THE PERFORMANCE COMPARISON OF A BRUTE-FORCE PASSWORD CRACKING ALGORITHM USING REGULAR FUNCTIONS AND GENERATOR FUNCTIONS IN PYTHON

Berker Tasoluk1 and Zuhal Tanrikulu2

1 Informatics, Istanbul University, Istanbul, Turkey
2 Management Information Systems, Bogazici University, Istanbul, Turkey

ABSTRACT

Python is used extensively in research, including algorithm testing. Python is a multi-paradigm programming language and supports both object-oriented programming and functional programming. In the functional side, it supports both regular functions and generator functions. This study tests both approaches in terms of usability cases and performance. A password-cracking algorithm is used for this tryout.

KEYWORDS

Python functions, generator functions, regular functions, regular functions, in-memory functions, iterator, generator expressions, time efficiency, memory efficiency, python programming, general programming, password cracking

1. INTRODUCTION

In academic research field, Python is used extensively, including algorithm testing. Machine Learning [1], data science [2], web scraping [3], text and language processing [4, 5]. Python is a multi-paradigm programming language and supports both object-oriented programming [6] and functional programming [7]. Functional side of python supports both regular functions and generator functions [8]. This study tests both approaches in terms of usability cases and performance. A password-cracking algorithm is used for this tryout. Passwords are used extensively in the authentication process, by users to identify themselves to the systems they want to access [9]. Each password is associated with a specific user id in the system like email address, username, or phone number, and by providing password, users prove that they are who they claim to be [10]. Passwords connects identification and authentication processes.

In this study, python regular functions and generators are tested for their memory usage and duration taken to try a certain number of passwords based on a given set of criteria. At the discussion part, practical usage advice is given for both regular and generator functions.

2. METHODS

In this study, our aim is to compare the two implementations of functions in the field, using a password cracking undertaking. We have a set of passwords, which eventually would be cracked,
which is called destination list passwords; and we have another set of passwords, which would be used to crack the destination list passwords, which we call source list passwords. Destination list passwords are composed of real-world revealed passwords. We try each entry in the source list with each entry in the destination list. When we find a match between source list entry and destination list entry, we suppose that the password is ‘broken’.

2.1. Destination (Target) List of Passwords

Destination list passwords, and the number of passwords included in them are as follows:

![Figure 1. Destination password list details](image1)

And the file size of each password file is as follows:

<table>
<thead>
<tr>
<th>Name</th>
<th>Date modified</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>alypaa.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>12 KB</td>
</tr>
<tr>
<td>carders.cc.txt</td>
<td>10/12/2011 10:24...</td>
<td>Text Document</td>
<td>17 KB</td>
</tr>
<tr>
<td>elitehacker.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>7 KB</td>
</tr>
<tr>
<td>facebook-pastebay.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>1 KB</td>
</tr>
<tr>
<td>facebook-phished.txt</td>
<td>10/12/2011 10:24...</td>
<td>Text Document</td>
<td>26 KB</td>
</tr>
<tr>
<td>faithwriters.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>71 KB</td>
</tr>
<tr>
<td>hak5.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>25 KB</td>
</tr>
<tr>
<td>hotmail.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>86 KB</td>
</tr>
<tr>
<td>myspace.txt</td>
<td>10/12/2011 10:23...</td>
<td>Text Document</td>
<td>348 KB</td>
</tr>
<tr>
<td>porn-unknown.txt</td>
<td>10/12/2011 10:24...</td>
<td>Text Document</td>
<td>57 KB</td>
</tr>
<tr>
<td>singles.org.txt</td>
<td>10/12/2011 10:24...</td>
<td>Text Document</td>
<td>105 KB</td>
</tr>
<tr>
<td>tuscl.txt</td>
<td>10/12/2011 10:24...</td>
<td>Text Document</td>
<td>318 KB</td>
</tr>
</tbody>
</table>

![Figure 2. Size of password files](image2)
2.2. Source List of Passwods

Here, the brute force passwords are used for password cracking. Uppercase, lowercase ascii letters and digits are used to form the source lists. Every possible combination of the following character sets are used in creating the password files. Python string module properties are used for the character sets.

```python
>>> import string
>>> string.ascii_uppercase
'ABCDEFGHIJKLMNOPQRSTUVWXYZ' (26 different characters)
>>> string.ascii_lowercase
'abcdefghijklmnopqrstuvwxyz' (26 different characters)
>>> string.digits
'0123456789' (10 different characters)
```

1, 2, 3, 4 and 5-character lists are created in this manner.

The number of entries for each list is calculated as follows:

- 1 character list: \(62^1 = 62\) entries
- 2 character list: \(62^2 = 3,844\) entries
- 3 character list: \(62^3 = 238,328\) entries
- 4 character list: \(62^4 = 14,776,336\) entries
- 5 character list: \(62^5 = 916,132,832\) entries

And these numbers are also obtained experimentally, using `countline_dir()` python function written.

Figure 3. Experimentally calculated line counts

And the space occupied in disk for each list is as follows:

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
<th>Date modified</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>brutelist-1uld.txt</td>
<td>✔️</td>
<td>4/26/2020 11:32 AM...</td>
<td>Text Document</td>
<td>1 KB</td>
</tr>
<tr>
<td>brutelist-3uld.txt</td>
<td>✔️</td>
<td>4/26/2020 11:32 AM...</td>
<td>Text Document</td>
<td>1,164 KB</td>
</tr>
<tr>
<td>brutelist-5uld.txt</td>
<td>✔️</td>
<td>4/26/2020 11:46 AM...</td>
<td>Text Document</td>
<td>6,262,627 KB</td>
</tr>
</tbody>
</table>
The code ran on a machine with following specs:

- 16 GB Ram, Intel Core i7-8750H 2.20 Ghz processor, Windows 10 Enterprise LTSC Operating System, 512 GB SSD Disk

3. RESULTS

3.1. Regular Python Functions

Each destination list entry is tried to be matched with 1-char through 5-char lists; and the result is as follows:

1-char list took under 1 second to be completed, while 2-char list took under 1 min, 3-char list completed under 1 hour. 4-char list took 1.8 days to be completed, and 5-char list couldn’t be completed due to memory error. This memory error is the reason behind re-writing the algorithm in python using generator functions.

If we look at the relationship between time taken to try all entries in the list (y), with the number of entries in the list (x), then we get approximately $y = (1/95)x$ linear relationship. From that; we may conclude that, if we had unlimited memory, it would take approximately 111 days for a 5-char password list to be completed.

![Figure 4. Password cracking tryout using regular python functions](image)

![Figure 5. Relationship between number of entries and time taken to crack passwords](image)
3.2. Generator Python Functions

Generator functions use iterators, and instead of getting the source list into the memory at once, they use iterator objects and performs every operation one by one. The old ones are cleared up from the memory, and the next entry in the iterator is loaded into the memory.

The results from the generator function is as follows:

![Figure 6. Password cracking tryout using generator python functions](image)

<table>
<thead>
<tr>
<th>List Name</th>
<th>In-Memory Generator</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>brutelist-1uld.txt</td>
<td>0.59</td>
<td>15.66</td>
</tr>
<tr>
<td>brutelist-2uld.txt</td>
<td>39.97</td>
<td>182.86</td>
</tr>
<tr>
<td>brutelist-3uld.txt</td>
<td>2,511.13</td>
<td>10470.93</td>
</tr>
</tbody>
</table>

![Figure 7. In-memory vs generator functions time comparison](image)

It took nearly 4 times as much time as it took in using all in-memory python functions. From this graph, it may be concluded that 4-char source list would take around 1 week (1.8 days x 4), and 5-char source list would take around more than a year (111 days x 4).

4. DISCUSSION

As we can see from the above experiments and calculations; for our case; if we have resources available, especially the memory; it is better to have the function written as a regular python function which works all in-memory; instead of using generators. In our scenario; using generators is efficient in terms of memory, but also slower. If we have time, and not the resources; then using generator functions may be needed; and sometimes it may be the only solution we have in hand.
REFERENCES