Selecting Secure Passwords

Eric Verheul
PricewaterhouseCoopers Advisory

&

Radboud University Nijmegen

VVSS 2007

Outline

• Password protection
• Mathematical model
• A new bound
• Application: selecting near optimal passwords
• Conclusion
Password protection: context

Main threats:
- Interception of passwords
- On-line guessing of passwords
- Stealing of passwords from server

Main controls:
- Simple: use SSL, difficult: use ‘Encrypted Key Exchange’
- Account lockout, minimal requirements on passwords.
- Use hashing of passwords

Password protection: Guessing attack

- On possession of password database, attacker can mount a Guessing Attack:
- Guess the password, pwd, for user X; most likely first
- Calculate \( H = \text{hash}(pwd) \)
- Validate if \( H \) occurs in record of user X

Hashed password database, i.e., records of type:
- \([\text{Username}, \text{hash(password)}] \)
Password protection: our context

• We assume proper hashing is used, and we restrict ourselves to the situation that the attacker possesses one hash $H$ of a password (and knows the hash function used). We further distinguish two types of guessing attacks on $H$:

**Complete attack**

• Attacker keeps on guessing until he has found a password that hashes to $H$
• Typically corresponds with powerful attacker

**Incomplete attack**

• Attacker is only willing to try a certain number of guesses for the password
• Typically corresponds with casual attacker only willing to let his PC guess for limited time, e.g. 24 hours ($\approx 2^{38}$ tries).

Password protection: informal description of the problem

• Finding a mathematical model for passwords, leading to passwords that are:
  — ‘Adequately’ secure (as acceptable by the user) against both complete and incomplete guessing attacks
  — On average have a length as ‘short’ as possible (given certain alphabet)

• Different from ‘easily memorized’ passwords, but relevant for one time used passwords (e.g., initial passwords, activation codes etc.)
Mathematical model: representation of passwords

- Passwords correspond to a finite variable $X$ with a discrete probability distribution $(p_1, p_2, p_3, \ldots, p_n)$ on $n$ points (number of passwords), i.e. $p_i \geq 0$ and they sum up to one.
- We assume throughout that $p_1 \geq p_2 \geq p_3, \ldots, \geq p_n \geq 0$.
- Evidently, $p_1 \geq 1/n$.

Mathematical model: measures of security

**Guessing entropy**: expected number of guesses in a complete off-line attack

$$\alpha = \sum_{i=1}^{n} i \cdot p_i$$

**Min entropy**: measure for resistance against an incomplete off-line attack

$$H_{\infty} = -\log_2 (p_1)$$

**Shannon entropy**: measure for average length of passwords

$$H = -\sum_{i=1}^{n} p_i \cdot \log_2 (p_i)$$

It simply follows that $H \geq H_{\infty}$.
Mathematical model: measures of security

**Guessing entropy:** expected number of guesses in a *complete* off-line attack

\[ \alpha = \sum_{i=1}^{n} i \cdot p_i \]

Example: Uniform Dist on n points

\[ (n+1)/2 \]

**Min entropy:** measure for resistance against an *incomplete* off-line attack

\[ H_{\infty} = -\log_2(p_1) \]

\[ \log_2(n) \]

**Shannon entropy:** measure for average length of passwords

\[ H = -\sum_{i=1}^{n} p_i \cdot \log_2(p_i) \]

\[ \log_2(n) \]

Mathematical model: problem formulation

- Given a value of guessing entropy \( \alpha \) and a upper bound \( \delta \) on \( p_1 \) (or equivalently a lower bound on the Min entropy): what is the minimal Shannon entropy possible?
- Efficiently find such distributions
- Efficiently generate such passwords

- Nist Special pub. 800-63: ‘electronic authentication guideline’ implicitly introduces this model, but does not pursues it.
Mathematical model: misconception

- sci.crypt crypto FAQ: \( 2^H \approx \alpha ? \)
  \[
  \frac{2 \log_2(n)}{n - 1} (\alpha - 1) \leq H \leq \log_2(e \cdot \alpha - 1)
  \]
  McEliece-Yu Massey

- [Massey]: there exists a sequence of distributions in \( n \) of fixed Guessing entropy with Shannon entropy converging to zero.
  Actually, in simulations with random distributions this inequality always seems to hold:
  \[
  \log_2(2\alpha - 1) \leq H \leq \log_2(e \cdot \alpha - 1)
  \]
A new bound: relaxing the condition

- Looking for the minimal value of the Shannon entropy on $C_{n,\alpha}$

$\{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} i \cdot p_i = \alpha \delta \geq p_1 \geq p_2 \ldots \geq p_n \geq 0\}$.

- First look for the minimal value the Shannon entropy $H$ takes on the set $C_{n,\alpha}$

$\{(p_1, \ldots, p_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} i p_i = \alpha, p_1 \geq p_2 \ldots \geq p_n \geq 0\}$

- That is, the set of probability distributions on $n$ points with a given guessing entropy $\alpha$.

A new bound: extreme points convex sets

- $C_{n,\alpha}$ is closed convex set
- Every point is convex combination of extreme points
A new bound: extreme points of $C_{n,\alpha}$

The extreme points of $C_{n,\alpha}$ take the form $X_{j,k,n}$ for integers $j, k$ satisfying $1 \leq j \leq 2\alpha - 1 \leq k \leq n$ and

$$
\begin{pmatrix}
  a_{j,k,n} & a_{j,k,n} & \cdots & a_{j,k,n} & b_{j,k,n} & \cdots & b_{j,k,n} & 0 & \cdots & 0 \\
 1, & 2, & \cdots & j, & j + 1, & \cdots & k, & k + 1, & \cdots & n,
\end{pmatrix}
$$

where

$$
a_{j,k,n} = \frac{-2\alpha + 1 + j + k}{j \cdot k}, \quad b_{j,k,n} = \frac{2\alpha - (j + 1)}{k(k - j)},
$$

and where we define $b_{j,k,n} = 1/(2\alpha - 1)$ for $j = 2\alpha - 1 = k$.

A new bound: extreme points of $C_{n,\alpha}$

$$
k = 2\alpha - 1 \\
k = 2\alpha \\
k = n
$$

These are vectors in $\mathbb{R}^n$. 

VVSS 2007 - Verification and Validation of Software Systems Symposium
A new bound: strong improvement of McEliece-Yu

- Fix number of passwords $n$

Let $H(j) := \text{Shannon entropy of } X_{j,n,n}$

$$k = n \quad X_{1,n,n} \quad X_{2,n,n} \quad \cdots \quad X_{2^{2\alpha-1},n,n}$$

Highest probabilities $\rightarrow a_{1,n,n} \quad a_{2,n,n} \quad \cdots \quad a_{2^{2\alpha-1},n,n}$

- the function $j \rightarrow a_{j,n,n}$ is decreasing
- find real $j$ such that $a_{j,n,n} = \delta$, then the Shannon entropy on $C_{n,\alpha,\delta}$ is $\geq \min(H(j), \log_2(2\alpha - 1))$.

- So extreme points are optimal distributions in ‘own’ $\delta$ class.

Application: selecting near optimal passwords

- Choose a Guessing entropy $\alpha$
- Choose an upper bound $\delta = 1/D$ on $p_1$
- Fix number of passwords $n$
- Choose extreme point $X_{D,n,n}$ in $C_{n,\alpha}$
- Generate passwords according to this distribution (easy)
- For $n \rightarrow \infty$, average password length $\downarrow \cdot \log_2(\delta)$
- Passwords come in two flavors:
  - length $D$ with probability $P_{\min}$ (should be large)
  - length $n-D$ with probability $P_{\max}$ (should be small)
- That is, some users get very long passwords
- Find trade-off between small average length and probability of bothering users with long passwords.
Application: selecting near optimal passwords

\[ \alpha = 2^{64} \]

<table>
<thead>
<tr>
<th>( \log_2(n) )</th>
<th>( \delta )</th>
<th>Average pwd length</th>
<th>Min length</th>
<th>Max length</th>
<th>( P_{\text{min}} )</th>
<th>( P_{\text{max}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>65.0</td>
<td>( 2^{-65.00} )</td>
<td>65.00</td>
<td>40.0</td>
<td>65.0</td>
<td>2.98E-08</td>
<td>1.00E+00</td>
</tr>
<tr>
<td>65.5</td>
<td>( 2^{-41.77} )</td>
<td>58.90</td>
<td>40.0</td>
<td>65.5</td>
<td>2.92E-01</td>
<td>7.07E-01</td>
</tr>
<tr>
<td>66.0</td>
<td>( 2^{-41.00} )</td>
<td>54.00</td>
<td>40.0</td>
<td>66.0</td>
<td>5.00E-01</td>
<td>5.00E-01</td>
</tr>
<tr>
<td>66.5</td>
<td>( 2^{-40.02} )</td>
<td>50.30</td>
<td>40.0</td>
<td>66.5</td>
<td>6.46E-01</td>
<td>3.53E-01</td>
</tr>
<tr>
<td>67.0</td>
<td>( 2^{-40.41} )</td>
<td>47.56</td>
<td>40.0</td>
<td>67.0</td>
<td>7.50E-01</td>
<td>2.50E-01</td>
</tr>
<tr>
<td>67.5</td>
<td>( 2^{-40.28} )</td>
<td>45.53</td>
<td>40.0</td>
<td>67.5</td>
<td>8.23E-01</td>
<td>1.76E-01</td>
</tr>
<tr>
<td>68.0</td>
<td>( 2^{-40.19} )</td>
<td>44.04</td>
<td>40.0</td>
<td>68.0</td>
<td>8.75E-01</td>
<td>1.25E-01</td>
</tr>
<tr>
<td>68.5</td>
<td>( 2^{-40.13} )</td>
<td>42.95</td>
<td>40.0</td>
<td>68.5</td>
<td>9.11E-01</td>
<td>8.83E-02</td>
</tr>
<tr>
<td>69.0</td>
<td>( 2^{-40.09} )</td>
<td>42.14</td>
<td>40.0</td>
<td>69.0</td>
<td>9.37E-01</td>
<td>6.25E-02</td>
</tr>
<tr>
<td>69.5</td>
<td>( 2^{-40.06} )</td>
<td>41.56</td>
<td>40.0</td>
<td>69.5</td>
<td>9.55E-01</td>
<td>4.41E-02</td>
</tr>
<tr>
<td>70.0</td>
<td>( 2^{-40.04} )</td>
<td>41.13</td>
<td>40.0</td>
<td>70.0</td>
<td>9.68E-01</td>
<td>3.12E-02</td>
</tr>
</tbody>
</table>

Conclusion