Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate

Marc Stevens, Alexander Sotirov, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik and Benne de Weger

Cryptanalysis of Hash Functions Seminar, Spring 2012

Presented by Sela Ferdman, University of Haifa

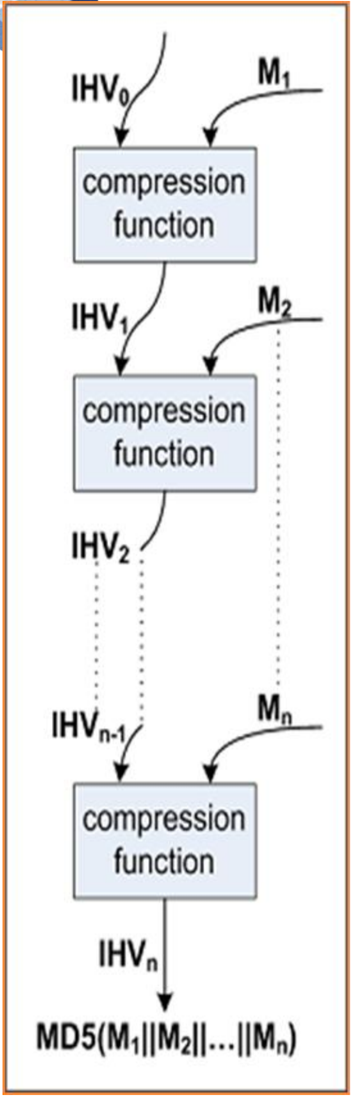
Outline

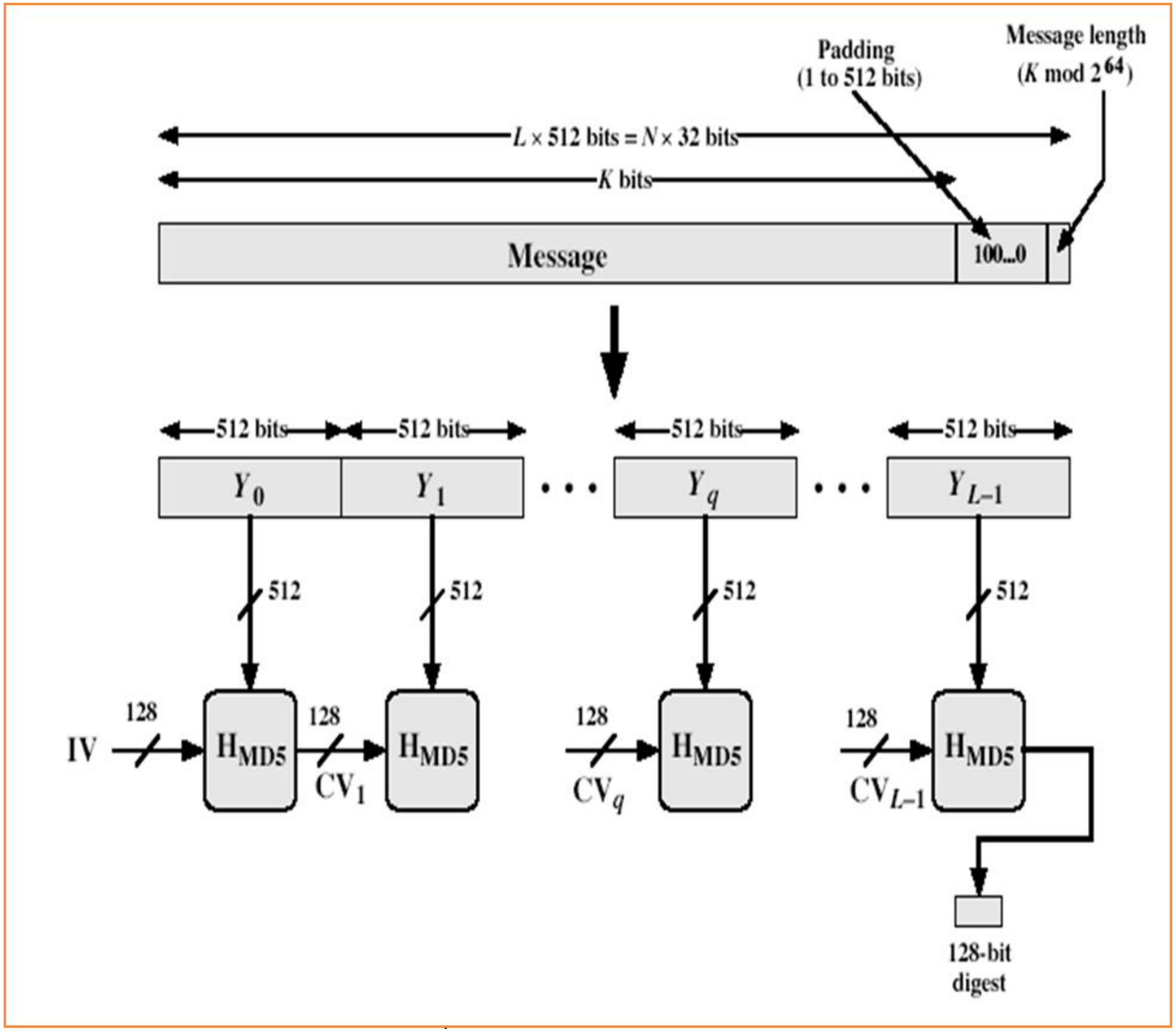
- **Introduction**
- **MD5 short overview**
- **MD5 Collision history**
  - 2004 – Identical-prefix collision attack [Wang,Yu]
  - 2006 – Chosen-prefix collision attack [Stevens, Lenstra, de Weger]
- **MD5 Short Chosen Prefix Collision Attack**
  - Outline of the attack
    - Certification Authority
    - Collision Constrains
  - Real vs. Rogue Certificate
- **Collision construction**
  - Overview
  - Details
  - Summary
- **Conclusions**

Introduction

- A new chosen-prefix construction for MD5 collision
  - For any two chosen message prefixes P and P', suffixes S and S' can be constructed such that the concatenated values P||S and P'||S' collide under MD5.
  - This allowed creation of a real rogue Certification Authority (CA) certificate, based on a collision with a regular end-user website certificate provided by a commercial CA.
  - The entire construction requires about 2^{49} MD5 compression function calls and took less than a day on 215 PlayStation 3 cluster.

MD5 Short Overview





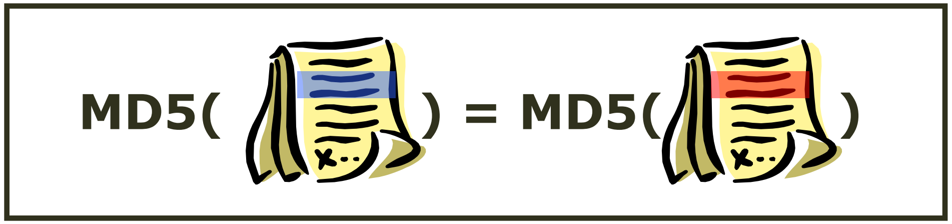
MD5 Collision history - IPC

2004: First collision for MD5 [Wang,Yu]:

- Two 128 byte messages with same MD5 hash value

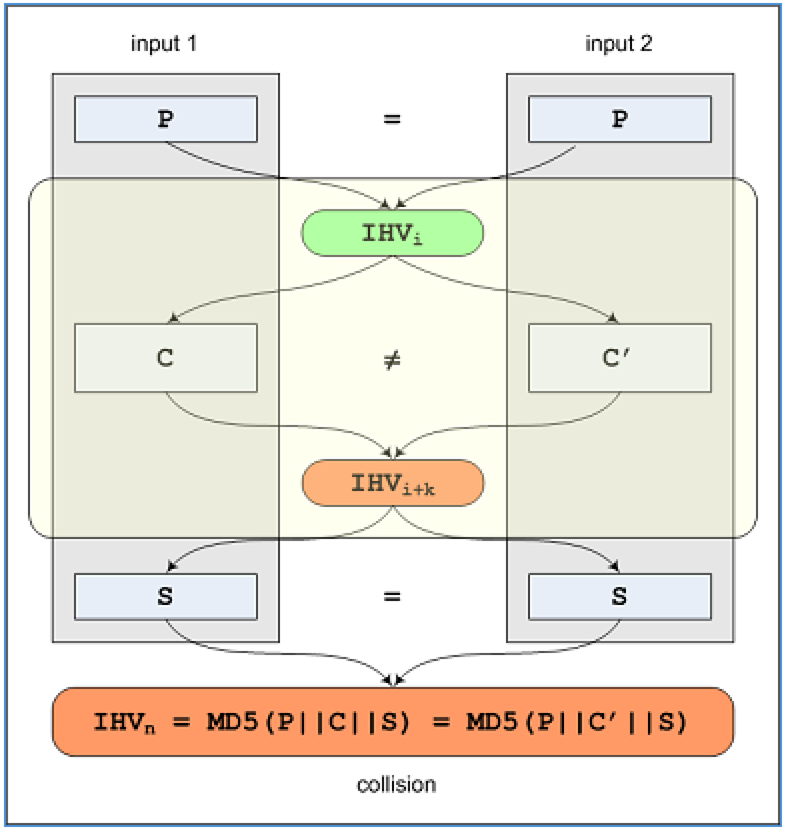
❖ *Identical prefix* collision (IPC) attack

- Messages differ only in 128 consecutive ‘random’ bytes
- Bytes before or after may not differ



MD5 Collision history - IPC

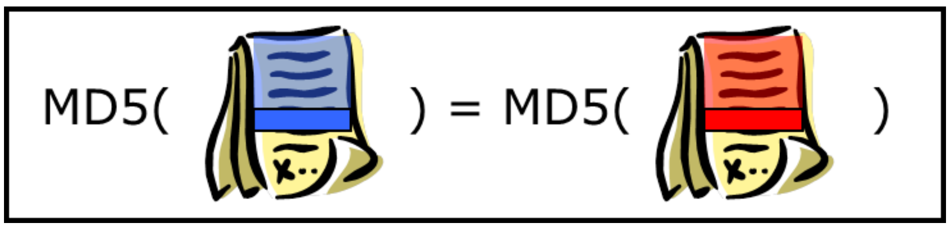
- For any given prefix P and any given suffix S a pair of "collision blocks" {C,C'} can be computed such that MD5(P||C||S) = MD5(P||C'||S).



MD5 Collision history - CPC

2006: *Chosen-prefix* collision (CPC) attack

- [Stevens, Lenstra, de Weger]
  - New stronger type of collisions
  - Choose two arbitrary files (same length)
  - Make them collide by appending 716 ‘random’ bytes



- Example:
  - Colliding certificates with <u>different identities</u>
- MD5 harmful for digital signatures

MD5 Collision history - CPC

set by the CA	serial number	chosen prefix (different)	serial number
	validity period		validity period
	"Arjen K. Lenstra"		"Marc Stevens"
	real cert RSA key 8192 bits	collision bits (computed)	real cert RSA key 8192 bits
	X.509 extensions	identical bytes (copied from real cert)	X.509 extensions
	valid signature		valid signature

MD5 Collision history

- … but CAs have continued to use MD5 to verify certificates since:
  - In ‘real life’ CA has final control of two fields of the to-be-signed part:
    - Serial number field
    - Validity period field
  - Current construction results in 8192-bit RSA moduli, while CA certificate has 2048-bit upper bound

MD5 Short Chosen Prefix Collision Attack - CAs

- Website digital certificates must be signed by a trusted Certificate Authority
- Browsers ship with a list of trusted CAs
- CAs' responsibilities:
  - Verify the identity of the requestor
  - Verify domain ownership for SSL certs
  - Revoke bad certificates

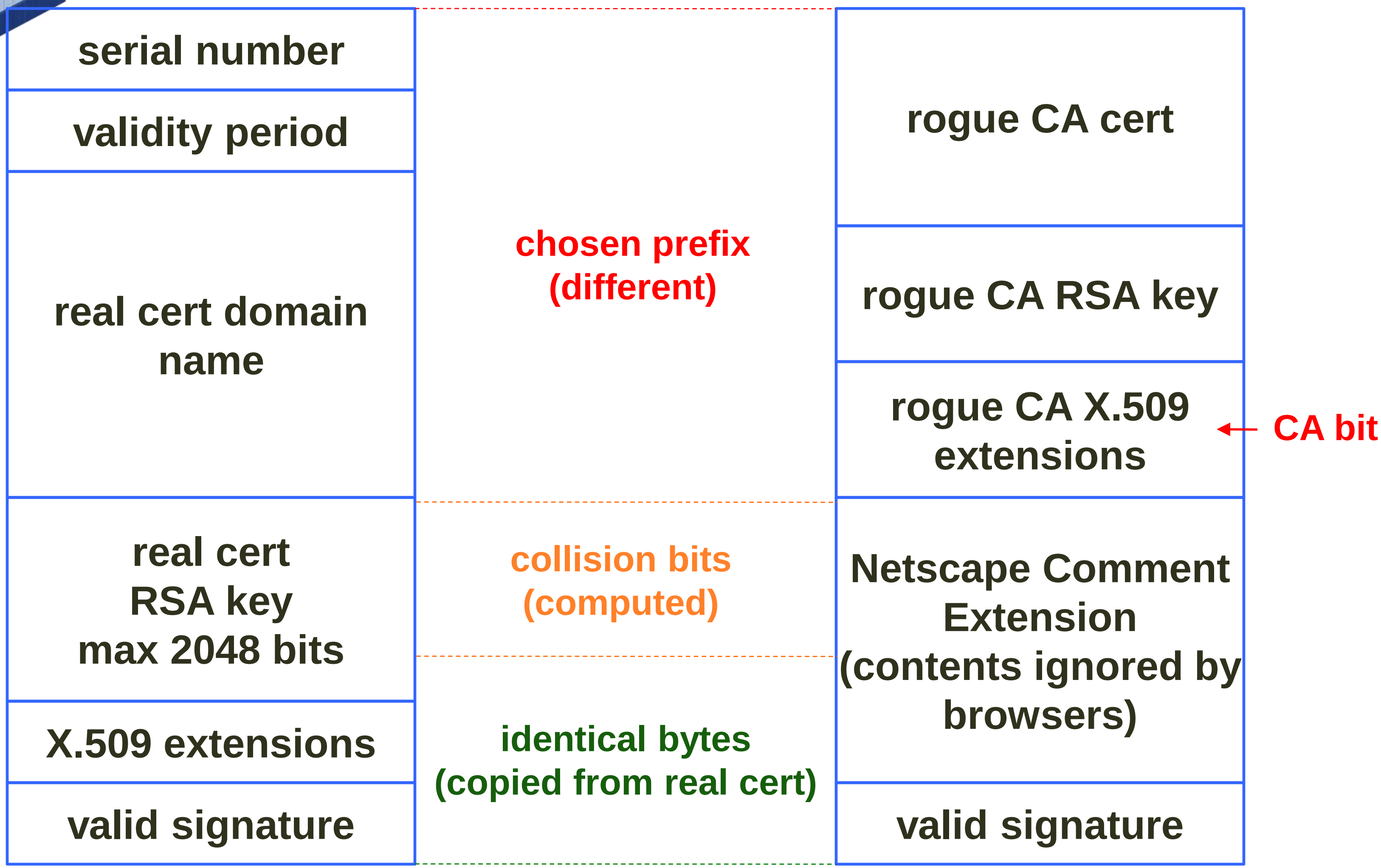
MD5 Short Chosen Prefix Collision Attack

- **We were able to create a sub-CA signed by a known trusted CA (RapidSSL)**
- **Same effect as subverting a known trusted CA**
- **Possible because one particular commercial CA**
  - used MD5 to create signatures
    - MD5 known to have significant weaknesses since 2004
  - had weaknesses in procedures

MD5 Short Chosen Prefix Collision Attack - Constrains

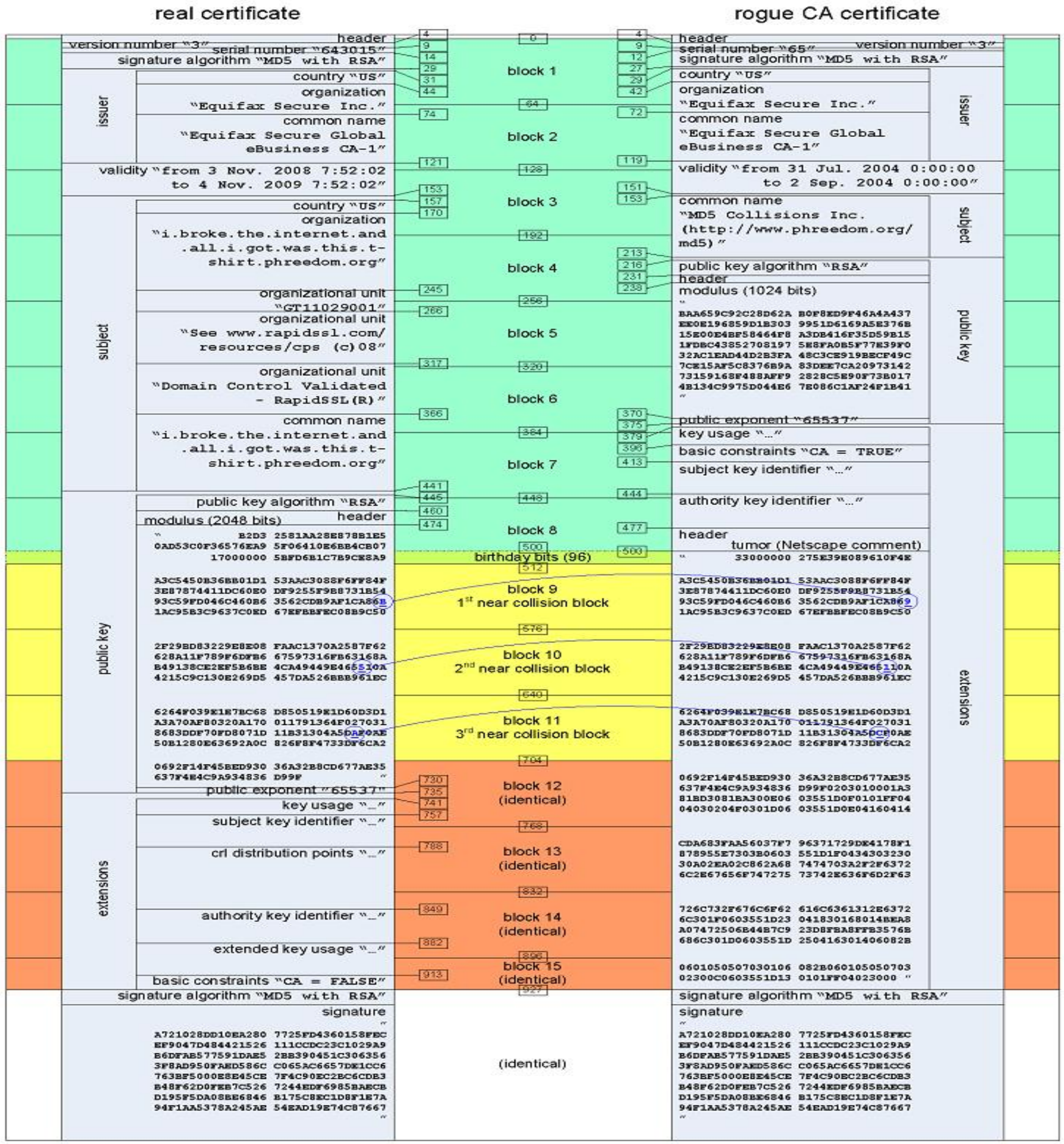
- Because the CA that is supposed to sign our (legitimate) certificate does not accept certification requests for RSA modulo larger than 2048 bits, each of our suffixes S and S` and their common appendage T must fit in 2048 bits. This implies that we can use at most 3 near-collision blocks. (each block 512 bits)
- Furthermore, to reliably predict the serial number, the entire construction must be performed within a few days.

MD5 Short Chosen Prefix Collision Attack – Real Vs. Rogue certificate



Collision construction – Overview

- Predict the serial number and validity period.
- Start calculating the collision block in a chosen-prefix collision, which consist of three consecutive parts:
  - padding bitstrings
  - birthday bitstrings
  - near-collision bitstrings
- Request a legitimate website certificate from a commercial Certification Authority trusted by all common browsers.
- Since the MD5 hashes of both the legitimate and the rogue certificates are the same, the digital signature obtained from the commercial CA can simply be copied into our rogue CA certificate and it will remain valid.


real certificate
rogue CA certificate
header
version number "3"
serial number "643015"
signature algorithm "MD5 with RSA"
issuer
country "US"
organization
"Equifax Secure Inc."
common name
"Equifax Secure Global eBusiness CA-1"
validity "from 3 Nov. 2008 7:52:02 to 4 Nov. 2009 7:52:02"
subject
country "US"
organization
"i.broke.the.internet.and.all.i.got.was.this.t-shirt.phreedom.org"
organizational unit
"GT11029001"
organizational unit
"See www.rapidssl.com/resources/cps (c)08"
organizational unit
"Domain Control Validated - RapidSSL(R)"
common name
"i.broke.the.internet.and.all.i.got.was.this.t-shirt.phreedom.org"
public key
public key algorithm "RSA"
header
modulus (2048 bits)
public exponent "65537"
extensions
key usage "..."
subject key identifier "..."
crl distribution points "..."
authority key identifier "..."
extended key usage "..."
basic constraints "CA = FALSE"
signature algorithm "MD5 with RSA"
signature
block 1
block 2
block 3
block 4
block 5
block 6
block 7
block 8
birthday bits (96)
block 9
1st near collision block
block 10
2nd near collision block
block 11
3rd near collision block
block 12
(identical)
block 13
(identical)
block 14
(identical)
block 15
(identical)
(identical)
header
serial number "65"
version number "3"
signature algorithm "MD5 with RSA"
issuer
country "US"
organization
"Equifax Secure Inc."
common name
"Equifax Secure Global eBusiness CA-1"
validity "from 31 Jul. 2004 0:00:00 to 2 Sep. 2004 0:00:00"
subject
common name
"MD5 Collisions Inc. (http://www.phreedom.org/md5)"
public key
public key algorithm "RSA"
header
modulus (1024 bits)
public exponent "65537"
key usage "..."
basic constraints "CA = TRUE"
subject key identifier "..."
authority key identifier "..."
header
tumor (Netscape comment)
extensions
signature algorithm "MD5 with RSA"
signature

Collision construction – Details

Predicting the serial number

- RapidSSL uses sequential serial numbers:
  - Nov 3 07:44:08 2008 GMT 643006
  - Nov 3 07:45:02 2008 GMT 643007
  - Nov 3 07:46:02 2008 GMT 643008
  - Nov 3 07:47:03 2008 GMT 643009
  - Nov 3 07:48:02 2008 GMT 643010
  - Nov 3 07:49:02 2008 GMT 643011
  - Nov 3 07:50:02 2008 GMT 643012
  - Nov 3 07:51:12 2008 GMT 643013
  - Nov 3 07:51:29 2008 GMT 643014
  - Nov 3 07:52:02 2008 GMT ?

Predicting the validity period

- RapidSSL uses a fully automated system
- Certificate issued exactly 6 seconds after clicking
- Valid for one year + one day

Collision construction – Details

- **Padding bitstrings**
  - Given two arbitrarily chosen messages, we first apply padding to the shorter of the two, if any, to make their lengths equal.
  - And so that the birthday bitstrings end on the same 512-bit block border.

Collision construction – Birthday bitstrings

Birthday bitstrings

- Find a pair of k-bit values that, when appended to the last incomplete message blocks, results in a specific form of difference vector between the IHVs.
- The specific form of difference vector between the IHVs that is aimed for during the birthday search is such that the difference pattern can relatively easily be removed by further appending to the messages a sequence of *near-collision blocks*.

Collision construction – Birthday bitstrings

Birthday search

- A birthday search on a search space V is generally performed by iterating a properly chosen deterministic function $f: V \to V$.
- After approximately $\sqrt{\pi|V|/2}$ iterations one may expect to have encountered a collision.
- Let *p* be the probability that a birthday collision satisfies additional conditions (like number of near collision blocks) that cannot be captured by V or f, then on average $1/p$ birthday collisions have to be found in cost of $\sqrt{\pi|V|/(2p)}$.
- In this paper, a variable birthday search was introduced, permitting flexible choice of search space between 64 and 96 bits.

Variable Birthday search

- Example: $|V| = 2^{96}$, $\delta IHV = (\delta a, \delta b, \delta c, \delta d), \delta a = 0, \delta b = \delta c = \delta d$ and 3 near collision blocks $\Rightarrow 2^{57.33}$ MD5 compressions, which takes 50 days on 215 PS3 cluster.
- Interpolating between 64 and 96 bits space searches, while taking advantage of a new family of differential paths that was presented in this paper, gives the desired results of collision construction cost less than one day on the PS3 cluster.

Near collision bitstrings

- We managed to generalize the known differential paths construction to an entire family of differential paths.
- As a result, more bits can be eliminated per pair of near-collision blocks.

Collision construction – Time-Memory tradeoff

r – # near collision blocks

w – a larger value allows elimination of more differences in δIHV per near-collision block.

k – (64+k)-bit birthday space search

k = 8 and w = 5 was chosen.

The overall chosen-prefix collision construction takes on average less than a day on the cluster of PS3s.

Birthday complexities and memory requirements for $r = 3$

	$w = 3$		$w = 4$		$w = 5$	
k	C_{tr}	M	C_{tr}	M	C_{tr}	M
0					$2^{48.17}$	231GB
2					$2^{49.10}$	210GB
4			$2^{50.43}$	330GB	$2^{49.29}$	68GB
6	$2^{51.33}$	287GB	$2^{50.54}$	96GB	$2^{49.69}$	30GB
8	$2^{51.98}$	177GB	$2^{50.74}$	32GB	$2^{49.99}$	11GB
10	$2^{52.43}$	82GB	$2^{51.24}$	16GB	$2^{50.44}$	5GB
12	$2^{52.44}$	22GB	$2^{51.64}$	7GB	$2^{50.90}$	3GB
14	$2^{52.76}$	9GB	$2^{52.01}$	3GB	$2^{51.38}$	2GB
16	$2^{53.13}$	4GB	$2^{52.48}$	2GB	$2^{51.96}$	675MB
18	$2^{53.59}$	2GB	$2^{53.02}$	733MB	$2^{52.61}$	418MB
20	$2^{53.96}$	673MB	$2^{53.46}$	340MB	$2^{53.13}$	215MB
22	$2^{54.43}$	324MB	$2^{54.01}$	182MB	$2^{53.73}$	123MB
24	$2^{54.92}$	160MB	$2^{54.59}$	102MB	$2^{54.33}$	71MB
26	$2^{55.52}$	92MB	$2^{55.25}$	64MB	$2^{55.04}$	47MB
28	$2^{56.11}$	52MB	$2^{55.95}$	42MB	$2^{55.83}$	36MB
30	$2^{56.74}$	32MB	$2^{56.68}$	29MB	$2^{56.61}$	26MB
32	$2^{57.27}$	17MB	$2^{57.27}$	17MB	$2^{57.27}$	17MB

Collision construction - Summary

- **Perform birthday search** (birthday bitstrings)
  - Find δIHVs of specific form
    e.g. δHV=(0,x,x,y)
  - Extend search to lower # near-collision blocks
- **Appends 64 to 96 bits to prefixes** (variable search space)
- **Iteratively eliminate differences in δIHV** (near-collision bitstrings)
- **Till δIHV=(0,0,0,0)**

Collision construction - Summary

Chosen prefixes

Lenghts of $P||A$ *and* $P||A'$ *are equal*

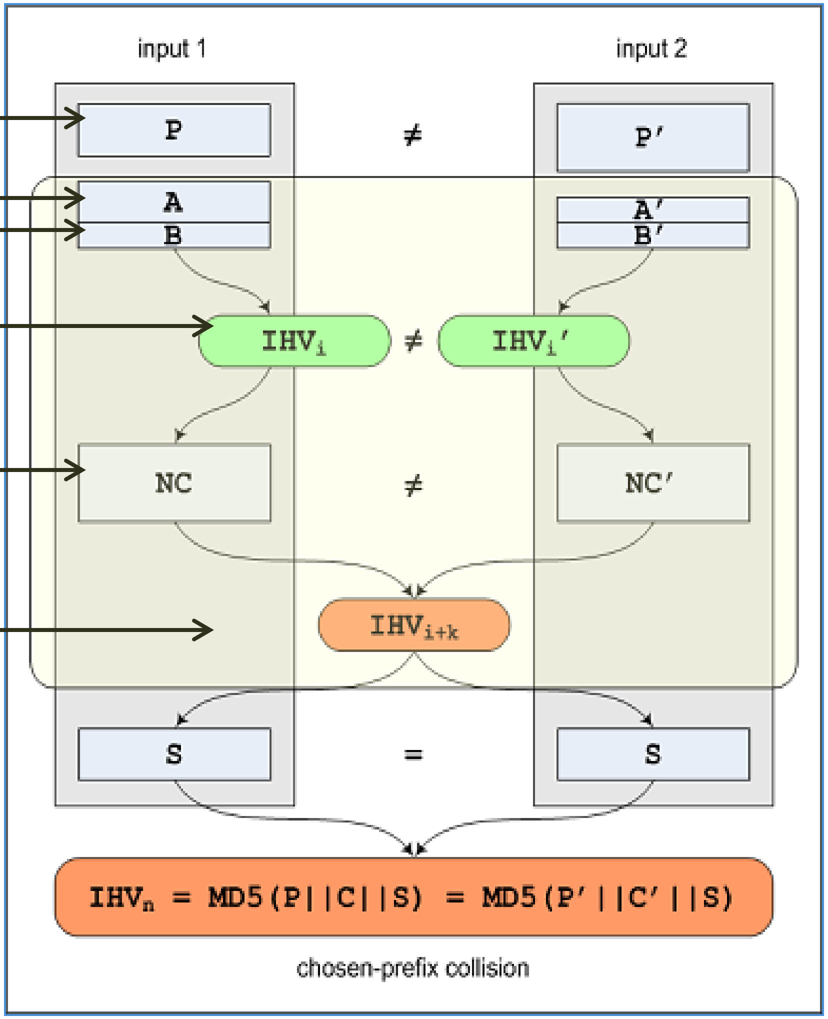
Padding bitstrings

Birthday bitstrings

δIHV *has a prespecified structure*

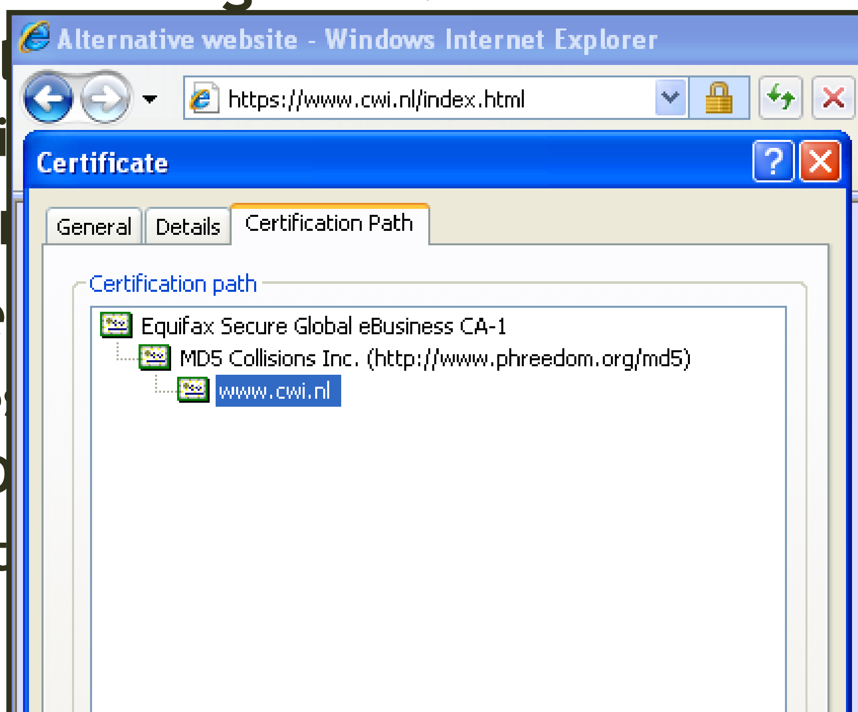
Near collision blocks

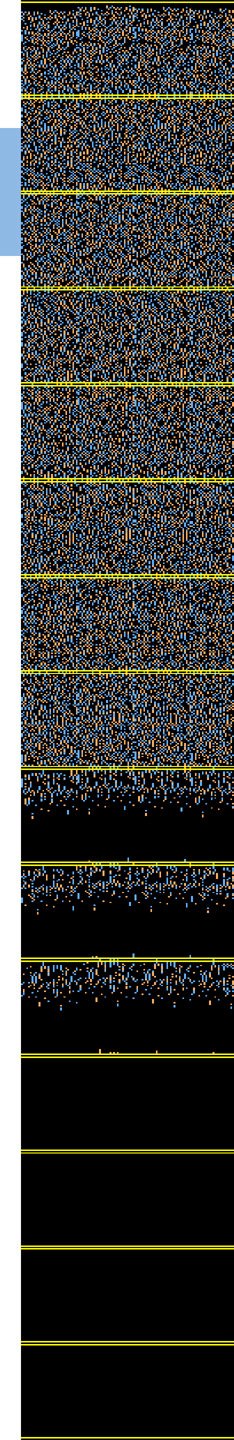
Collison



Results

- Success at 4th attempt
  - Generated CA signature for real cert also valid for rogue CA cert
- Explicit safeguards:
  - Val 4
  - Pri
- Majo were
  - Re
  - MD aft





Conclusion

- Collision attacks on MD5 form a real threat

Another applications

- **Hash based commitments**
  - The Nostradamus attack
    - Correctly predicted the outcome of the 2008 US presidential elections.
    - Using John Kelsey and Tadayoshi Kohno diamond structure and current chosen-prefix collisions construction.
- **Software integrity checking**
  - Colliding executables
    - Takes less than 2 days to create two different Windows executables with the same MD5 hash.
- **Colliding documents**
  - PDFs images

References

- Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate, Crypto 2009 (pp. 55-69)
- MD5 considered harmful today (http://www.win.tue.nl/hashclash/rogue-ca/)
- Saffi Keisari (http://www.eng.tau.ac.il/~yash/infosec-seminar/2009/Short%20Chosen-Prefix%20Collisions%20for%20MD5%20final.ppt)

Thank you