KEYSTROKE DYNAMICS

Evaluation of suitability as a biometric authentication method

---

Patryk Świder, u1971957
Web Programming with Cyber Security
University of Huddersfield
December, 2020

Table of Contents

Introduction

The world is moving towards digitalisation faster than ever. A number of information systems and cyber criminals is rapidly increasing. Therefore, researchers constantly seek for improvements in the field of authentication methods to raise levels of both security and usability. Despite the provided higher level of security, users are impatient and tedious about the topic. A combination of multiple authentication methods - two factor authentication, increases security at the expense of convenience. Biometric authentication methods fulfil the usability criterium, however due to their nature, some are lacking in terms of high security standards.

This report focuses on one of the biometric authentication mechanisms - keystroke dynamics. Everyone writes passwords in a certain way - has a unique typing rhythm, a bit like DNA. At the moment of entering the password, various timings between keystrokes are collected and processed (feature sets), then compared with an earlier generated model, which evaluates whether to give access or not. The evaluation process is based on the Manhattan distance measure and thresholds.

The main objective is to evaluate the method and understand relationships between different password characteristics and accuracy. Chosen passwords are 6, 8, 10 and 12 characters long - two passwords of each length. Additionally, each length has two passwords with substitution of uppercase, numeric, special and combination of all three, which is a total of 40 various passwords. A full list of passwords is available in the appendices section.

Analysis shows that the accuracy is between 89-95% and average of 89.89%. The most accurate were short passwords with no substitutions, which is contrary to a strong password policy.

Analysis suggests that keystroke dynamics could be suitable as a biometric authentication mechanism, however under several conditions only. Due to its nature, this method has some security flaws, although it is convenient. Different passwords, hardware setups, times of day/week, health condition and other factors significantly affect the accuracy, therefore it is rather not relevant for high risk systems.

Analysis

The dataset is analysed in terms of accuracy, best feature sets and typing duration with a division into groups by length and substitution. The data is acquired from 64 subjects, who typed the 40 passwords 32 times each over a period of 8 weeks.

Accuracy examination is based on the average accuracies of passwords in subsequent weeks, which results in 320 values (40 passwords * 8 weeks). The best feature set analysis is based on the data of the individual subjects. Due to incompleteness, data of 11 subjects had to be removed - resulting in 16960 best feature sets (53 subjects * 40 passwords * 8 weeks).

To increase efficiency and reduce the risk of error, Python was used. It was most useful in the case of the best feature sets and typing duration analysis, as hundreds of files had to be processed.

Accuracy

Table 1 shows that accuracy is between 89-95%, with an average of 89.89%. The highest accuracy was achieved by the password "action" with a result of 94.80%, the worst by both "gard£n" and "rec$ives" with a accuracy of 89.07%. As for the average accuracy by weeks, displayed in figure 1, the best accuracy was reached in week 1 (90.29%) and the worst in week 8 (89.54%). In case of

individual passwords by weeks, the best accuracy was achieved by "action" (97.08%, Week 4) and the worst by "d|ameter" (88.42%, Week 8).

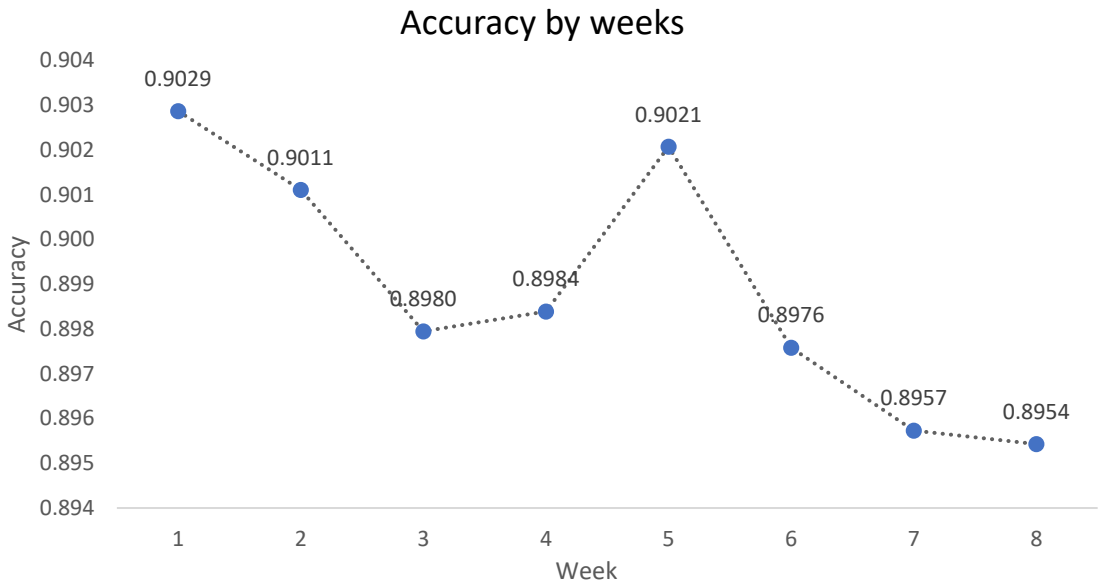

Figure 1, average accuracy across 8 weeks.

To examine a relationship between accuracy and various substitutions, passwords have been divided by substitution type into 5 groups: no substitution, uppercase, numeric, special and "combo" (all three substitutions). As evident in the table 2, the best accuracy was achieved by passwords with no substitution (91.46%) and the worst accuracy is represented by passwords with a special character substitution (89.19%).

Table 3 displays passwords grouped by length: 6, 8, 10 and 12 characters. The highest accuracy achieved passwords of 6 characters (90.44%) and the lowest accuracy have passwords of 12 characters (89.61%).

The accuracy difference (max - min), in the case of substitution is 2.27%, and in the case of various lengths it is 0.83%, which indicates, that accuracy is mostly affected by substitution, and then only slightly by the length of the password, which is also well visualised by figure 2. Figures 3 and 4 were an attempt to find anomalies in accuracies between the different password types, however it appears that most passwords maintain similar trend in given weeks and their accuracy either grows or decreases, with a general downward trend.

Substitution	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Average
No sub.	0.918	0.9182	0.9163	0.9172	0.9177	0.9108	0.9089	0.9095	0.9146
Uppercase	0.9032	0.9056	0.8987	0.8962	0.9049	0.8951	0.8944	0.8971	0.8994
Numeric	0.8986	0.8984	0.8933	0.8936	0.8954	0.8938	0.8933	0.8924	0.8948
Special	0.8971	0.8899	0.891	0.8909	0.8949	0.892	0.8912	0.8884	0.8919
Combo	0.8975	0.8935	0.8905	0.8941	0.8975	0.8962	0.8908	0.8897	0.8937

Table 2, accuracy across 8 weeks, grouped by substitution type

Feature sets

As appears in both table 4 and figure 5, the best feature set is “Tri Press to Press” (29%), followed by “Tri Release to Release” (20%), “Press to Press” (16%) and “Release to Press” (16%). The undeniably worst feature set is "Press to Release", which happened to be the best feature set only once ("action", subject 0, week 1) out of almost 17k attempts - accounted for 0.01%.

Furthermore, as in the case of accuracy evaluation, the passwords are grouped by length and substitution in order to identify dependencies within various password polices. Table 5 provides a percentage share of best feature sets for all substitutions. For passwords with no substitution, the best method is "Full Timing" (55%), for uppercase it’s "Release to Press" (26%), although there are 3 other feature sets in the range of 7%. In terms of numeric and special substitutions, the best method is "Tri Press To Press" with a share respectively of 43% and 42%, combo substitution – “Tri Release to Release” (37%).

Table 6 displays the percentage share of each best feature set within various password lengths, showing that length has very little to no effect in resulted best feature sets, which is indicated by very close distribution of best feature sets over different lengths.

Substitution	Full	Press to Press	Press to Release	Release to Press	Release to Release	Tri Press to Press	Tri Release to Release
No substitution	55.16%	10.94%	0.03%	7.96%	4.39%	14.56%	6.96%
Uppercase	0.00%	21.64%	0.00%	26.12%	11.91%	21.55%	18.78%
Numeric	0.00%	13.27%	0.00%	13.15%	12.12%	43.48%	17.98%
Special	0.00%	13.06%	0.00%	21.37%	3.15%	41.83%	20.58%
Combo	0.00%	22.49%	0.00%	11.76%	5.48%	23.35%	36.91%

Table 5, best feature sets, grouped by substitution

Table 4 demonstrates that the best feature sets are changing significantly from week to week, furthermore, at the level of different password lengths and substitutions (table 7 and 8) the variety is even more drastic. Hence, no relationships or patterns have been observed between the different password characteristics, weeks and accuracies.

Typing duration

The typing duration analysis is based on files from the "testing/Full/" folder. The last value in each row, is a password typing length in milliseconds. Table 9 (and figure 6 as a visual representation) shows that the average typing duration, similarly to accuracy, has decreased from week to week. The longest average typing duration in week 1 with a result of 3.5sec, the shortest in week 8 with a time of 2.6sec and the average of 2.9sec.

In terms of substitution, the shortest typing time was for passwords with no substitution (2.2sec) and the longest for passwords with special substitution (4.2sec) with a loss of over a 1sec to the next password group (combo, 3sec) - observed in all weeks.

In the case of different lengths, typing duration of 6 characters passwords (3.1sec) is only a slightly shorter than 12 characters (3.2sec), while the fastest are passwords with 8 characters (2.6sec).

On average, passwords were typed 25% faster by week 8. The biggest difference was in the case of special substitution (31%), which were also the longest typed passwords, and the smallest decrease in the case of no substitution (the shortest typed) and combo substitution by 18% each.

Results and discussion

Key findings

- The best accuracy is achieved by passwords with no substitution (91.46%), in particular "action" (95%) and "return" (93%) – nearly outliers in comparison with rest of the passwords being at 89%-91% (figure 2). The worst are passwords with special substitution (89.19%), specifically "gard£n" and "rec$ives" (89.07%).

- The best and worst accuracies were achieved by short passwords. Also, the difference between min and max accuracy of various lengths is 0.93%, while of various substitutions it is 2.27%. Both facts indicate, that accuracy is mainly dependent on substitution and to a lesser extent on password length.

- Average accuracy drops from week to week, being the highest in week 1 (90.29%) and the lowest in week 8 (89.54%), with an average of 89.89%. The difference between min and max values is 0.75%, which is quite low, meaning that accuracy is stable.

- Overall best feature set is "Tri Press To Press" (29%) and the worst is "Press to Release". (0.01%). Despite the fact, that relationships such as "Full Timing" being the best feature for passwords with no substitution (55%) or "Tri Press To Press" being the best for special and numeric substitutions (42%) can be observed (table 5), splitting these into individual weeks (table 7) show significant discrepancies. As an example, for passwords with combo substitution, the best feature in weeks 1, 2, 5 and 6 is "Tri Release to Release" with share of 85%, 45%, 70% and 59% respectively, while being nearly the worst in weeks 3, 4, 7 and 8 with share of 0.5%, 14%, 9% and 12% respectively. Moreover, such a large variety in distribution of best feature sets can be observed within all password characteristics.

- "Full Timing" is best feature set for 55% of passwords (attempts) with no substitution, which is the highest individual result for any pair of feature set and substitution type, yet it is not best for any other substitution type even once.

- Relationship between accuracy and substitution is evident. However, a reason for the significant variation of best feature sets in certain weeks has not been identified. Interestingly, despite the variety of best feature sets, accuracy remains relatively unchanged. This suggests, that differences in accuracy between the various feature sets is negligible, however it was not examined and is just a speculation.

- This analysis used the same passwords and methods, as a research[1] undertaken in 2020, but different dataset. While the accuracy results are similar, the best feature sets are much different, which reaffirms the difficulty in determining the best feature set.

- Typing duration, similarly to accuracy, decreases from week to week. This seems logical because subjects are getting familiar with passwords. Despite the strong negative correlation (r = -0.69) between typing duration and accuracy, typing duration does not explain the accuracy jump in

week 5. In addition, figure 6 shows that regardless of the type of division, all lines representing particular groups are nearly parallel to each other, indicating that the type of password has no effect on the variation of typing durations.

- No substitution passwords are at the same time typed the fastest and most accurate, while special substitution passwords have the worst accuracy and are typed for the longest. However, the difference between min-max accuracy is 2% and in the case of typing length it is nearly 100%, so it is probably more of a coincidence than a relationship.

- Despite the above mentioned relations between various password characteristics and accuracy, it is not possible to determine precisely which of the factors influence the small fluctuations of accuracy in the given weeks the most. While there is a general downward trend, it is difficult to find an explanation for the jump of accuracy in the week 5. Probably, the difference of 0.75% between the worst and best week is so small that the pursuit of a particular reason is pointless.

Conclusion

It can be concluded, that in reality, average accuracy would be negatively affected by other factors. For example, the two best passwords do not meet any of the basic requirements of strong password policies, which realistically excludes them. Moreover, identification of the best feature set for a specific sample requires simulation. In case of deployed systems this is not possible. Difficulty in choosing the best feature set for a specific password characteristic is due to the lack of a significant leader across the feature sets, combined with drastic changes of the best feature sets (and thresholds) in the subsequent weeks. Ultimately, leading to the use of not necessary most effective feature sets for majority of attempts.

An accuracy of 89% and EER of 11%, is rather unacceptable, especially within high risk environments. Meaning that, any system relying on keystroke biometrics, where the scoring system is based on the Manhattan distance measure, being the only authentication method, alongside the password itself, is unlikely to be deployed.

Accuracy improvements

Notwithstanding the weaknesses of keystroke biometrics, a number of ways to increase the security of the method exist.

Related research[2], performed in 2013 has shown, that a keystroke biometrics based on GMM-UBM or DBN have achieved an EER of 5.5% and 3.5% respectively, resulting in a significant improvement over this analysis. However, it should be noted that, only 1 password was used (".tie5Roanl.") - typed by 51 subjects, 400 times each, differing radically from the dataset and methodology used in this analysis. Although, in case of the mentioned research, the Manhattan distance measure was evaluated too, resulting in EER of 9.6% - similarly to this analysis. Therefore, it is fair to assume that use of alternative scoring system would result in a substantial increase in accuracy.

Another solution is to design an additional security layer supporting the keystroke biometrics authentication. In many existing systems, a lock after several unsuccessful login attempts is quite popular. Moreover, a use of username/email as a second evaluation phrase in combination with the above could create a chance for the keystroke dynamics to be usable as a biometric authentication method.

Existing implementation

The fact that keystroke dynamics could be useful is proven by the typingdna.com, which offers authentication API using this method. It is a successful company with high profile customers. They claim that their system is suitable for a financial sector, which is a high risk industry.

The way it works is similar to the analysed method. During registration, a user enrols samples of password and email. Regardless of the length of the phrase, each sample is a 320 feature vector[3], which is much higher than in the case of feature sets used in this analysis.

Apart from typingdna.com, there are many other companies offering similar services.

Limitations

This analysis, like any other, would greatly benefit from a larger volume of data - number of participants and samples collected. In addition, while the passwords vary quite a bit, the participants are somewhat homogeneous, as students of the same technical module being of a similar age. Therefore, a greater diversity of subjects, in terms of age and background, would bring additional value to this analysis. However, this is not a major problem, as some studies related to this subject have brought a lot of new information with much less data.

In case of the best feature set analysis, the data of 11 users had to be deleted due to missing rows. However, this did not affect the evaluation of accuracy, which was measured of weekly averages and not by individual participants.

This analysis would also benefited from the examination of the accuracy of individual subjects, however for the reasons described in a section below ("Confusion") this did not happen.

Previously mentioned typingdna.com, forms samples consisting of 320 values. In this analysis, accuracy is analysed using 7 different feature sets. The most extensive feature set - "Full Timing", contains twice as many values as the length of the password (10 characters - 20 values), which is a small number in comparison with the above. Analysis of the data using a feature set containing more information, even like a combination of "Full Timing" with "Tri Press to Press" and "Tri Release to Release", could provide an additional information about the analysed authentication method.

Confusion

(this section is more on a side note, because it could be a result of my misunderstanding)

One of the planned methods of accuracy analysis was to check the accuracy of individual subjects and see how it changes. Additionally, it could be divided by the typing speed. The main accuracy analysis of this report is based on the "bestfeature.csv" files, in which there are accuracy results for individual passwords. In order to check the accuracy for specific subjects, it is necessary to analyse the files in the folders "Participant data".

It seems that the average accuracy of individual subjects should be slightly higher than of passwords, because each accuracy was calculated on the basis of the best feature set for each participant individually and not on the basis of one "average" best feature set for a given password. Surprisingly,

the average accuracy of all subjects is between 81.7%-82.4%, the best result in the first week and the worst in the last week (figure 7) – similar trend to the earlier analysis.

Reaction to this result was to look for an error in the program. Although, a quick manual check of the files, showed that in fact a majority of the individual accuracies are between 80%-83%. Such accuracy of individual subjects seems not possible if the overall accuracy, based on the "bestfeature.csv" files, is around 90%. Therefore, either I am missing some important fact/detail/knowledge and the significantly smaller accuracy values are fine and explainable, or something went wrong in the pre-processing. There was a mistake in the dataset earlier, I do not know what was specifically wrong, however I believe that I got the latest version that was available in Brightspace.

I hope that I am not making an ignorant of myself, but I cannot explain that. Hence, apart from checking the average accuracy of all individual subjects, which is demonstrated in figure 7, I did not analyse this data further as I do not understand these figures. It would prevent me from making any meaningful conclusions.

Appendices

Password	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Average
Average	**0.9029**	**0.9011**	**0.8980**	**0.8984**	**0.9021**	**0.8976**	**0.8957**	**0.8954**	**0.8989**
action	0.9286	0.9439	0.9547	0.9708	0.9662	0.9411	0.9345	0.9440	**0.9480**
return	0.9254	0.9273	0.9256	0.9414	0.9263	0.9217	0.9178	0.9180	**0.9254**
bacteria	0.9167	0.9083	0.9124	0.9071	0.9111	0.9076	0.9002	0.9061	**0.9087**
football	0.9128	0.9145	0.9120	0.8949	0.9122	0.9000	0.9074	0.8993	**0.9066**
calculated	0.9162	0.9157	0.9050	0.9070	0.9121	0.9074	0.9090	0.9029	**0.9094**
automotive	0.9171	0.9153	0.9066	0.9043	0.9012	0.9020	0.9027	0.9032	**0.9066**
professional	0.9139	0.9193	0.9124	0.9083	0.9065	0.9038	0.9040	0.9029	**0.9089**
technologies	0.9131	0.9014	0.9013	0.9037	0.9061	0.9028	0.8960	0.8996	**0.9030**
Filter	0.9019	0.9097	0.8970	0.9010	0.9074	0.8997	0.8972	0.9024	**0.9020**
docTor	0.9076	0.9062	0.9067	0.8989	0.9084	0.8978	0.8922	0.8996	**0.9022**
coMputer	0.9072	0.9052	0.9053	0.8967	0.9066	0.8943	0.8980	0.8966	**0.9012**
clickiNg	0.8988	0.9074	0.8918	0.8982	0.9016	0.8957	0.8903	0.8962	**0.8975**
conDitions	0.9070	0.9057	0.9051	0.8961	0.9067	0.8943	0.8972	0.9018	**0.9017**
Conference	0.8993	0.9078	0.8913	0.8968	0.9094	0.8956	0.8894	0.8969	**0.8983**
disappointeD	0.9028	0.9020	0.8946	0.8887	0.8968	0.8953	0.8937	0.8899	**0.8955**
inflaMmation	0.9010	0.9007	0.8980	0.8933	0.9019	0.8878	0.8968	0.8934	**0.8966**
brok3n	0.9057	0.9034	0.8965	0.8896	0.8990	0.8976	0.8943	0.8905	**0.8971**
cr1sis	0.9006	0.9029	0.8917	0.8929	0.9053	0.8919	0.9018	0.8960	**0.8979**
deliv3ry	0.9043	0.9014	0.8959	0.8965	0.8959	0.8958	0.8920	0.8996	**0.8977**
ann0ying	0.9036	0.9011	0.8948	0.8962	0.8936	0.8939	0.8924	0.8987	**0.8968**
underst0od	0.8944	0.8933	0.8960	0.8966	0.8929	0.8945	0.8927	0.8862	**0.8933**
addressin9	0.8908	0.8987	0.8922	0.8926	0.8908	0.8907	0.8888	0.8959	**0.8926**
headqu4rters	0.8948	0.8932	0.8897	0.8859	0.8929	0.8923	0.8925	0.8864	**0.8910**
pr3scription	0.8946	0.8929	0.8897	0.8982	0.8924	0.8938	0.8920	0.8862	**0.8925**
fr!end	0.8976	0.8944	0.8871	0.8907	0.8992	0.8974	0.8972	0.8891	**0.8941**
gard£n	0.8921	0.8873	0.8919	0.8895	0.8916	0.8891	0.8901	0.8940	**0.8907**
d|ameter	0.8987	0.8957	0.8896	0.8913	0.8935	0.8913	0.8915	0.8842	**0.8920**
rec$ives	0.8922	0.8901	0.8933	0.8868	0.8917	0.8907	0.8911	0.8896	**0.8907**
univers!ty	0.9017	0.8887	0.8920	0.8853	0.8965	0.8896	0.8908	0.8881	**0.8916**
de\ivering	0.8982	0.8876	0.8912	0.8946	0.8954	0.8880	0.8891	0.8868	**0.8914**
bre$thtaking	0.8981	0.8877	0.8914	0.8945	0.8957	0.8948	0.8896	0.8874	**0.8924**
embarrassin?	0.8980	0.8878	0.8917	0.8947	0.8958	0.8952	0.8900	0.8879	**0.8926**
F@st3r	0.8985	0.8958	0.8903	0.8917	0.8996	0.8926	0.8923	0.8851	**0.8933**
pOl!c3	0.8957	0.8934	0.8882	0.8943	0.8969	0.8962	0.8893	0.8900	**0.8930**
sc!enC3	0.8958	0.8932	0.8878	0.8939	0.8967	0.8958	0.8891	0.8897	**0.8927**
he4Ven|y	0.8953	0.8929	0.8879	0.8932	0.8965	0.8957	0.8887	0.8894	**0.8925**
|ndig3nOus	0.8966	0.8942	0.8901	0.8948	0.8977	0.8977	0.8902	0.8913	**0.8941**
in5ul@tIon	0.8976	0.8953	0.8918	0.8964	0.8993	0.8991	0.8919	0.8926	**0.8955**
aSynchr0#ous	0.8958	0.8922	0.8899	0.8950	0.8978	0.8974	0.8901	0.8908	**0.8936**
cat@s7Rophic	0.9047	0.8907	0.8977	0.8934	0.8959	0.8954	0.8951	0.8887	**0.8952**

Table 1, accuracy of all passwords across all weeks

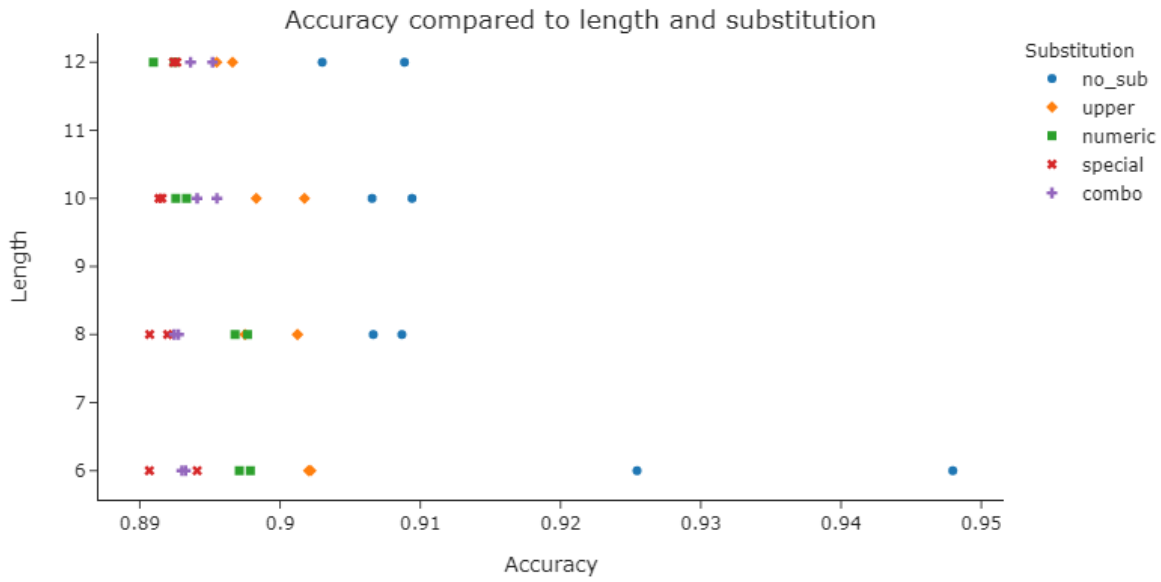

Figure 2, accuracy compared to length and substitution

Length	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Average
6 char	0.9054	0.9064	0.903	0.9061	0.91	0.9025	0.9007	0.9009	0.9044
8 char	0.9025	0.901	0.8971	0.8955	0.8999	0.8961	0.8941	0.8949	0.8976
10 char	0.9019	0.9002	0.8961	0.8964	0.9002	0.8959	0.8942	0.8946	0.8974
12 char	0.9017	0.8968	0.8956	0.8956	0.8982	0.8959	0.894	0.8913	0.8961

Table 3, accuracy across all weeks, grouped by number of characters

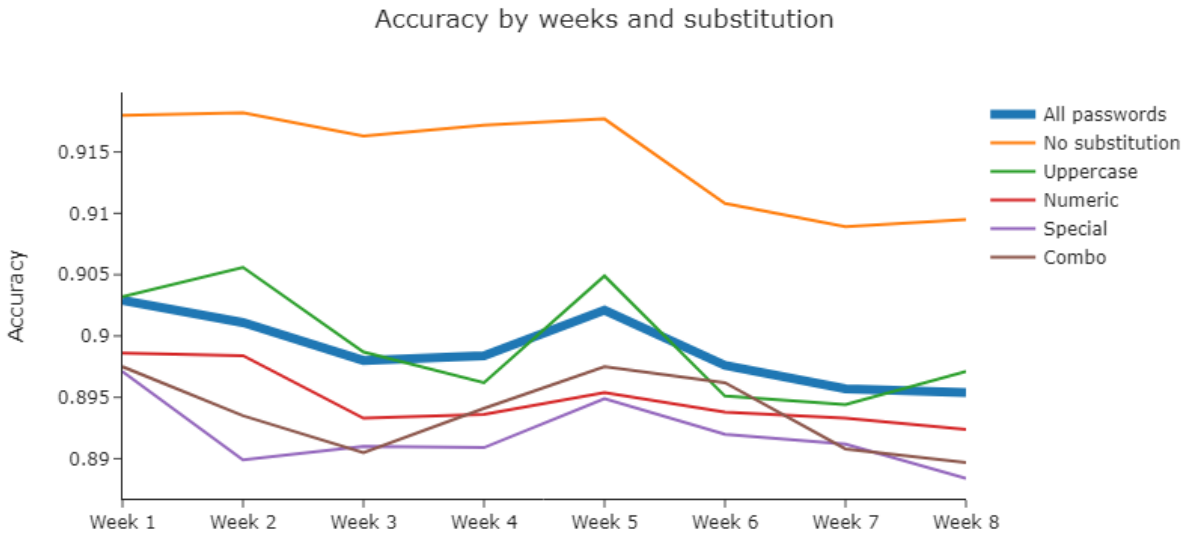

Figure 3, accuracy by weeks and substitution

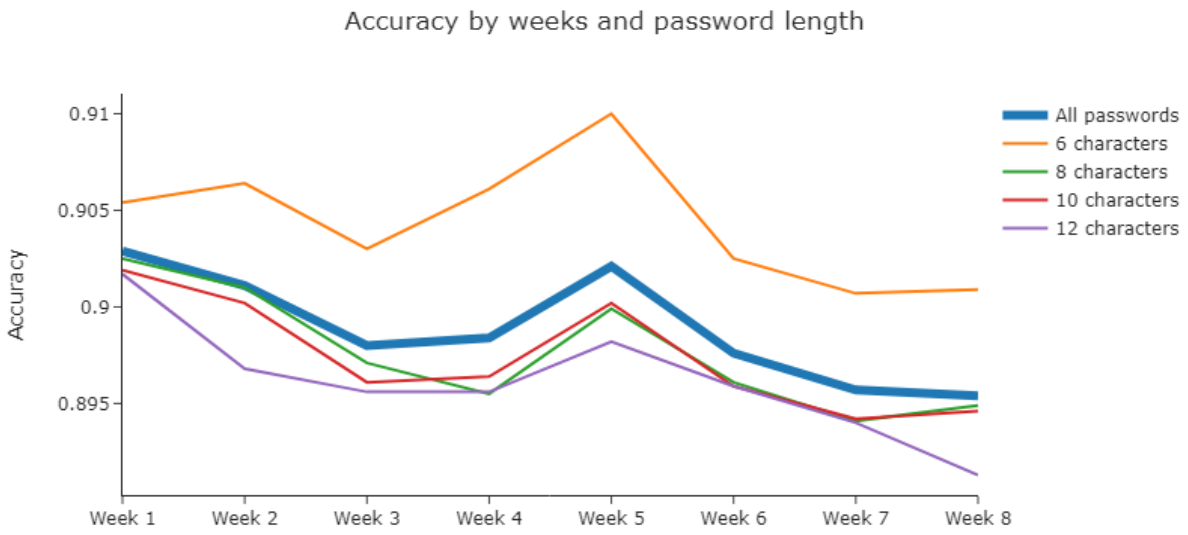

Figure 4, accuracy by weeks and password length

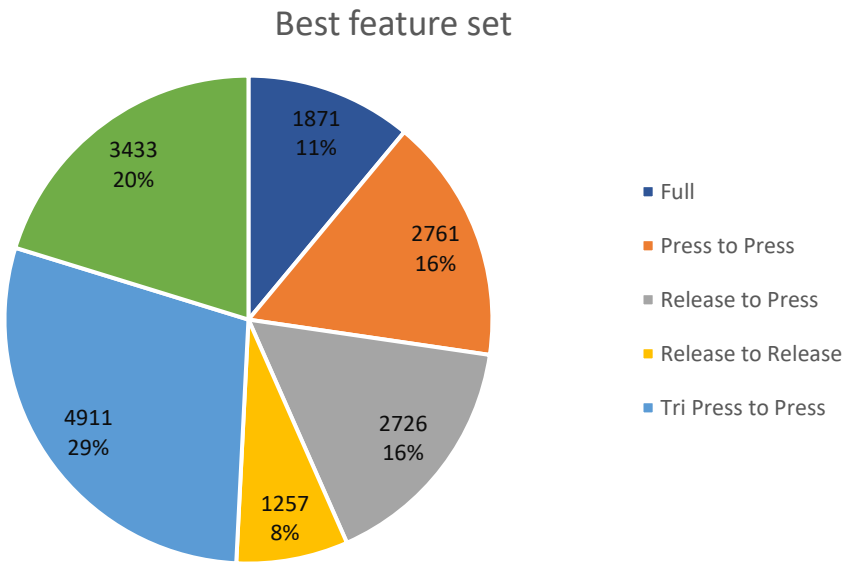

Figure 5, best feature set, combined all weeks/passwords/subjects

Feature set	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Total
Full	6.04%	11.46%	7.45%	13.58%	13.82%	11.27%	10.90%	13.73%	11.03%
Press to Press	15.14%	8.77%	20.00%	15.94%	5.52%	26.46%	30.33%	8.07%	16.28%
Press to Release	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
Release to Press	13.82%	12.74%	21.18%	28.63%	24.91%	8.02%	1.79%	17.50%	16.07%
Release to Release	8.35%	6.32%	11.46%	2.92%	14.29%	0.80%	7.97%	7.17%	7.41%
Tri Press to Press	25.33%	40.90%	37.50%	19.01%	17.50%	33.44%	23.68%	34.29%	28.96%
Tri Rel. to Rel.	31.27%	19.81%	2.41%	19.91%	23.96%	20.00%	25.33%	19.25%	20.24%

Table 4, best feature sets, combined all passwords/subjects

Length	Full	Press to Press	Press to Release	Release to Press	Release to Release	Tri Press to Press	Tri Release to Release
6 char	10.68%	15.21%	0.02%	17.81%	12.74%	23.75%	19.79%
8 char	12.05%	18.49%	0.00%	14.55%	3.33%	27.76%	23.82%
10 char	14.03%	13.58%	0.00%	17.03%	3.40%	32.12%	19.83%
12 char	7.36%	17.83%	0.00%	14.91%	10.19%	32.19%	17.52%

Table 6, best feature sets, grouped by password length

No substitution	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Full	30.19%	57.31%	37.26%	67.92%	69.10%	56.37%	54.48%	68.63%
Press to Press	19.34%	10.14%	7.08%	12.97%	6.60%	14.86%	11.32%	5.19%
Press to Release	0.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Release to Press	10.14%	3.54%	23.82%	0.94%	2.59%	4.01%	8.96%	9.67%
Release to Release	9.91%	0.00%	12.03%	0.24%	3.30%	0.24%	8.49%	0.94%
Tri Press to Press	23.58%	23.11%	12.50%	10.85%	14.62%	14.62%	8.49%	8.73%
Tri Release to Release	6.60%	5.90%	7.31%	7.08%	3.77%	9.91%	8.25%	6.84%

Uppercase sub.	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Press to Press	2.36%	1.65%	5.19%	38.68%	2.59%	60.61%	59.20%	2.83%
Release to Press	0.94%	52.83%	37.50%	14.39%	59.43%	18.40%	0.00%	25.47%
Release to Release	13.92%	1.42%	37.74%	6.37%	18.63%	0.00%	7.08%	10.14%
Tri Press to Press	46.46%	32.55%	15.33%	16.51%	4.48%	17.92%	16.04%	23.11%
Tri Release to Release	36.32%	11.56%	4.25%	24.06%	14.86%	3.07%	17.69%	38.44%

Numeric sub.	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Press to Press	49.53%	19.81%	12.03%	8.73%	3.54%	5.90%	4.95%	1.65%
Release to Press	6.13%	0.00%	20.28%	50.00%	16.75%	8.49%	0.00%	3.54%
Release to Release	14.62%	15.33%	3.54%	2.12%	34.20%	3.77%	9.43%	13.92%
Tri Press to Press	28.30%	37.03%	64.15%	4.72%	45.52%	81.84%	30.90%	55.42%
Tri Release to Release	1.42%	27.83%	0.00%	34.43%	0.00%	0.00%	54.72%	25.47%

Special sub.	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Press to Press	2.36%	7.55%	4.95%	17.45%	13.44%	27.12%	5.42%	26.18%
Release to Press	49.29%	0.00%	13.92%	4.01%	45.75%	9.20%	0.00%	48.82%
Release to Release	1.65%	7.55%	0.71%	4.25%	4.25%	0.00%	4.25%	2.59%
Tri Press to Press	19.34%	76.42%	80.42%	54.25%	5.42%	35.85%	53.77%	9.20%
Tri Release to Release	27.36%	8.49%	0.00%	20.05%	31.13%	27.83%	36.56%	13.21%

Combo sub.	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Press to Press	2.12%	4.72%	70.75%	1.89%	1.42%	23.82%	70.75%	4.48%
Release to Press	2.59%	7.31%	10.38%	73.82%	0.00%	0.00%	0.00%	0.00%
Release to Release	1.65%	7.31%	3.30%	1.65%	11.08%	0.00%	10.61%	8.25%
Tri Press to Press	8.96%	35.38%	15.09%	8.73%	17.45%	16.98%	9.20%	75.00%
Tri Release to Release	84.67%	45.28%	0.47%	13.92%	70.05%	59.20%	9.43%	12.26%

Table 7, best feature sets, grouped by substitutions and weeks

6 characters	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Full	6.23%	8.11%	10.94%	15.47%	13.40%	13.21%	9.62%	8.49%
Press to Press	9.43%	8.87%	16.79%	14.34%	8.87%	30.00%	23.96%	9.43%
Press to Release	0.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Release to Press	17.36%	16.04%	37.74%	1.89%	20.19%	22.64%	2.26%	24.34%
Release to Release	20.19%	9.43%	11.13%	3.40%	19.81%	0.19%	19.06%	18.68%
Tri Press to Press	24.91%	39.06%	18.87%	17.36%	10.75%	32.26%	22.26%	24.53%
Tri Release to Release	21.70%	18.49%	4.53%	47.55%	26.98%	1.70%	22.83%	14.53%

8 characters	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Full	6.23%	16.79%	4.72%	18.30%	14.15%	10.19%	10.57%	15.47%
Press to Press	4.72%	10.57%	26.60%	26.60%	0.57%	29.43%	46.04%	3.40%
Release to Press	4.72%	13.96%	18.30%	38.87%	29.81%	1.70%	1.70%	7.36%
Release to Release	3.58%	1.32%	12.64%	1.70%	2.08%	0.00%	3.96%	1.32%
Tri Press to Press	41.51%	26.98%	33.58%	9.06%	23.96%	27.92%	9.25%	49.81%
Tri Release to Release	39.25%	30.38%	4.15%	5.47%	29.43%	30.75%	28.49%	22.64%

10 characters	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Full	11.70%	20.00%	13.77%	9.43%	12.26%	16.98%	9.81%	18.30%
Press to Press	22.64%	0.00%	20.94%	18.49%	1.32%	19.43%	25.66%	0.19%
Release to Press	11.32%	9.25%	2.83%	44.53%	34.15%	0.57%	1.13%	32.45%
Release to Release	0.38%	1.70%	17.92%	5.09%	0.94%	0.00%	1.13%	0.00%
Tri Press to Press	12.45%	54.53%	44.53%	19.81%	27.92%	32.45%	32.45%	32.83%
Tri Release to Release	41.51%	14.53%	0.00%	2.64%	23.40%	30.57%	29.81%	16.23%

12 characters	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Press to Press	23.77%	15.66%	15.66%	4.34%	11.32%	26.98%	25.66%	19.25%
Release to Press	21.89%	11.70%	25.85%	29.25%	15.47%	7.17%	2.08%	5.85%
Release to Release	9.25%	12.83%	4.15%	1.51%	34.34%	3.02%	7.74%	8.68%
Tri Press to Press	22.45%	43.02%	53.02%	29.81%	7.36%	41.13%	30.75%	30.00%
Tri Release to Release	22.64%	15.85%	0.94%	23.96%	16.04%	16.98%	20.19%	23.58%

Table 8, best feature sets, grouped by password lengths and weeks

Password type	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Average
All passwords	3476	3234	3008	2782	2679	2725	2691	2603	2900

Password type	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Average
No substitution	2503	2393	2306	2128	2042	2126	2116	2042	2207
Uppercase	2850	2678	2498	2279	2214	2231	2193	2132	2384
Numeric	3355	3054	2823	2642	2482	2572	2492	2432	2731
Special	5279	4799	4314	3986	3840	3836	3824	3643	4190
Combo	3394	3248	3100	2874	2818	2859	2830	2766	2986

Password type	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Average
6 characters	3737	3482	3205	2978	2910	2948	2900	2806	3121
8 characters	3092	2823	2671	2460	2337	2370	2356	2303	2551
10 characters	3197	3050	2858	2669	2538	2633	2570	2507	2753
12 characters	3878	3582	3299	3020	2933	2948	2939	2797	3175

Table 9, typing duration in milliseconds, average of all + grouped by length, substitution and weeks

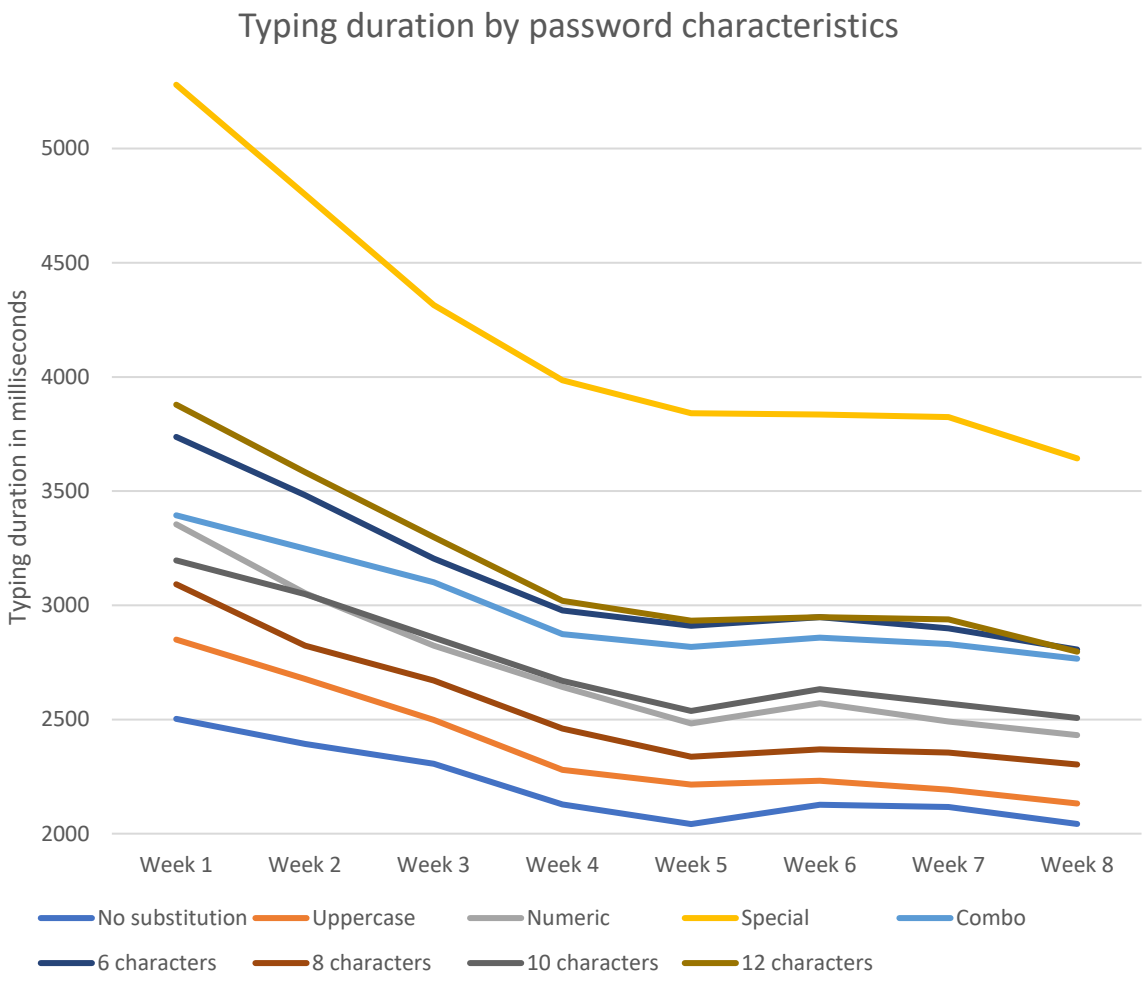

Figure 6, typing duration in milliseconds, grouped by length, substitution and weeks

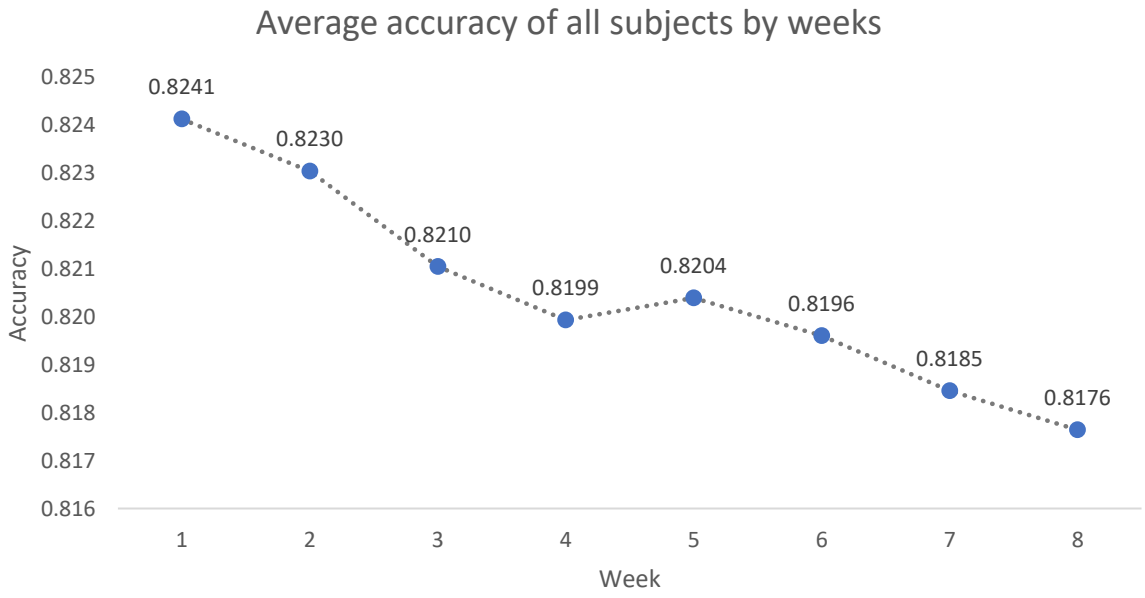

Figure 7, accuracy by weeks based on data of individual subjects

Password list

1. action	11. coMputer	21. underst0od	31. bre$thtaking
2. return	12. clickiNg	22. addressin9	32. embarrassin?
3. bacteria	13. conDitions	23. headqu4rters	33. F@st3r
4. football	14. Conference	24. pr3scription	34. pOl!c3
5. calculated	15. disappointeD	25. fr!end	35. sc!enCe
6. automotive	16. inflaMmation	26. gard£n	36. he4Ven|y
7. professional	17. brok3n	27. d|ameter	37. |ndig3nOus
8. technologies	18. cr1sis	28. rec$ives	38. in5ul@tIon
9. Filter	19. deliv3ry	29. univers!ty	39. aSynchr0#ous
10. docTor	20. ann0ying	30. de\ivering	40. cat@s7Rophic

References

1. Simon Parkinson , Saad Khan , Andrew Crampton , Qing Xu, Weizhi Xie, Na Liu, Kyle Dakin, “Password Policy Characteristics and Keystroke Biometric Authentication”, 2020
2. Yunbin Deng, Yu Zhong, “Keystroke Dynamics User Authentication Based on Gaussian Mixture Model and Deep Belief Nets”, table 1, 2013
3. TypingDNA.com. *Frequently asked questions*. https://www.typingdna.com/authentication-api.html