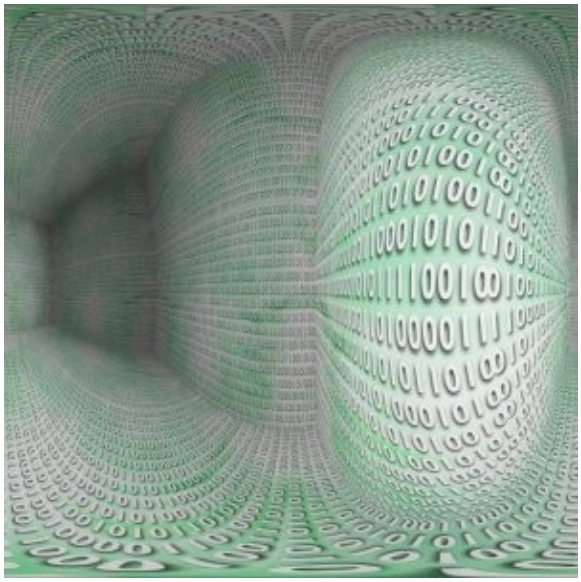
16/02/2006
News

The dangers of cracking hash



Cryptography needs a better theory, according to an article by Susan Landau from Sun Microsystems Laboratories, which will appear in the March issue of Notices of the American Mathematical Society. It's the security of electronic communications that Landau is worried about, and with good reason: last year cryptographers showed that the newest and most widely used security algorithm is vulnerable to attack. Although the code is still secure, it is only a matter of time and increased computing power for someone to actually break it. And then everything from password protected websites to sensitive documents, electronic credit card transactions and email could become open to abuse.

The code in question is called SHA−1, SHA standing for "secure hash algorithm". It was developed and is endorsed by the US National Security Agency, and is at the core of most internet security applications. SHA−1 uses mathematical algorithms to turn information, be it a long document or a single short password, into a string of 0s and 1s that has a fixed length of 160 bits (a bit is one digit; either a 0 or a 1). It does this by mixing information coming from the document with random bits and then boiling down this mixture to the string of prescribed length. This string, called the *digital digest*, is like a digital fingerprint of the document and is used to authenticate documents, passwords and digital signatures.

There are two main points that ensure the security of SHA−1 and other hash algorithms. Firstly, it should be incredibly difficult to guess the original information from its digest, in other words the algorithm should be

very hard to reverse. Secondly, it should be incredibly difficult to find two documents that give rise to the same digest, something cryptographers call a *collision*. Take passwords, for example: if the password for my on−line bank account was transmitted over the internet as clear text, then someone who manages to intercept it could log into my account from any computer terminal and wreak havoc with my finances. Worse, the intruder might even be lucky enough to find that I use the same password to protect other information. Sent in digested form, however, the information is useless as long as no−one can work out the original password or replace it by one with the same digest.

The same applies to sensitive documents: say I send a message to my bank instructing it to transfer £100 to your account. To be sure that you haven't tampered with the message and changed £100 to £10,000, the bank's software only needs to compare the digest of the received message with the digest of the sent message, because any tampering would have changed the digest. Hash algorithms are also used to generate *digital signatures*, which make sure that a document for example an agreement by me to pay you £100 really does originate from the person it says it does.

To cryptographers, "incredibly difficult" means "computationally unfeasible": it should take an unrealistically large amount of computing power to reverse the hash algorithm or to find two colliding documents. But if you can find a way to reduce the computing power needed, then you are en route to breaking the algorithm, and this is exactly the approach a group of cryptographers took to attack SHA−1.

Xiaoyun Wang from Tsinghua University in Beijing, together with various colleagues, announced last year that she had drastically reduced the number of guesses it would take to find two documents that collide: rather than the 2^{80} guesses that cryptographers expected, she would only need 2^{63}. This means that a network of powerful computers could find a collision within about a month.

Wang achieved this amazing result through sheer endurance and patience. By watching the many steps in which the algorithm encodes information, she developed a knack for how it's done and converted this into far more educated guesses than a brute−force approach would entail. This isn't the first time Wang has broken something important: she previously found ways to bring about collisions in SHA−1's predecessor SHA−0, and in another algorithm called MD5, which is still used in some older applications.

So far, no−one has actually found two colliding documents, and it is unlikely that malicious code breakers could muster up the computing power to do this. Wang's results just show how you could start going about it in theory. Also, luckily, no−one has as yet found a way of guessing a document or password from its hash, something that would have equally daunting security implications. But still, cryptographers are worried. Even

though there are ways of patching up SHA−1 so that Wang's methods are rendered useless, it is likely that code breakers will only ever follow one step behind. As Microsoft researcher Niels Ferguson told the magazine New Scientist, SHA−1 "is a wounded fish, and the sharks are circling." Short−term patches won't do it's the mathematical theory underlying hash algorithms that needs some serious scrutiny.

Further reading

Plus has the following articles on cryptography:

- Safety in numbers,
- Cracking codes,
- Cracking codes, part II and
- Exploring the Enigma.

Marianne Freiberger

Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the NRICH and MOTIVATE sites.