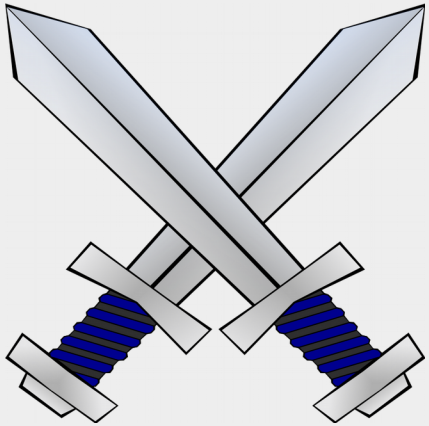
Zgostitveni algoritmi in zagotavljanje integritete digitalnih dokazov



Infosec-seminar.si, april 2018

Matej Kovačič

Institut Jožef Stefan

Center za prenos znanja na
področju informacijskih tehnologij

Laboratorij za umetno inteligenco

Forenzični zaseg podatkov

Cilj forenzičnega zasega je zagotoviti, da bodo zajeti podatki ohranili integriteto (istovetnost) in s tem **dokazno vrednost** na sodišču.

- Vprašanje lastništva podatkov.
- Vprašanje zasebnosti (tudi na delovnem mestu).
- Varstvo osebnih podatkov.
- Varstvo tajnih podatkov.
- Odločba ustavnega sodišča Up-106/05 -> ZKP-J.
- **Zagotavljanje integritete zajetih podatkov.**
- Varstvo zajetih podatkov.

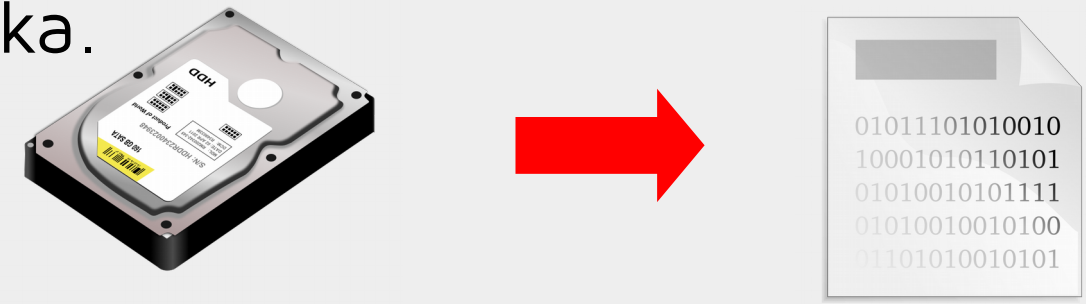
Integriteta digitalnih podatkov

Istovetnost podatkov na celotnem nosilcu podatkov (disk, USB ključek,...) - vključuje tudi tim. "prazen prostor".

Istovetnost podatkov na particiji (razdelku) - vključuje tudi tim. "prazen prostor".

Istovetnost vsebine (in metapodatkov) datoteke – ne pa tudi imena datoteke.

Kopiranje slike (ang. *image*) nosilca podatkov ali razdelka.



Nosilci podatkov, razdelki, datotečni sistemi...

Nosilec podatkov (disk,...)

- Disk0, Disk1,...
- /dev/hda, /dev/sda

Razdelki (particije):

- C:, D:,...
- /dev/hda1, /dev/sda1
- posamezen nosilec podatkov lahko vsebuje en sam razdelek;
- posamezen nosilec podatkov je lahko razdeljen na več razdelkov;
- posamezen razdelek se lahko razteza čez več nosilcev podatkov;

Datotečni sistem:

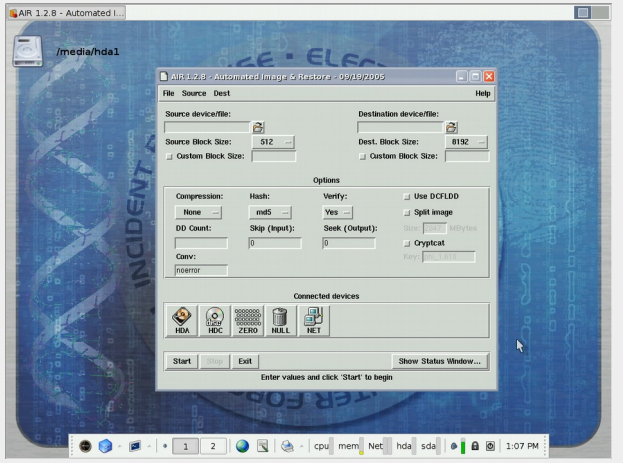
- FAT, NTFS, ext3, ext4, ReiserFS, zfs, Btrfs...

Skriti deli diskov:

- Host Protected Area, definiran s ATA-4 standardom leta 1998.
- Device Configuration Overlay, definiran s ATA-6 standardom leta 2002.

Postopek forenzičnega kopiranja

Priklop nosilca podatkov v posebno napravo ali prilagojen računalnik (strojno onemogočanje pisanja s posebnimi kabli ali vmesniki).



AIR Imager – programski zaseg podatkov

Delna rešitev: uporaba »živega CD-ja«, ki ne priklaplja »swap« razdelka na sistemu.

Razdelka **ne priklapljamo v sistem** (niti kot read-only), pač pa kopiramo sliko razdelka (*image*) in **kasneje** forenzično obdelamo to sliko.

Forensic Bridge, vir in avtorstvo: Tableau.com

Zakaj? Zato.

datotečni sistem	bralni način	pisalni način
FAT16	✔	✔
FAT32	✔	✔
NTFS	✔	*
ext2	✔	*
ext3	✔	*
ReiserFS	*	*

✔ - do spremembe ne pride
* - do spremembe pride!

Pri datotečnem sistemu NTFS pride do spremembe šele ko podatke preberemo iz za pisanje priključenega diska.

Zgostitveni algoritmi

Zgostitveni algoritmi (ang. *hash algorithms*, včasih tudi *hash values*, *hash codes*, *hash sums*, *checksums*, *message digests* ali *fingerprints*): poljubno dolg niz znakov preslikajo v število fiksne dolžine.

Izračunajo tim. prstni odtis (ang. *fingerprint*) oz. kontrolno vsoto (*hash*) tega niza znakov, kar je osnova za digitalni podpis oziroma za **zagotovilo, da so podatki ohranili integriteto**.

Zgostitveni algoritmi

Primeri zgostitvenih algoritmov: MD5, SHA-0, SHA-1, SHA-256, SHA-512, WHIRLPOOL, SHA-3...

MD5:

75222cee3990e39e9fb48fa7ca6a733b

SHA-1:

1f149834675ab2ae6d076ee3cbaa9158b6864ee1

SHA-256:

3226338fb2c35ca40d39de77a0735779b1c0886f39a3762de2b502901567d39e

e9a23cbc455158951716b440c3d165e0

c7931bbead86523571b02d5cf795a79d

Zgostitveni algoritmi

Zgostitveni algoritmi morajo biti:

- **enosmerni** (iz kontrolne vsote ni mogoče nazaj izračunati originalnih podatkov),
- (v praksi) ne sme priti do **kolizije** (ne smeta obstajati dva različna niza podatkov, ki bi vrnila isto kontrolno vsoto; to je sicer odvisno tudi od tim. »bitnosti kontrolne vsote«).

Dobri zgostitveni algoritmi imajo tim. "avalanche efekt" - če se vhod malenkost spremeni, se bo izhod drastično spremenil.

Uporaba

- Za zaščito gesel (hramba v hash obliki);
- kot pseudonaključni generator števil, pri generiranju naključnih imen datotek (npr. http://www.dnevnik.si/uploads/image_cache/305d4e28924252f4251b2baadb6dbc6a.jpeg);
- za preverjanje integritete pri prenosu datotek (npr. v P2P omrežjih);
- za preverjanje integritete arhivskih in drugih datotek (tim. *checksum*; npr. orodje Tripwire);
- pri implementaciji digitalnega podpisa, časovnega žigosanja, overovitvi certifikatov s strani CA;

Uporaba

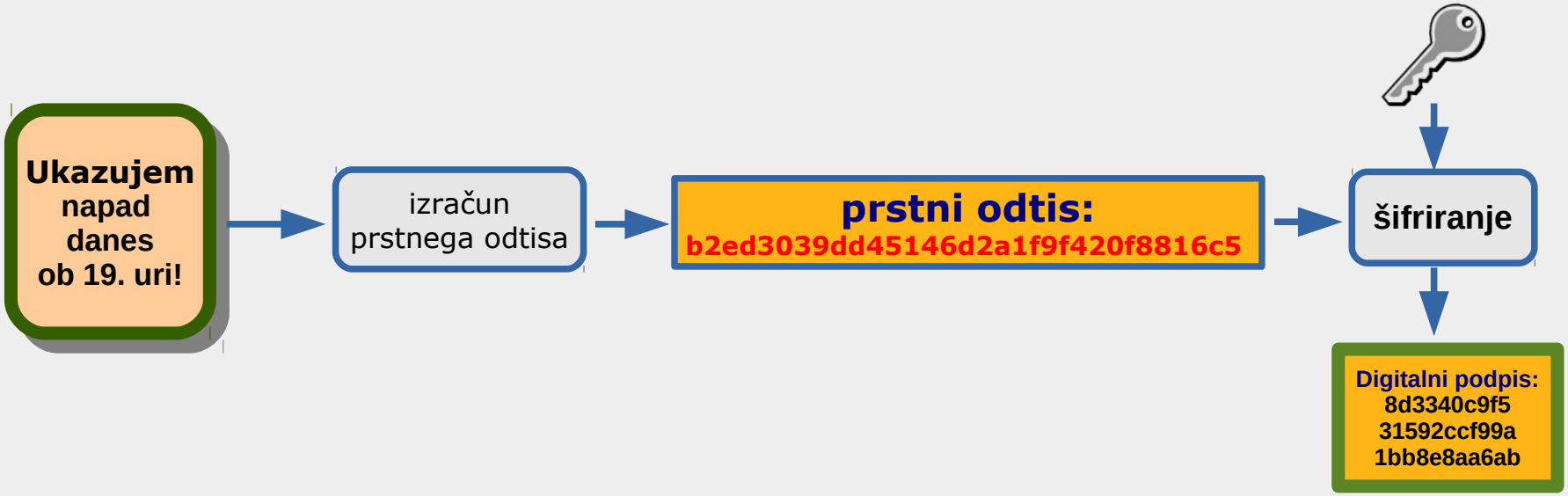
- za digitalno podpisovanje datotek, npr. gonilnikov;
- za zagotavljanje integritete podatkov pri digitalni forenziki;
- za prepoznavanje datotek (npr. pri antivirusnem programju, znanih "slabih" in znanih "dobrih" datotek v digitalni forenziki: NSRL hash set (*National Software Reference Library*), *HashKeeper*,...).

Digitalni podpis

Digitalni podpis zagotavlja integriteto sporočila (da se vsebina sporočila ni spremenila med prenosom).

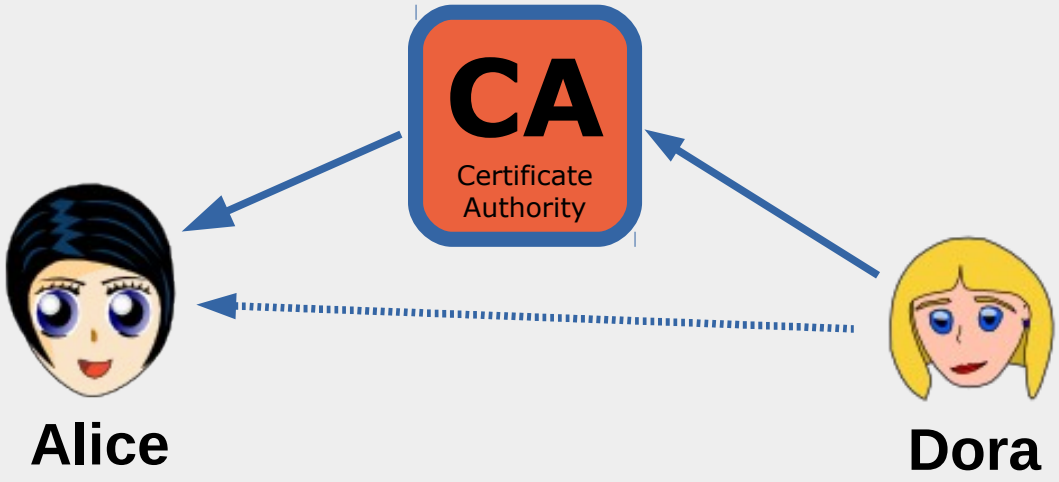
Pošiljatelj kontrolno vsoto sporočila zašifrira s svojim šifrirnim ključem.

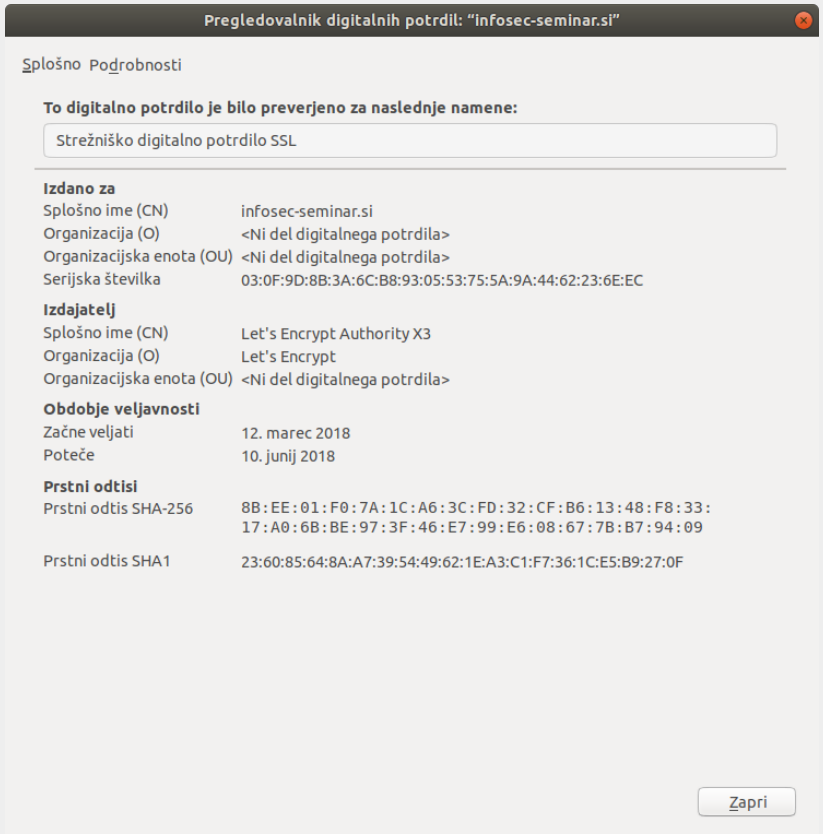
Prejemnik kontrolno vsoto dešifrira ter jo primerja s kontrolno vsoto dejansko prejetega sporočila.



Overovitev ključev

(PKI – Public Key Infrastructure)





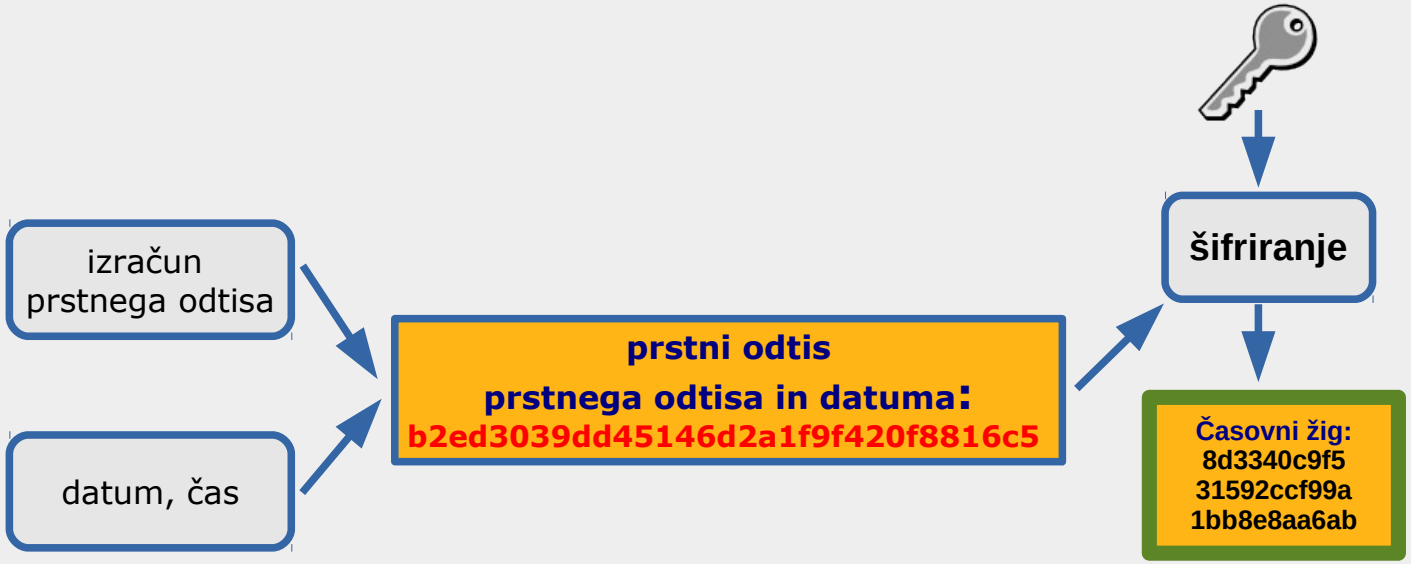
- CA je preveril identiteto in nato digitalno podpisal njen ključ.
- Dora zaupa, da ključ res pripada Alice, ker zaupa CA.

Časovno žigosanje

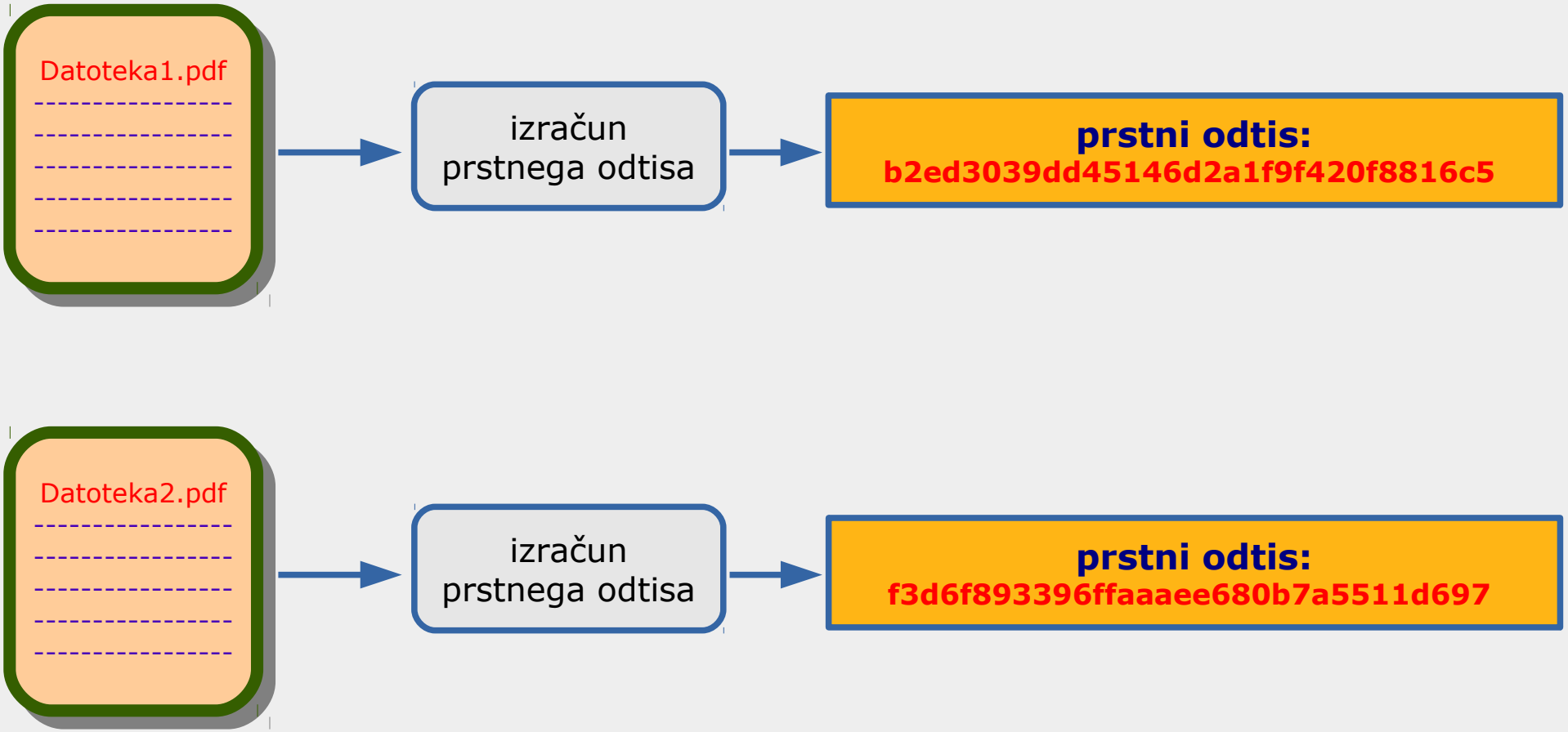
TSA - *Time Stamping Authority*, poseben zaupanja vreden strežnik, ki kontrolni vsoti dokumenta doda podatek o času nastanka in nato oba podatka digitalno podpiše.

Časovno žigosanje omogoča preverjanje časa nastanka (oz. žigosanja) danega dokumenta, časovnega žiga pa ne more spremeniti niti lastnik dokumenta.

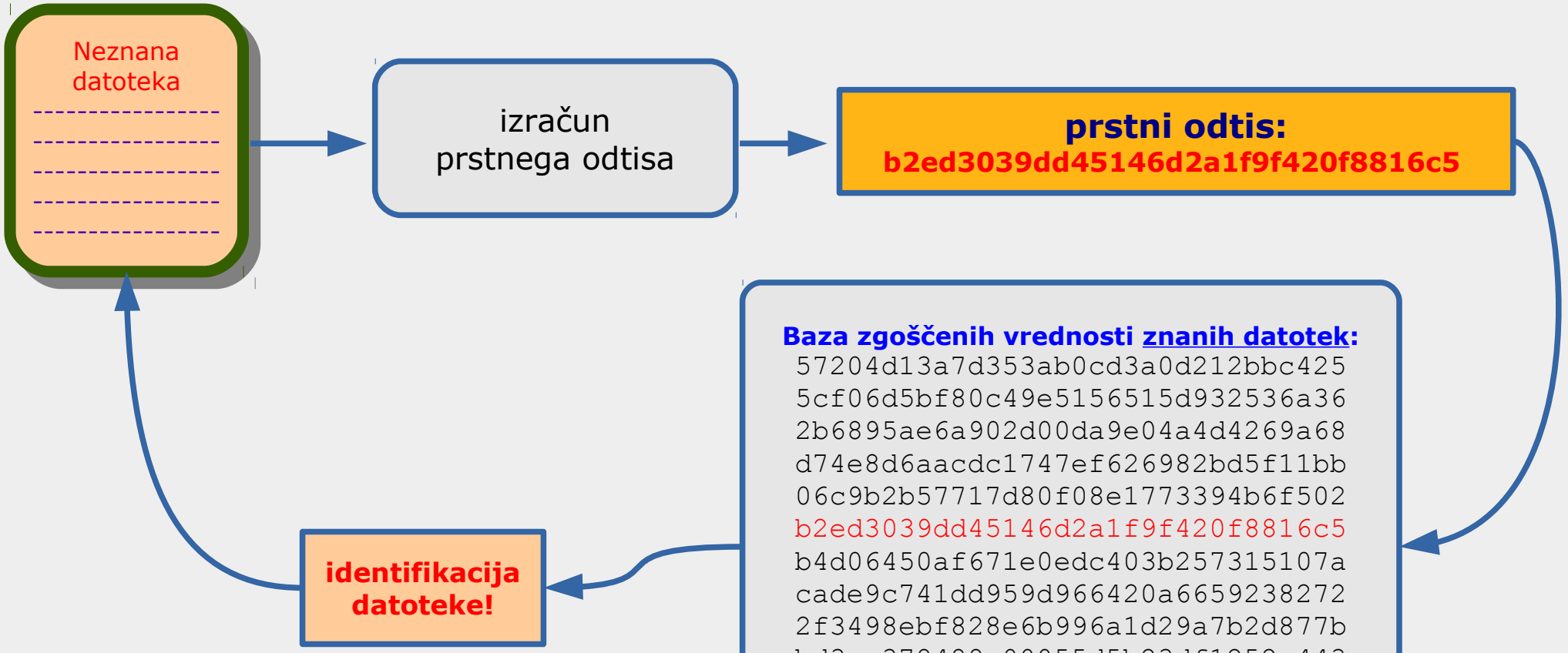
OpenTSA.



Identifikacija datotek

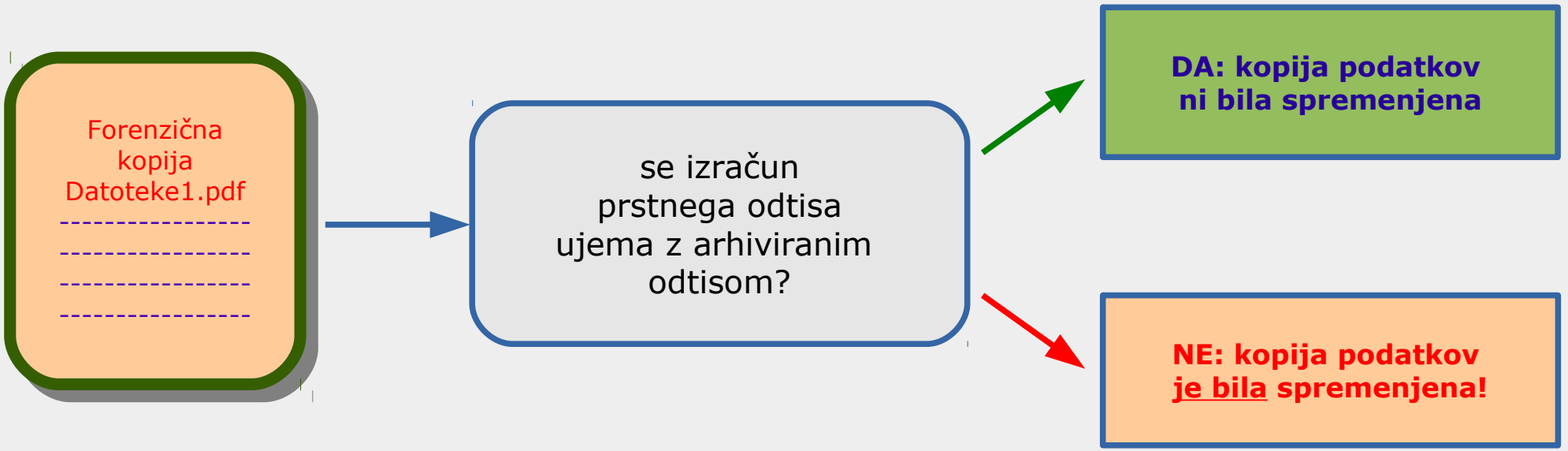


Identifikacija datotek



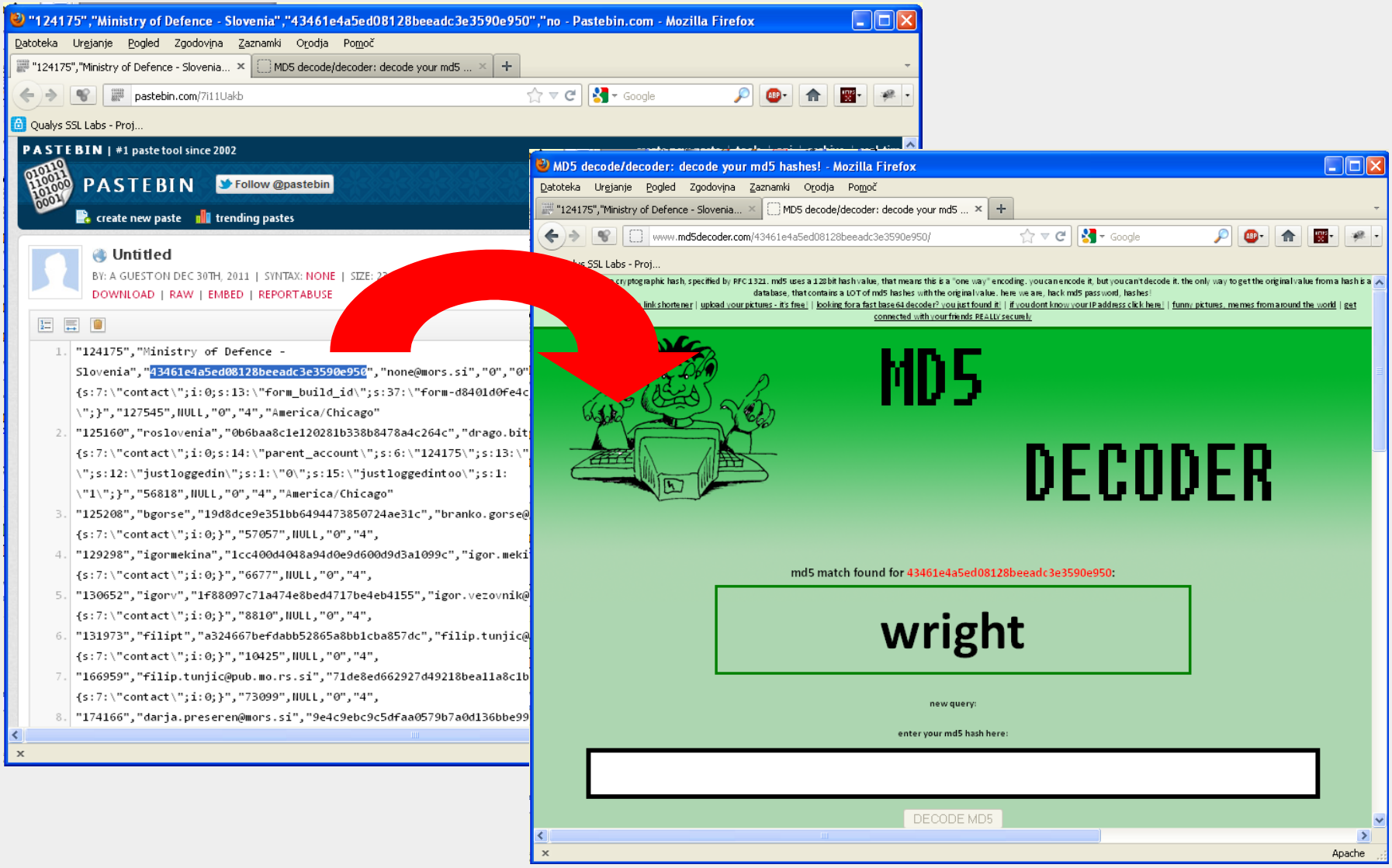
Pozor: razsodba *United States v. Crist, 2008 WL 4682806 (M.D.Pa. October 22 2008) (Kane, C.J.)* - izračunavanje digitalnih prstnih odtisov datotek predstavlja preiskavo.

Zagotavljanje integritete podatkov



Enosmernost?

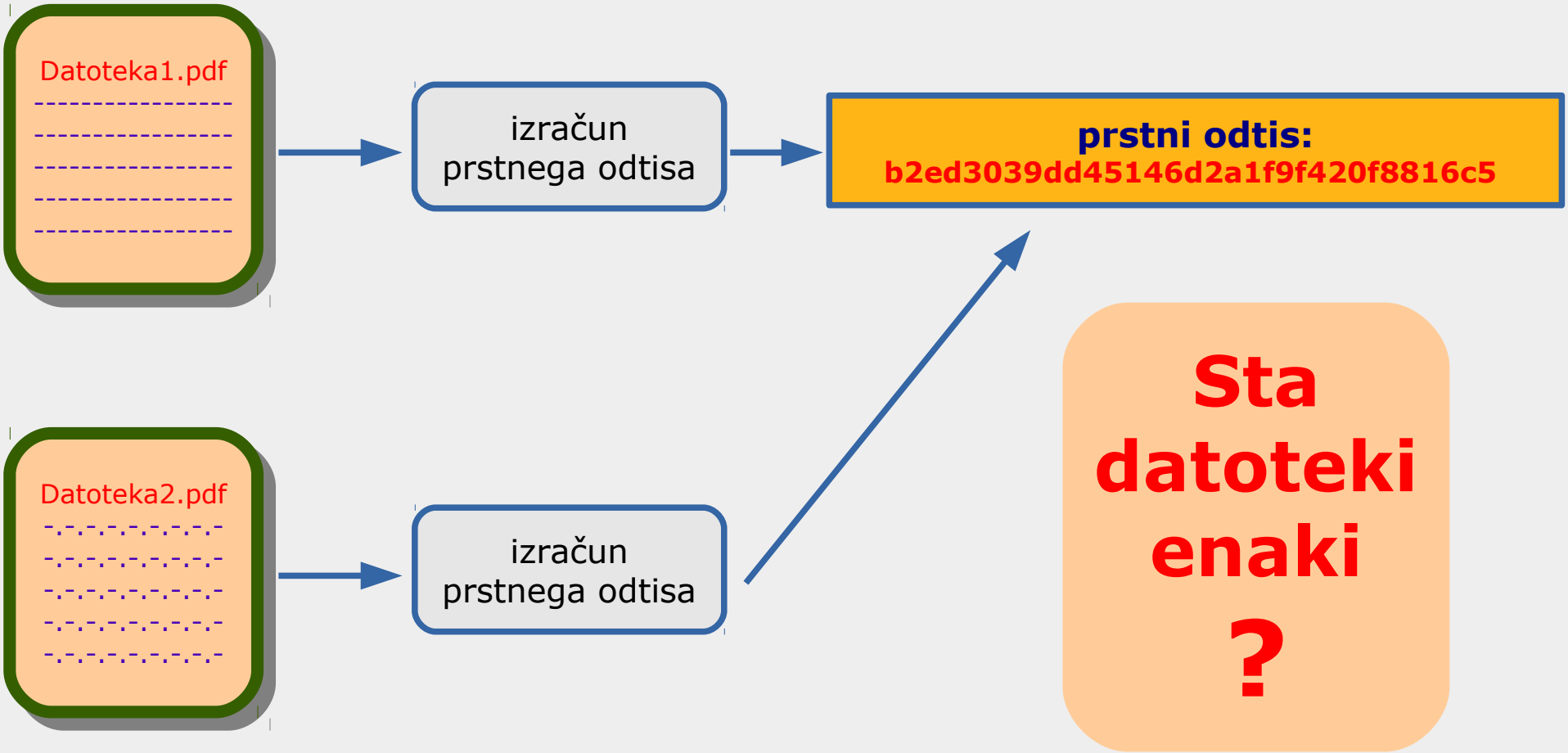
Mavrične tabele in predizračunane tabele zgoščenih vrednosti



Problem kolizije

Ne smeta obstajati dva
različna niza znakov,
ki bi vrnila
isto kontrolno vsoto,
sicer nastopi tim. kolizija.

Problem kolizije



Kolizijski napadi na MD5

MD5 se je v preteklosti uporabljal pri overjanju digitalnih potrdil ter pri zagotavljanju integritete podatkov v digitalni forenziki. Razvili so ga leta 1991.

Leta 1993 sta Den Boer in Bosselaers našla prvo "psevdo-kolizijo".

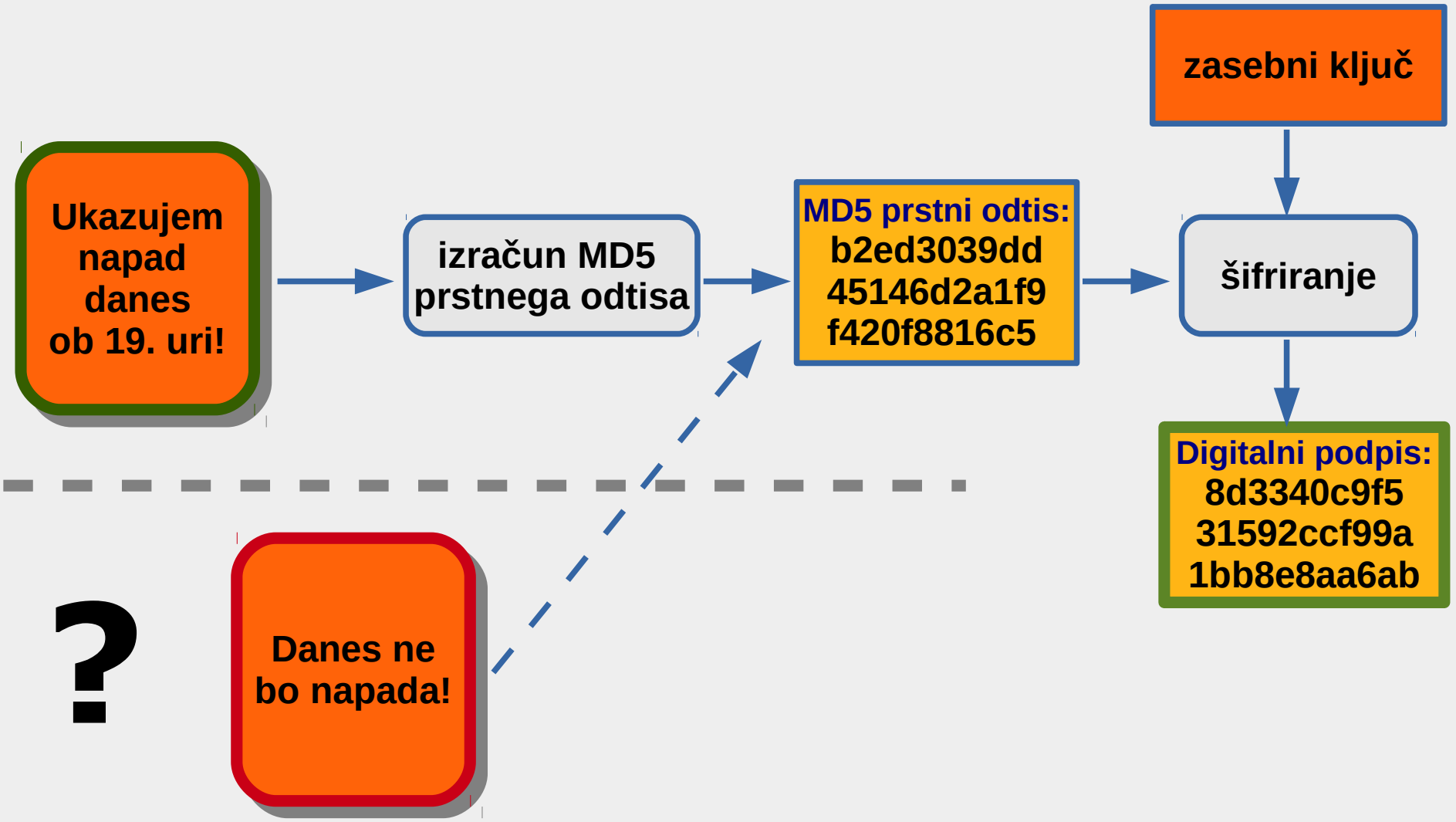
Leta 1996 je Dobbertin našel kolizijo v kompresijski funkciji MD5.

Leta 2004 so zagnali distribuirani projekt MD5CRK. Xiaoyun Wang, Dengguo Feng, Xuejia Lai in Hongbo Yu so pokazali, da je napad na MD5 mogoče izvesti v eni uri (na IBM p690 računalniški gruči).

Leta 2005 so Arjen Lenstra, Xiaoyun Wang in Benne de Weger prikazali izdelavo dveh X.509 certifikatov z različnimi javnimi ključi in isto MD5 zgoščeno vrednostjo.

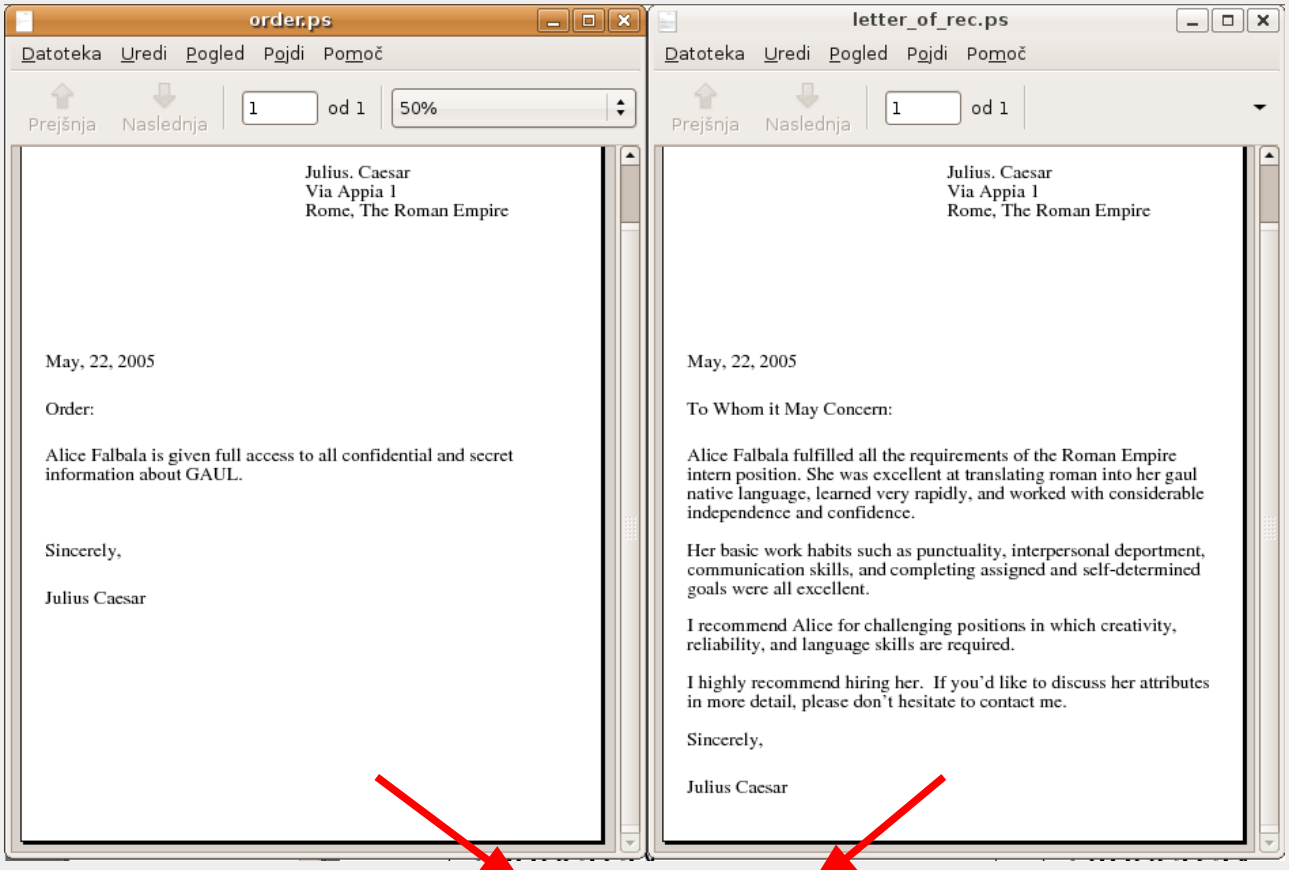
Leta 2006 je Vlastimil Klima objavil algoritem, ki je zmogel poiskati kolizijo na prenosniku v eni minuti.
Danes so znani napadi, ki omogočajo izračun kolizije na podlagi dveh poljubnih nizov vhodnih podatkov v nekaj urah.

Kolizija in digitalni podpis



[DEMO: dve PS datoteki]

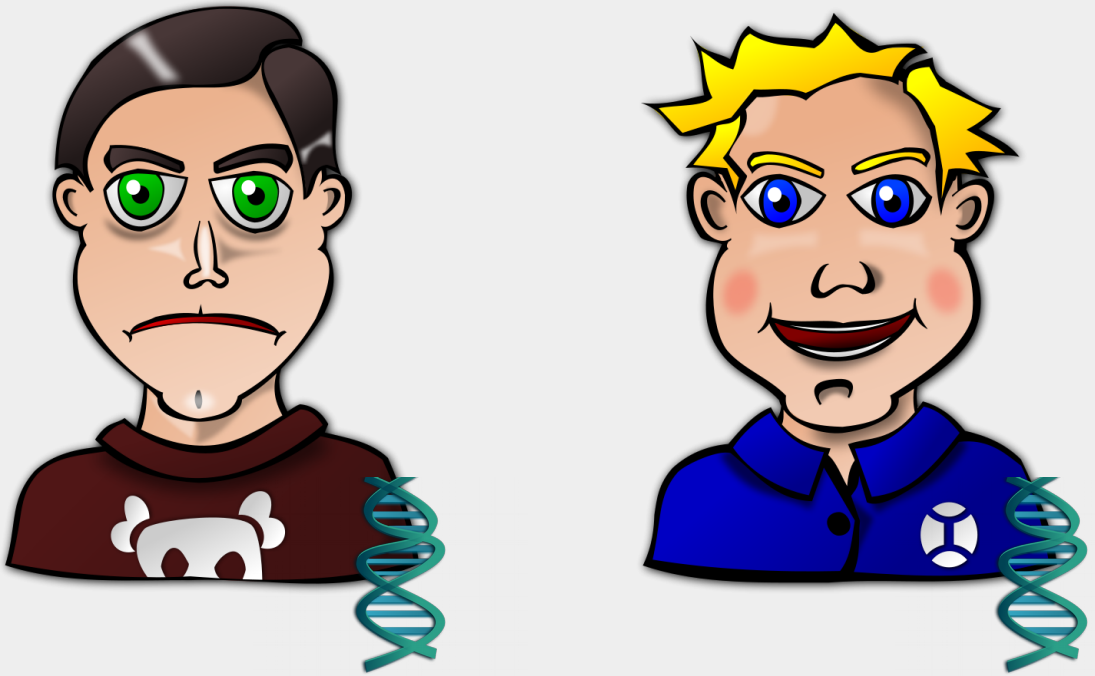
MD5 kolizija in digitalni podpis



MD5: 5421a523481fdc6a2a1c832e72c7b8a5

Vir: Magnus Daum in Stefan Lucks: The Story of Alice and her Boss: Hash Functions and the Blind Passenger Attack (Eurocrypt 2005, http://th.informatik.uni-mannheim.de/People/lucks/HashCollisions/rump_ec05.pdf).

MD5 kolizija in digitalni podpis



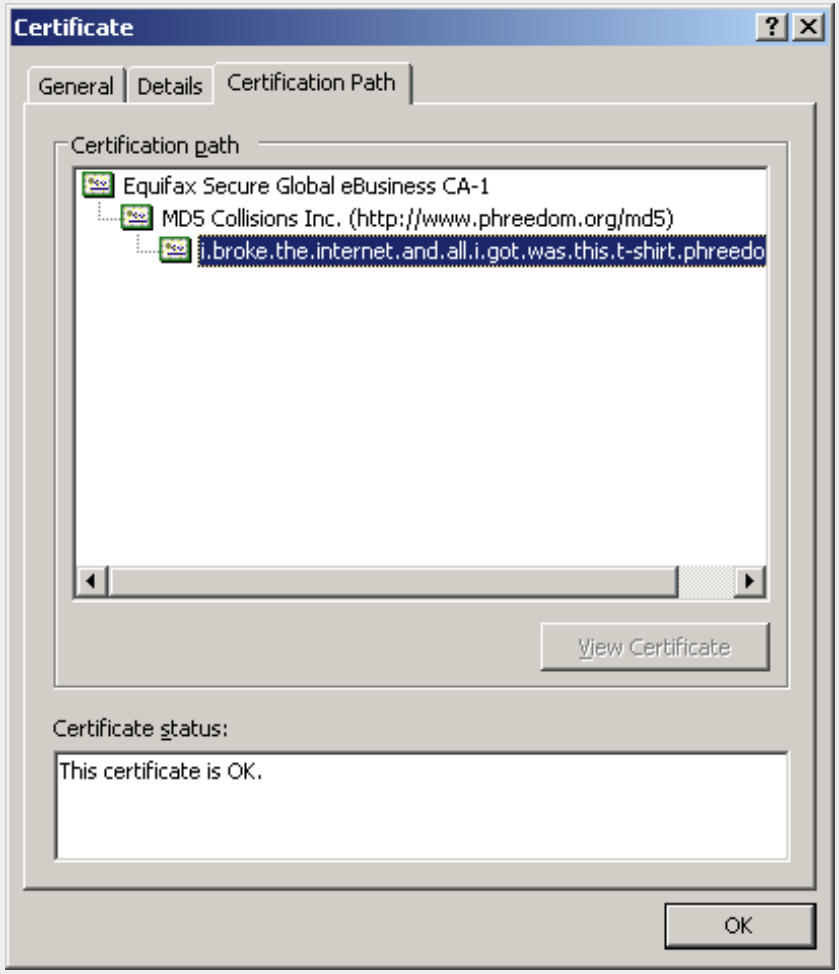
V primeru (MD5) kolizije gre za podobno situacijo, kot če bi imeli dve popolnoma **različni** osebi z **isto** DNK!

MD5 kolizija in digitalni certifikati

Alexander Sotirov, Marc Stevens, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, Benne de Weger, 2008.

"The most computationally intensive part of our method required about 3 days of work with over 200 game consoles, which is equivalent to 32 years of computing on a typical desktop computer."

Napad zahteva 1 do 2 dneva na gruči 200 PS 3 igralnih konzol, oziroma 8000 računalnikov z enojedrnim procesorjem, oziroma 20.000 USD na Amazon EC2.



MD5 in prepoznava / digitalno podpisovanje datotek

```
matej@cryptoloop:~$ md5sum hello
cdc47d67015geef60g16ca03agd4a007  hello

matej@cryptoloop:~$ ./hello
Hello, world!
(press enter to quit)q

matej@cryptoloop:~$ md5sum erase
cdc47d67015geef60g16ca03agd4a007  erase

matej@cryptoloop:~$ ./erase
This program is evil!!!
Erasing hard drive...1Gb...2Gb... just kidding!
Nothing was erased.
(press enter to quit)q
```

Didier Stevens je leta 2009 pokazal kako je mogoče dva različna programa podpisati z enakim Authenticode digitalnim podpisom, kar omogoča digitalno podpisovanje zlonamernih gonilnikov.

Napačna prepoznava virusov...

Viri: Peter Selinger, 2006, http://www.mscs.dal.ca/~selinger/md5collision/,
Didier Stevens, 2009, http://blog.didierstevens.com/2009/01/17/playing-with-authenticode-and-md5-collisions/

[DEMO: dve exe datoteki]

Problem kolizije v računalniški forenziki

Računalniška forenzika je proces identificiranja, zavarovanja, analiziranja in predstavljanja dokazov v elektronski obliki na način, ki je zakonsko sprejemljiv. To je postopek, ki se od običajnega, strogo tehničnega pregleda nosilcev dokaznega gradiva v elektronski obliki razlikuje v tem, da so na ta način pridobljeni dokazi **veljavni na sodišču**.

Cilj forenzičnega zasega je zagotoviti, da bodo zajeti podatki **ohranili integriteto** in s tem **dokazno vrednost** na sodišču.

Ključno je torej zagotavljanje integritete podatkov, s čimer se prepreči podtikanje ali nepooblaščeno brisanje dokazov.

Problem MD5 kolizije v računalniški forenziki

Problem kolizije ne vpliva (bistveno) na forenzične tehnike prepoznave znanih datotek...

... povsem drugače pa je pri **zagotavljanju integritete podatkov**.

Izračun kolizije nad spremenjenimi podatki

IHV	real certificate	rogue CA certificate
IHV0	0123456789ABCDEFFEDCBA9876543210	0123456789ABCDEFFEDCBA9876543210
IHV1	0584847F07A36382AAECF2DFE207A2	713F764E78B5C9B03F8878F7A440551B
IHV2	D52743425C3DAC23A9E62C6C9670622E	2AC9681DDB3B72D29A1422A515C9E4F4
IHV3	7789E58E3B45621A3E46A64CA9D7AC3A	104DD09F9F651E554C528578AC1F6885
IHV4	CDA2CB5673D3D3209 2C7F1EF80CE5729	15ADC95447929A2AC0EACF9E618E14EB
IHV5	F08E24604482508B959A0B5762207A3F	D6D6E59C0BDB1F701CB04C29A0573EA0
IHV6	A83EA6CCCC50B41A4BFADBC6D856B338	3AAB0CE98F1E9B2AC270A5A2C60FF605
IHV7	0B42EAAB4258AACA8C30BDB8192A1BC0	DE3CCC11526732CA0FD8B9F5992A7673
IHV8	D21CED8CC56726B6BF2AE4A93D742C3A	D21CED8CCE3D7EB4C82ADCA94674243A
IHV9	DC1EDBFFF3C3E9E7BCEB3F9E2D0705BD	DC1EDBFFF49941E8BDEB379E2E07FDBC
IHV10	F0D655805A71A74EF8A6A630D11977D8	F0D655805A479F4EF8A69E30D119 6FD8
IHV11	9808B5471E7130CC5A30A2ABF2BE4B4D	9808B5471E7130CC5A30A2ABF2BE4B4D
IHV12	AA1F57B21A8732130CB0CAEF4BB9C746	AA1F57B21A8732130CB0CAEF4BB9C746
IHV13	151754FA2FCC5914E72B71B4300B6485	151754FA2FCC5914E72B71B4300B6485
IHV14	271EECDC4DAC9E9C471C34C833917E26	271EECDC4DAC9E9C471C34C833917E26
IHV15	9ED7B966BD815C141B899DC64B528564	9ED7B966BD815C141B899DC64B528564
MD5	9ED7B966BD815C141B899DC64B528564	9ED7B966BD815C141B899DC64B528564

Vir: Alexander Sotirov, Marc Stevens, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, Benne de Weger, 2008, http://www.win.tue.nl/hashclash/rogue-ca/.

Cena?

Sotirov et. al.: leta 2008 je napad stal 20.000 USD na Amazon EC2.

Od novembra 2010 Amazon EC2 ponuja *Cluster GPU Instances*.

Ocena za leto 2012: izračun 150 milijonov MD5 na sekundo, ob kompleksnosti 2^{50} in ceni 2,1 USD na uro: ~4400 USD.

Thomas Roth, 2010, Cracking Passwords In The Cloud: Amazon's New EC2 GPU Instances, <http://stacksmashing.net/2010/11/15/cracking-in-the-cloud-amazons-new-ec2-gpu-instances/>.

Marc Stevens, Arjen Lenstra in Benne de Weger, 2007,
Chosen-prefix Collisions for MD5 and Colliding X.509 Certificates for Different Identities,
<http://www.win.tue.nl/hashclash/EC07v2.0.pdf>.

Posledice MD5 kolizije

Nekatere aplikacije za shranjevanje gesel v podatkovnih zbirkah še vedno uporabljajo MD5.

Problem prehoda iz MD5 na drugo funkcijo: gesla ne moremo dekodirati in ponovno zakodirati z novim algoritmom (saj ga nimamo).

Rešitev: *SHA over MD5*.

Za razliko od digitalne forenzike daljši izračun pri preverjanjih gesel ni pomemben, saj je podaljšanje časa za končnega uporabnika neopazno.

MD5 kolizija v računalniški forenziki

Ali so digitalne kopije pred leti zaseženih podatkov (npr. slike diskov osumljencev), katerih integriteta temelji na MD5 kontrolnih vsotah še verodostojne?

Ali so MD5 kontrolne vsote onkraj razumnega dvoma (ang. *beyond reasonable doubt*)?

Ostali napadi na MD5

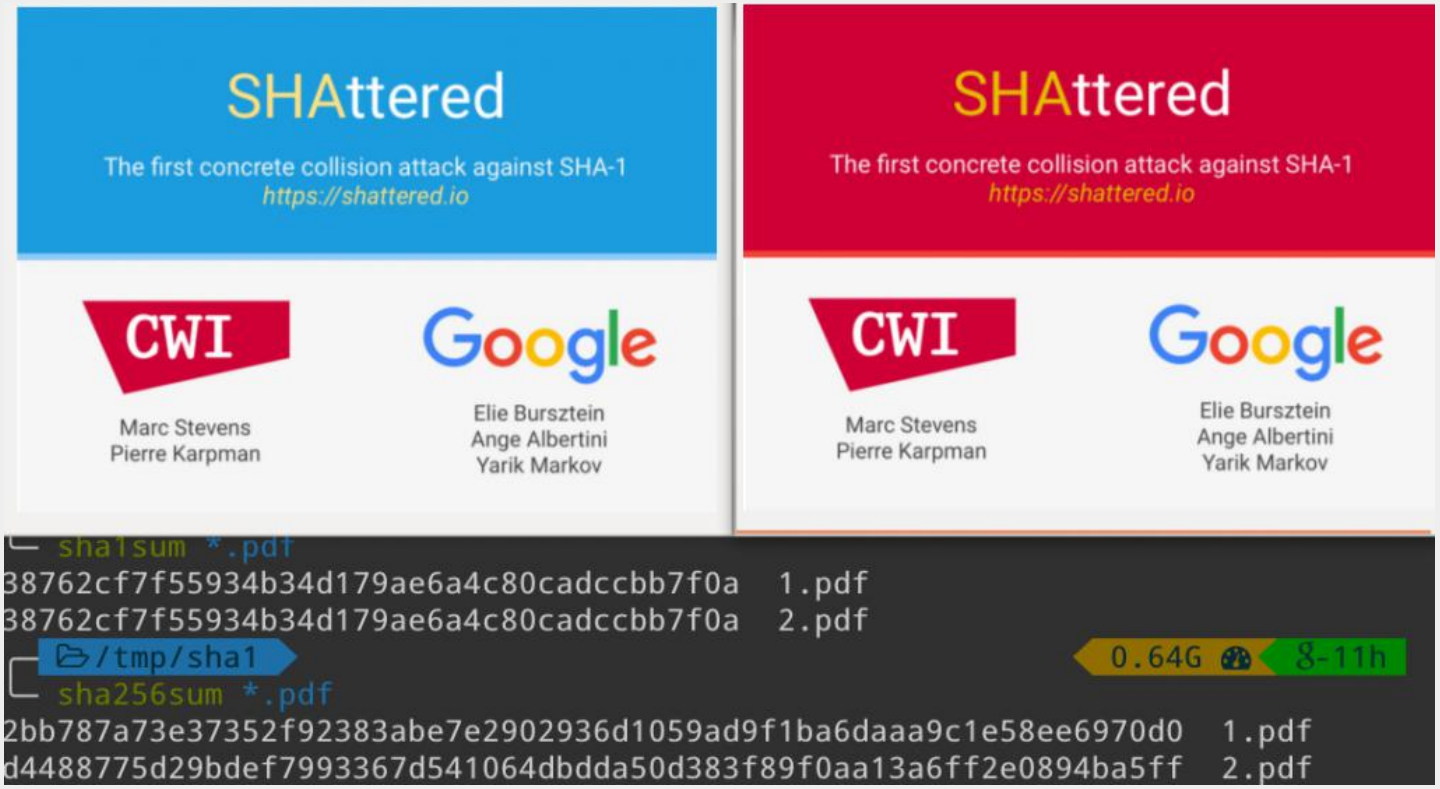
"Preimage napadi" (iskanje izvornega sporočila na podlagi zgoščenih vrednosti – napad na enosmernost funkcije):

Yu Sasaki, Kazumaro Aoki (2009-04-16). Finding Preimages in Full MD5 Faster Than Exhaustive Search. Springer Berlin Heidelberg. <http://www.springerlink.com/content/d7pm142n58853467>.

Mavrične tabele in predizračunane tabele zgoščenih vrednosti:

- Online Password Cracking na podlagi predizračunanih vrednosti: http://www.md5decrypter.co.uk/
- razbijanje SHA-1 zgoščenih vrednosti s CUDA-Multiforce (Amazon EC2): dolžina znakov od 1 do 6 v 49 minutah (cena: < 2,1 USD).

Kaj pa SHA-1?



SHA-0: opuščena zaradi resne pomanjkljivosti kmalu po uvedbi leta 1993.

SHA-1: teoretično kolizijo so našli leta 2005, praktično 2017.

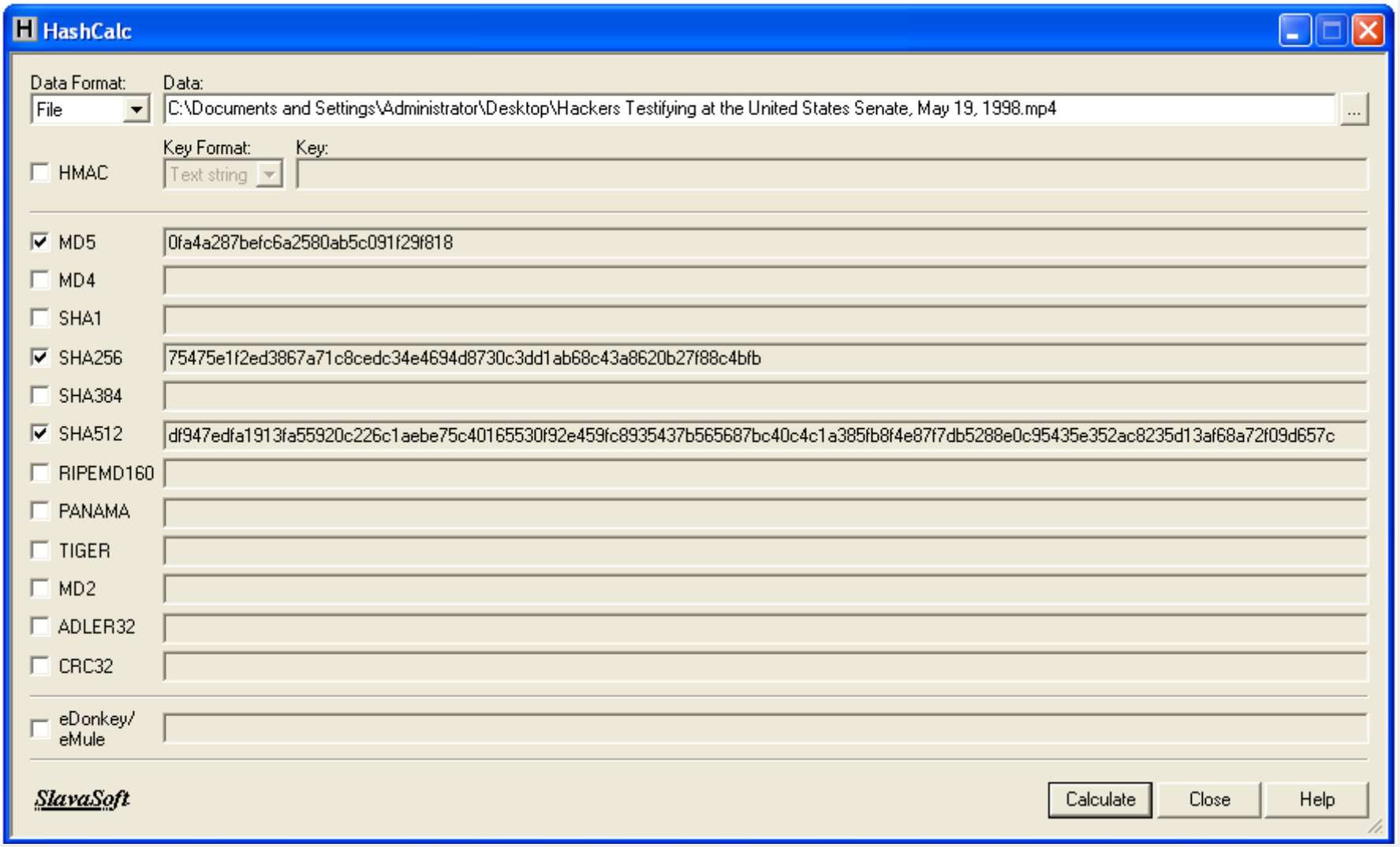
Rešitev?

Uporaba SHA-2 ali SHA-3 funkcij, opcijsko uporaba več različnih funkcij za zagotavljanje integritete podatkov:

- SHA-2 (SHA-224, SHA-256, SHA-384, SHA-512): kolizije še niso našli, a je izračun počasnejši kot pri MD5;
- SHA-3: od leta 2015, uporablja precej drugačno logiko kot MD5 in SHA-1 ter SHA-2, odporen pa je tudi na kvantne napade.

Večja varnost -> počasnejši izračun.

Rešitev?


HashCalc
Data Format:
File
Data:
C:\Documents and Settings\Administrator\Desktop\Hackers Testifying at the United States Senate, May 19, 1998.mp4
...
HMAC
Key Format:
Text string
Key:
MD5
0fa4a287befc6a2580ab5c091f29f818
MD4
SHA1
SHA256
75475e1f2ed3867a71c8cedc34e4694d8730c3dd1ab68c43a8620b27f88c4bfb
SHA384
SHA512
df947edfa1913fa55920c226c1aebe75c40165530f92e459fc8935437b565687bc40c4c1a385fb8f4e87f7db5288e0c95435e352ac8235d13af68a72f09d657c
RIPEMD160
PANAMA
TIGER
MD2
ADLER32
CRC32
eDonkey/
eMule
SlavaSoft
Calculate
Close
Help

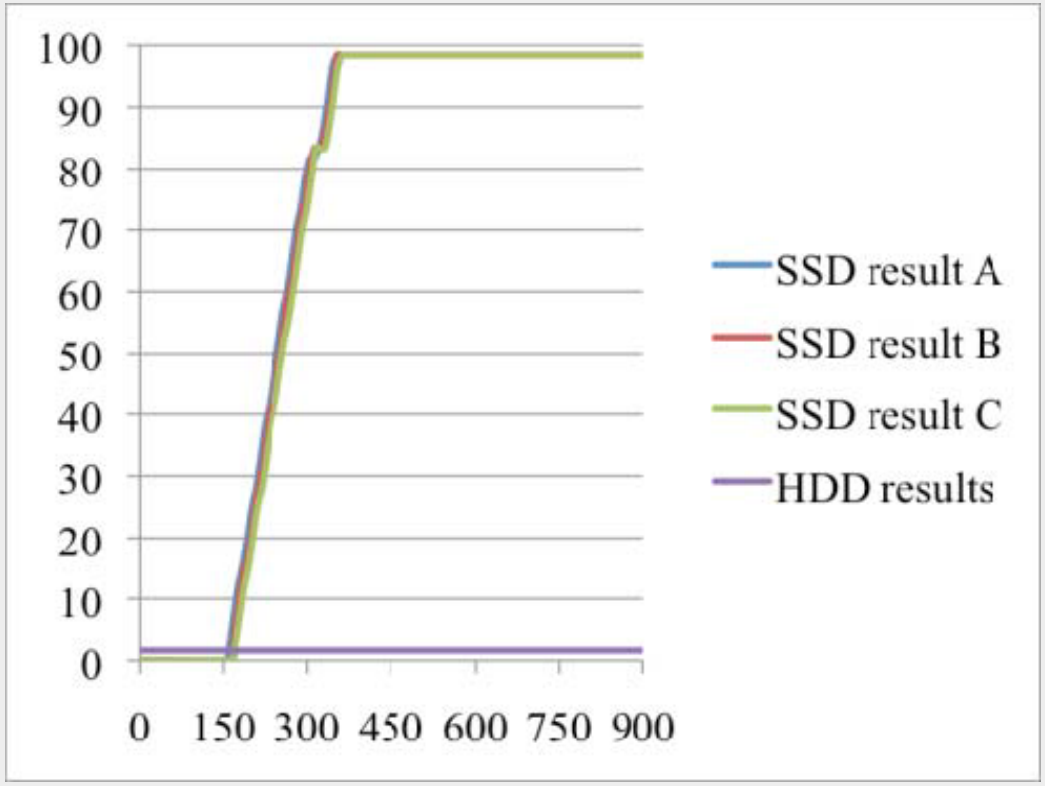
Vendar pa...

SSD diski, ki podpirajo funkcijo TRIM izbrisane podatke nepovratno uničijo – ob tim. hitrem formatiranju podatke uničijo v samo treh minutah, proces uničevanja podatkov pa poteka tudi v primeru, ko je bil disk priključen na blokator pisanja (tim. write blocker).

V primeru, ko je bil disk priključen na blokator pisanja, je bilo po 20 minutah nepovratno uničenih kar 19% izbrisanih datotek. Pri klasičnih trdih diskih je bilo mogoče obnoviti vse (izbrisane) datoteke, ne glede na pretečeni čas.

Graeme B. Bell in Richard Boddington. 2010. Solid State Drives: The Beginning of the End for Current Practice in Digital Forensic Discovery? V The Journal of Digital Forensics Security and Law. <http://ojs.jdfsl.org/index.php/jdfsl/article/viewFile/21/45>.

Vendar pa...



Na x-osi je čas, na y-osi pa delež uničenih podatkov. TRIM funkcija prične brisati podatke približno 200 sekund po zagonu diska. Po 510 sekundah so podatki skoraj povsem pobrisani. Trdi disk (HDD) podatkov ne briše.

Vprašanja...

Matej Kovačič
matej.kovacic@ijs.si

https://infosec-seminar.si

https://telefoncek.si