Graphical Passwords: will your doodle keep the hackers away?

Computer security experts often tell us not to choose an easy to guess word as a password – for example, the user's name or date of birth.

The Graphical Passwords team have been working on a new password system that is more secure. Their tests suggest that the Background Draw A Secret (BDAS) graphical passwords are up to 40 per cent stronger (and yet easier for the user to remember) than a typical eight character password made up of letters, numbers and symbols.

Just a simple smiley face password, as above, can have a strength of 55 bits

For example, traditional passwords such as ***C4hjy!89*** would have a strength of about 53 bits. The very simple smiley face above would have a similar strength. But a slightly more complex picture would be much stronger.

So how do computer scientists work out how 'strong' passwords can be?

In 1948, in a world where secure communication and code-breaking were vitally important, an American electrical engineer and mathematician called Claude Shannon published a ground-breaking book. "**The Mathematical theory of Communication**" established a new branch of applied mathematics now known as **information theory** which enabled mathematicians to quantify information. In particular, he came up with a way of measuring how much information there is in a message, called **information entropy**.

Entropy is often discussed in physics and chemistry and is a gauge of the disorder or chaos in a system. Similarly, **information entropy** can measure how much redundancy there is in a message, or, conversely, how much information it carries. A high entropy means there is a lot of uncertainty, so the message carries a lot of new information. So the higher the entropy, the more difficult a password is to guess.

Produced by the Mathematics Promotion Unit
at the London Mathematical Society and the Institute of Mathematics & its Applications
in conjunction with the Graphical Passwords project and the Royal Society
www.mathspromotion.org.uk

For example, a message telling you: "You are at the 2008 Royal Society Summer Exhibition" doesn't tell you anything you don't know – you knew that would happen when you set out this morning. But telling you: "You are at the Graphical Passwords stand" does give you new information.

The overall entropy of an event X is calculated as the sum of each possible outcome x within X. Let's write the probability of a particular outcome as $p(x)$. So if $p(x) = 1$, then we know that that particular outcome will definitely happen – in information terms, this means a content of 0.

Putting in some other conditions, it turns out that taking the logarithm - ie $\log(p(x))$ - is the best way to handle this measure of information. Logarithms can be calculated in any base, but if the chosen base is 2 then we say that the information is in **bits**, ideal for when you are talking about computer information.

So the **information entropy** is computed mathematically as:

$$H(X) = -\sum_{\text{all possible } x} p(x)\log_2 p(x)$$

Now consider another situation where two events are equally likely to happen, such as tossing a fair coin. So if we consider tossing a coin, the entropy is

$$-(\tfrac{1}{2}\log_2(1/2) + \tfrac{1}{2}\log_2(1/2)) = \mathbf{1\ bit}$$

In general, the more equally likely events n there are, $\log(n)$ increases and the entropy goes up as the amount of information you get is greater.

Suppose you are choosing an eight letter password from the 26 lower case letters. The **theoretical entropy** is

$$8 \times \log_2 26 \approx 8 \times 4.7 \approx \mathbf{37.6\ bits}$$

But choosing the password from digits and upper and lower case letters and other symbols, gives 95 characters to choose from. So the entropy is

$$8 \times \log_2 95 \approx 8 \times 6.6 \approx \mathbf{53\ bits}$$

However, the **practical strength** of commonly used 8-character passwords is far less than 53 bits since people often choose memorable words and names. A modern desktop computer can search through about 2^{40} (around 1,000 billion) passwords in 24 hours, and the speed of computers is fast increasing. So passwords with a bit-strength of less than 40 bits can be easily guessed, when hackers are able to make repeated tries to verify their guess.

In the BDAS system, it isn't just different characters that make up passwords and affect their strength. BDAS password strength comprises:
- **number** of strokes (a stroke is complete when the pen is lifted up)
- **length** or number of grid cells a drawing cuts through
- **size** of the grid and the grid size
- **order** of strokes.

The team calculated the total number of passwords of a length smaller than 13 on a 5x5 grid is more than 2^{53} - the total number of 8-character text passwords! The smiley face overleaf has a stroke count of 5 and a password length of 17 - the total number of passwords of such or less complexity in BDAS is around 2^{55} for a 5x5 grid.

The team found that the average strength of the BDAS passwords created by participants in their experiments could be as high as 60-70 bits, and 95% users were able to recreate their passwords within three attempts one week later.