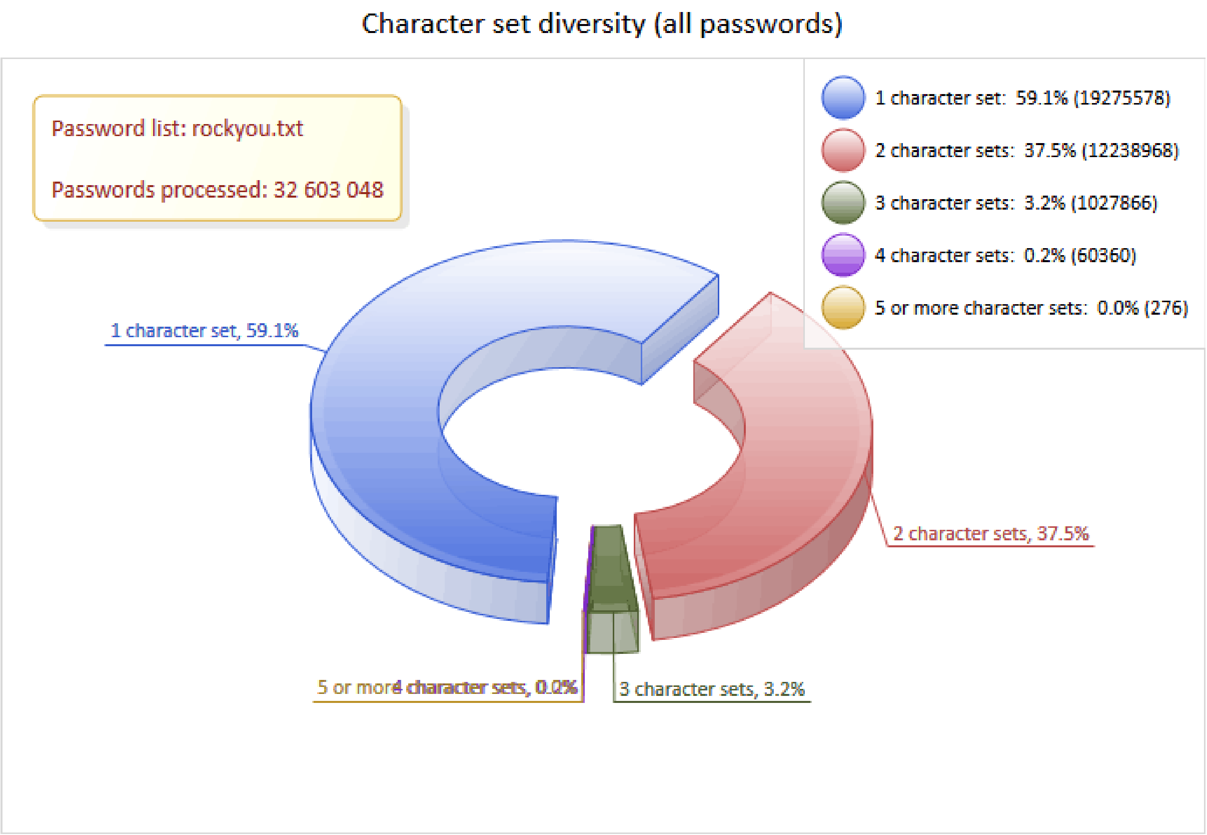
A password's length is an important factor of its resistance against attacks. You can say the longer the password, the harder it is to break it. But don't forget that this unstable equilibrium can be easily lost if the long password is "successfully" forgotten by the user after a lively party with lots of drinks. Some users put sticky notes with passwords right onto their monitors, but the majority are even less sophisticated and prefer easily remembered passwords consisting of at most 7 or 8 characters.

As you can see on the chart, the most popular passwords are **6, 8, or 7** characters in length. That's more than **65** percent of all passwords. It means that 2 of 3 passwords can be easily brute-forced.

Curiously enough, some users choose obscenely long passwords consisting of more than 20 characters, word combinations, or phrases. Here are some of these wonderful passwords (did you really think that nobody knows what you are typing?):

Hahaithinkilovejessebutthenagainmaybenotcuzheisadiknob
Lets you come back for your Countdown Timer
me plus food equals more sleep each night
tommmmmmmmmmmmmmmmmmmmmmmmmmmmm
truongcaodangcongdonghaiphong
icantbelievethisshit.12345
banditbanditbandit1bandit1bandit1banditbandit
1111111111111111111111111111111111111111111
aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
FUCKSCHOOLANDALLTHETEACHERSINIT
ilovepalmermyfuturehusband
1delightyourselfinthelord!
Imaprincessbecausemyfatheristheking
iluvanjabisset4evashesmawebaexxx
stuartandchrisrmybestmatesforeva
thisismypasswordyoullnevergetit

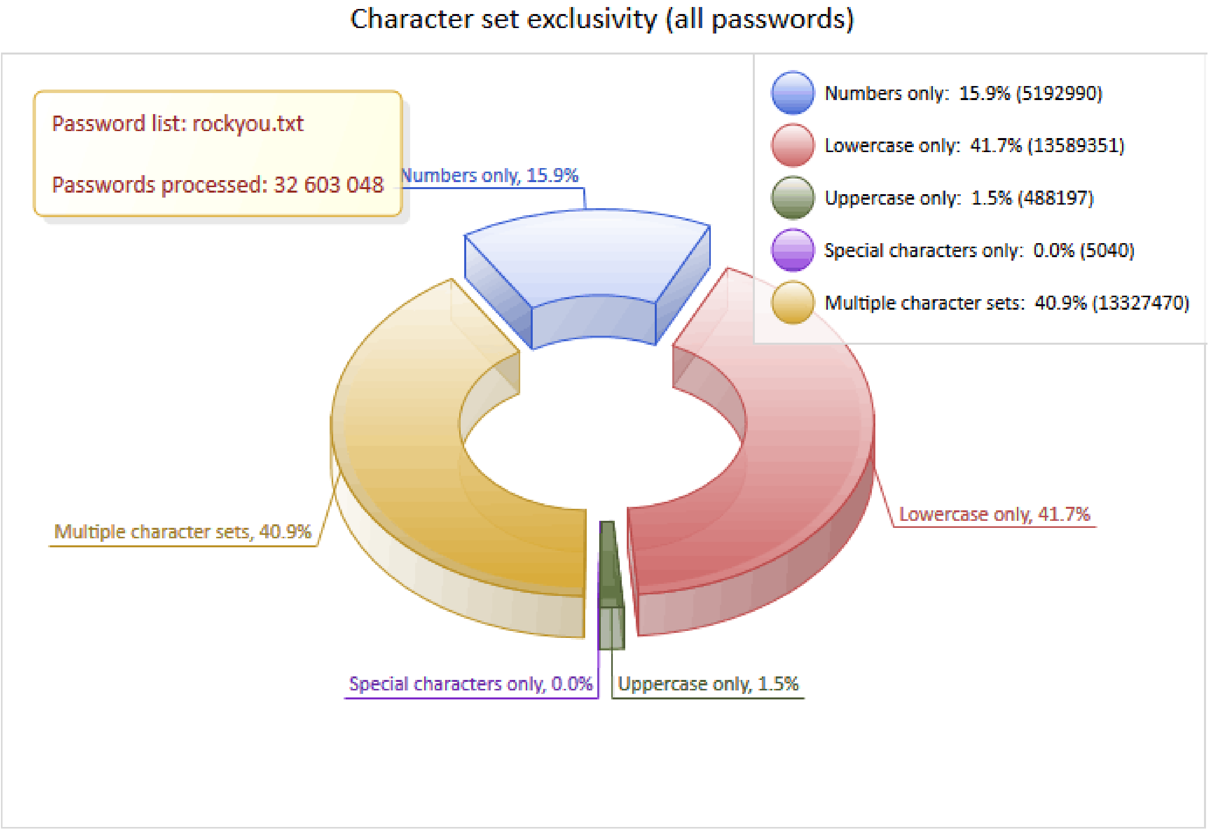
1.4 Character set diversity



Another important factor affecting the password strength is the diversity of characters. For example, the 7-character password "**aB1~cde**" is more than 10 times as strong as the 9-character password "**abcdefghi**".

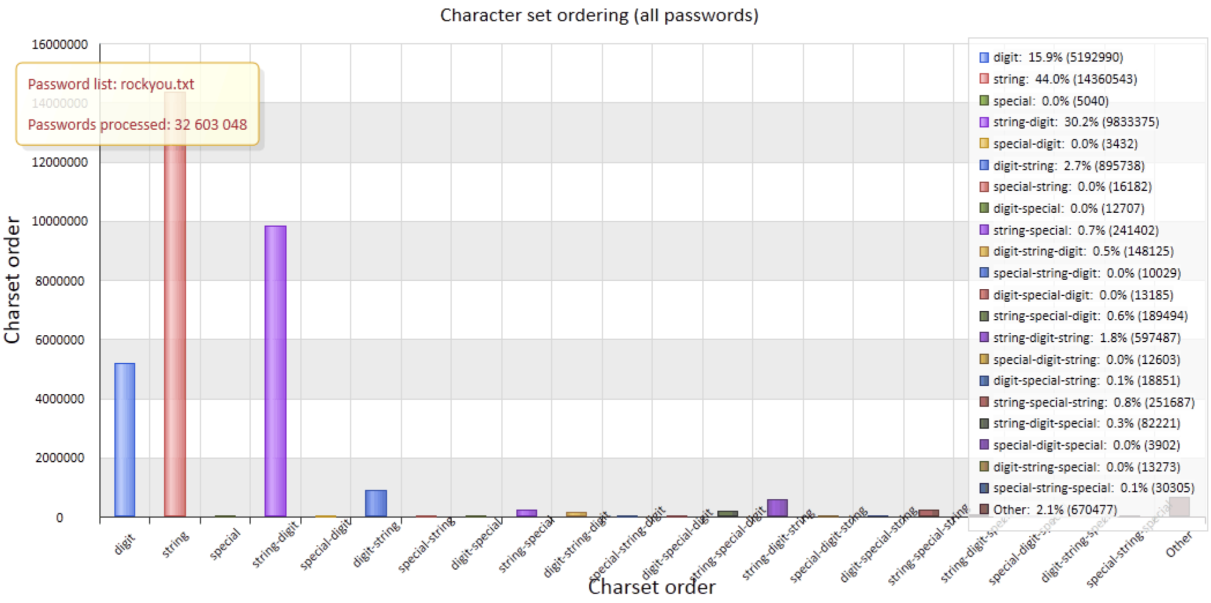
Over a half of all users have only one character set in their passwords. For example, they use only digits or only lower-case letters. More than 96.5 percent of users have only one or two character sets, and less than 3.5 percent use three or more charsets. The charts of password length and diversity clearly show that users' preferences in choosing passwords have barely changed in the last two decades. Our conclusion: We have successfully debunked the myth that you can force users to use strong passwords (at least, in this case).

1.5 Preferences in choosing character sets



The general statistics shows that when choosing passwords based on one character set, users mostly prefer words consisting of lower-case letters only (more than 41 percent of all passwords), of digits only (almost 16 percent), or upper-case letters only (1.5 percent). Passwords consisting of **special characters only are used very rarely**.

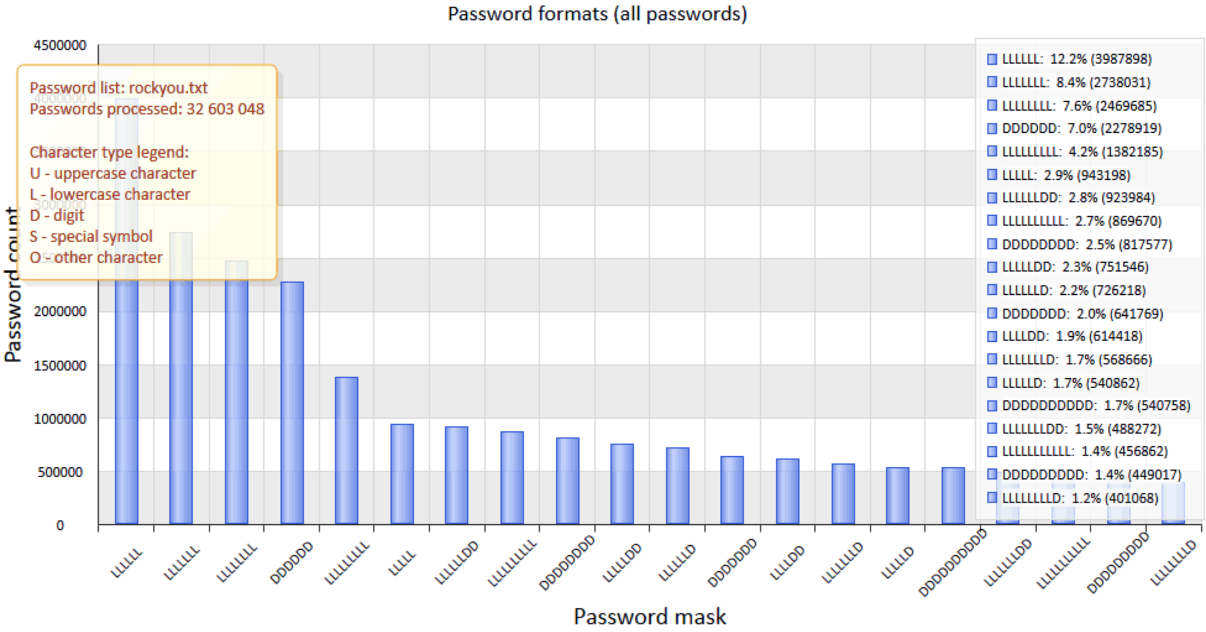
1.6 Character set order in passwords



If you take a look at this chart, you may be surprised at the fact that the vast majority of passwords consisting of two or more character sets are **string-digits** or **digits-string** combinations. An attacker skilled in social engineering can use this observation to crack such passwords.

Our conclusion: The "**iloveyou12345**" password is as weak as the "**iloveyou**" one.

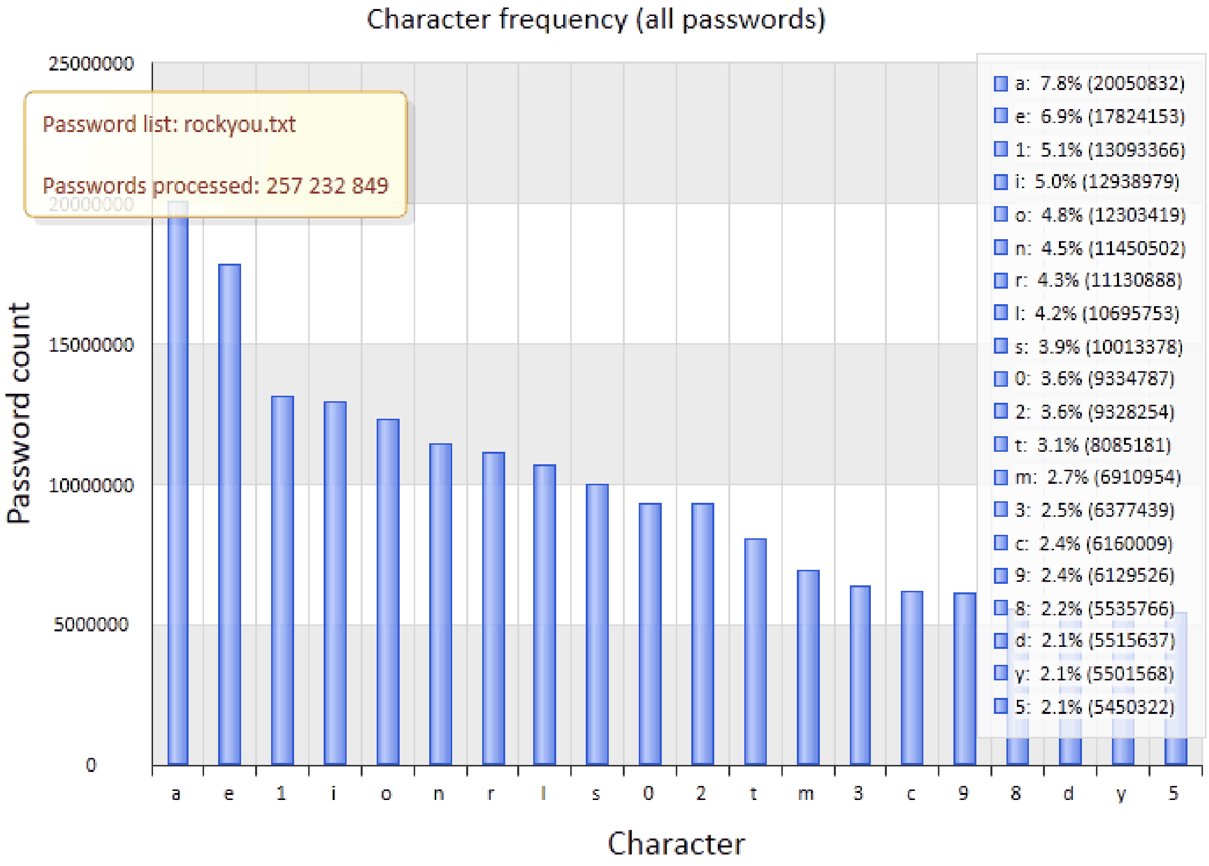
1.7 Password formats



Now let's consider the password mask, character by character. The most popular passwords consist of 6, 7, or 8 characters and use only one character set. Predictable, isn't it? In the more sophisticated passwords, users usually put one or more digits at the tail of a word.

Our conclusion: To speed up the process of password guessing, an attacker can do a very simple analysis of password formats and create special templates (masks).

1.8 Character frequency



This character frequency chart shows the characters most frequently used in the password list. Further, these statistics can be used for mathematical analysis, such as constructing Markov chains. You might think that such statistics are utterly useless, right? Alas, that's not the case. If you analyze all records in rockyou.txt for any passwords consisting of the 20 most frequently used characters, you'll find that there are as many as 4,789,597 of them in the database, or 14.7 percent of the total number. In theory, such analysis can be used for making perfect character sets for better and faster brute-forcing of passwords.

1.9 Conclusion

Recently we've been observing a very worrying trend: The performance of today's computers continues to grow following Moore's law, and the opportunity to use graphics processors (GPUs) for brute-forcing passwords is even more threatening. Even a low-end quad-core CPU can check more than 100,000,000 Windows NT passwords per second, and the performance of an inexpensive GPU is an order of magnitude higher. For example, using a GPU lets you brute-force a 7-character password in just a few seconds! It looks like software vendors have to find alternative data protection methods or keep encryption algorithms in password-protected systems up-to-date.

Simple password cracking methods, such as brute-forcing or dictionary search, are being replaced by more advanced tools based on statistical processing, mathematical analysis, templates, social engineering, new types of attacks, and so on.

The above shouldn't be that surprising. One ancient Greek said that nothing stays the same in the world around us (except the man's stupidity). Here's our version of his saying: Nothing stays the same but people's carefree approach to their own security. If you follow me, a strong password is much better than something like "123456" or "qwerty" if you write it down in your notebook or on a Post-it note (even if you stick it on your monitor). It's hard to make ordinary users care more about their security simply because they already have too many other problems on their mind. That's why software vendors should provide better security not by making it the user's burden but by developing and implementing in their products more advanced password protection algorithms that better meet today's needs.