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

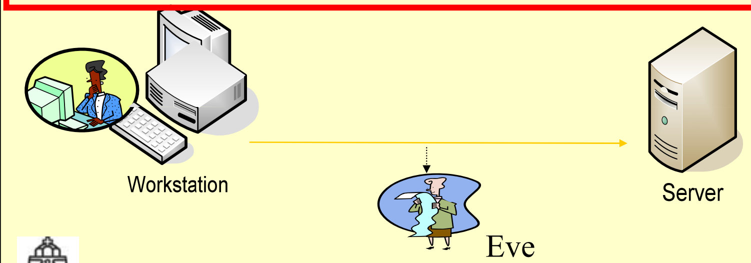
Password protection: context

Main threats:

- Interception of passwords
- On-line guessing of passwords
- Stealing of passwords from server

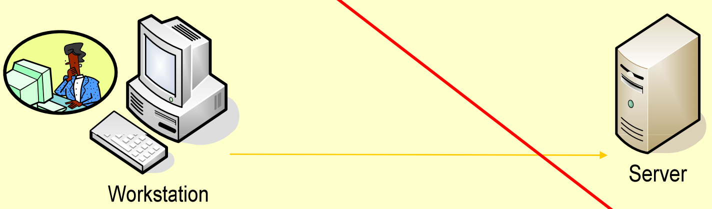
Main controls:

- 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 <u>Guessing Attack:</u>
- Guess the password, *pwd*, for user X; most likely first
- Calculate *H = hash(pwd)*
- Validate if *H* occurs in record of user X



Hashed password database,
i.e. records of type:
•[Username, 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 (≈ 2^{36} 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 ,……, p_n) on *n* points (number of passwords), i.e. $p_i \geq 0$ and they sum up to one.
- We assume throughout that $p_1 \geq p_2 \geq p_3 ,\ldots\ldots \geq p_n \geq 0$.
- 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

Example: Uniform Dist on n points

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

(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 α and a upper bound δ 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

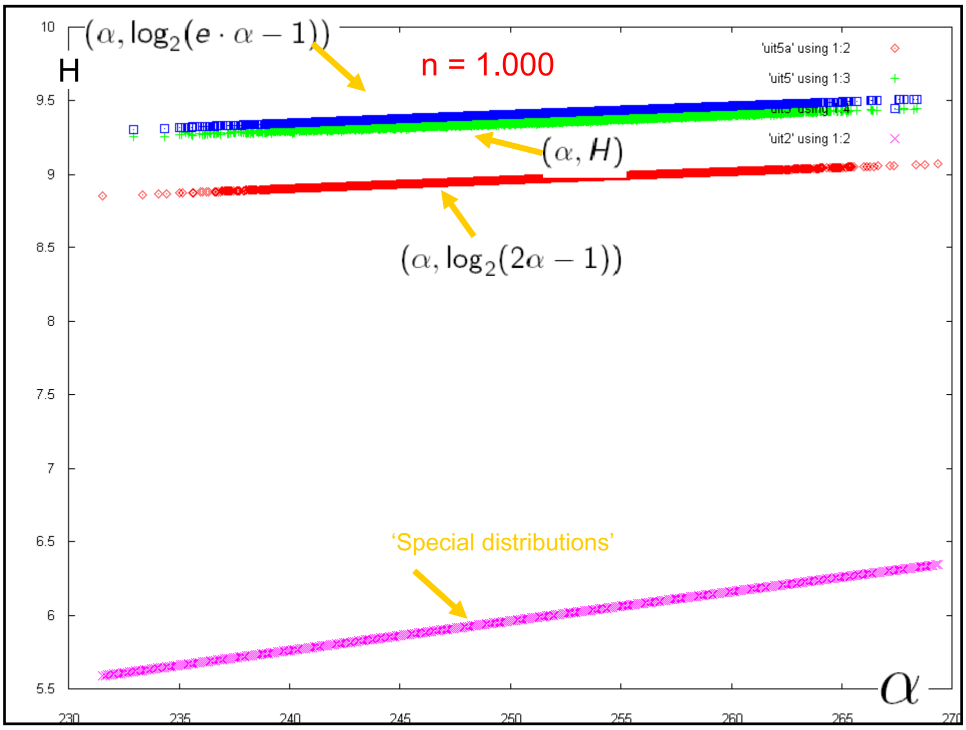
- sci.crypt crypto FAQ: $2^H \approx \alpha?$

$$\frac{2 log_2(n)}{n-1}(\alpha - 1) \leq H \leq \log_2(e \cdot \alpha - 1)$$

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 $C_{n,\alpha}$

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

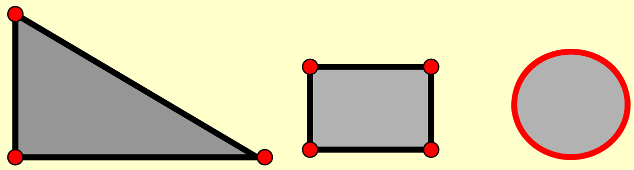
- First look for the minimal value the Shannon entropy H takes on the set $C_{n,\alpha}$

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

- That is, the set of probability distributions on *n* points with a given guessing entropy α.

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

$$
\begin{array}{ccccccccccc}
(& a_{j,k,n}, & a_{j,k,n}, & \cdots & a_{j,k,n}, & b_{j,k,n}, & \cdots & b_{j,k,n}, & 0, & \cdots & 0) \\
 & \uparrow & & & \uparrow & \uparrow & & \uparrow & \uparrow & & \\
 & 1, & 2, & \cdots & j, & j+1, & \cdots & k, & k+1, & \cdots & n,
\end{array}
$$

where

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

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

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

$k = 2\alpha - 1$	$X_{1, 2\alpha -1, n}$	$X_{2, 2\alpha -1, n}$	…………………	$X_{2\alpha -1, 2\alpha -1,}$
$k = 2\alpha$	$X_{1, 2\alpha, n}$	$X_{2. 2\alpha. n}$	…………………	$X_{2\alpha -1, 2\alpha, n}$
⋮				
$k = n$	$X_{1, n, n}$	$X_{2, n, n}$	…………………	$X_{2\alpha -1, n, n}$

These are vectors in R^n

A new bound: strong improvement of McEliece-Yu

- Fix number of passwords *n*

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

k = n $X_{1,n,n}$ $X_{2,n,n}$ ………………… $X_{2\alpha-1,n,n}$

Highest probabilities $a_{1,n,n}$ $a_{2,n,n}$ ………………… $a_{2\alpha-1,n,n}$

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

- So extreme points are optimal distributions in 'own' δ class.

Application: selecting near optimal passwords

- Choose a Guessing entropy α
- Choose an upper bound δ = *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 → ∞, average password length ↓ - $\log_2(\delta)$
- 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

$\alpha = 2^{64}$

$\log_2(n)$	δ	Average pwd length	Min length	Max length	P_{min}	P_{max}
65.0	$2^{-65.00}$	65.00	40.0	65.0	2.98E-08	1.00E+00
65.5	$2^{-41.77}$	58.90	40.0	65.5	2.92E-01	7.07E-01
66.0	$2^{-41.00}$	54.00	40.0	66.0	5.00E-01	5.00E-01
66.5	$2^{-40.62}$	50.30	40.0	66.5	6.46E-01	3.53E-01
67.0	$2^{-40.41}$	47.56	40.0	67.0	7.50E-01	2.50E-01
67.5	$2^{-40.28}$	45.53	40.0	67.5	8.23E-01	1.76E-01
68.0	$2^{-40.19}$	44.04	40.0	68.0	8.75E-01	1.25E-01
68.5	$2^{-40.13}$	42.95	40.0	68.5	9.11E-01	8.83E-02
69.0	$2^{-40.09}$	42.14	40.0	69.0	9.37E-01	6.25E-02
69.5	$2^{-40.06}$	41.56	40.0	69.5	9.55E-01	4.41E-02
70.0	$2^{-40.04}$	41.13	40.0	70.0	9.68E-01	3.12E-02

Conclusion