

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

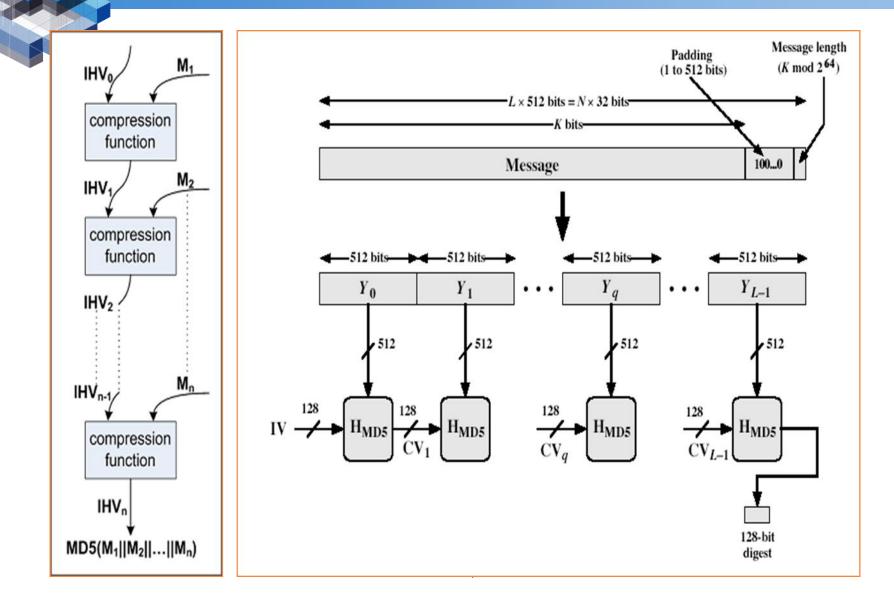
Conclusions

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 2⁴⁹ 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

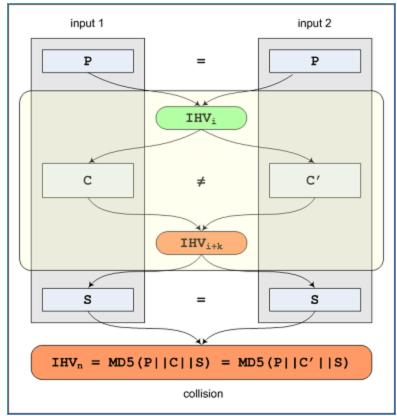
Identical prefix collision (IPC) attack

- 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

set by the CA	serial number		serial number	
	validity period	chosen prefix	validity period	
	"Arjen K. Lenstra"	(different)	"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 - CPC

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-besigned 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

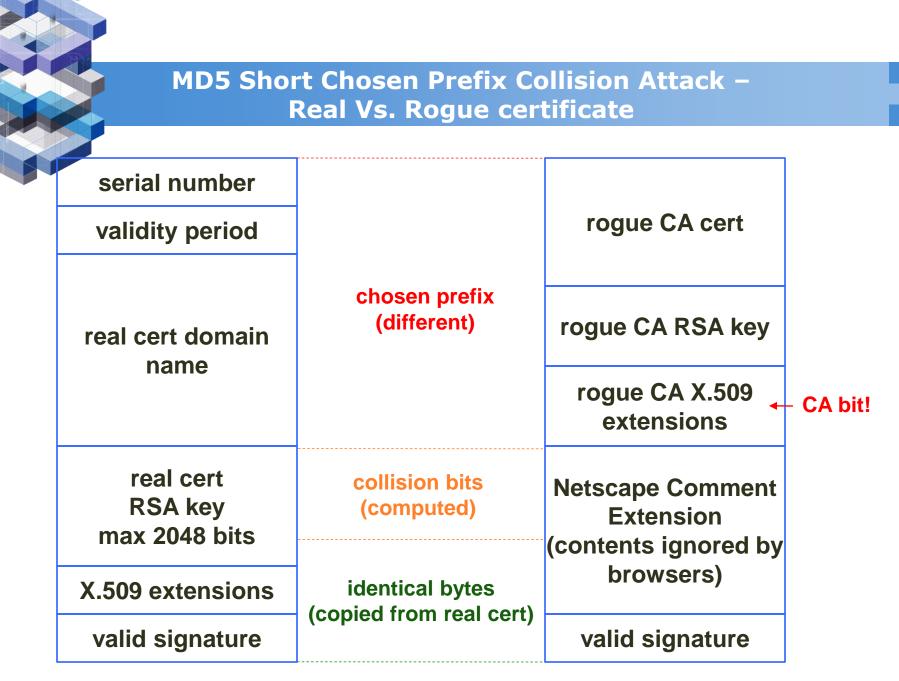
We were able to create a sub-CA signed by a known trusted CA (RapidSSL)

- 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.



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.

rogue CA certificate

real certificate

	version n si	umber "3" serial number "643015"	9	0	9	Version nul	mber "3"	
	si	rension number **3" serial number **643015"			12	serial number "65" version number "3" signature algorithm "MD5 with RSA"		1
		signature algorithm "MD5 with RSA"		block 1	27		A″	-
		country "US"	29	DIOCK	29	country "0s"		
		organization	44		42	organization		
	Issuer	"Equifax Secure Inc."	74	64	72	"Equifax Secure Inc."	issuer	
	iss	common name	74			common name	ē	
		"Equifax Secure Global		block 2		"Equifax Secure Global		
		eBusiness CA-1"				eBusiness CA-1"		
	validity "from 3 Nov. 2008 7:52:02		121	128	[119]	validity "from 31 Jul. 2004 0:	00:00	1
	to 4 Nov. 2009 7:52:02"		151		to 2 Sep. 2004 0:00:00"			
-			153	block 3	153	common name		+
		country "US"	170	DIOCK 5		"MD5 Collisions Inc.	22	
		organization				(http://www.phreedom.org/ md5) "	subject	
		"i.broke.the.internet.and		192			ect	
		.all.i.got.was.this.t-			213			
		shirt.phreedom.org"		block 4	216	public key algorithm "RSA"		
			Correct 1		231	header		
		organizational unit	245	256	[200]	modulus (1024 bits)		
		"GT11029001"	266	[200]		BAR659C92C28D62A B0F8ED9F46A4A437	70	
	**	organizational unit				EE0E196859D1B303 9951D6169A5E376B		
	je.	"See www.rapidssl.com/		block 5		15E00E4BF58464F8 A3DE416F35D59B15 1FDBC43852708197 5E8FA0B5F77E39F0	public key	
	subject	resources/cps (c)08"	-			32AC1EAD44D2B3FA 48C3CE919BECF49C	(e)	
		organizational unit	317	320		7CE15AF5C8376B9A 83DEE7CA20973142 73159168F488AFF9 2828C5E90F73B017	-	
		"Domain Control Validated				73159168F488AFF9 2828C5E90F73B017 4B134C9975D044E6 7E086C1AF24F1B41		
		- RapidSSL(R)"		block 6		"		
			366		370			
		common name			375	public exponent ~65537"	-	1
		"i.broke.the.internet.and		384	379	key usage ""		
		.all.i.got.was.this.t-			396	basic constraints "CA = TRUE"		
		shirt.phreedom.org"		block 7	413	subject key identifier ""		
			-441					
		public key algorithm "RSA"	445	448	444	authority key identifier ""		
			460	and the second		during noy identifier		
		modulus (2046 bits)	474	block 8	477			
		" B2D3 2581AA28E878B1E5 0AD53C0F36576EA9 5F06410E6BE4CB07		(SOO)		header tumor (Netscape comment)		
		17000000 5BFD6B1C7B9CESA9		birthday bits (96)	-500	" 33000000 275£39£089610F4£		
				512				
		A3C5450B36BB01D1 53AAC3088F6FF84F		block Q		A3C5450836880101 53AAC3088F6FF84F		
		3E87874411DC60E0 DF9255F9B8731B54 93C59FD046C460B6 3562CDB9AF1CA86B	-	block 9 1 st near collision block		3E87874411DC60E0 DF9255F988731854 93C59FD046C46086 3562CD89AF1CA869		
		1AC95B3C9637C0ED 67EFBBFEC08B9C50		T near collision block		1AC95B3C9637C0ED 67EFEBFEC08B9C50		
	2			[576]				
	public key	2F29BD83229E8E08 FAAC1370A2587F62		[2002]		2F29ED83229E8E08 FAAC1370A2587F62		
	-iei	628A11F789F6DFB6 67597316FB63168A		block 10		628A11F789F6DFE6 67597316FE63168A		
	Ē	B49138CE2EF5B6BE 4CA49449E465510A	2 nd near collision block	B49138CE2EF5B6BE 4CA49449E465110A				
		4215C9C130E269D5 457DA526BBB961EC				4215C9C130E269D5 457DA526BBB961EC	extensions	
-				640			sus	
		6264F039E1E7EC68 D850519E1D60D3D1	301		6264F039E1E7EC68 D850519E1D60D3D1 A3A70AF80320A170 011391364F027031 8683DDF70FD8071D 11B31304A50CF0AE	9		
		A3A70AF80320A170 011791364F027031 8683DDF70FD8071D 11831304A5DAF03E	3A70AF80320A170 011791364F027031 block 11				w	
		50B1280E63692A0C 826F8F4733DF6CA2		3 rd near collision block	•	50B1280E63692A0C 826F8F4733DF6CA2		
				[704]				
		0692F14F45BED930 36A32B8CD677AE35		- Contraction of the second se				
		637F4E4C9A934836 D99F "	730	block 12		0692F14F45BED930 36A32B8CD677AE35 637F4E4C9A934836 D99F0203010001A3		
		public exponent "65537"	735	(identical)		81BD3081BA300E06 03551D0F0101FF04	4	
		key usage "_"	757			04030204F0301D06 03551D0E04160414		
		subject key identifier ""		768	-			
			788	6		CDA683FAA56037F7 96371729DE4178F1		
		crl distribution points ""	700	block 13		878955E7303B0603 551D1F0434303230		
	SUC			(identical)		30x02Ex02c862x68 7474703x2F2F6372 6c2E67656F747275 73742E636F6D2F63		
	isio			12020		10/11/00/00/00/00/00/00/00/00/00/00/00/0		
	extensions			832				
	ex	authority key identifier ""	849	block 14		7260732F67606F62 61606361312E6372 60301F0603551D23 041830168014FF88		
				(identical)		6C301F0603551D23 041830168014BEA8 A07472506B44B7C9 23D8FBA8FFB3576B		
		extended key usage ""	882			686C301D0603551D 250416301406082B		
		extended key usage ""	6 8	896				
			913	block 15		0601050507030106 0828060105050703 0230000603551D13 0101FF04023000 "		
	basic constraints "CA = FALSE"			(identical)				
	si	gnature algorithm "MD5 with RSA"		1		signature algorithm "MD5 with RS.	A″	1
		signature				signature		
	A721028DD10EA280 7725FD4360158FBC EFF047D436421526 111CCDC23201029A9 B6D7FAB577591DA55 2B83904513036356 39FBAD9507FABD586C C065A26657DE01CC6 763BF50008E8425CC 7F4C90C22EC6CDE3					A721028DD10EA280 7725FD4360158FEC EF9047D484421526 111ccDc23C1029A9		I
								I
				(identical)		B6DFAE577591DAE5 2BB390451C306356 3F8AD950FAED586C C065AC6657DE1CC6		I
								1
		B48F62D0FEB7C526 7244EDF6985BAECE				763BF5000E8E45CE 7F4C90BC2BC6CDB3 B48F62D0FEB7C526 7244EDF6985BAECB		1
		D195F5DA08BE6846 B175C8EC1D8F1E7A 94F1AA5378A245AE 54EAD19E74C87667				D195F5DA08BE6846 B175C8EC1D8F1E7A 94F1AA5378A245AE 54EAD19E74C87667		1
		FALLADJ / DAZADAE 54EAD19E74C87667				WELAADS TOAZODAE SGEAD19E76C87667		1



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 $f: V \rightarrow V$.
- 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.

Collision construction – Variable Birthday search

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.

Collision construction – Near collision bitstrings

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 nearcollision 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.

w = 3w = 4w = 5 C_{tr} kM C_{tr} M C_{tr} M648.17231GB 0 $2^{49.10}$ 210GB $\mathbf{2}$ $2^{50.43}$ 330GB $2^{49.29}$ 68GB 4 $2^{51.33}$ 287GB $2^{49.69}$ $2^{50.54}$ 96GB 30GB 6 051.98 050.74 049.99 77GB 32GB 11GB $2^{50.44}$ $2^{52.43}$ 82GB $2^{51.24}$ 16GB10 5GB $2^{52.44}$ $2^{51.64}$ $2^{50.90}$ 12 22GB7GB3GB $2^{52.01}$ $2^{52.76}$ $2^{51.38}$ 9GB 14 3GB 2GB $2^{52.48}$ 253.13 $2^{51.96}$ 164GB2GB675MB $2^{53.59}$ 253.02 $2^{52.61}$ 18 2GB733MB 418MB $2^{53.46}$ $2^{53.96}$ $2^{53.13}$ 20673MB 340MB 215MB $2^{54.01}$ $2^{53.73}$ $2^{54.43}$ 324MB 22182 MB123MB $2^{54.59}$ $2^{54.33}$ $2^{54.92}$ 24160MB 102MB 71MB $2^{55.25}$ $2^{55.52}$ $2^{55.04}$ 2692MB64MB 47MB $2^{55.95}$ $2^{55.83}$ $2^{56.11}$ 2852MB42MB36MB $2^{56.68}$ $2^{56.74}$ $2^{56.61}$ 30 32MB29MB26MB $2^{57.27}$ $2^{57.27}$ $2^{57.27}$ 17MB17MB17MB

Birthday complexities and memory requirements for r = 3

Collision construction - Summary

Perform birthday search (birthday bitstrings)

Find δIHVs of specific form

e.g. δHV=(0,x,x,y)

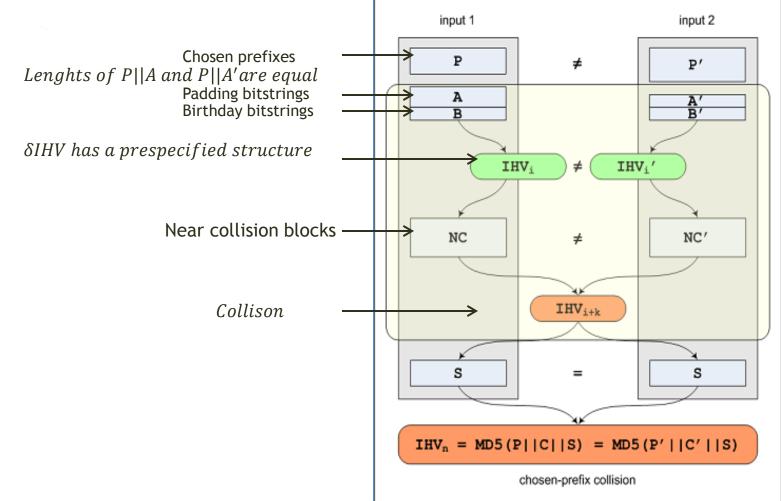
- Extend search to lower # near-collision blocks
- Appends 64 to 96 bits to prefixes (variable search space)

Iteratively eliminate differences in δIHV (near-

collision bitstrings)

Till δIHV=(0,0,0,0)

Collision construction - Summary

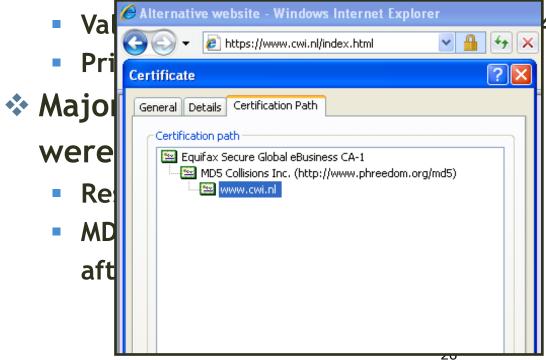


Results

Success at 4th attempt

 Generated CA signature for real cert also valid for rogue CA cert

Explicit safeguards:



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 (<u>http://www.win.tue.nl/hashclash/rogue-ca/</u>)
- Saffi Keisari (<u>http://www.eng.tau.ac.il/~yash/infosec-seminar/2009/Short%20Chosen-</u>Prefix%20Collisions%20for%20MD5%20final.ppt)



Thank you