

Grundläggande lösenordsanalys

Daniel Bosk*

pwdanalysis.tex 753 2013-03-13 22:12:47Z danbos

Innehåll

1	Val av lösenord	1
2	Att angripa lösenord	2
2.1	Råstyrkemethoden applicerad på Exempel 2	3
2.2	Forskning på området	3

1 Val av lösenord

Vi ska nu titta på hur detta kan användas för att undersöka säkerheten hos lösenord. Vi inleder med att definiera styrkan hos ett lösenord.

Definition 1. *Styrkan* hos ett lösenord av längd n som väljs från ett alfabet A är $|A|^n$, det vill säga antalet tecken i vårt alfabet upphöjt till längden av lösenordet.

Anledningen till att vi väljer Definition 1 är för att detta är antalet möjliga lösenord tillika antalet gissningar som krävs, i värsta fall, för att gissa på rätt lösenord. Det vill säga, desto fler antal gissningar ett lösenord kräver ju längre tid tar det att gissa och alltså är det säkrare. Rent statistiskt sett kommer antalet gissningar som krävs att vara hälften av vårt mått för lösenordsstyrkan eftersom att i medel kommer vi att behöva gå igenom hälften av antalet lösenord.

Det finns flera angreppssätt för att skapa lösenord, exempelvis genom att välja en kombination av tecken; då har vi bokstäver, siffror och specialtecken som alfabet. Det går också att skapa lösenord genom att slumpmässigt välja några ord som kombineras till ett lösenord; i detta fall utgör ordlistan som vi väljer från vårt alfabet.

Vi börjar med det första fallet, där vi skapar ett lösenord genom att kombinera tecken.

Exempel 1. Om vi ska skapa ett lösenord som är fem tecken långt och får innehålla bokstäverna A-Z, både gemener och versaler, siffrorna 0-9 samt specialtecknen " ! @ . % & ", då kan vi se att vårt alfabet

$$A = \{A, B, \dots, Z, a, b, \dots, z, 0, 1, \dots, 9, !, @, ., \%, \&\}.$$

*Detta verk är tillgängligt under licensen Creative Commons Erkännande-DelaLika 2.5 Sverige (CC BY-SA 2.5 SE). För att se en sammanfattning och kopia av licenstexten besök URL <http://creativecommons.org/licenses/by-sa/2.5/se/>.

Vårt alfabet innehåller således $|A| = 26 \cdot 2 + 10 + 5 = 67$ tecken. Styrkan hos ett sådant lösenord är följaktligen $67^5 = 1350125107$.

Nu fortsätter vi med att titta på fallet med att välja lösenord genom att slumpmässigt välja några ord.

Exempel 2. Vi bestämmer oss för att använda ett lösenord med fyra slumpmässigt valda ord från svenska språket. Detta ger oss ett alfabet

$$A = \{\text{apa, banan, hej, dator, \dots}\},$$

där A alltså innehåller samtliga ord i Svenska Akademiens Ordlista [10]. Enligt Svenska Akademien innehåller ordlistan ungefär 125000 ord, detta ger oss $|A| = 125000$. Enligt Definition 1 får vi då att styrkan för ett lösenord av denna typ är

$$|A|^4 = 125000^4 = (2^3 \cdot 5^6)^4 = 2^{12} \cdot 5^{24} = 2^{12} \cdot 25^{12} = (2 \cdot 25)^{12} = 50^{12}, \quad (1)$$

det vill säga 24414062500000000000.

Det senaste fallet kan också ses ur det första fallets perspektiv. Om vi tittar på det sista ledet i likheten i ekvation (1) ser vi direkt att detta skulle exakt motsvara ett alfabet med 50 tecken och en lösenordslängd av 12 tecken.

2 Att angripa lösenord

Den egentliga skillnaden mellan Exempel 1 and Exempel 2 är de möjliga angreppssätten. Exempelvis går det att angripa Exempel 2 på ett sätt som gör att dess styrka blir större.

För att kunna knäcka ett lösenord måste en gissning kunna testas. Detta kan göras på flera sätt. Exempelvis kan vi skicka en gissning till inloggningsprogrammet, exempelvis inloggningsgränssnittet för en webbmail. Detta tar oftast lång tid och gör systemet i fråga varse om ett angreppsförsök. Alternativet är att vi har tillgång till lösenordsdatabasen och kan direkt testa mot den. Detta är inte ovanligt [jmf. 6; 9; 7; 5]. I många system lagras inte själva lösenorden utan ett hashvärde av lösenordet. Det vi gör för att testa vår gissning är att vi beräknar hashvärdet för gissningen och jämför detta med det värde som finns i lösenordsdatabasen. Notera att eftersom att många användare återanvänder sina lösenord i flera system behöver vi inte nödvändigtvis ha lösenordsdatabasen för det system vi är intresserade av att knäcka, det räcker med att det finns användare som har konton i båda systemen.

Det finns huvudsakligen tre olika former av lösenordsknäckning. Dessa är

- råstyrkemethoden (brute-force attack),
- ordlistemetoden (dictionary-attack), och
- i övrigt kvalificerade gissningar.

Social engineering är egentligen inte en lösenordsknäckningsstrategi en teknik att ta sig förbi åtkomstkontroll [1, s. 18], men eftersom att åtkomstkontroll vanligtvis implementeras med lösenordsskydd är den värd att nämna i sammanhanget.

Råstyrkemetoden innebär att vi låter ett program gå igenom alla möjliga teckenkombinationer i vårt alfabet för att finna lösenordet i fråga. Med denna metod är vi garanterade att finna lösenordet, men det kan dock ta väldigt lång tid.

Ordlistemetoden är en effektivare metod, med denna metod använder vi en ordlista med de vanligast förekommande lösenorden och vi låter dessa vara våra gissningar. Det är vanligt att lösenordslistor skapas av publicerade hackade databaser, vilket är fallet i analyserna gjorda av Cubrilovic [6]; Oberheide [9]; Hunt [7]; Cluley [5]. Detta ger färre antal gissningar och går således snabbare, men om lösenordet vi försöker att knäcka inte finns med i ordlistan kommer vi aldrig att kunna knäcka det med denna metod.

2.1 Råstyrkemetoden applicerad på Exempel 2

Låt oss anta att medellängden av orden i Svenska Akademiens ordlista är fem tecken och att dessa tecken enbart är gemener från det svenska språket. Det innebär att Exempel 2 ger oss en lösenordsstyrka på

$$29^{5 \cdot 4} = 29^{20} = 176994576151109753197786640401. \quad (2)$$

Med hjälp av logaritmen kan vi enkelt se vilken av dessa som är störst. Vi har att $\log(29^{20}) \approx 29$ medan $\log(50^{12}) \approx 20$. Det är alltså mycket starkare sett ur detta perspektiv.

Hur spelar denna representation någon roll, vad betyder skillnaden mellan ekvation (1) och ekvation (2)? Det första som bör påpekas är att i ekvation (2) tas även teckenkombinationer som ej är svenska ord med. Detta eftersom att valet var att välja fyra uppsättningar av fem tecken. Betydelsen av detta är att om vi låter en dator bara slumpa fram 20 bokstäver (gemener), då kommer det att resultera i antalet gissningar från ekvation (2). Men det är inte ens säkert att datorn kommer att hitta rätt lösenord om vi råkar välja fyra långa ord som alla var minst sex bokstäver långa. Detta är ett problem med denna uppskattning.

Om vi däremot använder Svenska Akademiens ordlista som alfabet när vi tillämpar råstyrkemetoden då kommer antalet gissningar att maximalt bli de från ekvation (1) och vi kommer garanterat att finna lösenordet.

2.2 Forskning på området

Angreppsmetoder mot och hur användare väljer lösenord och -fraser är ett aktivt forskningsområde. Metoderna blir alltmer avancerade och exempelvis undersöker Bonneau och Shutova [4] hur lingvistikens påverkar valet av lösenord bestående av flera ord. Bonneau och Shutova finner att användarna inte väljer slumpmässiga ord utan föredrar att välja dem anpassade efter naturligt språk. Exempelvis XKCD:s ”correct horse battery staple”¹ föredras framför ”horse correct battery staple” på grund av att det första alternativet är mer grammatiskt korrekt.

Kuo m.fl. [8] gjorde en undersökning av hur användare skapar lösenord som är lätta att komma ihåg. Kuo m.fl. undersökte styrkan hos frasbaserade lösenord.

¹URL: <http://xkcd.com/936/>.

Det vill säga lösenord som skapas utifrån en mening, exempelvis Googles exempel ”To be or not to be, that is the question”² som ger lösenordet ”2bon2btitq”. Det visade sig då i undersökningen att denna typ av lösenord är lite säkrare än medellösenordet, men användare väljer fortfarande lösenord som är lätta att gissa.

Utöver de ytliga analyserna av läckta lösenordsdatabaser som gjordes av Cubrilovic [6]; Oberheide [9]; Hunt [7]; Cluley [5] har Bonneau [3] gjort en mer formell och djupgående analys. Bonneau har tittat på lösenorden hos nästan 70 miljoner Yahoo!-användare och har då kunnat undersöka skillnader mellan olika demografiska grupper. I artikeln utvecklas en metrik för styrkan hos lösenord, en mer formell och ingående än den mycket enkla metrik som ges i Definition 1.

Bonneau [2] har skrivit sin avhandling om tillvägagångssätt för att gissa hemligheter valda av människor, där lösenord är en självklar sådan hemlighet. I avhandlingen presenteras en matematisk modell för mänskligt val och en metrik för att modellera motståndskraften mot olika gissningsattacker.

Referenser

- [1] Anderson, Ross J. *Security engineering : a guide to building dependable distributed systems*. Wiley, Indianapolis, IN, 2 utgåvan, 2008. ISBN 978-0-470-06852-6 (hbk.).
- [2] Bonneau, Joseph. *Guessing human-chosen secrets*. Doktorsavhandling, University of Cambridge, may 2012. URL <http://www.cl.cam.ac.uk/~jcb82/doc/2012-jbonneau-phd-thesis.pdf>.
- [3] Bonneau, Joseph. The science of guessing: analyzing an anonymized corpus of 70 million passwords. I: *IEEE Symposium on Security and Privacy*, 2012. URL http://www.cl.cam.ac.uk/~jcb82/doc/B12-IEEEESP-analyzing_70M_anonymized_passwords.pdf.
- [4] Bonneau, Joseph och Shutova, Ekaterina. Linguistic properties of multi-word passwords. I: *USEC*, 2012. URL http://www.cl.cam.ac.uk/~jcb82/doc/BS12-USEC-passphrase_linguistics.pdf.
- [5] Cluley, Graham. The worst passwords you could ever choose exposed by Yahoo Voices hack, jul 2012. URL <http://nakedsecurity.sophos.com/2012/07/13/yahoo-voices-poor-passwords/>.
- [6] Cubrilovic, Nik. RockYou Hack: From Bad to Worse, dec 2009. URL <http://techcrunch.com/2009/12/14/rockyou-hack-security-myspace-facebook-passwords/>.
- [7] Hunt, Troy. A brief Sony password analysis, jun 2011. URL <http://www.troyhunt.com/2011/06/brief-sony-password-analysis.html>.
- [8] Kuo, Cynthia, Romanosky, Sasha, och Cranor, Lorrie Faith. Human Selection of Mnemonic Phrase-based Passwords. Teknisk rapport 36, Institute of Software Research, 2006. URL <http://repository.cmu.edu/isr/36/>.

²URL: <http://www.lightbluetouchpaper.org/2011/11/08/want-to-create-a-really-strong-password-dont-ask-google/>.

- [9] Oberheide, Jon. Brief analysis of the gawker password dump, dec 2010. URL <https://blog.duosecurity.com/2010/12/brief-analysis-of-the-gawker-password-dump/>.
- [10] Svenska Akademin. SAOL på nätet. URL http://www.svenskaakademien.se/svenska_spraket/svenska_akademiens_ordlista/saol_pa_natet. Hämtad den 24 februari 2012.