FEATURE 2

HEARING A PIN DROP

David Harley
Small Blue-Green World/Mac Virus, UK

One thing has become painfully obvious in the light of the recent spate of attacks and data leakages, not to mention various LulzSec nautical naughtiness. Clearly, people are continuing to use highly stereotyped password strategies: in other words, many, many people are using a very, very small selection of passwords. But while there's plenty of research on password use and re-use – mostly derived from the analysis of known collections of exposed passwords [1] to see which are the most commonly used – there is virtually no equivalent research concerning purely numerical passcodes such as PINs (Personal Identification Numbers). While there are no high-profile and publicly available repositories of leaked PIN data allowing empirical analysis (scraping underground forums presents both practical and ethical problems), there is one recent instance [2] of research based on analysis of smartphone passcodes, though it's not the result of a LulzSec-type breach.

Daniel Amitay has been marketing an app [3] called Big Brother (social networking meets reality TV?) – intended to take photos of anyone using an *iPhone* or *iPod Touch 4* without permission (i.e. without entering a passcode). A recent update to the app added code that captures (completely unidentifiably, he promises) the passcodes entered during set-up of the app. This enabled Amitay to run some analysis on a sample set of 204,508 gadgets. These particular iGadgets offer a choice of passcode modes for screenlocking: off, simple four-digit passcode, or a more complex passcode. While we cannot assume that a choice of passcode for Big Brother would reflect either screenlocking passcode selection or PIN selection practice, it seems reasonable to assume that, given the size of the sample, there is likely to be some correlation. (*Apple* clearly thought so, since it has removed the app from the *App Store* and insisted that the passcode-recording code be removed before it is reinstated.)

Here are some preliminary thoughts based mostly, like Amitay's analysis, on the ten highest scoring passcodes. (Since he has kindly shared his data with me, I plan to do a lot more work in the near future on the strategies people employ and on how they might be improved.)

It turns out that 15% of the collected passcodes could be found in the top ten, which consists of the following:

1. 1234
2. 0000
3. 2580
4. 1111
5. 5555
6. 5683
7. 0852
8. 2222
9. 1212
10. 1998

The *iPhone* gives you ten chances to try an activated four-digit screenlock passcode before locking you out. As Amitay et al. have suggested [4, 5], this gives an intruder a disconcertingly good chance of getting in using only the top ten. Other security applications are less forgiving, but selection strategies still bear closer examination.

MNEMONIC LOGIC

The mnemonic logic behind the top ten numbers is more obvious in some cases than in others.

It's hard to think of a more memorable (or obvious) passcode than 1234, for the same reason that 12345 and 123456 regularly appear in password top ten lists – the latter is usually right at the top.

However, any sequence of four *identical* digits is likely to be almost as popular: in this instance, we have 0000, 1111 and 2222 all in the top ten. My guess is that while 3333, 4444 etc. don't feature in the top ten, they're probably not far behind. While 0000 is particularly easy to remember (and therefore to guess), it seems likely that people might choose a different single number according to some rule that makes it more memorable for them, then repeat it as necessary – just as some people use aaaaaaaaa, ggggggggg or zzzzzzzzz or a similar alphabetical sequence for passwords. (However, positioning and accessibility on the keypad may also have a bearing.) Of course, the length of a same-character string may vary according to the requirements of the service demanding the password/passcode: however, that makes very little difference to the ease with which it can be guessed. In the course of a 'dictionary attack' where passwords, passcodes or passphrases are tried according to an ordered list of possibilities, these are likely to be tried very early in the attack.

But why are the other sequences apparently so popular? Figure 1 shows a fairly standard keyboard layout for a basic cellphone/feature phone. Virtual numeric keypads for making phone calls from a smartphone generally follow the same pattern, but have a virtual QWERTY keypad for other kinds of data input, while some feature phones and smartphones have a miniature (hardware) QWERTY keyboard.

Some sequences can be explained by pattern. The middle column of the keypad in descending order gives you 2580, the third most popular choice according to the top ten list.

Figure 1: Standard keyboard layout for a basic cellphone/ feature phone.

Going up the other way – which is just as easy but perhaps a little less intuitive – gives you 0852, the seventh most popular choice.

The middle column is the only one that gives you four digits – the most common length for a PIN – so that probably explains the popularity of these two pattern/code pairs. Other vertical choices in combination with the 0 character are possible, but apparently less popular: 1470 is the 51st most popular choice, while 3690 is the 68th most popular choice. Curiously, given that keyboard patterns are an acknowledged mnemonic aid [6], the reverse patterns did not occur within the sample.

What about 5683? Amitay suggests, convincingly enough, that this is less random than it seems. On the basic phone keypad in Figure 1, you'll see that the letters associated with the number 5683 provide a simple mnemonic using the word LOVE:

(5) JKL (6) MNO (8) TUV (3) DEF

However, that particular association wouldn't work on devices with a QWERTY keyboard like that found on a *BlackBerry* (Figure 2).

Thanks to the single letter-to-number pairing on most *BlackBerry* keypads, there are relatively few four-letter words that conveniently match the nine available letters (as shown in Figure 2: w, e, r, s, d, f, z, x, c). So the single letter-to-number pairing on this type of keypad militates against this particular memorization strategy. The more traditional layout in Figure 1, meanwhile, allows the use of the full alphabet and thus the use of real words and other alphabetical strings as a memory aid, even where passcodes and PINs are limited to four digits.

Figure 2: BlackBerry with QWERTY keyboard.

What about 1212? That would be easy for me to remember, because I'm old enough to remember when New Scotland Yard was the headquarters of London's Metropolitan Police and its very famous telephone number was Whitehall 1212. Later generations might simply use it because a simple [n; n+1; n; n+1] sequence is almost as easy to remember as [n; n; n; n;].

And 1998? Amitay suggests that people use four-digit sequences relating to years that have special significance for them, such as their date of birth or date of graduation. One of the nice things about being my age is that you have a lot of memorable dates behind you – if you can remember them, of course, and are confident they aren't too public to be safe.

What does this tell us about other digital passcodes? Telephone keypads are not always the same as ATM keypads, most significantly in that while even antique rotary telephone dials have letters as well as numbers [7] (though the matching of letters to numbers hasn't always been consistent), not all ATM keypads do. While many modern keypads have the same layout as the telephone keypad shown in Figure 1, some use the common calculator configuration shown in Figure 3:

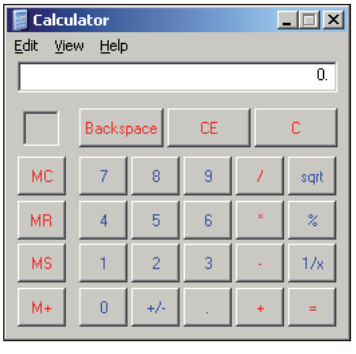

Figure 3: Common calculator configuration.

Many (perhaps even most) ATM keypads and many of the digital safes you find in hotel rooms also use this layout, so the stratagems relating to memorable dates or number sequences are probably also commonly used for ATM PINs.

Some keypads (including numeric keypads for computers and many models of calculator) use almost the same layout but in reverse.

This is a layout that has been used for fast data entry for many years in business. It may not be so frequently encountered in contexts commonly associated with numerical passcodes but nonetheless, it's still worth noting that in contexts where a keypad like this *is* in use, a similar strategy would probably result in the use of 7410 and 0147 rather than 0852 and 2580, and the off-centre positioning of the 0 may further lessen the likelihood of its use in combination with the other columns.

Rasmussen and Rudmin [6] offer a list of mnemonic strategies:

1. Learning by rote
2. Remembering by keypad patterning
3. Code re-use
4. Code with personal meaning
5. Code written down and kept separately
6. Code stored in mobile phone
7. Code concealed in a phone number
8. Numbers paired with letters
9. Written down and kept in proximity
10. Written down but rearranged
11. Notated as a transform of the code.

This data gives an opportunity to confirm to some extent the degree to which these strategies are used. More importantly, perhaps, while articles on the best and worst strategies for choosing passcodes are not in short supply, the data gives us a better starting point for evaluating the entropic efficacy of these strategies as the basis for better recommendations to end-users. And that's a topic I certainly plan to return to.

REFERENCES

[1] Harley, D. Good passwords are no joke. SC Magazine, 2011. http://www.scmagazineus.com/good-passwords-are-no-joke/article/204675/.

[2] Amitay, D. Most Common iPhone Passcodes. 2011. http://amitay.us/blog/files/most_common_iphone_passcodes.php.

[3] Amitay, D. Big Brother Camera Security. 2011. http://amitay.us/projects/big%20brother.php.

[4] Harley, D. Passcodes and Good Practice. Mac Virus, 2011. http://macviruscom.wordpress.com/2011/06/15/passcodes-and-good-practice/.

[5] Cluley, G. The top 10 passcodes you should never use on your iPhone. Naked Security, 2011. http://nakedsecurity.sophos.com/2011/06/14/the-top-10-passcodes-you-should-never-use-on-your-iphone/.

[6] Rasmussen, M.; Rudmin, F. W. The coming PIN code epidemic: A survey study of memory of numeric security codes. Electronic Journal of Applied Psychology. 6(2):5-9 (2010). http://ojs.lib.swin.edu.au/index.php/ejap/article/viewPDFInterstitial/182/220.

[7] http://www.zyra.info/phonedial.htm.