# Password security in practice

Magnus Andersson Elias Faxö
*Email: {magan472,elifa571}@student.liu.se*
Supervisor: Viiveke Fåk, {viiveke@isy.liu.se}
Project Report for Information Security Course
*Linköpings universitet, Sweden*

## Abstract

*Authentication using passwords is not only one of the most popular methods today but also one of the most insecure [7]. In order to address these security problems regarding passwords the classical approach has been to enforce a password policy putting certain demands on the quality of the passwords used. These policies, however, seldom take the social aspects into consideration, such as how users actually respond to them, and still places a great deal of trust into the users' ability to select a secure password. This report aims to explore these social aspects in a quantitative way and merge with the classical combinatorial aspect on password security in order to present a more correct interpretation of the security provided by enforcing classical password policies.*

## 1. Introduction

Password security and management is a virtual jungle in terms of different theories and best practices in which it is a difficult task to navigate and find any true answers regarding the actual impact or effect of enforcing a strict password policy in terms of complexity and resilience against social engineering attacks. Moreover the demands on secure passwords and secure password handling has changed much lately, during the past years there has been a great increase of phishing attacks reaching an all time high in august 2009 [1]. These types of threats against security is not considered in established password policies [4] and fundamentally changes the considerations that should be taken when securing a system.

### 1.1 Problem definition

Keeping social considerations in mind is important due to the increasingly large number of passwords handled by the average person [2] which is reaching a critical limit. In addition, the enforcement of classical policies requiring users to change their passwords at given intervals or use a mixing of capitalized and non-capitalized letter, numbers and special characters in an attempt to make passwords harder to guess, makes matters even worse in terms of memorability. The concept of memorability might seem just a smaller inconvenience, but might in fact indirectly compromise the integrity of the whole system and will be included as an important part in the scope of the report as a basis for discussions regarding the actual security impact of different methods used today.

An additional social aspect that on a fundamental level affects the security of passwords is that it is the users that own the password. There is no way to know what the users might do with the password or what kind of password they might choose to use. If the users are free to choose passwords, many tend to choose an easily remembered password such as a sequence, something that they can relate to the service or simply names of people they know [5]. This, which is also covered in "Consumer Password Worst Practices" [5], poses a great threat to security enabling attackers to brute-force guess the passwords fairly easy, but what is not covered in the article is how enforcing a more elaborate classical password policy would affect the security of the passwords and is instead just assumed – on unclear grounds – to strengthen the resilience against such attacks. Based on a survey performed amongst students at Linköping university and a combinatorial analysis of the implications from different policies on password guessing, these issues will be addressed in an attempt to quantify the effective result on a per policy basis.

These problems and analyses are addressed thoroughly in the report next to the fact that users tend to share their passwords between multiple services [3], thus forcing an undesirable trust relationship between the services, a problem that in many ways shares the complications and implications with the phishing issue. This phenomenon or problem also is based on the fact that the users own their password – and thus are free to do whatever they want with it – and some reasoning concerning how this problem can be mitigated by enforcing policies and practices is included in the analysis as well.

### 1.2 Question formulations

- What effect does enforcing classical password policies have from a combinatorial viewpoint?
- What effect does enforcing classical password policies have from a social viewpoint?
- How does password policies affect users choice of passwords?
- Can it be proven that enforcement of classical password polices increases security?

### 1.3 Report structure

The background section will present some general assumptions taken regarding passwords and how they are used to define the scope of the report on a more detailed level. The theories and methods used in the survey and the combinatorial analysis will also be presented, including motivations to the choices regarding the conduction of the survey and to the choice of combinatorial methods.

In the analysis section the result and analysis of a user survey on how users manage and chose their passwords will be presented, and further discussed with respect to the problems presented in the introduction of this report. The combinatorial analysis will also be presented with results concerning complexity and brute-force possibilities.

Other authors´ work on this subject will be presented under the related work section including a brief evaluation and comparison in relation to this report as well as amongst themselves.

At the end of the report you will find the conclusion section containing a summary of the conclusions drawn described in a non-technical fashion. The conclusion will also contain suggestions of secure best practices regarding password policies and management on an organizational as well as individual level. There will also be a brief summary on what could not be covered in this report and suggestions on further work to be performed on the subject.

## 2. Background

The definition of a password as used in this report refers to a string of characters that the personal owner is keeping as a secret and is used for authentication to gain access to a resource, typically on the Internet. Furthermore, a password is assumed to be able to contain uppercase and lowercase alphabetical characters, of which it exists 26 of each in the English language, numerical characters of which it exists 10 and special characters of which it exists 24 (._!-*@#?/$,\&+=)(';<%]:[ that are used in more than 0.02% of all passwords, see Appendix A: Frequency table). These figures totaling at 86 will be used in the combinatorial analysis performed in the analysis section of this report. The term password policy refers to the set of rules that can be applied to a password to require certain properties on the passwords used. Classical password policies include restrictions as minimum password lengths, maximum password age and/or complexity requirements on the password.

**Table 1. Identified sets of characters**

| Collection | Cardinality |
|---|---|
| Uppercase alphabetical | 26 |
| Lowercase alphabetical | 26 |
| Numerical characters | 10 |
| Special characters | 24 |

### 2.1 Theoretical Methods

The report will present methodologies from a number of classical combinatorial branches such as number theory, set theory and probability. These methods were used to develop further insight into the actual effect on brute force complexity by different policy properties. Amongst those utilized most frequently are the principle of inclusion and exclusion, which is a way of avoiding double – or even triple – counting when calculating the intersection of different sets, developed by Abraham de Moivre [8] which in turn is a generalization made on the Venn visualization of all hypothetically possible logical relations between collection [18]. In the case of this report the collections used were the grouped sets of characters (numerical, special character, uppercase and lowercase alphabetical) which all had their predetermined cardinality (see Table 1).

The actual complexity has been calculated by determining the number of permutations possible in the different subsets of relations in the Venn diagrams. The number of permutations containing 8 characters possible on a set of 24 characters are $24^8$ and on a 9 character word there are $24^9$ and so on.

### 2.2 Practical Methods

In addition to the theoretical methods presented the data used in the analysis was gathered through a survey and a set of interviews. The survey was designed with nine questions and with the aim of gathering quantitative categorical data about the usage of passwords and how people respond to various password policies in their choice of passwords. The survey was sent out as a digital form to various groups of students at Linköping university. As an addition to the survey a set of interviews were conducted in order to complement the survey. These interviews targeted the same groups of people as the survey.

The purpose of these methods was to gather quantitative as well as qualitative data to aid in analysis in regard to the problems presented in the introduction with the qualitative data gathered from the interviews as the base of the hypotheses.

### 2.3 Context and limitations

The research done for this report is limited to students at Linköping university. This restriction is due to the limitations in time, as this project is only a smaller part of a course in information security given at Linköping university. This restriction should be taken into account when utilizing the results presented in this report, as they cannot be considered feasible in a global scope or in any other context, where security awareness differs from the subjects in the survey performed in conjunction with this report.

## 3. Solution and Analysis

In order to answer the questions formulated in the introduction a reference policy has been developed based on parts of the NASA "Weak Passwords" recommendation checklist [6] where a password policy usually is used to enforce the requirements. This resulted in the following reference policy which was used in the analysis.

- Passwords must contain at least eight characters.
- Passwords must contain a mix of four different types of characters (lowercase and uppercase alphabetical, numeric, special characters).
- If there is only one special, numeric or alphabetical character, it must not be either the first nor the last character in the password.

### 3.1 Combinatorial analysis

As in all good mathematical proofs, let's start from the beginning. Password complexity has – looking at it from a combinatorial viewpoint – the beauty of, at least at the first glance, being infinitely complex. This, however, is far from the truth since the infinite complexity is based on that a password can – theoretically – be infinitely long. In reality passwords seldom reach above ten characters [5], and proceeding in this analysis the focus will be reduced to 93% of all passwords and thus assume that all passwords are below or equal to ten characters in length.

This assumption, keeping 93% of all passwords in the target surface, reduces the complexity from the baffling infinite to a mere – at least in comparison – $22.39 \times 10^{18}$ (Equation I) making this analysis a bit more interesting.

**Equation I**
$$86^{10} + 86^{9} + .. + 86^{1} = 22390512687494871810$$

Now, let's consider the previously declared reference policy and start to recalculate these figures. Starting with the first rule regulating the length of the password (Equation II) this gives us a small reduction of complexity by roughly $1{,}57 \times 10^{-4}$ percent.

**Equation II**
$$86^{10} + 86^{9} + 86^{8} = 22390477485385800448$$

Moving on to the second property of the reference policy demanding a mix of all four character types to be used in the password further reduction of the complexity can be achieved. Logical reasoning and the principle of inclusion and exclusion [8] based from the Venn diagram presented in Figure 1 gives a set of new equations to calculate the effective complexity $C_{8,9,10}$.

**Equation III**
$$C_8 = 86^8 - 62^8 - 76^8 - 60^8 + 34^8 + 50^8 + 50^8 + 36^8 + 36^8 + 52^8 = 16.318555202772 \times 10^{14}$$

$$C_9 = 86^9 - 62^9 - 76^9 - 60^9 + 34^9 + 50^9 + 50^9 + 36^9 + 36^9 + 52^9 = 15.607198370856 \times 10^{16}$$

$$C_{10} = 86^{10} - 62^{10} - 76^{10} - 60^{10} + 34^{10} + 50^{10} + 50^{10} + 36^{10} + 36^{10} + 52^{10} = 14.606552131178 \times 10^{18}$$

$$C_{8,9,10} = C_8 + C_9 + C_{10} = 14.764255970407 \times 10^{18}$$

The new complexity of $14.76 \times 10^{18}$ possible passwords is in fact a reduction of roughly 36.02 percent from the previous complexity, however, the complexity can still be considered high but one property still needs to be explored.

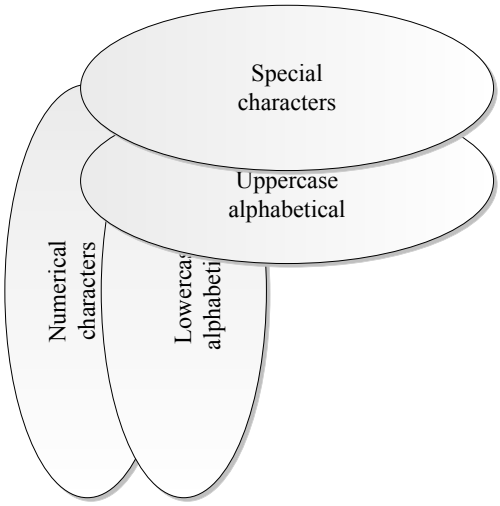


**Figure 1. Venn diagram of the character types.**

The third and the last property of the reference policy denotes that, if there is only one numerical, alphabetical or special character, it is not allowed to be the first nor the last in the password. The impact of this property is solved through a similar approach as the previous property starting from a Venn diagram (Figure 2) and resulting in the distinct passwords containing only one instance of a particular character type. The calculation consists of four very similar

steps of which only one will be presented along with the overall result.

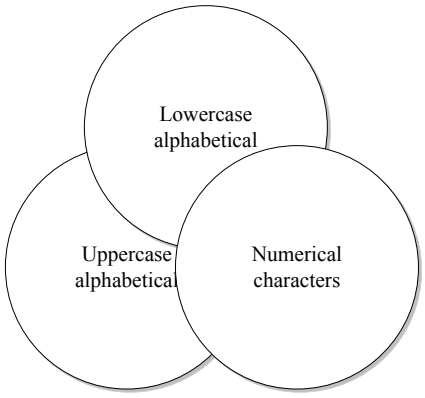


**Figure 2. Venn diagram of password composition when a special character is assumed to be the first or the last.**

The Venn diagram (Figure 2) visualizes the possible compositions of a password assuming that the first or the last character is a special character and none of the others are. This is further explored mathematically in Equation IV and Equation V, where the principle of inclusion and exclusion is utilized once again.

**Equation IV**

$$U \cup L \cup N - U \cup L - L \cup N - N \cup U + U + L + N$$

Inserting the numerical values into the equation and then multiplying with the number of permutations that comes from the assumption of a special character either as the first or the last character gives the sought values.

**Equation V**

$$\begin{aligned}&(62^x - 52^x - 36^x - 36^x + 26^x \\ &+ 26^x + 10^x) \times 48 = \langle x = \{7,8,9\}\rangle \\ &= 10.726929556212 \times 10^{15} \times 48 \\ &= 55.780033692305 \times 10^{16}\end{aligned}$$

Repeating this for all four character classes ends up with the conclusion that another $52.118141765158 \times 10^{17}$ passwords will be rejected by the policy summing up to a reduction of another 35.30 percent. The grand total is $9.19 \times 10^{18}$ possible passwords which is a total reduction of 58.96 percent from the initial set. This, however, is - despite the substantial reduction - far from enough to exploit a system. It would still typically take over 150 milion years to breach passwords of this strength assuming that an attacker can make 1000 attempts per second.

## 3.2 Social analysis

Picking up where the combinatorial analysis left off the social perspective of passwords and how they are used, chosen and remembered will help in reducing the complexity further. Even though complexity was found to be considered beyond brute forcibility, the fact that 84 percent of computer users consider memorability to be the most important attribute of a password [10] opens up for new ways of cracking passwords.

As derived from the survey performed in conjunction with this report, 52 percent (see Figure 3) of the subjects use a word as the baseline for their password. This, however, does not – as is made apparent in the survey – mean that they use a plain dictionary word but merely a variation of one. Considering the fact that the English language consists of approximately 171476 words in current use [9] and the reference policy, the 52 percent of the passwords that where based on a word has a real complexity that should be significantly less than previously concluded in the combinatorial analysis.

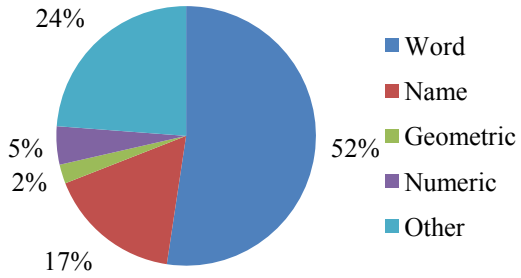


**Figure 3. Type distribution of passwords.**

This conclusion is based on the variation of what is considered a word specified in the survey, namely a word appended or prepended by one or several numerical characters or letters substituted with numbers and with mixed case. This of course extends the number of possible words beyond the 171476 in a English dictionary. The exact number of possibilities following these rules and using the substitutions presented in Table 2 is hard to define, however, so an approximation will be made later on in the analysis, when the combinatorial perspective and the social results will be combined.

But first things first. The question whether the enforcement of a policy can make the users choose more secure passwords still must be answered. According to the survey most users, 38 percent, facing a policy forcing them to have numerical characters in their password would simply append a number to their original password if it did not already contain enough to satisfy the policy, while 36

percent would insert a numerical character at an unspecified (not the beginning nor the end) location in the password. When forced to add a mixture of uppercase and lowercase letters to the password 49 percent claim they would change the very first letter in the password to uppercase and according to the informal interviews conducted policies such as our reference policy would probably lead to the first two letters being uppercase.

The most interesting result derived from the survey is regarding the users´ response to getting forced to change their password at a given interval. 34 percent claims that they would just make a minor modification to their existing password rather than change it entirely, as is visualized in Figure 4. This negates the positive effects of this policy attribute to a high extent for over $1/3$ of the users, bringing up the discussion whether its security increasing effect adequately compensates for the low social acceptance derived from the interviews regarding this policy property and the discussions on poor cost-benefit in [11].

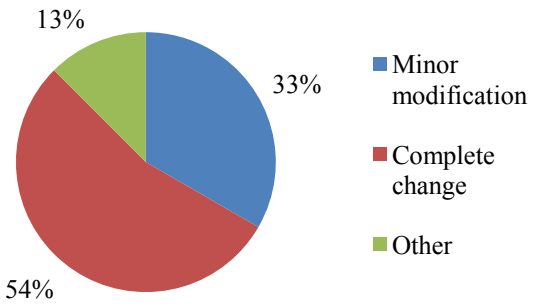


**Figure 4. Responses to periodically forced changes.**

### 3.3 Back to the math

Now, as previously promised, the impact of these responses in relation to our reference policy will be calculated and used to estimate the real complexity of user passwords. However, not all types of password will be analyzed due to limitations in time and space so the focus is on word based passwords from here on, reducing the target surface to approximately 49 percent of all passwords. Which is based on the survey result combined with the length distribution analysis presented in [5] for passwords below 10 characters in length.

The next task is to reduce the number of words to those between 8 and 10 letters, but allowing the possibility of two numerical or special character appended or prepended the calculation is of those between 6 and 10 letters in length.

Assuming there is a normal (Gaussian) distribution of the length of words and the average length is 5,1 [17] the number of words exceeding 6 letters and below 10 letters in length should be approximately 33 percent or 56587 (Equation VII) with the standard deviation $\sigma$ set to 3 (Equation VI). Where the deviation is calculated from the 4484 words of this report (in writing) excluding names, numerical values and words in figures and tables.

**Equation VI**

$$\sqrt{\sum_{i=1}^{4484} W_i \Big/ 4484} = 2.9907991527033$$

Where $W_i$ is the absolute deviation of the word length from the mean.

**Equation VII**

$$\int_{-\infty}^{10} f(x) - \int_{-\infty}^{6} f(x) = \left\langle f(x) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right\rangle = 0.330889$$

**Table 2. Alphabetical substitution table.**

| Substitutee | In [16] | Substitut |
|---|---|---|
| A | 8,12% | 4 |
| I | 7,31% | 1 |
| L | 3,98% | 1 |
| E | 12,08% | 3 |
| O | 7,68% | 0 |
| B | 1,49% | 8 |
| S | 6,28% | 5 |

The contents of Table 2 is not statistically based but should contain the most used substitutions according to the experience of the authors. Table 2 also contains the typical frequency of the letters in a word which by calculation increases the 56587 words presented earlier to 83148 (Equation VIII) possible words, not considering that one word can have several of the substitutable letters.

**Equation VIII**

$$56587 + 56587 \times (0.0812 + 0.0731 + 0.0398 + 0.1208 + 0.0768 + 0.0149 + 0.0628) \approx 83148$$

This figure is further increased when considering the possibilities of two numerical letters or two special characters before and after the password (Equation IX) resulting in 112416096 distinct passwords, observe that this is an approximation not taking into account that some of the words would get longer than 10 letters when adding these.

**Equation IX**

$$83148 \times (100 \times 2 + 576 \times 2) = 112416096$$

Actually these two properties should not just be considered but required based on the reference policy, assuming that no English words contain numbers or special characters originally, meaning that the original $56587$ should be removed from the set. The result now sums up to $112359509$ and the time has come to consider the distinction between upper and lowercase letters.

Based on the reference policy a maximum of 6 to 8 characters can be alphabetical depending on the total length of the password. At least one of these has to be of different case from the others. However, to simplify the calculation no consideration is taken to how many characters the word initially had and they are, in contradiction to the previous assumption assumed to be distributed equally rather than Gaussian.

**Equation X**

$$\begin{aligned} &\frac{2^6 \times 112359509}{3} \\ &+\frac{2^7 \times 112359509}{3} \\ &+\frac{2^8 \times 112359509}{3} \\ &\approx 2397002858 \\ &+ 4794005717 \\ &+ 9588011434 \\ &= 16779020009 \end{aligned}$$

This very positive – in regards to assumptions and simplifications – analysis concludes that there are approximately $16779020009$ passwords following the definition of a word presented in the survey, a definition 49 percent of all passwords seem to follow. Passwords of this complexity would take no more than 194 days to break making the same assumption regarding attack performance as before.

## 4. Related work

The subject of password policies and their actual effect on security as well as on other aspects of business have been attended in many studies besides this. Even though none has been found by the authors that has the exact same approach, however, there is much work that complements the work presented in this report in a beneficial way.

In a study conducted by Microsoft, principal researcher Cormac Herley discusses the economical aspects of password policies and concludes that "Most security advice simply offers a poor cost-benefit trade-off to users" [11] alongside the fact that most users are aware of this and therefore tend to ignore or find shortcuts around the policies as a way of saving time and effort. The study complements the work presented in this report with the economical impacts of enforcing policies and further strengthen the conclusion presented here as to the users actual response to policies from a security perspective.

In the article "Security Nightmare: How do you maintain 21 different passwords?" by Graham Hayday [10] there are some complementing statistics to the ones presented from the survey in this report. In the survey from NTA Monitor they present that 81 percent of all users select a common word as their password, supposedly as a result of 84 percent considering memorability the most important attribute of a password.

As to what solutions there are to the problem of memorability versus complexity, there are many theories and opinions circling around. One of the strongest opinions comes from Jesper Johansson, security expert at Microsoft, who claims that there is less a risk in writing down your password then to reuse the passwords for multiple services [12]. This opinion is backed by Bruce Schneider, Chief Security Technology Officer of BT [13] and can be part of a solution to keeping passwords memorable, complex and unique for each service at the same time.

Statistics compiled by Verzion Business RISK Team shows that hacking was responsible for 64 percent of the data breaches in 2008 and of these 64 percent roughly 10 percent where due to stolen credentials while almost 25 percent where due to shared credentials [14]. This further emphasizes the importance of not sharing passwords between different services and confirms it as a greater threat then both phishing and keeping copies of the passwords for memorability.

The works discussed here is just a small selected portion of the work available on the subject and were chosen as they were considered good complements to the work presented in this report, taking other aspects of password security into consideration.

## 5. Conclusions

From the initial combinatorial analysis a couple of key conclusions can be drawn. Firstly, which of the properties actually affects the theoretical complexity. This is visualized in Figure 5, and as is made clear the most substantial reduction originates from the second property of the reference policy denoting that all four sets of characters must be present in all passwords. This property alone decreases complexity by 36 percent, making a substantial impact on overall complexity and brute force resilience. The second largest decrease in complexity originates from the third property while the first, forcing the minimum length to be eight characters has nearly no effect on complexity at all. While these findings may be interesting when designing policies the most important conclusion drawn from the combinatorial analysis is simply that employing a combinatorial approach to password cracking will not take you far – enough. Even with the total reduction of 58,96

percent, the passwords should be considered impossible to crack using a brute force approach.

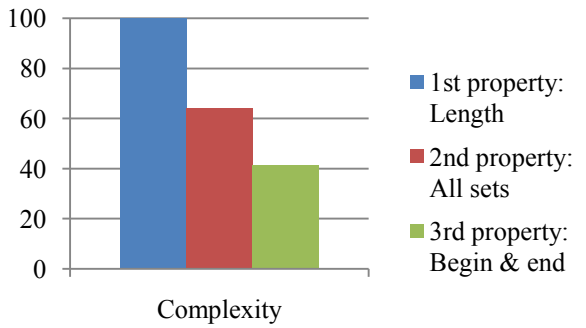


**Figure 5. Complexity reduction from policy properties.**

The social analysis provided new hope after the initial combinatorial analysis proposing new ways of looking at password cracking. The mere fact that 52 percent of all passwords where words or variations thereof opens up for much improvement to the result from the combinatorial approach.

The most obvious conclusion that can be drawn from the social analysis is how users tend to work against policies rather than with the policies in finding ways to promote memorability rather than security. This issue does not seem to be mitigated by the enforcement of classical password policies as preferred, and awakes the question if focus should not be put on promoting awareness of the users rather than simply enforce a policy which the users tend to negate. This discussion will not, however, be more elaborated in this report other then this mentioning.

Lastly, the combined social and combinatorial analysis of the word based passwords provides a set of conclusions regarding both policies and the practical password security in many systems. The definition of a word used in this report is formulated to open up for much modifications, such as appending numbers, replacing letters etc. As is shown in chapter 3.3 this is a source of escalading complexity increasing the initial 56587 words possible to fulfill the reference policy to approximately 16779020009 possible passwords based on these words. The observation of this escalation leads to a question of how much impact an increment of the initial word base would affect the total complexity. Approximate calculations that are not presented in detail here based on Equation VIII, Equation IX and Equation X, where the reference policy has not been taken into consideration, concludes that the number of possible passwords will increase to 69698633347 or a 415 percent increase of complexity from previously.

The conclusions drawn from this reasoning is not just that it is fully possible to brute force password based system by considering how users actually choose their password but also, and probably more importantly, that when reasoning with the assumption that a word is used as the base of the password, then strict policies reduce rather than increase complexity leaving the system more vulnerable to attacks.

Further, there is a fair assumption to make that this approach would be equally successful when reasoning around name based and geometric based passwords. However, the number of names can be assumed to far outnumber the number of words especially considering company and product names.

As far as answering the question if classical policies can be proven beneficial to security the answer would be no, however, there are policies that probably would succeed in increasing security based on the findings in this report. Since the by far largest source of vulnerabilities seems to be the social aspect, or the users, a solution would be to simply remove control from the users and instead let the system choose the password in a non-predictable way. This solution would off course have its issues regarding memorability but as is mentioned in [12] and [13], there are a lot of things worse than writing down the password on a piece of paper. This conclusion goes out to the users as well, prefer a safe password over memorability to the highest extent possible.

The conclusions draw in this report are based on approximations leaving much to be explored further and on a more detailed level, such as what policy properties can be used with acceptable impact on complexity, and how users respond to policies through a case study rather than a hypothetical survey. These topics are both subjects for further studies and potential sources of error in this report.

## 6. References


[1] "Phishing Activity Trends Report 4th quarter 2009", APWG.

[2] Yuki Noguchi, "Access denied", The Washington Post, 09-23-2006.

[3] Anna Pickard, "Are you suffering from password pressure?", The Guardian, 01-17-2008

[4] "Creating a Strong Password Policy", Microsoft, 09-30-2009.

[5] "Consumer Password Worst Practices", The Imperva Application Defense Center, 2010.

[6] "Weak Passwords", NASA, 1998.

[7] Bruce Schneider, "Secrets and Lies", ISBN 0-471-25311-1, 2000.

[8] Weisstein, Eric W. "Inclusion-Exclusion Principle", MathWorld, 2010.

[9] Jessica Stone, "How many words are there in the English language?", 12-09-2009.

[10] Graham Hayday, "Security Nightmare: How do you maintain 21 different passwords?", Silicon.com, 12-11-2002.

[11] Cormac Herley, "So Long, And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users", Microsoft Research, 2009.

[12] Munir Kotadia, "Microsoft security guru: Jot down your passwords", CNET News, 05-23-2005.
[13] Bruce Schneider, "Write down your password", Schneider on Security, 06-17-2005.
[14] "2009 Data Breach Investigation Report", Verzion Business RISK Team, 2010
[15] "Basic English combined wordlist", Wikipedia, 2010.
[16] "English Letter Frequency", Cornell University Math Explorer's Project.
[17] Patrick Hall, "Languages by Average word length", Hacklog, 06-26-2007.
[18] J. Venn, "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings", Philosophical Magazine and Journal of Science, Series 5, vol. 10, No. 59, July 1880.

## 7. Appendix A: Frequency table

This frequency table is based on the data from the breach of the RockYou.com password database in 2009 containing 32603388 passwords and the figures are calculated based on this data rounded to 5 significant figures.

| Chr | Total | Unique | In |
|---|---|---|---|
| a | 20969579 | 15227221 | 46.70441% |
| e | 18551202 | 14145337 | 43.3861% |
| i | 13489542 | 11475764 | 35.19807% |
| 1 | 13093321 | 10038459 | 30.78962% |
| o | 12815713 | 9998821 | 30.66804% |
| n | 11941319 | 10000129 | 30.67205% |
| r | 11629448 | 10211248 | 31.31959% |
| l | 11196140 | 9090648 | 27.88253% |
| s | 10495004 | 8676775 | 26.61311% |
| 0 | 9334773 | 6168635 | 18.92023% |
| 2 | 9328236 | 7335381 | 22.49883% |
| t | 8458405 | 7211768 | 22.11969% |
| m | 7283040 | 6354985 | 19.49179% |
| c | 6476021 | 5793658 | 17.77011% |
| 3 | 6377410 | 5299064 | 16.25311% |
| 9 | 6129514 | 4591014 | 14.0814% |
| d | 5807157 | 5148126 | 15.79016% |
| y | 5746936 | 5331834 | 16.35362% |
| h | 5543670 | 5100001 | 15.64255% |
| 8 | 5535756 | 4309671 | 13.21848% |
| 5 | 5450309 | 4376304 | 13.42285% |
| 4 | 5409308 | 4504058 | 13.81469% |
| u | 5312268 | 4835459 | 14.83115% |
| b | 5161941 | 4201896 | 12.88791% |
| 6 | 4956688 | 4031099 | 12.36405% |
| 7 | 4649633 | 3745538 | 11.48819% |
| k | 4265694 | 3894369 | 11.94468% |
| g | 4033890 | 3589523 | 11.00966% |
| p | 4033501 | 3437191 | 10.54244% |
| j | 2545146 | 2377799 | 7.2931% |
| v | 2474159 | 2371128 | 7.27264% |
| f | 2216339 | 1959117 | 6.00894% |
| w | 1898445 | 1781263 | 5.46343% |
| z | 1363825 | 1186305 | 3.63859% |
| x | 931975 | 800482 | 2.45521% |
| q | 389649 | 367616 | 1.12754% |
| . | 301996 | 226990 | 0.69622% |
| _ | 211764 | 187093 | 0.57385% |
| ! | 211480 | 179678 | 0.5511% |
| - | 155560 | 127099 | 0.38983% |
| * | 150389 | 95405 | 0.29262% |
| | 126501 | 92273 | 0.28302% |
| @ | 117398 | 104335 | 0.32001% |
| # | 65058 | 60025 | 0.18411% |
| ? | 63964 | 24279 | 0.07447% |
| / | 58621 | 37850 | 0.11609% |
| $ | 42851 | 31501 | 0.09662% |
| , | 33907 | 27735 | 0.08507% |
| \ | 31091 | 9008 | 0.02763% |
| & | 30947 | 28814 | 0.08838% |
| + | 30096 | 24001 | 0.07362% |
| = | 26391 | 18745 | 0.05749% |
| ) | 20469 | 18362 | 0.05632% |
| ( | 18400 | 16583 | 0.05086% |
| ' | 17738 | 16142 | 0.04951% |
| ; | 16442 | 14414 | 0.04421% |
| < | 13905 | 11858 | 0.03637% |
| % | 12932 | 11385 | 0.03492% |
| " | 12683 | 3214 | 0.00986% |
| ] | 12230 | 10735 | 0.03293% |
| ~ | 9145 | 5824 | 0.01786% |
| : | 8809 | 7243 | 0.02222% |
| [ | 8706 | 7725 | 0.02369% |
| ^ | 7528 | 5863 | 0.01798% |
| ` | 6232 | 5006 | 0.01535% |
| ñ | 5543 | 5319 | 0.01631% |
| > | 3872 | 2766 | 0.00848% |
| { | 1135 | 1059 | 0.00325% |
| } | 1041 | 943 | 0.00289% |
| ç | 883 | 779 | 0.00239% |
| £ | 786 | 599 | 0.00184% |
| \| | 731 | 506 | 0.00155% |
| é | 601 | 499 | 0.00153% |
| ö | 514 | 446 | 0.00137% |
| ü | 508 | 452 | 0.00139% |
| Ñ | 373 | 360 | 0.0011% |

| | | | |
|---|---|---|---|
| ´ | 368 | 331 | 0.00102% |
| á | 341 | 320 | 0.00098% |
| ä | 322 | 283 | 0.00087% |
| à | 252 | 166 | 0.00051% |
| º | 193 | 128 | 0.00039% |
| ó | 157 | 153 | 0.00047% |
| è | 148 | 129 | 0.0004% |
| ø | 129 | 121 | 0.00037% |
| í | 97 | 93 | 0.00029% |
| å | 94 | 81 | 0.00025% |
| ¡ | 94 | 76 | 0.00023% |
| ß | 84 | 76 | 0.00023% |
| § | 81 | 59 | 0.00018% |
| | 72 | 17 | 5.0E-5% |
| æ | 69 | 61 | 0.00019% |
| ã | 66 | 66 | 0.0002% |
|  | 59 | 57 | 0.00017% |
| ò | 54 | 30 | 9.0E-5% |
| ¿ | 53 | 50 | 0.00015% |
| ¨ | 50 | 49 | 0.00015% |
| · | 42 | 39 | 0.00012% |
|  | 42 | 37 | 0.00011% |
| Ç | 40 | 40 | 0.00012% |
| ¦ | 36 | 8 | 2.0E-5% |
| ù | 34 | 31 | 0.0001% |
| ì | 32 | 32 | 0.0001% |
| ú | 31 | 30 | 9.0E-5% |
| ° | 31 | 15 | 5.0E-5% |
| µ | 30 | 22 | 7.0E-5% |
| ª | 30 | 17 | 5.0E-5% |
|  | 28 | 21 | 6.0E-5% |
| ¸ | 27 | 10 | 3.0E-5% |
|  | 26 | 24 | 7.0E-5% |
| ¶ | 26 | 6 | 2.0E-5% |
| ô | 23 | 23 | 7.0E-5% |
|  | 23 | 10 | 3.0E-5% |
| ¬ | 19 | 17 | 5.0E-5% |
| Ö | 19 | 15 | 5.0E-5% |
| â | 18 | 18 | 6.0E-5% |
| « | 17 | 10 | 3.0E-5% |
| ý | 15 | 15 | 5.0E-5% |
| Ä | 15 | 15 | 5.0E-5% |
| Á | 15 | 14 | 4.0E-5% |
| Ü | 14 | 13 | 4.0E-5% |
| ¤ | 14 | 8 | 2.0E-5% |
| ¢ | 14 | 7 | 2.0E-5% |
| î | 12 | 10 | 3.0E-5% |
| É | 11 | 11 | 3.0E-5% |
| ê | 11 | 10 | 3.0E-5% |
|  | 11 | 5 | 2.0E-5% |
| õ | 10 | 10 | 3.0E-5% |
| ² | 10 | 9 | 3.0E-5% |
| ® | 10 | 9 | 3.0E-5% |
| » | 10 | 7 | 2.0E-5% |
| Ã | 9 | 9 | 3.0E-5% |
| © | 8 | 7 | 2.0E-5% |
|  | 8 | 5 | 2.0E-5% |
|  | 8 | 5 | 2.0E-5% |
| ð | 7 | 6 | 2.0E-5% |
| ¹ | 7 | 6 | 2.0E-5% |
|  | 7 | 5 | 2.0E-5% |
| | 6 | 6 | 2.0E-5% |
| Ó | 6 | 6 | 2.0E-5% |
| ë | 6 | 6 | 2.0E-5% |
| þ | 6 | 4 | 1.0E-5% |
| ÿ | 5 | 5 | 2.0E-5% |
| Æ | 5 | 4 | 1.0E-5% |
| ¥ | 5 | 4 | 1.0E-5% |
| û | 5 | 4 | 1.0E-5% |
| ¯ | 5 | 3 | 1.0E-5% |
| Í | 4 | 4 | 1.0E-5% |
| À | 4 | 4 | 1.0E-5% |
| ½ | 4 | 3 | 1.0E-5% |
|  | 4 | 3 | 1.0E-5% |
|  | 4 | 2 | 1.0E-5% |
| Ò | 3 | 3 | 1.0E-5% |
| Å | 3 | 3 | 1.0E-5% |
| Ê | 3 | 3 | 1.0E-5% |
| Ø | 3 | 3 | 1.0E-5% |
|  | 3 | 2 | 1.0E-5% |
|  | 2 | 2 | 1.0E-5% |
| Đ | 2 | 2 | 1.0E-5% |

| | | | |
|---|---|---|---|
| ± | 2 | 2 | 1.0E-5% |
| Ô | 2 | 2 | 1.0E-5% |
|  | 2 | 2 | 1.0E-5% |
| Ë | 2 | 2 | 1.0E-5% |
| ³ | 2 | 2 | 1.0E-5% |
| Ì | 2 | 2 | 1.0E-5% |
| | 2 | 1 | 0% |
| Â | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
| Î | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
| | 1 | 1 | 0% |
| □ | 1 | 1 | 0% |
| | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
| ï | 1 | 1 | 0% |
| × | 1 | 1 | 0% |
| Õ | 1 | 1 | 0% |
|  | 1 | 1 | 0% |
| Ï | 1 | 1 | 0% |
| ÷ | 1 | 1 | 0% |
|  | 1 | 1 | 0% |