

The logo for CWI (Centrum Wiskunde & Informatica) is a red trapezoidal shape with the letters 'CWI' in white, bold, sans-serif font.

Centrum Wiskunde & Informatica

Real World Cryptanalysis

Marc Stevens
Cryptology Group
CWI Amsterdam

The word 'SHATTERED' is written in a stylized, fragmented font. The letters are black with red highlights and white outlines, giving it a shattered or broken appearance.

Real World Cryptanalysis

Real World Cryptanalysis

Cryptographic Standards crucial for secure Internet



Gain confidence in security over time through extensive scrutiny
(Before & After Standardization)



Real World Cryptanalysis

Cryptographic Standards crucial for secure Internet



Occasionally leap in cryptanalysis exposes unknown weaknesses



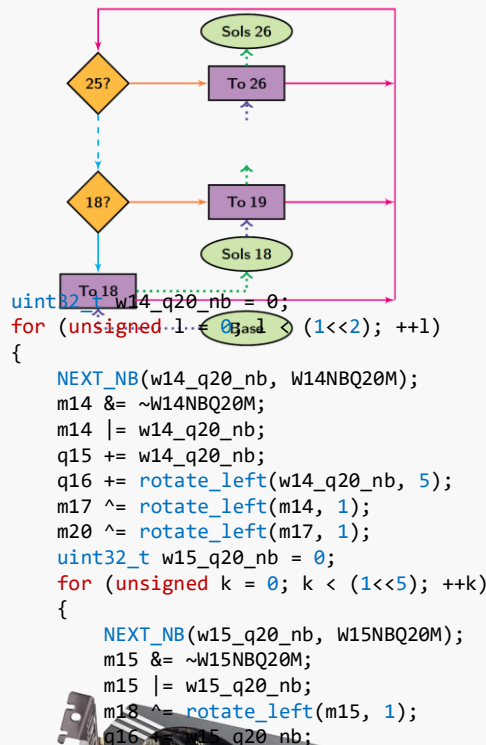
Real World Cryptanalysis

Theory
Analyze

Practice
Build

Demonstrate
Apply

t	Bits Q_t : $b_{31} \dots b_0$	#
-3	10001010 11111100 01010110 11011110	32
-2	11000100 10011010 01100010 -0-10110	32
-1	01111101 01010011 01101110 -0-11110	32
0	11101011 00011111 000010-+ +0-11010	32
1	..0..0.. ..11.. 0-0-..01 -1-.....	13
2	.1.!0+.. ..1+.. -0++..00 --+.....	15
3	.!1.01.. ..0..+.. 1-0-.... --+.....	14
4	!-..-1.. ..0..+..! +1++.... +.....	13
5	!-00-.00 ^-001-.0 101+0000 1+000000	30
6	!+11-011 ++11-01 1.+ -1111 1.111111	30
7	!1...00 ..00.^-..! -.01... ..1^..	15
8	!1...+... 10!-... -0-0-... ..+0+..	15
9	..!1... ..010.. -..+0.. ..!01^0	14
10	00.!-010 00.1..10 ..00!+0 ..0!+1-1-	25
11	110.-111 1100^011 01110+01 001-000+	31
12	.11^00+1 0010+1^ 00^1111. 1-0-0+0	30
13	^1+----0 1-0+0+0- ++++++1+ +-+----0	32
14	--1110+ +++++0+1 00000010 +-+0---	31
15	1+1+1-1- 011-1+10 0000000- 011-.10.	30
16	01...00+ 10111+1. ..+..1. ..100-^01.	21
17	..0.^+1 .1.^+.. ..1.^ ..0.0.0.	13
181 .+.....+.....+.....+.....1.1.1.	8
19	0.....^+0-... ..-.....-.....	8
20	1...0... ..0..-1-... 0.....^ ..1.....0	9
21	+...1... ..0..0-0.. 1.^.....0.....0...	11
22+..... ..1.^ ..+..... ..1.....+	6
230 ..0-... ..1.....+ ..+0.....	8
24^1 ..10..... 0.....0.....1.....^	8
25- ..+..... ..-..... ..-.....	5
26	..0..... ..0.....+..... ..-.....	4
27	..1..... ..1.....+..... ..-.....	7
28	..+..... ..0.....+..... ..-.....	4
29	..0..... ..0..... ..-.....	2
30-..... ..1..... ..-.....	3
31-..... ..-..... ..-.....	1
32-..... ..-..... ..-.....	0
33!..... ..-..... ..-.....	1
34-60-..... ..-..... ..-.....	0
61-..... ..-..... ..-.....	
62+..... ..-..... ..-.....	
63+..... ..-..... ..-.....	
64+..... ..-..... ..-.....	



```

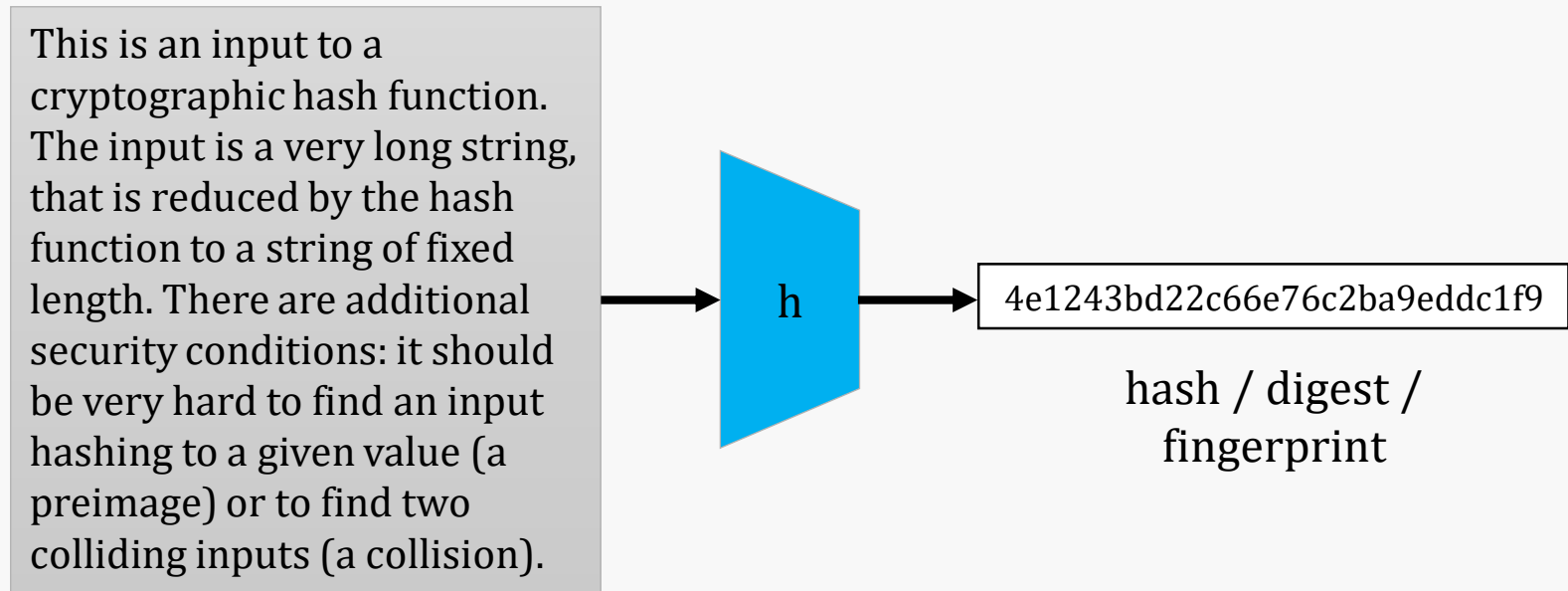
%PDF-1.3.%.....
.1 0 obj.<</Width
h 2 0 R/Height 3
0 R/Type 4 0 R/
SubType 5 0 R/Fi
lter 6 0 R/Color
Space 7 0 R/Leng
th 8 0 R/BitsPer
Component 8>>.st
eam.....$SHA-1
f is dead!!!!!/.
#9u.9...<L...
F.....;.....V.
K.g....K.Ly.++=
.m.i...kE.S...
8.rr/.r..I..F.
OW.....+5
    
```

Collision blocks	Identical prefix	Suffix
000: 2550 4446 2431 2a33 0a25 e2a3 c143 0a0a	%PDF-1.3.%.....	PDF header
010: 0a31 2030 206f 626a 0a3c 3c2f 5769 6a74	.1 0 obj.<</Width	Image object declaration
020: 6820 3220 3020 522f 4865 0967 6874 2033	h 2 0 R/Height 3	
030: 2030 2062 2164 7970 6820 3420 3020 522f	0 R/Type 4 0 R/	JPG header and comment declaration
040: 8375 6274 7970 6820 3520 3020 522f 4669	SubType 5 0 R/Fi	
050: 6c74 6872 2036 2030 2062 2543 6f6c 6f72	lter 6 0 R/Color	same hash at this point
060: 5370 6163 6820 3720 3020 522f 4c65 6e67	Space 7 0 R/Leng	
070: 7468 2036 2030 2062 2142 6974 7350 6872	th 8 0 R/BitsPer	first image data
080: 436f 6470 6f6a 656a 7420 383a 3a0a 7374	Component 8>>.st	
090: 7265 6164 0aff 48ff	eam.....\$SHA-1	second image data (dupes)
0a0: 2069 7320 6465 6164	f is dead!!!!!/.	
0b0: 0923 3976 3c39 61a1 c83c 4c3f e1ff f4f1	#9u.9...<L...	PDF footer
0c0: 7246 dc93 a626 7e01 3602 9a8a 1a82 560b	F.....;.....V.	
0d0: 45ca 6746 88c7 f81b 8c4c 791f e02b 3d16	K.g....K.Ly.++=	
0e0: 14f8 6db1 8909 01c5 8945 c153 0a1e d7d7	.m.i...kE.S...	
0f0: 6038 4972 722f a7ad 728f 0a40 0a40 46c2	8.rr/.r..I..F.	
100: 3057 6f49 6413 99ab 412a f58c 942b 4335	OW.....+5	
110: 42a4 8026 98b5 4707 2a33 71ac 351a		
120: 4748 d00f 2c1c 4874 c00e 7830 5a21 866e		
130: 6130 8786 60b0 400f 3f58 cda1 0a46 29d1		
230: 0000 f1fe 012d 0000 0000 0000 0000 f1fe		
240: 0010 4a46 4946 0001 0101 0048 0048 0000	..JFIF...R.R..	
340: e946 4667 a7b0 7a65 1299 e394 39c0 c7ff9...	
3b0: 493a 2a2a 28ff e000 104a 6649 4600 0101JFIF...	
3c0: 0100 4800 4800 00ff 4b00 4300 0101 0101	..R.R.....C...	
440: 4b14 9717 7f39 1c07 f1ff 000a 656a 64739.....e	
4f0: 7472 6561 640a 656a 646f 626a 0a0a 3220	stream.endobj.2	
500: 3020 6f62 6a0a 380a 656a 646f 626a 0a0a	0 obj.&.endobj.	
840: 3a0a 0a73 7461 7274 7872 656a 0a61 3830	>.stateref.180	
850: 380a 2625 454f 460a	8.XREF.	

Cryptographic hash functions

Cryptographic hash functions

A hash function is a deterministic mapping
from arbitrary length inputs to a fixed length output



MDC-2
MD2, MD4, MD5
SHA-1
old & weak



RIPMD-160
SHA-2- $\{256,512\}$
secure



SHA-3- $\{256,512\}$
New Std in 2015
NIST: use SHA-2

Collision resistance

Collision resistance

Find $m \neq k$ such that $H(m) = H(k)$

Only max. $(n/2)$ -bit security!

128-bit hash \Rightarrow **64-bit security**

160-bit hash \Rightarrow **80-bit security**

256-bit hash \Rightarrow 128-bit security

512-bit hash \Rightarrow 256-bit security

Note:

Bitcoin network computes

- 2^{64} SHA-2 / sec
- $2^{80.5}$ SHA-2 / day
- 2^{84} SHA-2 / 12days
- 2^{89} SHA-2 / year

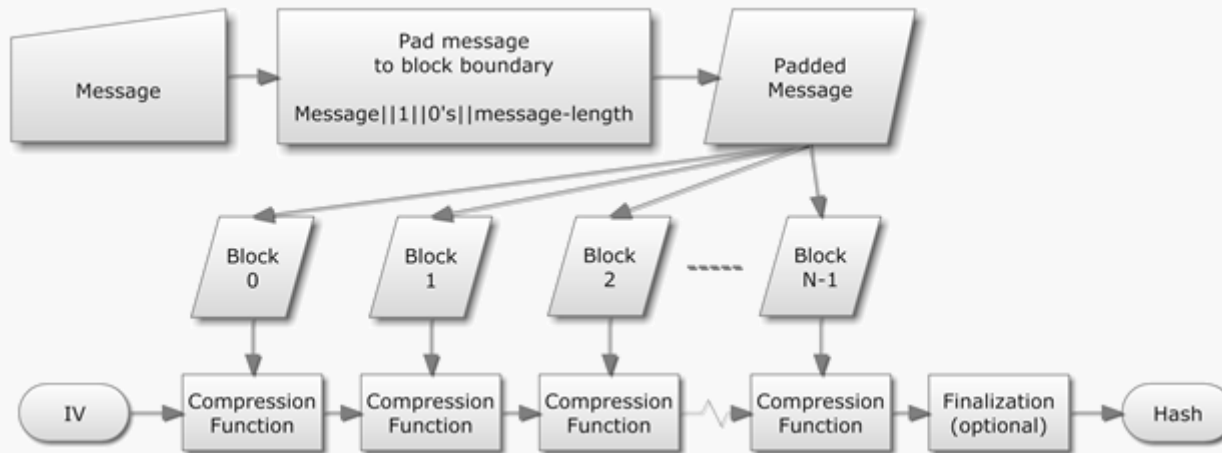
\Rightarrow

breaks **80-bit security**
brute-force in 1 day!

Design of hash functions

- **Merkle-Damgård Construction**

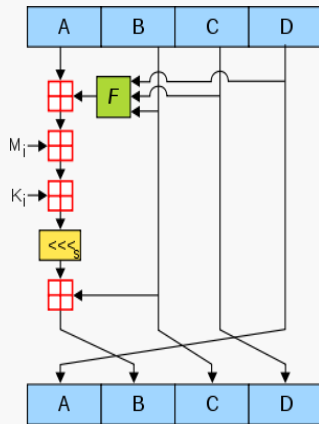
- Splits message into 512-bit **blocks**
- Processes blocks iteratively using **compression function**



- **Security reduction**

- collision hash function \Rightarrow collision compression function

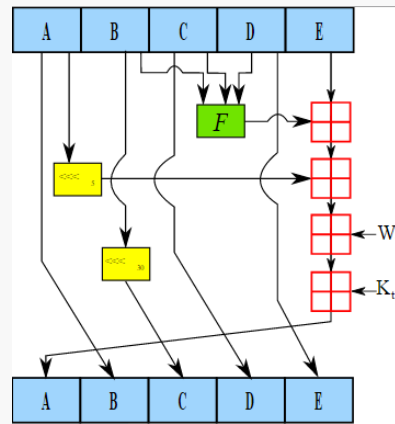
MD5 / SHA-1 / SHA-256 compression function



MD5

very weak

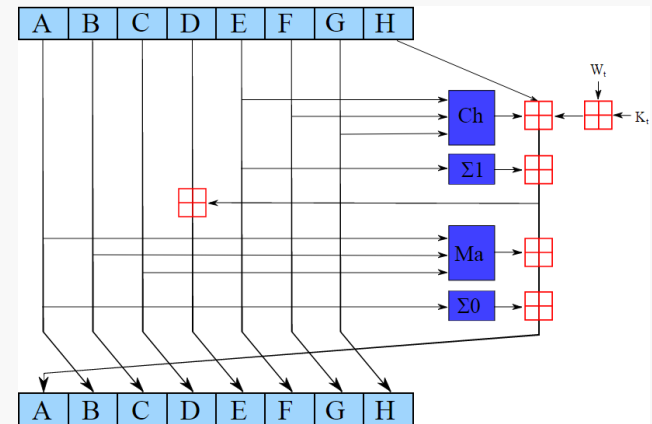
permutation



SHA-1

theoretically weak

linear recurrence



SHA-2

secure

non-linear

message expansion

$16 \times 32\text{-bit} \rightarrow \{64,80\} \times 32\text{-bit}$

Differential cryptanalysis

Differential cryptanalysis

- Consider two different instances

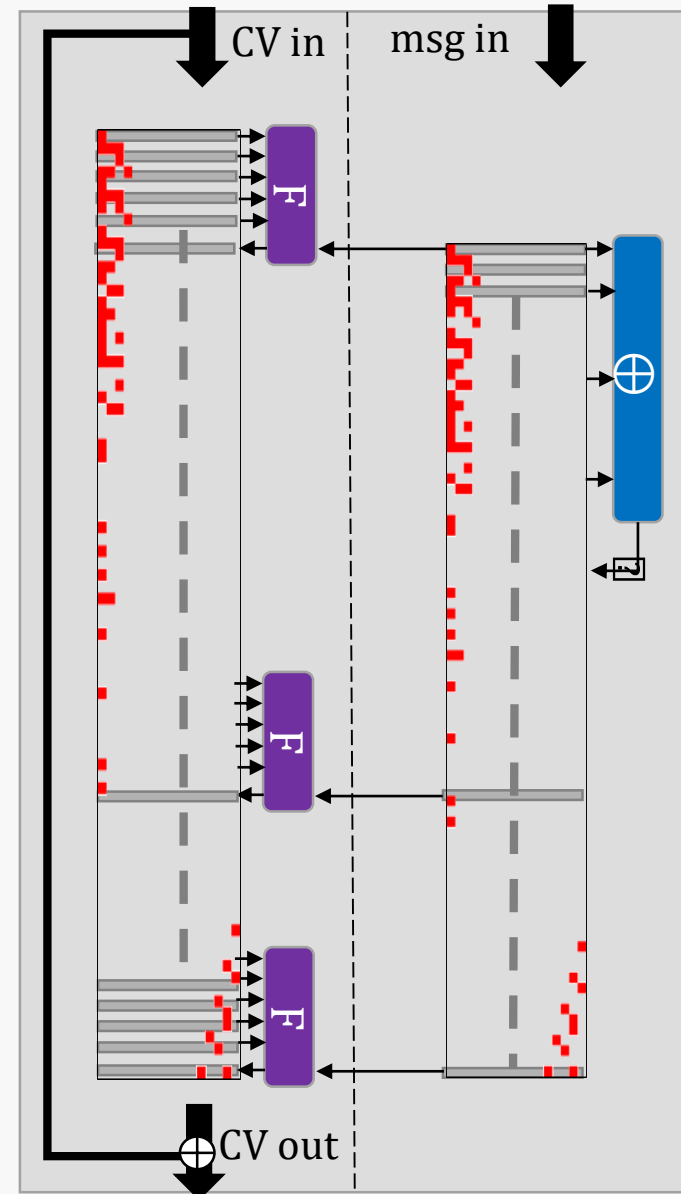
Compress(CV,M)

Compress(CV',M')

- Analyze differences

Differential path

- Precise description of all differences propagating through compression function
- Translate differential path into **system of equations** to solve to find M, M'



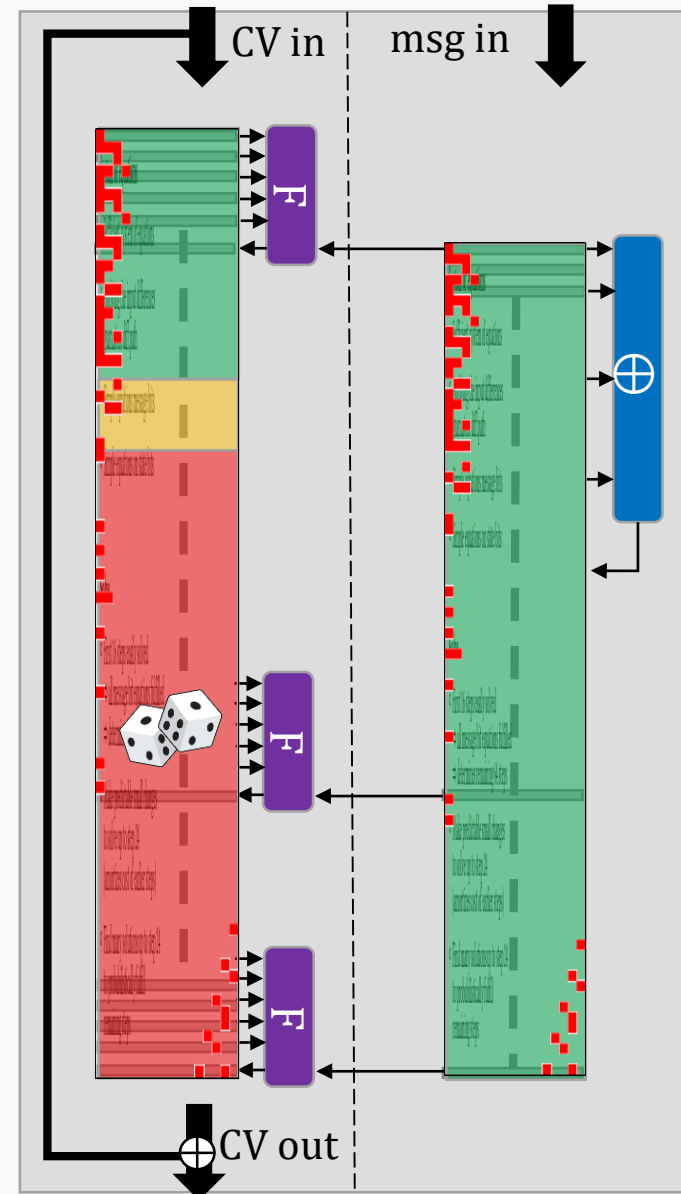
Differential cryptanalysis

System of equations

- Sufficient system of equations:
Applying the input differences guarantees diff.path
- Simple equations on message and state bits

Solve

- First 16 steps easily solved
⇒ all message bit equations fulfilled
- Make predictable small changes to solve up to step 24
(amortizes cost of earlier steps)
- Probabilistically fulfill remaining steps
(with many solutions up to step 24)



Deprecating MD5 in 2008

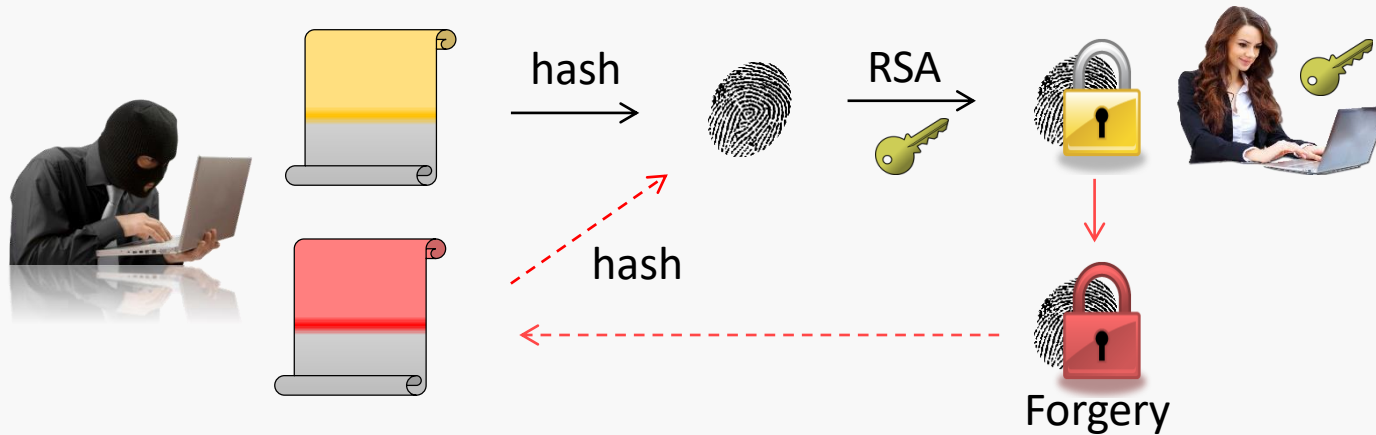
(known to be practically broken since 2004)

Joint work with:

Alexander Sotirov, Jacob Applebaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, Benne de Weger

2008 [SSALMOdW]: Breakthrough on MD5

- Practical *chosen-prefix collision* attack on MD5
- Arbitrary different prefixes made to collide



Example chosen-prefix collision between

- Our website:
<https://i.broke.the.internet.and.all.i.got.was.this.t-shirt.phreedom.org>
- A hand-crafted sub-C.A. certificate 🍆

Legitimate website certificate		Rogue CA certificate
Serial number 643015	chosen-prefixes:	Serial number 65
Commercial CA Equifax		Commercial CA Equifax
Validity period from 3 nov'08 7:52:02 to 4 nov'09 7:52:02	same length (500 bytes)	Validity period from 31 ju'04 0:00:00 to 2 sep'04 0:00:00
Website domain .i.broke.the.internet.and.all.i.got.was.this.t-shirt.name .phreedom.org		Sub-CA name MD5 collisions Inc. (http://www.phreedom.org/md5)
2048-bit RSA public key B2D32581AA28 E878B1E50...	different contents	1024-bit RSA public key BAA659C92C28 D62AB0F8E...
Extensions "CA = false"	collision bits	Extensions "CA = true"
Identity verified by Equifax	identical suffixes	Comment 33000000275E 39E089610...
	identical signatures	Identity verified by Equifax

- Valid signature for both from Verisign!

Using 200 PlayStation 3s



Rogue C.A.

- Realistic Man-in-the-Middle attack against any secure website



- Responsible disclosure
 - Pre-informed Browsers and C.A.s
 - Rogue C.A. purposely crippled: only valid in August 2004
- MD5 deprecated within hours
- Software released for research
 - Anyone can create chosen-prefix collisions
 - ≈ 1 day on quadcore machine
 - <https://github.com/cr-marcstevens/hashclash>

What happened since?

- 2009: CABforum: MD5 deprecated for signatures
- 2012: supermalware Flame uses forged MD5 signature to push fake Windows Updates
Discovery of yet-unknown variant MD5 collision attack
- 2016: SLOTH: Transcript collision attacks against TLS, IKE, SSH
- 2017: Oracle JRE rejects MD5 signatures
Originally planned for Januari, was postponed till April
- 2018: US SWGDE (Scientific Working Group on Digital Evidence) Publication *“explains that the use of the MD5 and SHA1 hash algorithms remains acceptable”*

Deprecating SHA-1 in 2017

(known to be weak since 2005)

Joint work with

Ange Albertini, Elie Bursztein, Pierre Karpman, Yarik Markov

Weaknesses

SHA-1 is not collision resistant

Collision attack with complexity 2^{69} (4M core-years) [WangYY 2005]

Later improved to 2^{61} (15,300 core-years) [Stevens 2012]

Projected costs of SHA-1 collisions [Schneier 2012]

\$2.77M in 2012

\$700K by 2015

\$173K by 2018 ⇒ “we can postpone 5 years..”

\$43K by 2021

(based on [Stevens 2012], Amazon EC2 rates & Moore’s Law)

Practical SHA-1 collision remained open problem

- [S13]: SHA-1 collision attack with complexity $\approx 2^{61}$
- \Rightarrow CPU attack: 15.3K coreyears
- [SPK16]: attack complexity $\approx 2^{62.2}$ on GTX-970
- \Rightarrow GPU attack: 112 GPUyears
- \approx \$100k renting fee (on Amazon EC2)
- $\times 7$ lower cost in 2015 than predicted earlier by Schneier
- Initiated collaboration with Google



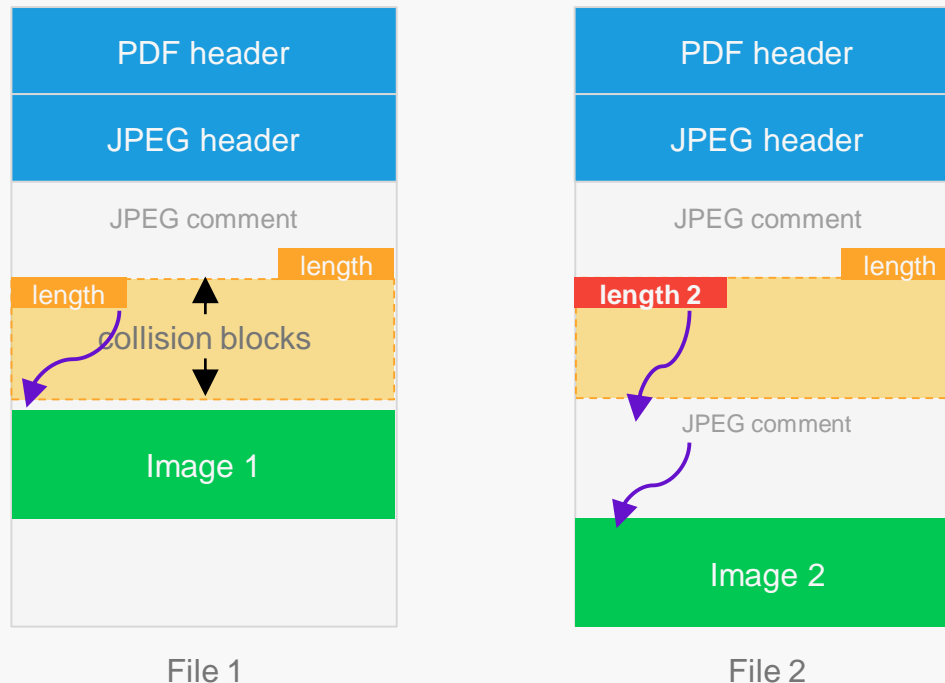
First SHA-1 Collision

- **Collaboration with Google [SBKAM17]**
 - Google Infrastructure:
Large heterogenous cluster of CPUs & GPUs
 - But: no direct access, proprietary Compile & Job system
 - ‘Blind’ adaptation source-code by Google
- **First near-collision attack**
 - Took 3583 core years $\approx 2^{60}$ SHA-1 compressions
 - Run on 100k+ PCs in several weeks
- **Second near-collision attack**
 - Tailored to 1st NC output
 - Using NVIDIA Tesla K20, K40, K80
 - Took ≈ 114 K20years ≈ 71 K80years $\approx 2^{62.8}$ SHA-1
 - Run on >3000 GPUs in just **8 calendar days**
- Collision on <https://shattered.io/>

Colliding PDFs

Reusable meaningful SHA-1 collision:

- 1 collision: infinite colliding PDF-pairs with distinct embedded JPGs
- Use JPG for page-content \Rightarrow arbitrary distinct page contents
- Use PDF image cropping \Rightarrow arbitrary distinct multi-page contents



- DIY: <https://github.com/nneonneo/sha1collider>

- Project Webpage, Google Drive & Gmail check for SHA-1 collisions
- Unexpectedly collision can break Subversion repositories
 - Webkit developer submitted test to prove WebKit resistant to SHA-1 collisions
 - Broke Webkit repository
 - Internal deduplication uses SHA-1 and keeps only 1 colliding file
 - MD5 is used to check integrity \Rightarrow will always fail on checkout
- Git started moving away from SHA-1
- Git & GitHub now using strengthened SHA-1 implementation by default
- CA/Browser Forum: Ballot 152
 - Extend issuance SHA-1 certificates up to 1 Jan. 2017 (before: 1 Jan. 2016)
 - (unaltered: deprecate SHA-1 certificates after 1 Jan. 2017)
 - Our recommendations on 8 Oct. ensured Ballot did not pass on 16 Oct.
- TLS 1.3 draft 9
 - Deprecated all uses of SHA-1 digital signatures

From attacks to toys



Instant Collisions

- Instant collision scripts for many file formats
 - Instant, re-usable and generic collisions
 - Take any pair of files, run script, get colliding files
- SHA-1: PDF, HTML
- MD5
 - PDF
 - PNG, JPG, JP2
 - MP4, GIF
 - PE (windows executable)

OUR CONTRIBUTIONS - 1/2

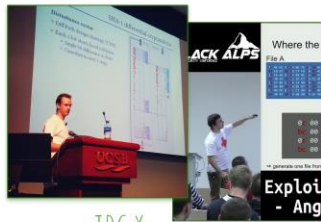
INSTANT MD5 COLLISIONS, WITH NO RECOMPUTATION
(COLLISION DATA IS PRE-COMPUTED)



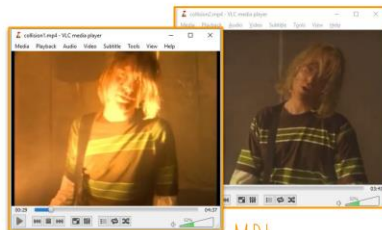
PDF



PNG*



JPG*



MP4

*SOME LIMITATIONS

OUR CONTRIBUTIONS - 2/2



GIF*

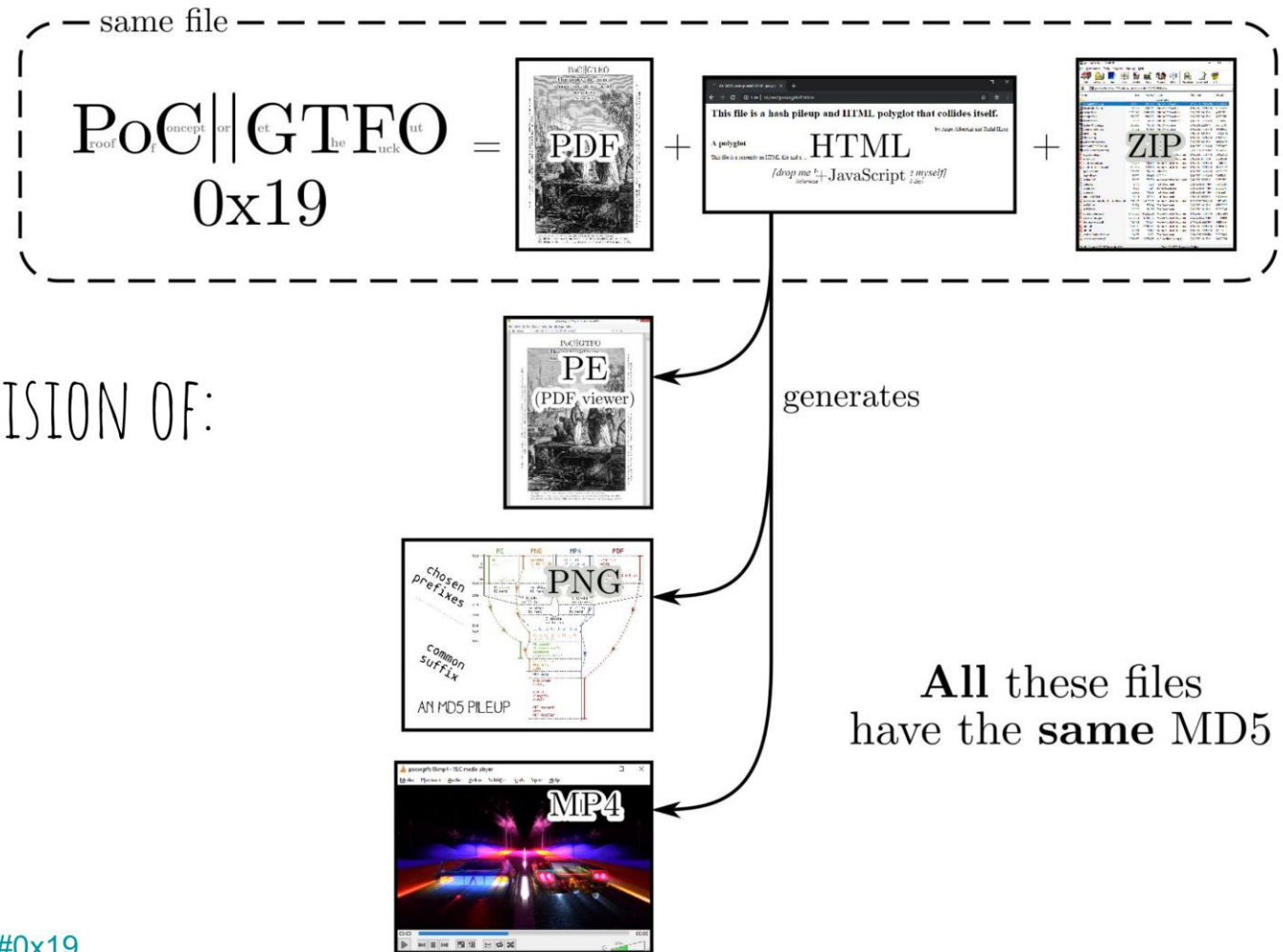


PE (WINDOWS EXECUTABLES)



JP2

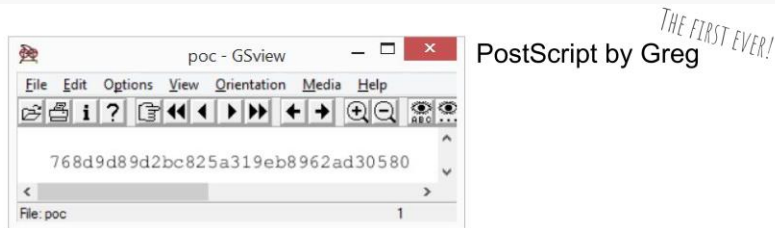
*SOME LIMITATIONS



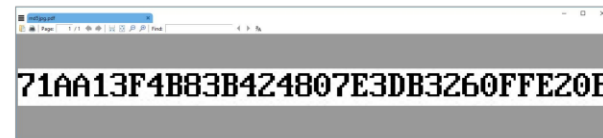
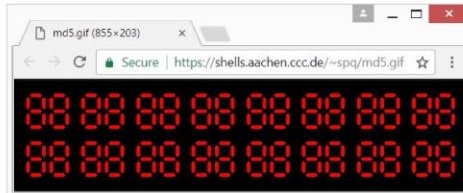
AN INSTANT COLLISION OF:

- A DOCUMENT
- AN EXECUTABLE
- AN IMAGE
- A VIDEO.

Hashquines: documents that show their own MD5 hash



ANIMATED
GIFs by spq



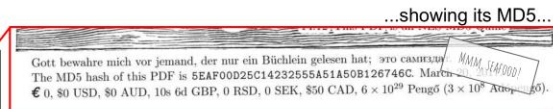
As images

PDFs by Mako

A LaTeX-generated
PDF...



(15x32=480 collisions)



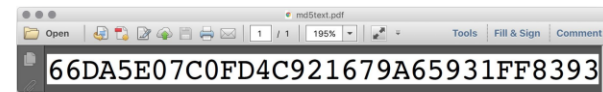
...showing its MD5...

...also a NES rom...



...showing the same MD5!

(4x32=128 collisions)



As text

```
$ pdftotext -q md5text.pdf -
66DA5E07C0FD4C921679A65931FF8393
$ md5sum md5text.pdf
66da5e07c0fd4c921679a65931ff8393 md5text.pdf
```

Thank you!