



Exploiting weaknesses in the MD5 hash algorithm to subvert security on the Web

Dr. Tim Moses,
Director of Advanced Security, Entrust Inc.,
Chair CAB Forum

January 2009

Entrust is a registered trademark of Entrust, Inc. in the United States and certain other countries. Entrust is a registered trademark of Entrust Limited in Canada. All other company and product names are trademarks or registered trademarks of their respective owners. The material provided in this document is for information purposes only. It is not intended to be advice. You should not act or abstain from acting based upon such information without first consulting a professional. ENTRUST DOES NOT WARRANT THE QUALITY, ACCURACY OR COMPLETENESS OF THE INFORMATION CONTAINED IN THIS ARTICLE. SUCH INFORMATION IS PROVIDED "AS IS" WITHOUT ANY REPRESENTATIONS AND/OR WARRANTIES OF ANY KIND, WHETHER EXPRESS, IMPLIED, STATUTORY, BY USAGE OF TRADE, OR OTHERWISE, AND ENTRUST SPECIFICALLY DISCLAIMS ANY AND ALL REPRESENTATIONS, AND/OR WARRANTIES OF MERCHANTABILITY, SATISFACTORY QUALITY, NON-INFRINGEMENT, OR FITNESS FOR A SPECIFIC PURPOSE.

© Copyright 2009 Entrust. All rights reserved.

Table of Contents

1.	Executive Summary	1
2.	The Web PKI	1
3.	Digital signatures	2
4.	Hash algorithms	3
5.	MD5 weakness.....	3
6.	Recommended precautions	4
7.	Going forward.....	4
8.	Conclusion.....	5
9.	About Entrust	5

1. Executive Summary

A group of renowned researchers has published some of the details of their exploitation of a vulnerability in the MD5 hash algorithm. The advance they describe would allow an attacker to create fraudulent Web-site certificates with which they could launch a phishing or man-in-the-middle attack on an eCommerce, eBanking or eGovernment Web-site, resulting in identity theft and/or financial loss for the site's users.

This advance has been anticipated for some time. And, it confirms, once and for all, that MD5 is no longer secure for use in signature applications, such as SSL certificates. Platform suppliers may, in the near future, eliminate the MD5 algorithm from their cryptographic suites, thereby causing site certificates that use MD5 to fail. It is, therefore, recommended to replace MD5 with an alternative algorithm with some urgency. Fortunately, secure alternatives are widely supported.

A user or site operator can tell what hash algorithms are in use in a certificate path for a particular site by clicking on the lock icon in the browser when visiting that site. If either the site certificate or intermediate certificate signature algorithm is found to be MD5WithRSAEncryption, then the certificate should be replaced. sha1RSA is a secure alternative. There is no danger if the root certificate signature algorithm uses MD5.

The Extended Validation Web is a region of the Web that is safer for users and site operators. Users are assured that they are in the EV Web by distinct indications in the frame of the browser, including the presence of the color green in the address bar and the presence of identity information for the site operator and certificate issuer. EV certificates are not vulnerable to this exploit. So, a simple remedy for this exploit is to switch to an EV certificate and give users the peace of mind that comes with EV.

2. The Web PKI

Security mechanisms on the Web, such as entity authentication and encryption, are provided by an overlay that is commonly referred to as SSL. SSL, in turn, depends upon certificates in a public-key infrastructure to protect the integrity and authenticity of the keys it uses in its security mechanisms.

The infrastructure comprises a number of major components (see Figure 1). Suppliers of platforms, such as browsers and operating systems, act as policy management authorities, qualifying certification authorities by embedding their root certificates in their products. While a CA's root certificate is formatted in the syntax of an X.509 certificate, and it includes a signature that verifies with the key contained in the certificate itself, it is not a certificate in the true meaning of the word: it is just a public key. The integrity of this key is not protected by the certificate's digital signature, but rather by various platform security mechanisms.

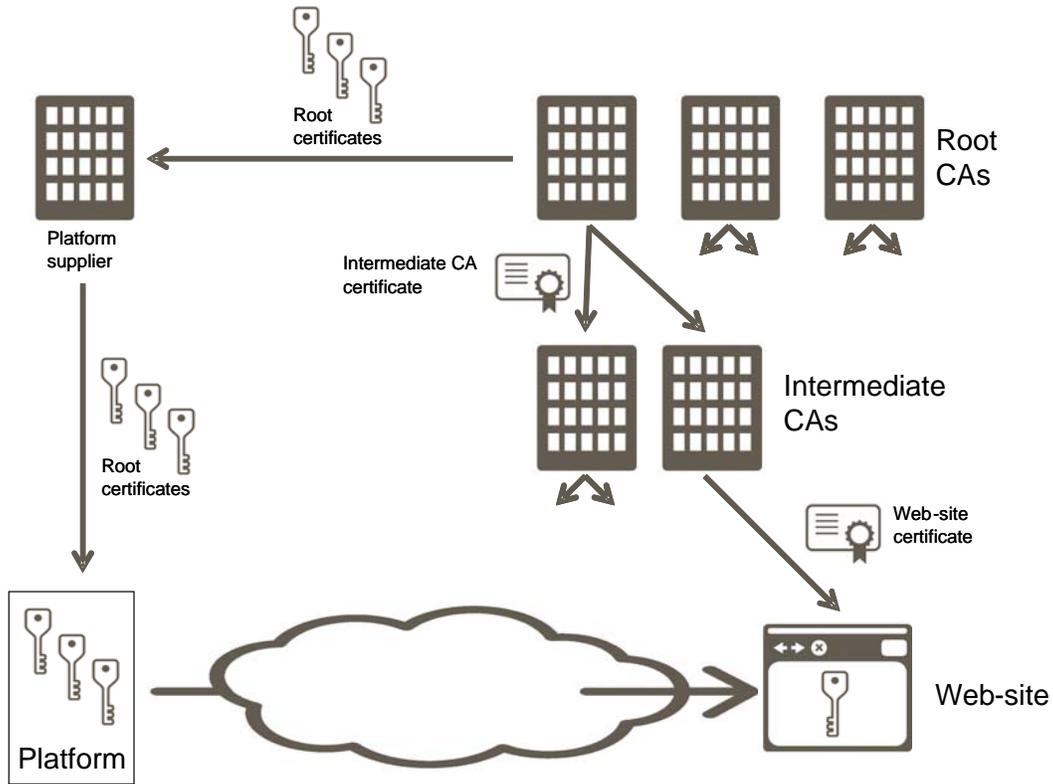


Figure 1 - The Web PKI

Certification authorities issue certificates to other certification authorities and Web sites, and the platform verifies the Web site certificate and uses the key it contains to establish the SSL security mechanisms.

The authenticity of certificates in the Web PKI is protected using digital signatures. The signature on an entity's certificate is applied by the certification authority immediately above it in the hierarchy and verified using the public key from that entity's certificate. This process repeats, going up the hierarchy, until a certificate is encountered that can be verified using one of the root certificates that is embedded in the platform.

The rules for evaluating the chain of certificates between the embedded key and the Web-site certificate and for using the Web-site certificate in the SSL protocol are defined by the Internet Engineering Task Force.

The platform suppliers, acting in the role of policy management authorities, set the rules by which CAs operate when processing a Web-site's application for a certificate and issuing the resulting certificate.

3. Digital signatures

Public-key algorithms involve operations on the members of a mathematical construct called a "group". A group is a finite set of objects (integers in the case of RSA) that possess certain mathematical properties. Because the size of a group is finite, inputs, outputs and keys for public-key algorithms must be represented by strings of fixed length.

Of course, many documents that have to be signed (including certificates) are not of a suitable length to be input directly to a public-key algorithm. So, they have to first be compressed. This is the job of the hash algorithm.

4. Hash algorithms

Hash algorithms work by taking the input document one block at a time, compressing each block to an intermediate hash value while mixing in the result from the previous block (see Figure 2).

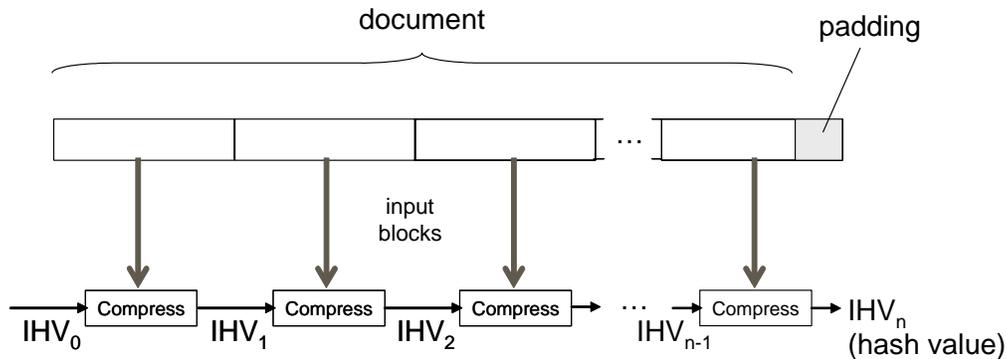


Figure 2 - Structure of a hash algorithm

Good cryptographic hash algorithms possess a number of properties. Of particular relevance to this discussion, it must be impossible in practice to generate a collision. Inevitably, there exist pairs of documents that compress to the same hash value and, therefore, would have identical signatures when signed by the same individual. Such documents are said to 'collide' with each other. If meaningful collisions can be produced, then it can lead to confusion and even fraud, particularly when a subscriber prepares a certification request for a certification authority to sign. It would become possible to produce a CA's signature on a certificate without the CA's knowledge and without compromising its private key.

For this reason it is an important property of cryptographic hash algorithms that it be impractical to find meaningful collisions.

5. MD5 weakness

Since 2004, it has been known to be possible to produce collisions for the MD5 hash algorithm. SHA-1 is also known to be weaker in this regard than it was originally believed to be. However, it has not yet been shown to be even theoretically practical to find collisions for SHA-1.

The 2004 attack requires manipulation of two input blocks. But, because of the iterative way in which hash functions work on large documents, these two collision blocks can be embedded within larger documents, provided that the portions of the documents that precede the collision block (known as the prefix) are identical, and the portions of the documents after the collision block (known as the suffix) are also identical, then the complete documents collide, and a signature on the first document can be attached to the second document in such a way that the signature will verify correctly, and the signer appears to have signed the second document, even though it has never even seen it.

In 2007, an advance was announced, called a "chosen prefix collision" attack. It is more powerful than the earlier attack. In this attack, the prefixes may differ, giving the attacker more freedom in

selecting prefixes, and making it easier to trick a signer into creating a signature for an innocuous document that can then be transferred to another document that is more useful to an attacker for fraudulent purposes. Suffixes still have to be identical. But, even if the signer chooses all or part of the suffix, the attacker may simply be able to copy the suffix from the signed document into the fraudulent document.

At the very end of the year 2008, an actual exploit of the weakness in MD5 was demonstrated. The demonstration involved getting a real root certification authority to create a certificate for an apparently benign end-entity. The certificate contents had been specially crafted so that its signature would be identical to one on another, far less benign, certificate. The fraudulent certificate was, in fact, that of a subordinate CA that, had the researchers chosen to, could have been used to sign apparently genuine Web-site certificates for malicious sites impersonating bona fide sites: the basis for phishing and man-in-the-middle attacks.

From this point onwards, there can be no question that MD5 is insecure in signature applications, and platform suppliers may eliminate it from their products in the near future, thereby causing SSL Web-sites whose certificates use MD5 to fail. Because of the way the attack works, however, documents and certificates signed using MD5 prior to announcement of the full details of the exploit are not likely to be fraudulent.

A point that appears to have caused much confusion relates to the signature on root certificates and the hash algorithm associated with that signature. As mentioned above, the integrity of a root key is not protected by its root certificate. So, the fact that a root certificate may include a signature created using MD5 is not a threat to the security of its root key.

6. Recommended precautions

A user or site operator can tell what hash algorithms are in use in a certificate path for a particular site by clicking on the lock icon in the browser when visiting that site. If either the site certificate or intermediate certificate signature algorithm is found to be MD5WithRSAEncryption, then the certificate should be replaced. sha1RSA is a secure alternative. There is no danger if the root certificate signature algorithm uses MD5.

The Extended Validation Web is a region of the Web that is safer for users and site operators. Users are assured that they are in the EV Web by distinct indications in the frame of the browser, including the presence of the color green in the address bar and the presence of identity information for the site operator and certificate issuer. EV certificates are not vulnerable to this exploit for two reasons. First of all, MD5 is not an allowed algorithm under the rules for issuing EV certificates. And, secondly, EV certificates cannot be issued using an automated process, and the exploit relies upon the predictability of automated issuance.

So, a simple remedy for the exploit is to switch to an EV certificate and give users the peace of mind that comes with the EV Web.

7. Going forward

The long-term solution to this flaw in the Web PKI may involve a new approach to hashing called 'randomized hashing'. But, defining a new hash algorithm and deploying it to a significant portion of all the platforms in use on the Web will take many years. The SHA-2 family of hash algorithms has no known weaknesses. But, while it is supported by all modern platforms, even it is not available in a significant number of legacy platforms. SHA-1, despite its short-comings, is still the best choice if interoperability with legacy platforms is a big consideration; which it definitely is in the case of the Web.

There are some additional safeguards being considered by standards bodies that could further protect against advances in the cryptanalysis of SHA-1. The exploit described above depends upon the attacker's ability to control or predict the contents of the first few fields of the certificate, some of which are chosen by the CA. The researchers' description contains a fascinating explanation of the steps they took to improve their chances of guessing these contents correctly. If CAs were to act less predictably in the assigning of values to these fields, then the task of the attacker becomes much more difficult; effectively impossible. So, it has been proposed that certificate serial numbers and validity dates contain a degree of randomness under the control of the CA. For various reasons, neither of these solutions is quite as straightforward as it may, at first, appear. But, while they may seem like somewhat inelegant solutions to the underlying problem of a weak hash algorithm, they do have merit as supplementary safeguards against further advances in the cryptanalysis of SHA-1.

These developments are some way off in the future. Today, the best answer is to ensure that MD5 is replaced with one of the common secure alternatives.

8. Conclusion

The EV Web PKI is not vulnerable to this and similar exploits. There are two main reasons for this: MD5 is not allowed in the EV Web PKI and EV certificates cannot be issued in an automated process.

The safest way for Web-site operators to ensure that their users cannot be defrauded by this exploit is to operate in the EV Web and to train customers to deal with them exclusively in the EV Web.

9. About Entrust

Entrust, Inc. [NASDAQ: ENTU] is a world leader in securing digital identities and information. More than 2,000 enterprises and government agencies in more than 60 countries use on Entrust solutions to help secure the digital lives of their citizens, customers, employees and partners. Our proven software and services help customers in achieving regulatory and corporate compliance, while helping to turn security challenges such as identity theft and e-mail security into business opportunities.

For more information visit <http://www.entrust.net>