Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

PIN selection policies: Are they really effective?

Hyoungshick Kim [a], *Jun Ho Huh* [b,*]

[a] *Computer Laboratory, University of Cambridge, UK*
[b] *Information Trust Institute, University of Illinois at Urbana-Champaign, 1308 West Main St, Urbana, Illinois 61801, United States*

ARTICLE INFO

Article history:
Received 14 October 2011
Received in revised form
12 January 2012
Accepted 3 February 2012

Keywords:
PIN selection policy
PIN distribution
Entropy
Memorability
Password

ABSTRACT

Users have conflicting sets of requirements when it comes to choosing Personal Identification Numbers (PINs) for mobile phones or other systems that use PINs for authentication: the conflict lies between the 'easy to remember' usability requirement and the 'hard to guess' security requirement. Users often ignore the security requirement and choose PINs that are easy to remember and reuse, making it also easy for attackers to guess and compromise them. Just as the password strength is controlled through various password policies, PIN selection policies may be used to help users choose stronger PINs and meet various security requirements. An example policy would not allow the use of the most commonly selected PINs.

An online user study was conducted to investigate the effectiveness of such PIN selection policies, requesting the participants to choose PINs under some carefully designed policies. The participants were also asked to record the memorability (remembrance difficulty) score of each PIN, indicating how easy/hard it was to remember the selected PIN. Based on the entropies calculated on the collected PINs and their memorability scores, this paper demonstrates that restricting some number of commonly used PINs (e.g. restricting the 200 most commonly used ones) is beneficial: this type of policy would significantly increase the randomness of PINs without incurring significant memorability overhead. Our results also showed that any PIN- or PIN-pattern-based blacklisting policy should be constructed with caution since the total PIN space may become too small, making it easier for attackers to guess PINs.

1. Introduction

Mobile phones used to be simple. One could simply make phone calls and send/receive text messages. With the emergence of smart phones, however, more and more people have also started using them as a digital-wallet, storing sensitive information like credit cards, identity cards, loyalty/gift cards, vouchers, and mobile banking tokens (Anderson, 2011). Just as one would try to safeguard a wallet full of cash and credit cards from strangers, a digital-wallet user also wants to protect its contents through strong user authentication mechanisms. Among many authentication mechanisms available, Personal Identification Numbers (PINs) are dominantly used. A PIN is a *numeric password* that the user must type into the mobile phone to authenticate its use.

Unlike biometric and smart card authentication, PINs are easy to implement and do not require extra hardware support. This is what attracts most of the mobile phone companies to use PINs as their primary authentication mechanism. However, PINs too have their own inherent limitations – namely, memorability and security. Problems arise because of the following two conflicting requirements:

* *Corresponding author.*
E-mail addresses: hk331@cl.cam.ac.uk (H. Kim), jhhuh@illinois.edu, etpfest@gmail.com (J.H. Huh).
0167-4048/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2012.02.003

1. usability – PINs should be easy to remember;
2. security – PINs should be secure, meaning they should be randomly distributed and difficult to guess; a user should change their PINs frequently, and use different PINs on different accounts.

In practice, it is difficult to satisfy all of these requirements. A PIN that is difficult to guess is also likely to be hard to remember. As one would imagine, many users choose PINs that are easy to remember without really paying close attention to the security implications. Trivial PINs like '1234' and '0000', users' birthdays or telephone numbers are often used. A recent study shows that among 204,508 recorded PINs, 15% of them were part of the top 10 most commonly used PINs (Amitay, Jun 2011). Similar trends are evident for passwords, which are a more general form of PINs (Vance, Jun 2010). Such a trend implies that the actual space of PINs used is much smaller than the theoretical space ($10^{\text{length of the PIN}}$), dramatically increasing the likelihood of an attacker compromising a PIN through brute-force type of attacks. One motivation of our work is to investigate the extent that this PIN space can be affected by helping the users choose stronger PINs.

Based on a large dataset of real PINs collected from an iPhone application (Amitay, Jun 2011), Fig. A.1 shows how frequent each button on the keypad was used. This clearly demonstrates a poor PIN selection practice: buttons '1', '2' and '0' were used much more often than '8', '6' and '7'. Such statistical information can be misused by attackers to make effective guesses for the PINs. To prevent users from using bad PIN selection practices and choosing weak PINs (that are easy to guess), devices/applications may enforce various PIN selection policies. These policies capture security requirements that must be satisfied upon selecting a PIN; an example policy might be that 'a PIN shall not have any duplicating number'. Such policies, in theory, should help users choose stronger PINs; but how do we know that they really work well in practice? Precisely predicting how the security and usability requirements stated above will be affected by different policies can be difficult. For instance, if a policy restricts the use of 10 most popular PINs, the next top 10 PINs will soon replace them, becoming the new 10 most popular PINs. Usability would definitely be affected by this policy, but have we really improved PIN security?

To answer these questions, an online survey was conducted, asking the 332 participants to select PINs while conforming with carefully designed PIN selection policies. To maximize consistency in the participants' attitude and perception towards choosing PINs, the scope of the study was set to focus on locking mobile phones – this information was made clear to the participants prior to starting the survey. By narrowing down the scope, we wanted the participants to have similar perception on the level of complexity required/necessary for their PINs. For instance, a participant's perception may be different when it comes to choosing PINs for banking purposes. Based on the survey results, the effectiveness of each policy was analysed and suggestions were made on how the policies should be designed. This paper contributes in the following areas: (1) an analysis of the characteristics of the PINs used on mobile phones, and (2) security and usability evaluation of the proposed PIN selection policies.

The following section explains the need for PIN selection policies and explores related work, mainly in the areas of password security and policies. Section 3 analyses the distribution of a sample PIN dataset for mobile phones that have been generated free from any PIN selection policies. This puts us in a position to describe the methodology of our own study in Section 4, and evaluate the effectiveness of different PIN selection policies in Section 5. Our conclusions and future work are in Section 6.

2. Related work

User authentication is an integral part of security-critical systems that manage sensitive information or provide personalised services. Some commonly employed user authentication technologies include passwords, PINs, digital certificates, physical tokens such as smart cards, one-time passwords, transaction profile scripts, and biometric identification. Among these, 'what users know' type of authentication – generally passwords or PINs – is still the dominant technology; this is due to its low implementation and deployment costs. For usability reasons, however, many users choose passwords/PINs that are easy to remember; such weak passwords/PINs are also easy to guess and vulnerable against brute-force and dictionary attacks.

To help users choose stronger PINs, 'PIN selection policies' may be defined and enforced. These policies define the security requirements for the PINs in terms of allowable number combinations that must be satisfied upon PIN selection. As for passwords, there have been rigorous debates about the effectiveness of such selection policies. Some have shown that the selection policies improve security of password selection (Kuo et al., 2006; Vu et al., 2007); Kuo et al. (2006), for instance, have shown that mnemonic phrase-based passwords are more resistant to brute-force attacks than those selected with some optional guidelines (e.g. a password should be at least eight characters in length, and a mixture of lower/upper case letters, numbers, punctuation, and special characters should be used). The selection policies – despite the usability penalties they are likely to impose – could fail to achieve stronger passwords if the users simply choose substitute passwords that are relatively easy to remember; the password distribution in this case will still be skewed. Similar implications may result from enforcing PIN selection policies on users; thus, their effectiveness and usability must be carefully evaluated before being used.

Numerous studies have already been carried out for password selection policies, examining how they affect the users: some of these studies were based on theoretical estimates (Burr et al., 2006; Shay et al., 2007; Shay and Bertino, 2009), some based on small-scale laboratory studies (Sasse et al., 2001; Brown et al., 2004; Yan et al., 2004; Gaw and Felten, 2006; Vu et al., 2007), and some based on two large-scale studies (Shay et al., 2010; Komanduri et al., 2011). Inglesant and Sasse (2010) have shown that many users, despite knowing that repeatedly using the same passwords is a bad security practice, rarely change their passwords. Through a survey conducted on a group of undergraduate students, Hart (2008) demonstrated that the majority of students are not

really concerned about the strength of their passwords. They also showed that user training has limited impact on improving the users' security awareness.

Vu et al. (2007) conducted a laboratory study, demonstrating that passwords chosen under strong selection policies are generally harder to compromise through the use of automated password-cracking tools; but these are also harder to generate and remember, affecting the overall usability. Kuo et al. (2006) showed that automated tools were less effective against mnemonic passwords than control passwords. Simulations performed by Shay et al. (2007) and Shay and Bertino (2009) have shown that stringent password selection policies can lead to users writing down the passwords and thereby jeopardizing their confidentiality. In a more recent study (Shay et al., 2010), they examined the users' behaviours and practices related to the password creation under a new, more strict policy. Users were annoyed by the transition of password policy from a less-constrained one to a stricter one, but felt more secure under the new policy. Some users struggled to comply with the new policy, taking longer to create passwords and finding it harder to remember them. Komanduri et al. (2011) investigated password strength, user behaviour, and user sentiment across numerous password selection policies, and recommended a 16-character policy as a policy that would achieve strong passwords without putting too much burden on the users. Our work is an extension of the studies described here: instead of passwords, the focus is on mobile PIN security and studying the effectiveness of PIN selection policies through an online survey.

3. What real world PINs look like

This section shows that the actual distribution of the real world PINs, generated free from any PIN selection policy, is quite different from the ideal uniform distribution. A large sample of PINs collected from an existing iPhone application called 'Big Brother Camera Security' (Amitay, Jun 2011) was used to show this: it anonymously collected PINs from 204,508 users that were used for locking the application. The users, through the end user licence terms of the application, have agreed that their information may be used to improve the products or provide services so long as it is in a form that does not personally identify them (Apple Inc.).

Although the statistics for these PINs is not exactly the same as those for passwords collected from rockyou.com (Vance, Jun 2010) (a different study), the overall distribution patterns are quite similar: many users felt strongly about the easy-to-choose and easy to remember requirements, and chose commonly used PINs or passwords like '1234' or '123456'. There was a high degree of overlap with such common PINs/passwords in both datasets.

3.1. Occurrence frequency of the PINs

First, we demonstrate that the occurrence frequency of the PINs (i.e. the number times each PIN appears in the dataset) follows a power law distribution (Clauset et al., 2009). Formally, the PIN occurrence frequency obeys a power law if it is drawn from a probability distribution

$$p(x) \propto x^{-\alpha} \quad \text{if } x \geq x_{\min}, \tag{1}$$

where $x_{\min}$ is a positive constant known as the *scaling region* and α is a constant parameter of the distribution known as the *scaling parameter*. In power law distributions, the scaling parameter α generally lies in the range between 2 and 3 (i.e. $2 < \alpha < 3$). The collected PINs were sorted by decreasing order of occurrence frequency, and the histogram of the PIN occurrence frequency, the cumulative distribution function (CDF), and the top 10 popular PINs was plotted (see Fig. A.2). In the histogram, the PIN occurrence frequency decreases dramatically around the 10th PIN, indicating that the PIN distribution is heavily skewed in favour of a small number of commonly used PINs. In graph (b), the dotted line that runs through the circles represents the best fit. An interesting observation from graph (c) is that the most popular PIN, '1234', alone, accounts for 4.3% of the total number of PINs.

Maximum Likelihood Estimation (MLE) (Clauset et al., 2009) was used to prove that the distribution of the PINs follows power law distribution, working out the scaling parameter α as 2.25 and the scaling region $x_{\min}$ as 10 (see graph (b), Fig. A.2). This indicates that the PINs are distributed according to power law with $\alpha \approx 2.25$, also implying that the occurrence frequency of the PINs decreases as a power function of their ranking. The most frequent PIN will approximately occur twice as often as the second most frequent PIN, three times as often as the third most frequent PIN, and so on. Table 1 shows that a small number of popularly used PINs account for a large proportion of the total number of PINs: for instance, the top 100 popular PINs account for about 29.3% of the total number of PINs.

The occurrence frequency of different digits in the PINs was then analysed. Fig. A.3 shows the occurrence frequency of each number (0–9) in the four different positions of the PINs. An interesting observation is that the frequency distribution of the first digit (see graph (a)) is skewed more towards the left compared to the other digits (see graphs (b), (c) and (d)). It would be relatively easier for an attacker to guess the first digit than the other digits. In contrast, the fourth digit is more

Table 1 – The proportion of the most popular PINs.

	Top 100	Top 200	Top 300	Top 400	Top 500
Proportion of combinations (A)	0.0100	0.0200	0.0300	0.0400	0.0500
Proportion of PINs (B)	0.2926	0.3537	0.3969	0.4307	0.4595
Ratio of B to A	**29.2600**	17.6850	13.2300	10.7675	9.1900

evenly distributed. The occurrence frequency of the commonly used subsequences of the digits in the PINs was then analysed, showing similar trends. The details of the analysis is presented in Appendix A.

3.2. PINs generated from dates and years

To ease the skewness of the PIN distributions, it is important to understand the different types of information that are used by the users in selecting PINs. Being aware that many users use memorial dates (e.g. birthdate) or year as their PINs, we examined the proportion of PINs that are likely to have been derived from such information. Table 2 shows that such PINs account for quite a large proportion of the total number of PINs. In particular, 50 PINs that represent the years ranging from 1951 to 2000 account for about 5.5% of the total distribution of the PINs. Moreover, the proportion of the PINs that are likely to have been derived from a date format (either Date(W) or Date(E) as explained in Table 2) is over 10.0%. Considering the significance of these numbers, PIN selection policies should be designed with such PIN characteristics in mind.

3.3. PINs generated through arithmetic operations

Some tech-savvy users may try to generate PINs using their own rules. Inspecting that some simple mathematical equations may be used to generate PINs, we calculated the proportion of PINs that could have been derived from arithmetic operations like addition, subtraction, multiplication and division. Our checking involved doing an arithmetic operation on the first and the second digits, and comparing the answer against the last two digits. Taking into consideration that these proportions can easily be overestimated by counting the '0000'PINs (which was the second most popular PIN), '0000' was not counted when calculating the proportions. The results are shown in Table 3. The results indicated that about 1.3% of the PINs could have been generated through addition, about 0.5% through subtraction, about 1.1% through multiplication, and about 0.2% through division. Since these numbers are relatively small, such PIN characteristics may be ignored when designing PIN selection policies.

3.4. PINs with close proximity

It is also possible for users to choose PINs that consist of numbers that have close 'proximity'. PINs like '0000' or '1111', for example, have very close proximity and are popular since they are easy to remember and use. The proportions of PINs that have such proximity characteristics were also calculated.

One reasonable measure of PIN proximity is the sum of the differences between all the consecutive numbers in a PIN. This sum of a PIN ρ is denoted using $D_n(\rho)$. An alternative measure is to calculate the physical distance between the keypad buttons. The physical distance between two buttons, $b_i = (x_i,y_i)$ and $b_j = (x_j,y_j)$, on a keypad, is derived from $\max\{|x_i - x_j|, |y_i - y_j|\}$. The sum of the physical distances between the consecutive digits in a PIN ρ is denoted using $D_p(\rho)$. Here is an example calculation for both measures: D_n('1234') = 3 and D_p ('1234') = 4. The proportion of the PINs was analysed using both D_n and D_p, where the proximity values are less than a series of constants varying from 2 to 10 and from 1 to 5, respectively. The results are shown in Tables 4 and 5, indicating that such PINs account for a large proportion of the total number of PINs. It is interesting to note that 10 PINs that are made up of a single digit account for about 7.4% of the total PINs, and 64 PINs with $D_n < 2$ account for about 8.6%.

To analyse the relationship between the occurrence frequency of PINs and D_n/D_p, the Pearson correlation coefficients between them were calculated. To avoid bias, the 200 most popular PINs were excluded in this analysis. The scatter plot graphs in Fig. A.4 show a negative correlation between the occurrence frequency of PINs and D_n/D_p, even though, this trend appears to be weaker with D_p. The correlation coefficients of the scatter plot graphs are −0.0602 and −0.0994, respectively; there is a weak negative correlation between the variables. These results indicate that the PINs with small D_n/D_p values are more frequently selected.

4. Methodology

Section 3 showed that the PIN distribution for mobile phones is highly skewed due to the tendency of the users to select weak PINs that are easy to remember. Such statistical information can be beneficial to an attacker whose goal is to compromise users' PINs. How can a better PIN distribution be achieved? As it was discussed in Section 2, one simple solution is to enforce PIN selection policies to ensure that users do not choose weak PINs (see Section 2). To examine the effectiveness of PIN selection policies, an anonymous online user study was conducted, requesting users to choose PINs under a number of carefully designed PIN selection policies.

Table 2 – The proportion of the PINs that are likely to have been derived from date or year information. 'Date(W)' and 'Date(E)' represent the set of PINs that are likely to have been derived from common western-style format, DD/MM, and eastern-style format, MM/DD, respectively. '$Y_{start} \sim Y_{end}$' represents the set of PINs that are likely to have been derived from a number between the two years, Y_{start} and Y_{end}.

	Date(W)	Date(E)	1901 ~ 1950	1951 ~ 2000	2001 ~ 2050
Proportion of combinations (A)	0.0366	0.0366	0.0050	0.0050	0.0050
Proportion of PINs (B)	0.1101	0.1460	0.0085	0.0548	0.0179
Ratio of B to A	3.0082	3.9891	1.7000	**10.9600**	3.5800

Table 3 – The proportion of the PINs that are likely to have been derived from arithmetic operations.

	Addition	Subtraction	Multiplication	Division
Proportion of combinations (A)	0.0099	0.0054	0.0099	0.0032
Proportion of PINs (B)	0.0132	0.0054	0.0106	0.0020
Ratio of B to A	**1.3333**	1.0000	1.0707	0.6250

The aim of our user study was to observe *within-subject* changes in the PIN selection behaviours before and after introducing a stricter selection policy. We note that it was not our primary intention to compare the performance of different selection policies. To this end, we designed a within-subject design user study that is more suitable for examining the changes in pretest and posttest performances. The questions were carefully ordered according to the strictness of the policies being applied – as the policies became stricter, more rules were being enforced on the participants. The participants were encouraged to try out the same PIN, that passed the less strict policies, on the stricter policies.

Partly explained in the introduction was the application domain that the study focused on. To achieve better ecological validity and consistency in the participants' perception towards selecting PINs, the application scope of the experiments were narrowed to authentication for mobile phones: the participants were asked to choose PINs for locking their mobile phones. This is how the consistency between the participants' perception on the complexity/strength required for their PINs was ensured.

Also, to achieve high accuracy, the participants were encouraged to enter PINs that are close representation of the PINs they are currently using on their mobile phones. To alleviate some of the participants' concerns about their PINs being exposed, we clarified the academic motivations behind this study and the fact that the experiments were anonymous – emphasising that no identification information like name and IP address (which may be used to correlate the PINs and participants) will be collected. To further assure the wary participants, the survey was hosted on one of the author's own domain (www-dyn.cl.cam.ac.uk).

4.1. Demographics of the participants

A large group of the authors' university colleagues and friends as well as several online communities were invited to participate in the survey, collecting results from a total of 332 respondents during a month period. The majority of the respondents were male in the age group of 18–49 with at least some college education. The detailed demographics of the participants are presented in Table 6.

4.2. PIN selection policy design

Our user study consisted of five different parts: for each, the participants were instructed to choose a PIN under a different PIN selection policy through a GUI representation of a typical keypad one would see on mobile phones (see Fig. A.5). The PIN selection policies were designed based on the results of the real PIN analysis that were presented in Section 3. Here are our polices.

1. `4-free` – participants shall choose a 4-digit PIN without any restriction.
2. `4-short` – participants shall choose a 4-digit PIN where the 200 most popular PINs from (Amitay, Jun 2011) (2%) are not allowed since the 200 most popular PINs account for a reasonable portion (about 35.4%) of the total number of PINs without the significant decrease of the usable PIN space. This is used as a representative of short PIN blacklist.
3. `4-long` – participants shall choose a 4-digit PIN where the 200 most popular PINs as well as PINs that contain consecutive numbers that are adjacently located on the keypad are not allowed; we disallowed 3205 (≈32%) PINs which comprise of the 200 most popular PINs and PINs with $D_p < 4$. We set $\lceil$(average physical distance between two buttons) $\times$ (the length of a PIN $-$ 1)$\rceil$ as the threshold to define a set of PINs that contain consecutive numbers that are adjacently located on the keypad. Since the average physical distance between two buttons is 1.4, the threshold is 4 for 4-digit PINs. This policy prevents the use of PINs that have $D_p(\rho)$ values that are less than or equal to the average D_p values. This is used as a representative of long PIN blacklist.
4. `6-free` – participants shall choose a 6-digit PIN without any restriction. This is used to show the effects of PIN length.

Table 4 – The proportion of the PINs that have sum of differences less than the given D_n.

	$D_n < 2$	$D_n < 4$	$D_n < 6$	$D_n < 8$	$D_n < 10$
Proportion of combinations (A)	0.0064	0.0500	0.1558	0.3194	0.5060
Proportion of PINs (B)	0.0855	0.2072	0.3121	0.4558	0.6123
Ratio of B to A	**13.3594**	4.1440	2.0032	1.4271	1.2101

Table 5 – The proportion of the PINs that have sum of physical distances less than the given D_p.

	$D_p < 1$	$D_p < 2$	$D_p < 3$	$D_p < 4$	$D_p < 5$
Proportion of combinations (A)	0.0010	0.0148	0.0964	0.3102	0.6046
Proportion of PINs (B)	0.0743	0.0956	0.1742	0.4307	0.7054
Ratio of B to A	**74.3000**	6.4595	1.8071	1.3885	1.1667

5. `6-long` – participants shall choose a 6-digit PIN where the PINs that contain consecutive numbers that are adjacently located on the keypad are not allowed; internally, we disallowed 399,994 (≈40%) PINs including the PINs with $D_p < 7$ for the same reason as `4-long`.

If the selected PIN failed to meet the requirements stated in a policy, the participant was informed as to why it failed and asked to enter another PIN. After choosing a PIN, the participant was also asked to rate how easy it was to remember this PIN (through a 5-point Likert scale question), ranging from 1 = `very easy to remember` to 5 = `very difficult to remember`. One of the purposes of recording this 'remembrance difficulty' was to remind the participant that they should also consider the usability (easy to remember) aspects of choosing PINs – encouraging the participant to choose more practical PINs. Another purpose was to compare and evaluate the usability of the PINs that were selected under different policies.

4.3. *Measuring the distribution of the PINs*

The entropy of the collected PINs were calculated to measure their randomness and study how their distributions would be affected under different PIN selection policies. Two different methods were used for calculating the entropy: Shannon's method (Shannon, 2001) and Massey's method (Massey, 1994) called the 'guessing entropy'. The distribution of PINs are considered to be a random variable X drawn from a finite distribution $P = \{p_1, p_2, \ldots, p_n\}$ with probability $p_i = P(X = x_i)$ for each possible answer x_i, where $i \in [1, n]$. The 'Shannon entropy' is a traditional estimator that measures information about X as follows (*log* is to the base of 2):

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \quad (2)$$

The Shannon entropy should be sufficient to show the amount of information that can be leaked. However, it may not be sufficient depending on what the attacker is capable of performing or finding out. For instance, if the attacker has access to a source that can confirm or refute PIN guesses, a better metric to use would be the 'average number of guesses' required to compromise a PIN. Here, the attacker's goal is to compromise the PIN with the lowest number of guesses. In practice, the attacker may get a second chance to guess even if the first guess was incorrect. Under this threat model, supposing that the attacker knows the distribution of X, the best strategy would be to start by guessing the most likely X and proceeding in a decreasing order of likelihood until the secret information is known. This entropy is well known as the guessing entropy introduced by Massey (1994), and can be calculated as follows:

$$G(X) = \sum_{i=1}^{n} p_i \cdot i \quad (3)$$

Without loss of generality, it is assumed that the probabilities are arranged as a monotonically decreasing distribution where $p_1 \geq p_2 \geq \ldots \geq p_n$.

Table 6 – The demographics of the participants.

Gender		
Male	79.82%	(265)
Female	20.18%	(67)
Age group		
under 17	3.92%	(13)
18–29	53.92%	(179)
30–49	36.75%	(122)
50–64	5.42%	(18)
Highest level of education completed		
High school	9.94%	(33)
College/University	51.51%	(171)
Graduate	34.64%	(115)
Others	3.92%	(13)

5. Results and recommendations

This section looks at how effective the policies designed in Section 4.2 can be in improving the PIN distribution for mobile phones. Based on the results collected from the online survey, this section studies (1) the number of participants who had to change the PINs they selected first in order to conform with the stricter PIN selection policy, (2) how the participants felt about their changed PINs, and (3) the impact of the stricter policies on the randomness of the PINs selected.

For `4-short`, about 39.5% of the participants had to change the PINs they selected under `4-free` since they were one of the blacklisted, 200 most popular PINs. For these PINs, the changes in the *remembrance difficulty* were measured. On average, the *remembrance difficulty* increased by about 0.36. Further, about 25.9% of the participants felt that the changed PINs are difficult to remember. For `4-long`, about 52.7% of the participants had to change the PINs that satisfied `4-short`. The `4-short` PINs that were disallowed on `4-long` must have had consecutive numbers that are adjacently located on the keypad (this would violate the rule that was added in `4-long`). Again, the differences in the *remembrance difficulty* between `4-short` and `4-long` were measured. On average, the *remembrance difficulty* increased by about 0.74. Also, about 49.0% of the participants felt that the PINs selected under `4-long` are more difficult to remember compared to the ones selected under `4-short`.

As for the 6-digit PINs, about 63.9% of the participants had to change the PINs they selected under `6-free` to satisfy `6-`

Table 7 – Entropy computation of the PINs.

Policy	Shannon entropy	Guessing entropy	Remembrance difficulty
4-free	8.1563	139.3012	2.0181
4-short	8.2945	153.9729	2.2560
4-long	8.2500	147.5964	2.6054
6-free	8.2896	155.7410	2.3916
6-long	8.3487	162.5392	2.9398
4-rand	8.3329	159.6476	—
6-rand	8.3750	166.5000	—

long. PINs that violated 6-long must have had consecutive numbers that were adjacently located on the keypad. By measuring the changes in the *remembrance difficulty*, we noticed a significant increase in the average *remembrance difficulty* of 1.53. Also, a surprising 85.0% of the participants felt that PINs they re-selected under 6-long are more difficult to remember than the original ones selected under 6-free.

To develop a more quantitative description, we calculated the entropy of the PINs that were chosen under the five different PIN selection policies, 4-free, 4-short, 4-long, 6-free and 6-long (see Section 4.2). To show the absolute upper bound entropy values, the entropy of randomly generated 4-digit PINs (4-rand) and 6-digit PINs (6-rand) was also calculated. These entropy values are shown in Table 7; also shown in this table is the average *remembrance difficulty* for each policy. The remembrance difficulty is used to compare the usability of the polices.

Not surprisingly, the PIN distribution has the highest entropy of 8.3487 (Shannon) and 162.5392 (guessing entropy) under the 6-long PIN selection policy. These entropy values are close to those of the ideal PIN distribution, represented through the randomly generated 6-rand. Considering how difficult it is to remember the 6-long PINs (see Table 7 and Fig. A.6), however, these would not be our top recommendation. In fact, all other PIN distributions, except for the distribution of 4-short, have significantly worse remembrance difficulty than the 4-free distribution (P-value < 0.001, two sample t-tests). Nevertheless, Table 8 shows that the 10 most PINs have changed significantly through enforcing these policies.

Table 8 – 10 Most popular PINs.

Rank	4-free	4-short	4-long	6-free	6-long
1	0000	0061	0061	123456	101010
2	1234	0217	0701	111111	030303
3	1111	0327	0273	314159	172839
4	0061	0623	0654	121212	001202
5	0217	1004	0821	990607	001260
6	0852	1324	0901	000000	001430
7	0930	1456	1004	000143	001725
8	1104	1579	1324	000273	002047
9	1122	2048	1436	000821	002048
10	1324	6507	2048	000857	003081

Our top recommendation would be to use the 4-short PIN selection policy as an alternative to the non-restrictive 4-free policy. The 4-short policy not only provided better entropy values (8.2945, 153.9729) that are closest to the values for 4-rand (8.3329, 159.6476), but also a remembrance difficulty not too far off from the 4-free policy (P-value $= 0.0053$, two sample t-tests). Using the 4-short policy seems like a suitable compromise between security and usability, which would involve enforcing a short blacklist of commonly used PINs. Unlike our expectations, the 4-long policy did not provide better entropy results (8.2500, 147.5964) than the 4-short policy (8.2945, 153.9729), while having a significantly worse remembrance difficulty than the 4-short (P-value < 0.001, two sample t-tests). This demonstrates that the weak PINs or PIN patterns to be blacklisted need to be selected with more caution: a poor use of a long blacklist may result in a PIN space that is too small, not easing the skewness of the PIN distribution (see Fig. A.7(a)).

The 6-free policy has significantly higher entropy (8.2896, 155.7410) than the 4-free policy (8.1563, 139.3012). This indicates that the length of the PINs is directly related to the strength of the PINs. This trend is clearly captured in Fig A.7 (b): only five PINs were chosen more than once under the 6-free policy. However, there is still a high likelihood of an attacker who already knows the user's 4 digit PIN to also compromise the 6 digit PIN. We mention this because by comparing the participants' PINs selected under 4-free and 6-free polices, it was found that 39.76% of the PINs selected under the 4-free policy were subsequences of the PINs selected under 6-free policy.

In addition, we observed how the occurrence frequency of each digit changed among the PINs selected under the five different policies. The graphs in Fig. A.8 are used to compare the occurrence frequency of each number (0–9) in each digit of the PINs. Our first observation is that the graphs for the 1st, 3rd and 5th digits are more skewed, implying that there is a tendency for the odd digits to have more uneven distribution of numbers. 4-short was more effective in easing the skewness of the 1st digit whereas 6-free did not make much difference. As for the 3rd digit, 4-short made a slight difference, whereas 6-free, again, was not too effective. Just looking at these graphs, however, it is not easy to say how effective the PIN selection policies would be in practice as the skewness (uneven distribution) still exists even when these policies are enforced. Again, this emphasises the importance of designing any kind of PIN blacklists with caution.

6. Conclusions and future work

When it comes to choosing PINs for mobile phones or any other system, users face conflicting set of requirements between security and usability: how easy is it to remember versus how hard is it for an adversary to guess. The reality is that, the users often ignore the security requirements and choose ones that are easy to remember and reuse. This provides opportunities for attackers to efficiently make guesses and compromise PINs.

To help users choose stronger PINs, PIN selection policies may be used to ensure that the PINs meet the minimum security requirements. This paper studied the effectiveness of some carefully designed PIN selection policies aimed at improving the PIN security, and investigated how such policies can affect the overall usability of PINs. An online user study was conducted, requesting the users to choose PINs under different PIN selection policies. By observing the entropies of the collected PINs and the remembrance difficulty (metrics that measure how easy it is to remember a PIN), we made recommendations to use a policy that restricts the user from selecting some number of commonly used PINs. It was also recommended that any type of PIN- or PIN-pattern-based blacklists as such should be constructed with caution as they could lead to a PIN space that is too small, making it easier for the attackers to guess.

Having relied on a relatively small sample pool (332 participants), there are some limitations in generalising our observations about the effectiveness of the PIN selection policies. For instance, people based in different geographical locations compared to the ones participated in the study may have different PIN selection behaviours and expectations. Another limitation comes from the way the PIN memorability was measured. The participants rated memorability based on their thoughts/feelings about the PINs based on 5-point Likert scales; these thoughts/feelings might be different to their actual ability to remember PINs. As an extension to this work, we plan to conduct a between-subject study in the future to perform a more comprehensive analysis of the policies and their performances. It is also our plan to study the PIN characteristics for banking applications and compare them with the results discussed in this paper. It would also be interesting to investigate how the users feel about having to change their PINs frequently and how the PIN selection policies should be designed and enforced in those scenarios.

Acknowledgements

We thank Daniel Amitay for sharing his collected PIN datasets with us. The authors would like to thank Ross Anderson, Mike Bond, and Katherine Kim for their careful attention and insightful comments. Not least, we would like to thank everyone who has participated in the user study.

Appendix A. Analysis of the occurrence frequency of commonly used PIN subsequences in the real PIN dataset

We examine the occurrence frequency of commonly used PIN subsequences from the population of real PINs in (Amitay, Jun 2011). Our assumption is that the subsequences would work in circles; i.e. the first digit would appear after the last digit. The results are shown in Fig. A.9 and A.10, both showing similar trends: the occurrence frequency of PIN subsequences decreases dramatically around the 10th rank. Again, we observe that a few popular PIN subsequences make up a large portion of the total set of PIN subsequences.

To measure the randomness of these PIN subsequences, we used the two entropy methods discussed in Section 4.3, Shannon's method and Massey's method that is referred to as the guessing entropy. We consider the distribution of PIN subsequences of length l, starting from a position s in a PIN, to be a random variable X drawn from a finite distribution $P = \{p_1, p_2, \ldots, p_n\}$ with probability $p_i = P(X = x_i)$ for each possible answer x_i, where $i \in [1, n]$.

The computation results are shown in Table A9, demonstrating similar trends in both entropy measures. An obvious observation is that the entropy of PIN subsequences increases as their length increases. It is natural; it is relatively easy to guess a digit correctly than two or three digits together. We also noted that the entropy of a PIN subsequence that contains the first digit is significantly lower than that of the other PIN subsequences. For example, the entropy with $l = 2$ and $s = 4$ is lower than the entropy with $l = 2$ and $s = 3$. Thus, the attacker's chance to correctly guess a PIN subsequence will increase if it contains the first digit.

Table A9 – Entropy computation of PIN subsequences that have length l and start from position s in a PIN.

l	s	Shannon entropy	Guessing entropy
1	1	2.9914	3.7041
	2	3.2664	4.7412
	3	3.2750	4.7953
	4	3.3042	5.0813
2	1	6.0878	28.5872
	2	6.3266	33.8842
	3	6.4232	36.6526
	4	6.2222	30.6322
3	1	8.8134	224.1422
	2	9.2238	285.6152
	3	9.0430	241.8040
	4	9.0521	250.2686
4	1	11.4165	1819.8000

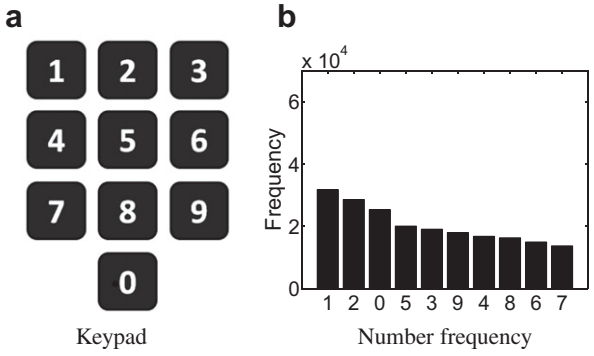

Fig. A.1 – A typical keypad layout for mobile phones and a skewed frequency distribution on numbers used for the PINs collected from (Amitay, Jun 2011).

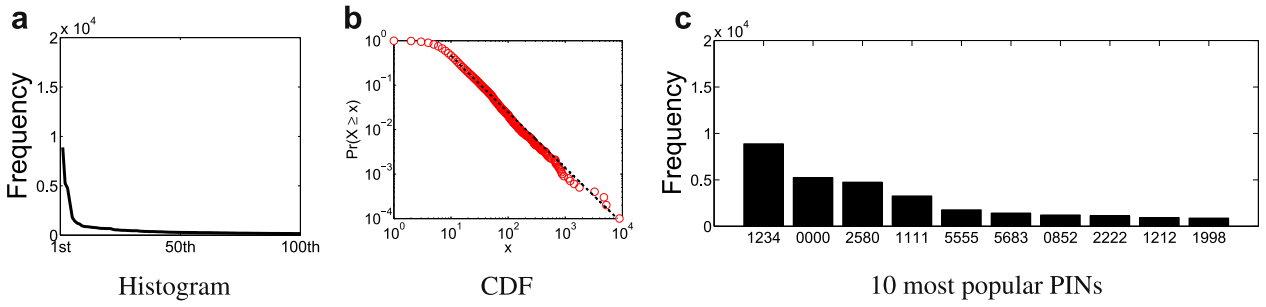

Fig. A.2 – Various characteristics of the PINs collected from (Amitay, Jun 2011) and the best fit function. (a) the occurrence frequency of the 100 most popular PINs; (b) the cumulative distribution function (CDF) of PINs (circles); (c) the 10 most popular PINs.

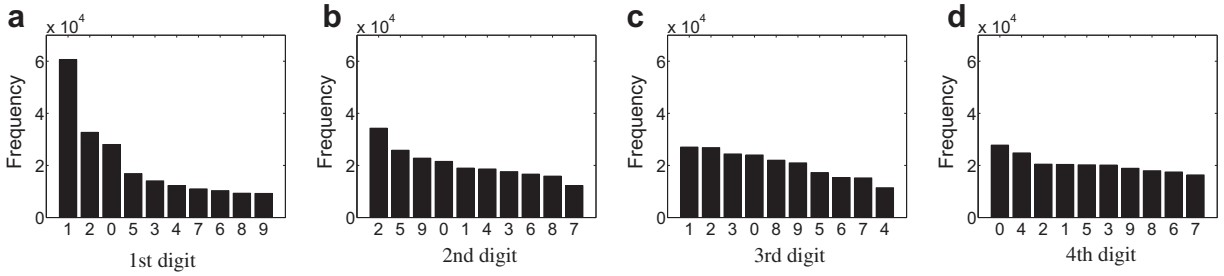

Fig. A.3 – Occurrence frequency of each number in four different positions of the PINs. Numbers are sorted in descending order by the occurrence frequency.

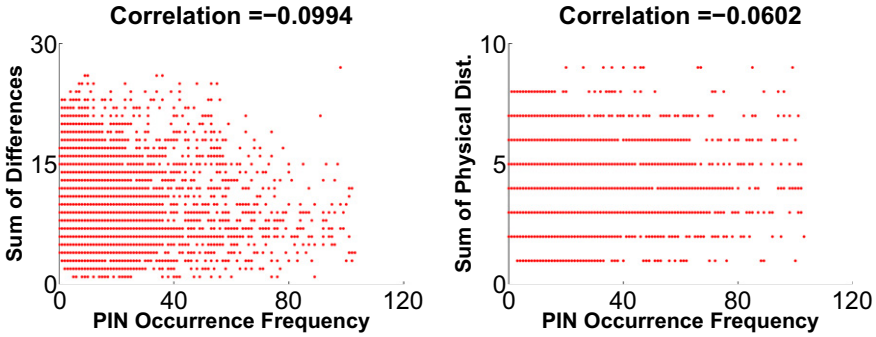

Fig. A.4 – Two scatter plots: (a) shows the correlation between the PIN occurrence frequency (*x*-axis) and the sum of the differences (*y*-axis); (b) shows the correlation between the PIN occurrence frequency and the sum of the physical distances.

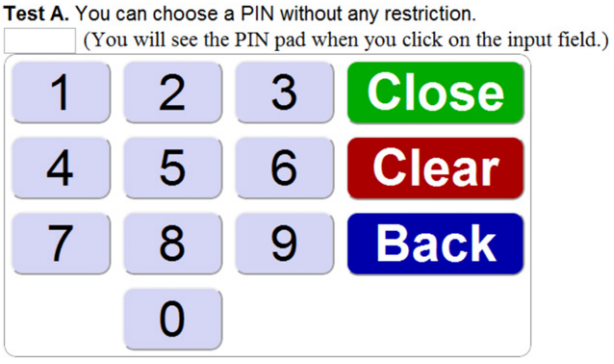

Fig. A.5 – GUI for entering the PIN. For each test, this keypad appeared when the participant clicked on the input text field.

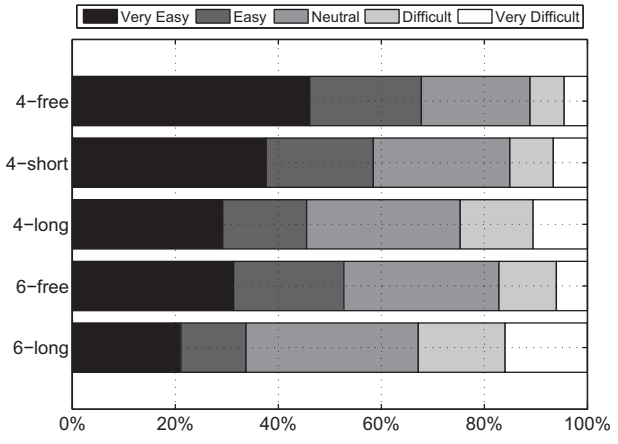

Fig. A.6 – The participants' memorability response to how difficult/easy it is to remember the selected PINs. 'Very Easy' represents very easy to remember, and 'Very Difficult' represents very difficult to remember.

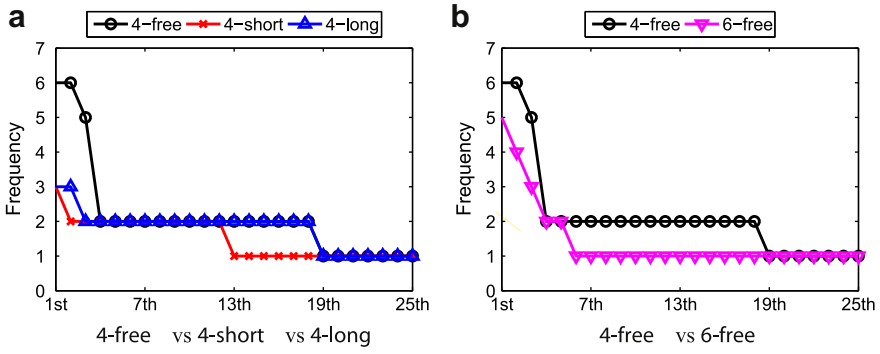

Fig. A.7 – Frequencies of occurrence of the most 25 popular PINs under different PIN selection policies.

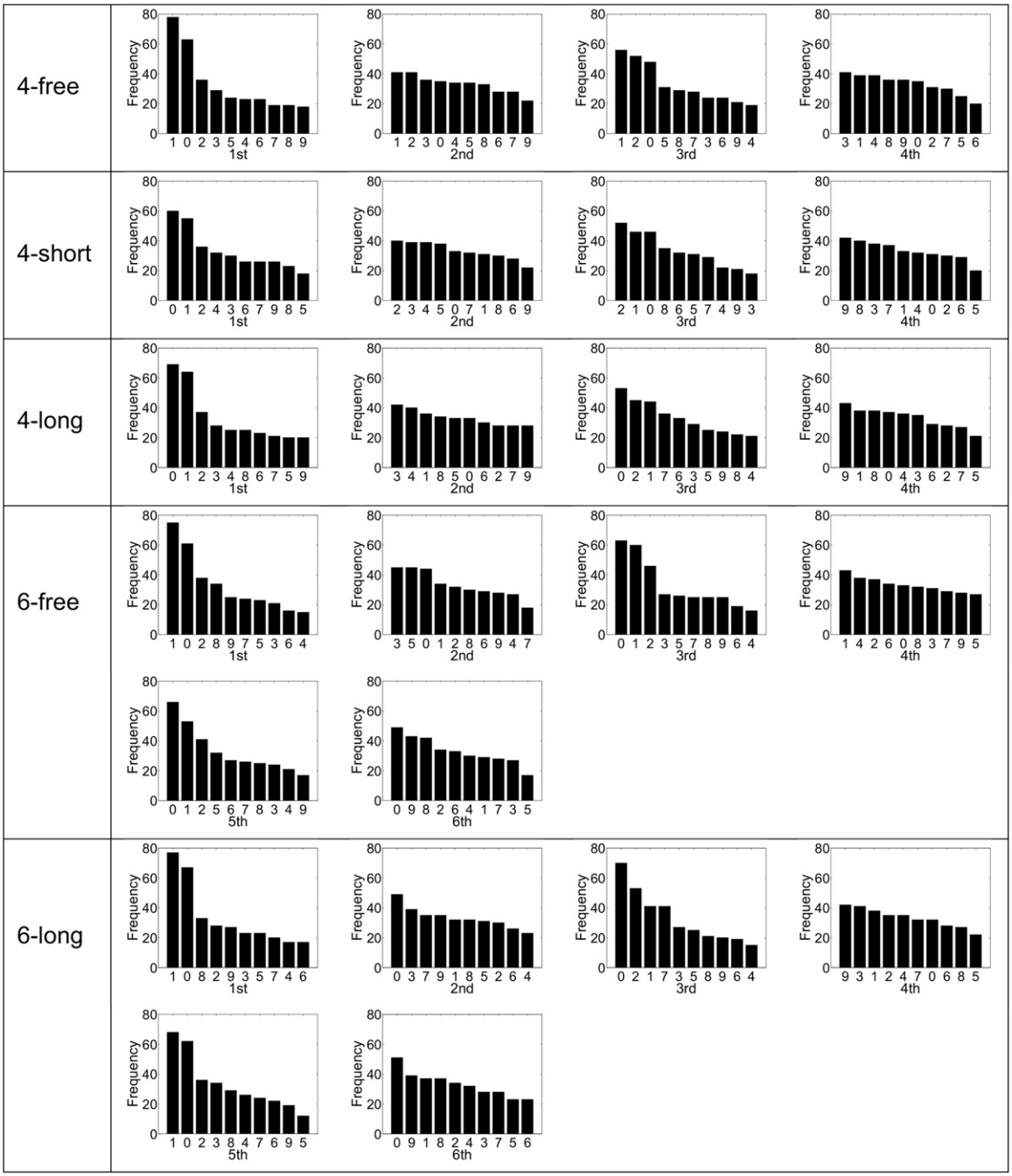

Fig. A.8 – Occurrence frequency of each digit of the PINs selected under five different policies. For each digit, the numbers (0–9) are sorted by their occurrence frequency in descending order.

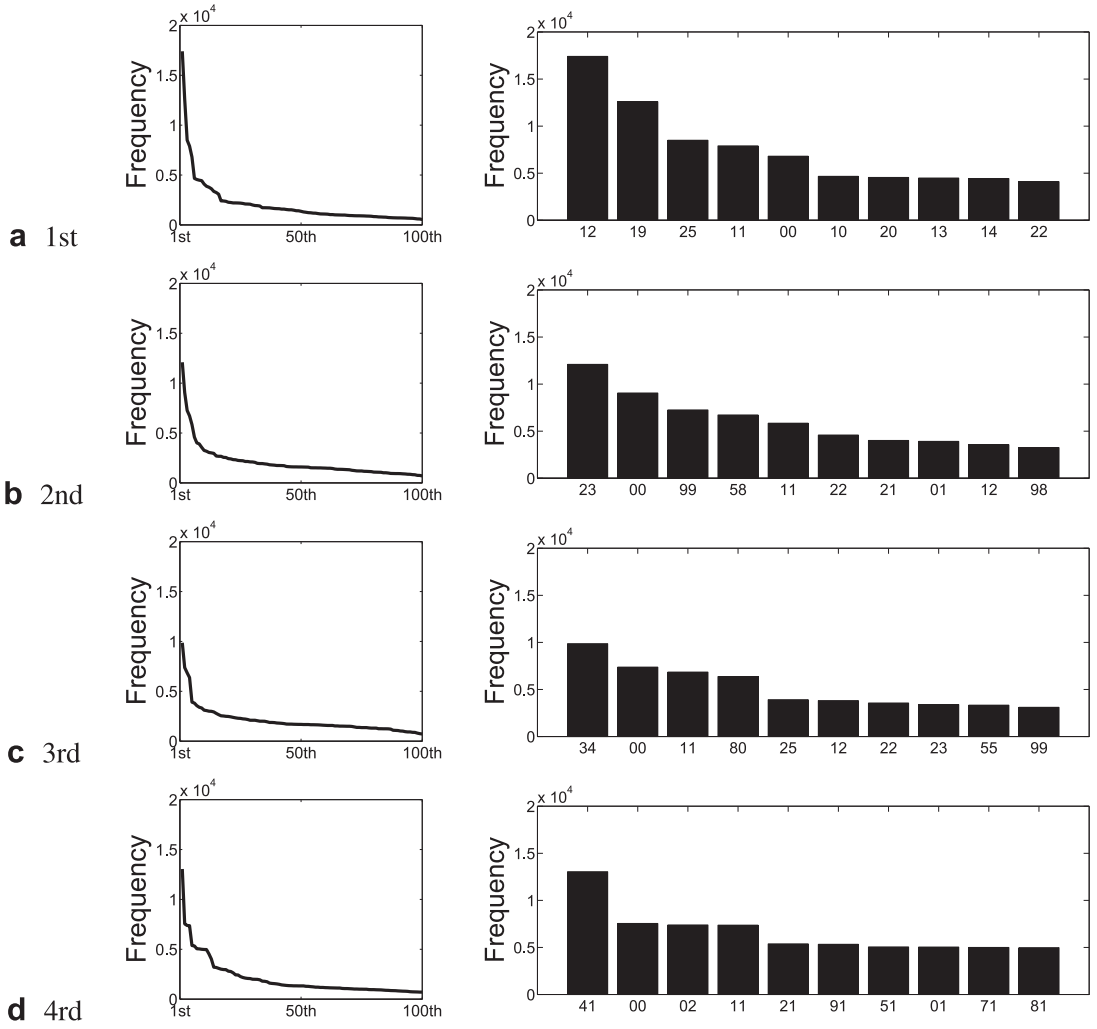

Fig. A.9 – Occurrence frequency of 100 most popular PIN subsequences that are of length 2, where each subsequence starts form four different positions of the PINs. PIN subsequences are sorted in descending order of their occurrence frequency. The 10 most popular PIN subsequences are represented on the right.

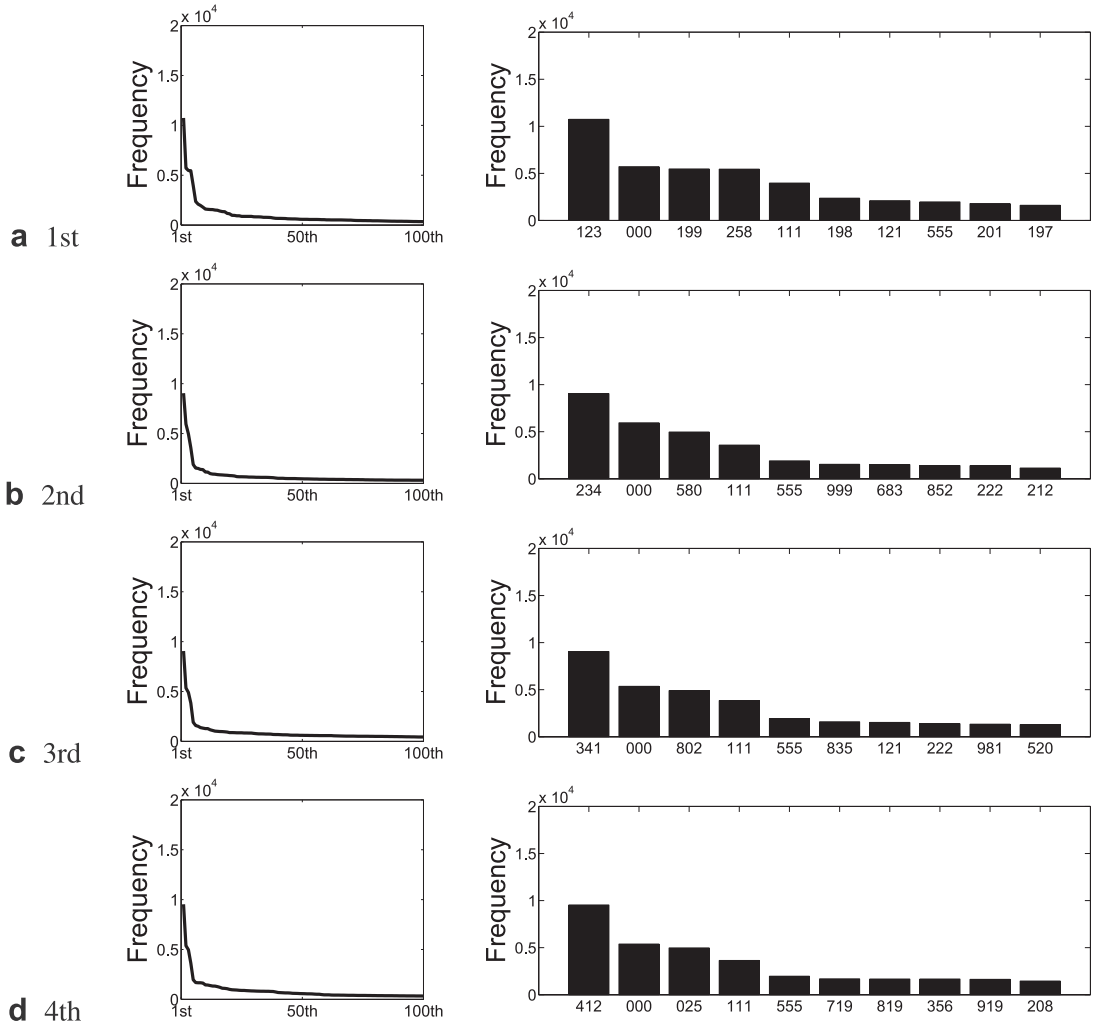

Fig. A.10 – Occurrence frequency of 100 most popular PIN subsequences that are of length 3.

REFERENCES

Amitay D. Most common iPhone passcodes, http://amitay.us/blog/files/most_common_iphone_passcodes.php; Jun 2011.

Anderson R. Can we fix the security economics of federated authentication?. In: SPW 2011, 19th international workshop on security protocols, London, UK; 2011.

Apple Inc. Licensed application end user license agreement, http://www.apple.com/legal/itunes/appstore/dev/stdeula/.

Brown AS, Bracken E, Zoccoli S, Douglas K. Generating and remembering passwords. Applied Cognitive Psychology 2004; 18(6):641–51.

Burr WE, Dodson DF, Polk WT. Electronic authentication guideline; 2006. Tech. rep.

Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. SIAM Review 2009;51:661–703.

Gaw S, Felten EW. Password management strategies for online accounts. In: Proceedings of the second symposium on usable privacy and security, SOUPS '06. New York, NY, USA: ACM; 2006. p. 44–55.

Hart D. Attitudes and practices of students towards password security. Journal of Computing Sciences in Colleges 2008;23: 169–74.

Inglesant PG, Sasse MA. The true cost of unusable password policies: password use in the wild. In: Proceedings of the 28th international conference on human factors in computing systems, CHI '10. New York, NY, USA: ACM; 2010. p. 383–92.

Komanduri S, Shay R, Kelley PG, Mazurek ML, Bauer L, Christin N, et al. Of passwords and people: measuring the effect of password-composition policies. In: Proceedings of the 2011 annual conference on human factors in computing systems, CHI'11. New York, NY, USA: ACM; 2011. p. 2595–604.

Kuo C, Romanosky S, Cranor LF. Human selection of mnemonic phrase-based passwords. In: Proceedings of the second symposium on usable privacy and security, SOUPS '06. New York, NY, USA: ACM; 2006. p. 67–78.

Massey JL. Guessing and entropy. In: Proceedings of the IEEE international symposium on information theory; 1994. p. 204.

Sasse MA, Brostoff S, Weirich D. Transforming the 'weakest link' – a human/computer interaction approach to usable and effective security. BT Technology Journal 2001;19:122–31.

Shannon CE. A mathematical theory of communication. Bell System Technical Journal 2001;27.

Shay R, Bertino E. A comprehensive simulation tool for the analysis of password policies. International Journal of Information Security 2009;8:275–89.

Shay R, Bhargav-Spantzel A, Bertino E. Password policy simulation and analysis. In: Proceedings of the 2007 ACM workshop on digital identity management, DIM '07. New York, NY, USA: ACM; 2007. p. 1–10.

Shay R, Komanduri S, Kelley PG, Leon PG, Mazurek ML, Bauer L, et al. Encountering stronger password requirements: user attitudes and behaviors. In: Proceedings of the sixth symposium on usable privacy and security, SOUPS '10. New York, NY, USA: ACM; 2010. pp. 2:1–2:20.

Vance A. If your password is 123456, Just make it HackMe; Jun 2010.

Vu K-PL, Proctor RW, Bhargav-Spantzel A, Tai B-LB, Cook J, Schultz EE. Improving password security and memorability to protect personal and organizational information. International Journal of Human-Computer Studies 2007;65(8): 744–57.

Yan J, Blackwell A, Anderson R, Grant A. Password memorability and security: empirical results, security privacy. IEEE 2004;2(5): 25–31.

Hyoungshick Kim is a Ph.D. candidate in the Computer Laboratory at the University of Cambridge as a PhD student. He received the B.S. degree from the Department of Information Engineering at Sungkyunkwan University in Korea and M.S. degree from the Department of Computer Science at KAIST in Korea, in 1999 in 2001, respectively. He previously worked for Samsung Electronics as a senior engineer from May 2004 to September 2008. He also served a member of DLNA and Coral standardization for DRM interoperability in home networks. His current research interest is focused on privacy and anonymity in complex networks and distributed systems.

Jun Ho Huh is a postdoctoral research associate in Information Trust Institute, University of Illinois at Urbana-Champaign. He received his Ph.D. degree from Oxford University, investigating new ways of applying Trusted Computing and virtualization to the design of trustworthy audit/logging systems. At ITI, he is currently involved in the design and development of a least-privilege access control system for DCS/SCADA systems.