SHA-1 is a Shambles

First Chosen-Prefix Collision on SHA-1 and Application to the PGP Web of Trust

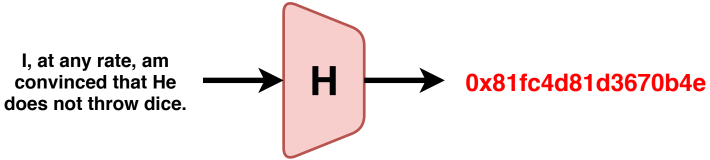
Gaëtan Leurent (INRIA - France)
Thomas Peyrin (NTU - Singapore)

USENIX 2020

Boston (USA) - August 14, 2020

https://sha-mbles.github.io/

What is a Hash Function ?



H maps an **arbitrary length input** (the message M) to a **fixed length n-bit output**.

Typically :

- $n = 128$ bits (MD5)
- $n = 160$ bits (SHA-1)
- $n = 256$ bits (SHA-256)

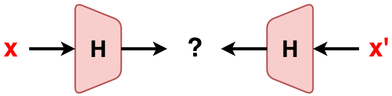
The cryptographic hash functions security goals

pre-image resistance :

2nd pre-image resistance :

collision resistance :

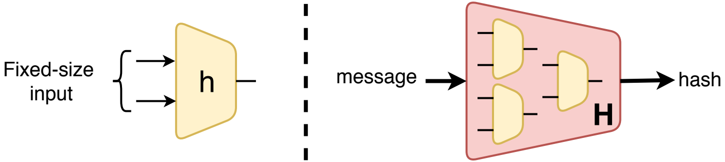
The attacker can not find two messages (x, x') such that $H(x) = H(x')$, in less than $\theta(2^{n/2})$ operations (generic birthday paradox attack).



General hash construction

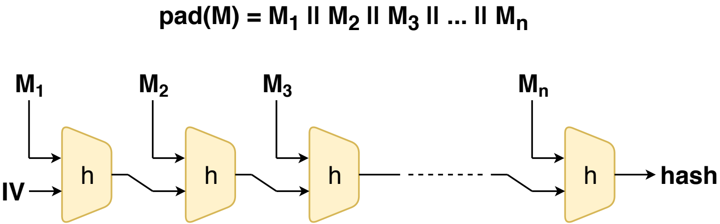
Most hash functions are composed of two elements :

- **a compression function** h **:**
  a function for which the input and output size is fixed.
- **a domain extension algorithm :** an iterative process that uses the compression function h so that the hash function H can handle inputs of arbitrary length.



The Merkle-Damgård domain extension algorithm

The most famous domain extension algorithm used is called the **Merkle-Damgård** [MD-CRYPTO89] iterative algorithm.



The compression function h now takes two fixed-size inputs, the incoming chaining variable CV_i and the message block M_i, and outputs a new chaining variable CV_{i+1}.

Current security of SHA-1

The (bad looking) current situation of SHA-1 :

- 1995 SHA-1 published (SHA-0 (1993) with a slight twist) [NIST-FIPS-180-1]
- 2005 **theoretical collision attack** on the full hash - 2^{69} [WYY-CRYPTO05]
- 2006-2011 lots of works computing collisions for reduced-round versions
- 2015 collision computed on the full compression function - 2^{57} [SKP-EUROCR.16]
- 2017 **computations of a collision** on the full hash (identical-prefix collision) - $2^{64.7}$ [SBK+-CRYPTO17]
- 2019 practical chosen-prefix collision attack on the full hash - $2^{67.2}$ [LP-EUROCR.19]
- **New** computation of a **chosen-prefix collision on the full hash** - $2^{63.7}$ **PGP/GnuPG key-certification forgery**

Motivations to study SHA-1

SHA-1 is not used anymore, right ? .... right ! ?

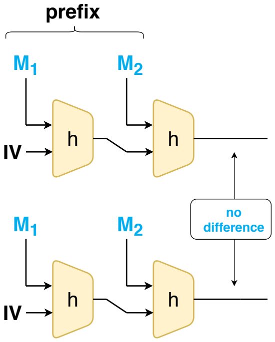
- SHA-1 **certificates** (X.509) still exists
  - CAs sell legacy SHA-1 certificates for legacy clients
  - Accepted by many non-web modern clients
  - ICSI Certificate Notary : 1.3% SHA-1 certificates
- PGP signatures with SHA-1 are still trusted
  - Default hash for key certification in GnuPGv1 (legacy branch)
  - 1% of public certifications (Web-of-Trust) in 2019 use SHA-1
- SHA-1 still allowed for in-protocol signatures in TLS, SSH (used by more than 3% of Alexa top 1M servers)
- HMAC-SHA-1 ciphersuites (TLS) still used by more than 8% of Alexa top 1M servers
- Probably a lot of more obscure protocols ... (EMV credit cards use weird SHA-1 signatures)

Another push is needed to accelerate the retirement of SHA-1

What are identical-prefix collisions ?

Identical-prefix collision attack

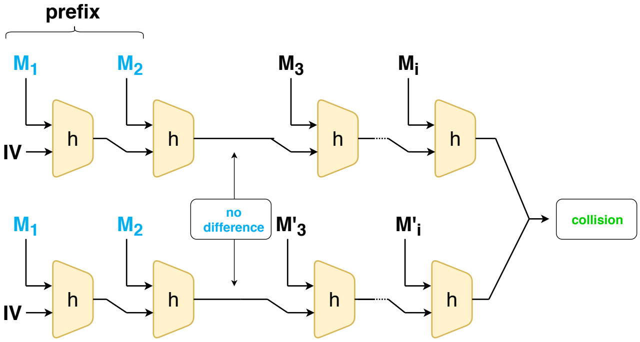
The attacker is first challenged with **one prefix** P and its goal is to compute two messages M and M' to create the **collision** $H(P||M) = H(P||M')$, where $||$ denotes concatenation



What are identical-prefix collisions ?

Identical-prefix collision attack

The attacker is first challenged with **one prefix** P and its goal is to compute two messages M and M' to create the **collision** $H(P||M) = H(P||M')$, where $||$ denotes concatenation



What are identical-prefix collisions ?

Identical-prefix collision attack

The attacker is first challenged with **one prefix** P and its goal is to compute two messages M and M' to create the **collision** $H(P||M) = H(P||M')$, where $||$ denotes concatenation

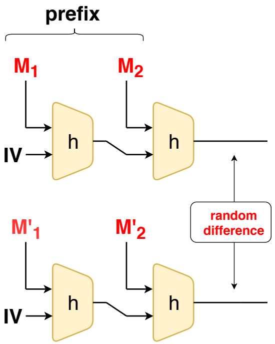
The colliding blocks will be almost random looking, but any prefix or suffix can be used (as long as no difference inserted)

- breaks integrity
- colliding PDFs (see SHAttered for SHA-1 [SBK+-CRYPTO17])

What are chosen-prefix collisions ?

Chosen-prefix collision attack

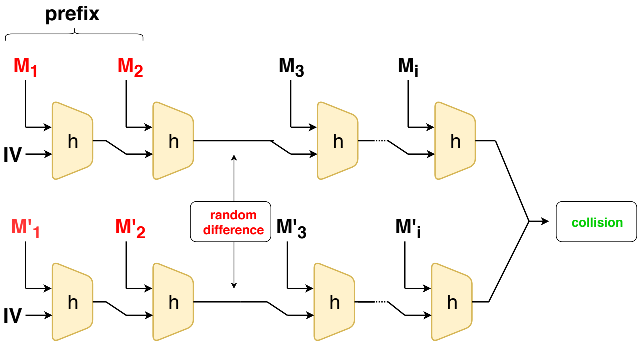
The attacker is first challenged with **two message prefixes** P and P', and its goal is to compute two messages M and M' to create the **collision** $H(P||M) = H(P'||M')$, where $||$ denotes concatenation



What are chosen-prefix collisions ?

Chosen-prefix collision attack

The attacker is first challenged with **two message prefixes** P and P', and its goal is to compute two messages M and M' to create the **collision** $H(P||M) = H(P'||M')$, where $||$ denotes concatenation



What are chosen-prefix collisions ?

Chosen-prefix collision attack

The attacker is first challenged with **two message prefixes** P and P', and its goal is to compute two messages M and M' to create the **collision** $H(P||M) = H(P'||M')$, where $||$ denotes concatenation

Much more powerful and **much harder** than an identical-prefix collision

- breaks certificates (Rogue CA [SSA+-CRYPTO09]
- breaks TLS, SSH (SLOTH attack [BL-NDSS16])

Our results

1 - Complexity improvements (factor 8 ∼ 10)

- identical-prefix collision from $2^{64.7}$ to $2^{61.2}$
  (11 kUS$ in GPU rental)
- chosen-prefix collision from $2^{67.1}$ to $2^{63.4}$
  (45 kUS$ in GPU rental)

2 - Record computation

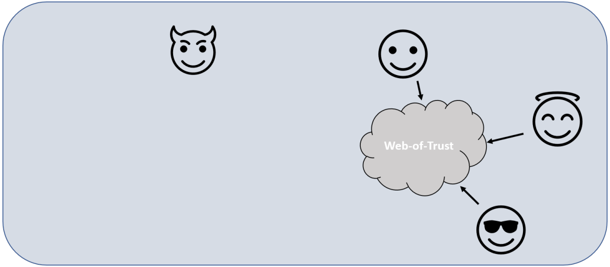
- implementation of the full (very technical) attack
- **2 months** of computation using 900 GPU (GTX 1060)

3 - PGP Web-of-Trust impersonation

- **2 keys with different IDs and colliding certificates**
- certification signature can be copied to the second key

Result 3 - PGP Web-of-Trust impersonation

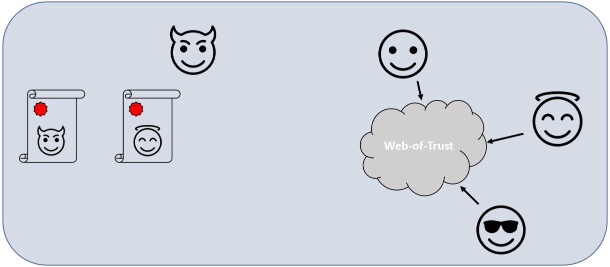
The **Web of Trust** is a trust model used for PGP that relies on users signing each other's identity certificate, instead of using a central PKI. For compatibility reasons the legacy branch of GnuPG (version 1.4) still uses SHA-1 by default for identity certification.



Result 3 - PGP Web-of-Trust impersonation

Idea :

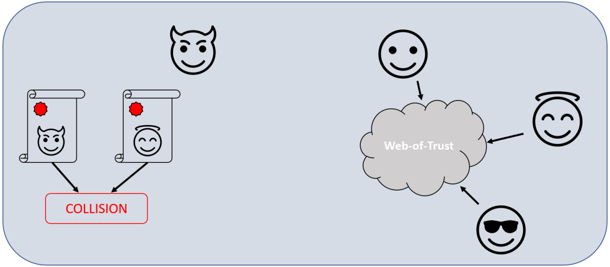
- create a pair of keys with two different UserIDs : victim name (A) and attacker name (B)
-
-
-



Result 3 - PGP Web-of-Trust impersonation

Idea :

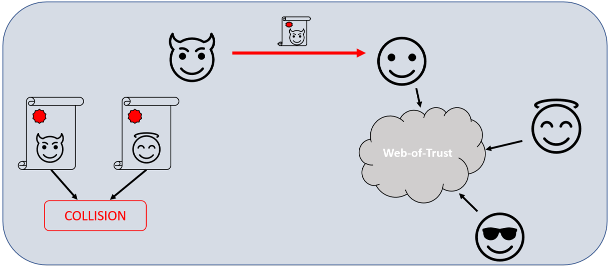
- create a pair of keys with two different UserIDs : victim name (A) and attacker name (B)
- using a chosen-prefix collision, we craft the keys such that the SHA-1 hash that is signed for the key certification is the same for both keys.
-
-



Result 3 - PGP Web-of-Trust impersonation

Idea :

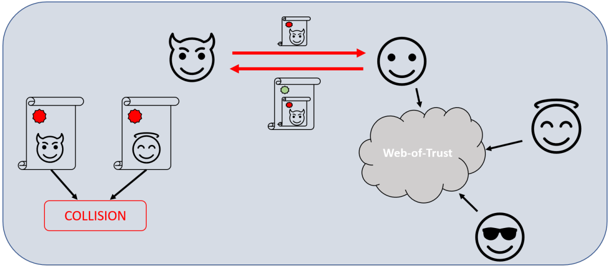
- create a pair of keys with two different UserIDs : victim name (A) and attacker name (B)
- collide key certifications
- the attacker asks for key certifications of key B : since he knows the corresponding secret key, and the UserID matches his official ID, he will collect trust-worthy signatures and integrate the web-of-trust.
-



Result 3 - PGP Web-of-Trust impersonation

Idea :

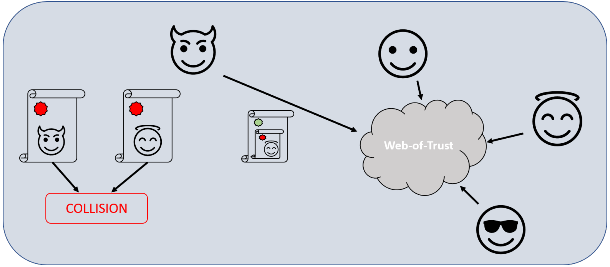
- create a pair of keys with two different UserIDs : victim name (A) and attacker name (B)
- collide key certifications
- the attacker asks for key certifications of key B : since he knows the corresponding secret key, and the UserID matches his official ID, he will collect trust-worthy signatures and integrate the web-of-trust.
-



Result 3 - PGP Web-of-Trust impersonation

Idea :

- create a pair of keys with two different UserIDs : victim name (A) and attacker name (B)
- collide key certifications
- integrate web of trust with UserID B
- since the hash of both keys collide, he can transplant the signatures to key A, creating a key with the UserID of the victim, trusted by the web-of-trust, and for which he controls the secret key. He can then sign messages pretending to be the victim.



Impact of our attack

GnuPG

CVE-2019-14855 : a countermeasure has been implemented since GnuPG version 2.2.18 (November 2019). SHA-1-based identity signatures created after 2019-01-19 are now considered invalid.

OpenSSL

Recent OpenSSL versions no longer allow X.509 certificates signed using SHA-1 at security level 1 (default configuration for TLS/SSL) and above

OpenSSH

Latest versions of OpenSSH (since 8.2) include a “future deprecation notice” explaining that SHA-1 signatures will be disabled in the near-future

... and more. Please check https://sha-mbles.github.io/

Conclusion

If you didn't know it already

DON'T USE SHA-1 ! Use SHA-2 or SHA-3 instead.

What about HMAC-SHA-1 ?

Our attack doesn't apply to HMAC-SHA-1, but we still advise to move to another hash function. SHA-1 has been dead for 15 years now, time to move on !

On security margin

Deprecating a cryptographic primitive is incredibly complex, long and painful : don't underestimate the importance of security margin in crypto designs.

64-bit security = no security

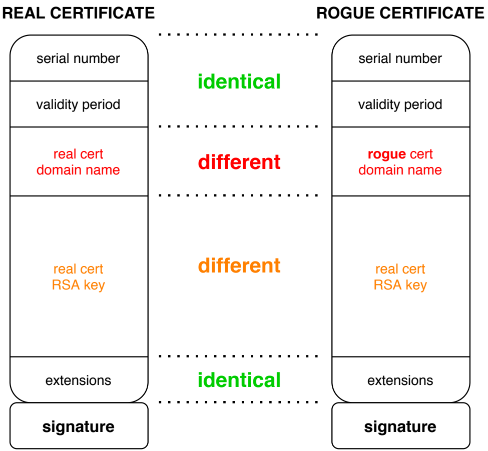
2^{64} **is now a feasible computation**, even if you are not the NSA or Google

Thanks for watching this presentation !

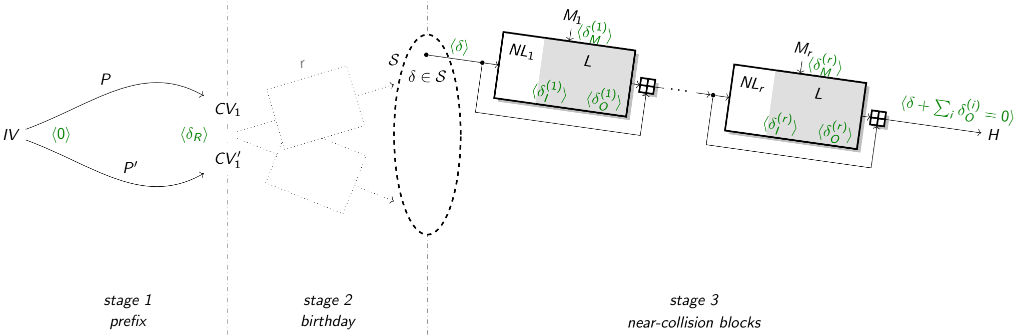
Contact :
gaetan.leurent@inria.fr
thomas.peyrin@ntu.edu.sg

Why chosen-prefix collisions are interesting ?

Colliding SSL certificates [SLW-EUROCR.07] :



Result 1 - Complexity improvements



1. **Prefix :** Compute $CV_1 = h(IV, P)$ and $CV_1' = h(IV, P')$
2. **Birthday phase :** Find M, M' such that $H(P \parallel M) - H(P' \parallel M') \in \mathcal{S}$
3. **Near-collision phase :** Erase the state difference, using near-collision blocks

Complexity improved from $\approx 2^{67}$ [LP-EUROC.19] to $2^{63} \sim 2^{64}$

Result 2 - Record computation

- Running the attack on Amazon/Google cloud GPU is estimated to cost 160 kUS$ (spot/preemptible instances)
- After cryptocurrency crash in 2018, cheap GPU farms to rent !
  - 👍 3–4 times cheaper
    45 kUS$ with current public prices on `gpuserversrental.com`
  - 👎 Gaming or mining-grade GTX cards (rather than Tesla)
  - 👎 Low-end CPUs
  - 👎 Slow internet link
  - 👎 No cluster management
  - 👎 Pay by month, not on-demand

Pricing fluctuates with cryptocurrencies markets, we didn't get optimal prices
(the actual computation costed us 75 kUS$)

Result 2 - Record computation

Pricing fluctuates with cryptocurrencies markets, we didn't get optimal prices
(the actual computation costed us 75 kUS$)

September 27 : The First SHA-1 Chosen-prefix Collision

- 416-bit prefix
- 96 birthday bits
- 9 near-coll. blocks

Message A	Message B
99040d047fe81780012000ff4b65792069732070617274206f66206120636f6c	99030d047fe81780011800ff50726163746963616c205348412d312063686f73
6c6973696f6e212049742773206120747261702179c61af0afcc054515d9274e	656e2d70726566697820636f6c6c6973696f6e211d276c6ba661e1040e1f7d76
7307624b1dc7fb23988bb8de8b575dba7b9eab31c1674b6d974378a827732ff5	7f076249ddc7fb332c8bb8c2b7575dbec79eab2be1674b7db34378b4cb732fe1
851c76a2e60772b5a47ce1eac40bb993c12d8c70e24a4f8d5fcdedc1b32c9cf1	891c76a0260772a5107ce1f6e80bb9977d2d8c68524a4f9d5fcdedcd0b2c9ce1
9e31af2429759d42e4dfdb31719f587623ee552939b6dcdc459fca53553b70f8	9231af26e9759d5250dfdb2d4d9f58729fee553319b6dccc619fca4fb93b70ec
7ede30a247ea3af6c759a2f20b320d760db64ff479084fd3ccb3cdd48362d96a	72de30a087ea3ae67359a2ee27320d72b1b64fecc9084fc3ccb3cdd83b62d97a
9c430617caff6c36c637e53fde28417f626fec54ed7943a46e5f5730f2bb38fb	904306150aff6c267237e523e228417bde6fec4ecd7943b44a5f572c1ebb38ef
1df6e0090010d00e24ad78bf92641993608e8d158a789f34c46fe1e6027f35a4	11f6e00bc010d01e90ad78a3be641997dc8e8d0d3a789f24c46fe1eaba7f35b4
cbfb827076c50eca0e8b7cca69bb2c2b790259f9bf9570dd8d4437a3115faff7	c7fb8272b6c50edaba8b7cd655bb2c2fc50259e39f9570cda94437bffd5fafe3
c3cac09ad25266055c27104755178eaeff825a2caa2acfb5de64ce7641dc59a5	cfcac0981252661 5e827105b79178eaa43825a341a2acfa5de64ce7af9dc59b5
41a9fc9c756756e2e23dc713c8c24c9790aa6b0e38a7f55f14452a1ca2850ddd	4da9fc9eb56756f2563dc70ff4c24c932caa6b1418a7f54f30452a004e850dc9
9562fd9a18ad42496aa97008f74672f68ef461eb88b09933d626b4f918749cc0	9962fd98d8ad4259dea97014db4672f232f461f338b09923d626b4f5a0749cd0
27fddd6c425fc4216835d0134d15285bab2cb784a4f7cbb4fb514d4bf0f6237c	2bfddd6e825fc431dc35d00f7115285f172cb79e84f7cba4df514d571cf62368
f00a9e9f132b9a066e6fd17f6c42987478586ff651af96747fb426b9872b9a88	fc0a9e9dd32b9a16da6fd16340429870c4586feee1af96647fb426b53f2b9a98
e4063f59bb334cc00650f83a80c42751b71974d300fc2819a2e8f1e32c1b51cb	e8063f5b7b334cd0b250f826bcc427550b1974c920fc280986e8f1ffc01b51df
18e6bfc4db9baef675d4aaf5b1574a047f8f6dd2ec153a93412293974d928f88	14e6bfc61b9baee6c1d4aae99d574a00c38f6dca5c153a834122939bf5928f98
ced9363cfef97ce2e742bf34c96b8ef3875676fea5cca8e5f7dea0bab2413d4d	c2d9363e3ef97cf25342bf28f56b8ef73b5676e485cca8f5d3dea0a65e413d59
e00ee71ee01f162bdb6d1eafd925e6aebaae6a354ef17cf205a404fbdb12fc45	ec0ee71c201f163b6f6d1eb3f525e6aa06ae6a2dfef17ce205a404f76312fc55
4d41fdd95cf2459664a2ad032d1da60a73264075d7f1e0d6c1403ae7a0d861df	4141fddb9cf24586d0a2ad1f111da60ecf26406ff7f1e0c6e5403afb4cd861cb
3fe5707188dd5e07d1589b9f8b6630553f8fc352b3e0c27da80bddba4c64020d	33e5707348dd5e1765589b83a7663051838fc34a03e0c26da80bddb6f464021d

Impact of our attack (2)

DNSSEC

SHA-1 remains used in DNSSEC, with 18% of top-level domains using SHA-1 signatures : anyone using a SHA-1 DNSKEY algorithm should upgrade - see related page from Tony Finch or IETF related discussions for more details

X.509 certificates

X.509 certificates could be broken (Rogue CA [SSA+-CRYPTO09]) **if some CAs issue SHA-1 certificates with predictable serial numbers**

TLS and SSH

TLS and SSH connections using SHA-1 signatures to authenticate the handshake could be attacked with the SLOTH attack [BL-NDSS16] **if the chosen-prefix collision can be generated extremely quickly** (within seconds or minutes)