Generating probable password candidates for the offline assessment of Dutch domain password hashes

By Tom Broumels
Supervisor: P. Campers
Introduction

• Use and abuse of passwords common practice today

• Passwords stored as hashes

• Cracking by trying: selecting likely password candidates

• NIST recommendation on using breach corpuses

Secura → 0DF335A49BD7B40BE674EEE80A6FBADD


Assessing password hashes at Secura

• Password hash strength assessment part of security assessments, e.g., red teaming exercises

• Improved assessment can lead to shorter lead times or more complete results:
  – Finding more passwords in total
  – Finding more passwords first 30 minutes

• Frequently assessing hashes for Dutch clients
Ethical considerations

• Using breached passwords
  - Realistic assessments
  - Removal of e-mail addresses

• Validating research on hashes of active user accounts
  - Secured environment
  - Password hashes only
Research question

How do different password guessing algorithms compare in selecting probable password candidates for assessing offline Dutch domain password hashes?
Important related work

• No publications available on Dutch passwords

• Human behaviour related to password generation

• Password candidate generation:
  – Markov, PCFG, OMEN, PRINCE
  – PassGAN, NeuralNetwork

• Combining approaches: TarGuess
Research method

1. Dutch domain password selection
2. Selecting different password cracking approaches
3. Comparing approaches using experiments
4. Selecting a well performing approach
Research method

1. Dutch domain password selection
2. Selecting different password cracking approaches
3. Comparing approaches using experiments
4. Selecting a well performing approach
Dutch domain password selection
Dutch domain passwords

- Unique email/password entries: 3,424,464
- Unique passwords: 2,355,739
- 31.2% duplicates
# Common Dutch domain passwords

<table>
<thead>
<tr>
<th>Password</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456</td>
<td>8795</td>
<td>0.26</td>
</tr>
<tr>
<td>welkom</td>
<td>3950</td>
<td>0.12</td>
</tr>
<tr>
<td>SKIFFY</td>
<td>3708</td>
<td>0.11</td>
</tr>
<tr>
<td>welkom1</td>
<td>2547</td>
<td>0.07</td>
</tr>
<tr>
<td>123456789</td>
<td>2524</td>
<td>0.07</td>
</tr>
<tr>
<td>qwerty</td>
<td>2304</td>
<td>0.07</td>
</tr>
<tr>
<td>welkom01</td>
<td>2220</td>
<td>0.06</td>
</tr>
<tr>
<td>wachtwoord</td>
<td>2177</td>
<td>0.06</td>
</tr>
<tr>
<td>geheim</td>
<td>1792</td>
<td>0.05</td>
</tr>
<tr>
<td>amsterdam</td>
<td>1568</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### All passwords

<table>
<thead>
<tr>
<th>Password</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welkom01</td>
<td>861</td>
<td>0.26</td>
</tr>
<tr>
<td>ka_dJKHJsy6</td>
<td>198</td>
<td>0.06</td>
</tr>
<tr>
<td>Welkom123</td>
<td>187</td>
<td>0.06</td>
</tr>
<tr>
<td>PPPrt30TA</td>
<td>152</td>
<td>0.05</td>
</tr>
<tr>
<td>Feyenoord1</td>
<td>139</td>
<td>0.04</td>
</tr>
<tr>
<td>P@ssw0rd</td>
<td>107</td>
<td>0.03</td>
</tr>
<tr>
<td>Amsterdam1</td>
<td>102</td>
<td>0.03</td>
</tr>
<tr>
<td>Hallo123</td>
<td>101</td>
<td>0.03</td>
</tr>
<tr>
<td>Wachtwoord1</td>
<td>94</td>
<td>0.03</td>
</tr>
<tr>
<td>Geheim01</td>
<td>76</td>
<td>0.02</td>
</tr>
</tbody>
</table>

### 8ULNS passwords
Research method

1. Dutch domain password selection
2. Selecting different password cracking approaches
3. Comparing approaches using experiments
4. Selecting a well performing approach
Approach 1 of 3: Human behaviour on password selection

• Alan S. Brown et al. (2004)
• Generating and remembering passwords
• Questionnaire 218 US students
• Common content of basewords (e.g., reference to self, relative, animal, personal interest, job)
• Common use of basewords (e.g., complete word used in 97% of the time)
Approach 2 of 3: Probabilistic Context-Free Grammars

- Matt Weir et al. (2009)
- Breaking up and recombining passwords
- Frequencies important

```
Welkom2020!
L_6 \Rightarrow \text{welkom}
D_4 \Rightarrow 2020
S_1 \Rightarrow !
Rule \Rightarrow L_6D_4S_1
```
Approach 3 of 3: Generative Adversarial Network

- Hitaj et al. (2019)
- Learning how to generate “passwords”.
- Machine learning based
Research method

1. Dutch domain password selection
2. Selecting different password cracking approaches
3. **Comparing approaches using experiments**
4. Selecting a well performing approach
Comparing approaches in an experiment

Dutch domain passwords

Unique passwords
Generate PassGAN candidates
Rulesets: Best64, T0XICv1, OneRule

All passwords
Generate PCFG candidates

Categorized word lists
NL Dictionary

Benchmark

Dictionaries: Rockyou, Weakpass2

Graphs
Results

A NTLM 8ULNS (1338), OneRule

- 64.8%
- 58.3%
- 38.5%
- 65.4%

Retrieved unique passwords vs Time (minutes)

- ALL
- GANL
- PCFGL
- WEAK
Results

A NTLM 8ULNS (1338), OneRule

- Retrieved unique passwords
- Time (minutes)

- 35.3% for NLCOMBO
- 35.8% for NLLISTS
- 41.6% for NLPASS
- 41.7% for NLWORDS
- 11.5% for ROCKYOU
Research method

1. Dutch domain password selection
2. Selecting different password cracking approaches
3. Comparing approaches using experiments
4. Selecting a well performing approach
## Combining approaches

<table>
<thead>
<tr>
<th>Hashes</th>
<th>Hash type</th>
<th>Unique hashes</th>
<th>Ruleset</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>NTLM</td>
<td>1338</td>
<td>(none)</td>
<td>24.7%</td>
<td>30.7%</td>
<td>33.1%</td>
<td>33.6%</td>
<td>33.9%</td>
</tr>
<tr>
<td>A</td>
<td>NTLM</td>
<td>1338</td>
<td>Best64</td>
<td>37.0%</td>
<td>41.6%</td>
<td>43.0%</td>
<td>43.7%</td>
<td>44.2%</td>
</tr>
<tr>
<td>A</td>
<td>NTLM</td>
<td>1338</td>
<td>T0XICv1</td>
<td>58.4%</td>
<td>62.0%</td>
<td>63.5%</td>
<td>64.3%</td>
<td>64.8%</td>
</tr>
<tr>
<td>A</td>
<td>NTLM</td>
<td>1338</td>
<td>OneRule</td>
<td>65.4%</td>
<td>69.1%</td>
<td>70.0%</td>
<td>70.3%</td>
<td>70.6%</td>
</tr>
</tbody>
</table>
Combining approaches

1. Removing duplicates (9.1%)
2. Merging dictionaries by “weaving”
### Results

After 30 minutes

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>NLWOVEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCFGL</td>
<td>52.32%</td>
<td>55.75%</td>
</tr>
<tr>
<td>ALL</td>
<td>25.71%</td>
<td></td>
</tr>
<tr>
<td>NLWOVEN</td>
<td>55.75%</td>
<td></td>
</tr>
</tbody>
</table>

![Graph showing retrieval of unique passwords over time for different categories: PCFGL, ALL, NLWOVEN, with percentages for 38.5%, 58.3%, 64.8%, and 70.1% at various points in time.](image-url)
Results

After 30 minutes

<table>
<thead>
<tr>
<th>PCFGL</th>
<th>ALL</th>
<th>NLWOVEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.79%</td>
<td>7.34%</td>
<td>32.9%</td>
</tr>
</tbody>
</table>

B NTLM 8ULNS (1702), OneRule

- 22.9%
- 40.2%
- 50.2%
- 50.1%
- 53.5%
Research question

How do different password guessing algorithms compare in selecting probable password candidates for assessing offline Dutch domain password hashes?
Conclusions

• Single approaches perform well on one or two aspects:
   - total amount, amount 30 mins. or guesses per crack
• Combining approaches can:
  - Increase the total amount of passwords found:
    +7.2% and +6.7%
  - Increase the amount of passwords found within 30 mins:
    +117% and +348%
• Rulesets increase the amount of hashes found for all the selected approaches.
Discussion & Future work

• Dirty data in breach compilations
• Tests on two sets of hashes
• Consider adding organisation specific information
• Consider iterative cracking by using cracked passwords as input for further cracking
Acknowledgement

Secura
P. Campers
E. Slangen
R. Moonen

Scattered Secrets
J. van Beek
Questions

Conclusions:
  - Single approaches perform well on one or two aspects
  - Combining approaches can:
    • Increase the total amount of passwords found
    • Increase the amount of passwords found within 30 mins.
  - Rulesets increase the amount of hashes found for all the selected approaches.
“When processing requests to establish and change memorized secrets, verifiers SHALL compare the prospective secrets against a list that contains values known to be commonly-used, expected, or compromised. For example, the list MAY include, but is not limited to:

- Passwords obtained from previous breach corpuses.
- Dictionary words.
- Repetitive or sequential characters (e.g. ‘aaaaaa’, ‘1234abcd’).
- Context-specific words, such as the name of the service, the username, and derivatives thereof.”
Benchmark results

<table>
<thead>
<tr>
<th>Hashes</th>
<th>Ruleset</th>
<th>GANL</th>
<th>PCFGL</th>
<th>NLISTS</th>
<th>NLWORDS</th>
<th>NLPASS</th>
<th>NLCOMBO</th>
<th>ROCKYOU</th>
<th>WEAK</th>
<th>ALL</th>
<th>NLOVEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>A NTLM 1338</td>
<td>(none)</td>
<td>4.6%</td>
<td>24.4%</td>
<td>0.0%</td>
<td>0.2%</td>
<td>3.1%</td>
<td>3.1%</td>
<td>3.1%</td>
<td>22.2%</td>
<td>21.3%</td>
<td>33.2%</td>
</tr>
<tr>
<td>A NTLM 1338</td>
<td>Best64</td>
<td>10.8%</td>
<td>33.6%</td>
<td>1.6%</td>
<td>7.2%</td>
<td>9.9%</td>
<td>12.4%</td>
<td>12.0%</td>
<td>37.0%</td>
<td>36.6%</td>
<td>43.4%</td>
</tr>
<tr>
<td>A NTLM 1338</td>
<td>T0XICv1</td>
<td>26.6%</td>
<td>52.5%</td>
<td>5.4%</td>
<td>22.0%</td>
<td>25.4%</td>
<td>30.1%</td>
<td>32.1%</td>
<td>58.4%</td>
<td>58.4%</td>
<td>63.8%</td>
</tr>
<tr>
<td>A NTLM 1338</td>
<td>OneRule</td>
<td>39.5%</td>
<td>58.3%</td>
<td>11.5%</td>
<td>35.3%</td>
<td>35.8%</td>
<td>41.6%</td>
<td>41.7%</td>
<td>64.8%</td>
<td>65.4%</td>
<td>70.1%</td>
</tr>
<tr>
<td>B NTLM 1702</td>
<td>(none)</td>
<td>1.6%</td>
<td>7.9%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.8%</td>
<td>0.8%</td>
<td>0.9%</td>
<td>9.0%</td>
<td>8.5%</td>
<td>13.1%</td>
</tr>
<tr>
<td>B NTLM 1702</td>
<td>Best64</td>
<td>3.8%</td>
<td>12.6%</td>
<td>0.4%</td>
<td>1.6%</td>
<td>2.4%</td>
<td>3.0%</td>
<td>4.0%</td>
<td>17.6%</td>
<td>17.3%</td>
<td>22.2%</td>
</tr>
<tr>
<td>B NTLM 1702</td>
<td>T0XICv1</td>
<td>13.6%</td>
<td>28.3%</td>
<td>1.4%</td>
<td>5.1%</td>
<td>7.6%</td>
<td>8.8%</td>
<td>12.3%</td>
<td>38.9%</td>
<td>38.7%</td>
<td>43.9%</td>
</tr>
<tr>
<td>B NTLM 1702</td>
<td>OneRule</td>
<td>22.9%</td>
<td>40.2%</td>
<td>4.0%</td>
<td>12.7%</td>
<td>12.8%</td>
<td>16.8%</td>
<td>19.9%</td>
<td>50.2%</td>
<td>50.1%</td>
<td>53.5%</td>
</tr>
</tbody>
</table>
# Combining 2 approaches

<table>
<thead>
<tr>
<th></th>
<th>WEAK</th>
<th>ROCKYOU</th>
<th>NLCOMBO</th>
<th>GANL</th>
<th>PCFGL</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEAK</td>
<td>64.8%</td>
<td>64.8%</td>
<td>64.8%</td>
<td>66.9%</td>
<td>67.5%</td>
<td>65.8%</td>
</tr>
<tr>
<td>ROCKYOU</td>
<td>64.8%</td>
<td>41.7%</td>
<td>47.5%</td>
<td>52.5%</td>
<td>59.2%</td>
<td>65.4%</td>
</tr>
<tr>
<td>NLCOMBO</td>
<td>64.8%</td>
<td>47.5%</td>
<td>41.6%</td>
<td>53.1%</td>
<td>60.6%</td>
<td>65.4%</td>
</tr>
<tr>
<td>GANL</td>
<td>66.9%</td>
<td>52.5%</td>
<td>53.1%</td>
<td>39.5%</td>
<td>60.7%</td>
<td>67.3%</td>
</tr>
<tr>
<td>PCFGL</td>
<td>67.5%</td>
<td>59.2%</td>
<td>60.6%</td>
<td>60.7%</td>
<td>58.3%</td>
<td>67.9%</td>
</tr>
<tr>
<td>ALL</td>
<td>65.8%</td>
<td>65.4%</td>
<td>65.4%</td>
<td>67.3%</td>
<td>67.9%</td>
<td>65.4%</td>
</tr>
</tbody>
</table>
## Common basewords used in Dutch domain passwords

<table>
<thead>
<tr>
<th>Category</th>
<th>Matching passwords</th>
<th>Matching unique passwords</th>
<th>Elements in wordlist</th>
</tr>
</thead>
<tbody>
<tr>
<td>First names</td>
<td>531337</td>
<td>267085</td>
<td>9348</td>
</tr>
<tr>
<td>Family names</td>
<td>203503</td>
<td>107539</td>
<td>9113</td>
</tr>
<tr>
<td>Pet names</td>
<td>159859</td>
<td>71978</td>
<td>646</td>
</tr>
<tr>
<td>Cities and townships</td>
<td>64118</td>
<td>30498</td>
<td>7120</td>
</tr>
<tr>
<td>Comic character names</td>
<td>43515</td>
<td>19593</td>
<td>774</td>
</tr>
<tr>
<td>Animals</td>
<td>32178</td>
<td>13445</td>
<td>4924</td>
</tr>
<tr>
<td>Payed soccer teams</td>
<td>28638</td>
<td>8130</td>
<td>310</td>
</tr>
</tbody>
</table>

...