NWD 2009

DIAMANT

Discrete Interactive and Algorithmic Mathematics Algebra and Number Theory

A mathematics cluster funded by NWO OCW and EZ

Voorspellen met een PlayStation3

Benne de Weger [1])

b.m.m.d.weger@tue.nl
www.win.tue.nl/hashclash/Nostradamus/
www.win.tue.nl/hashclash/rogue-ca/

6 februari 2009

TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

[1]) in samenwerking met:
Marc Stevens (CWI), Arjen Lenstra (EPFL)

onze website met voorspelling

Nostradamus - Windows Internet Explorer

http://www.win.tue.nl/hashclash/Nostradamus/

File Edit View Favorites Tools Help

Nostradamus

Predicting the winner of the 2008 US Presidential Elections using a Sony PlayStation 3

November 30, 2007

Marc Stevens — CWI — Centrum voor Wiskunde en Informatica

Arjen Lenstra — EPFL École Polytechnique Fédérale de Lausanne — Alcatel·Lucent — EPFL and Bell Labs

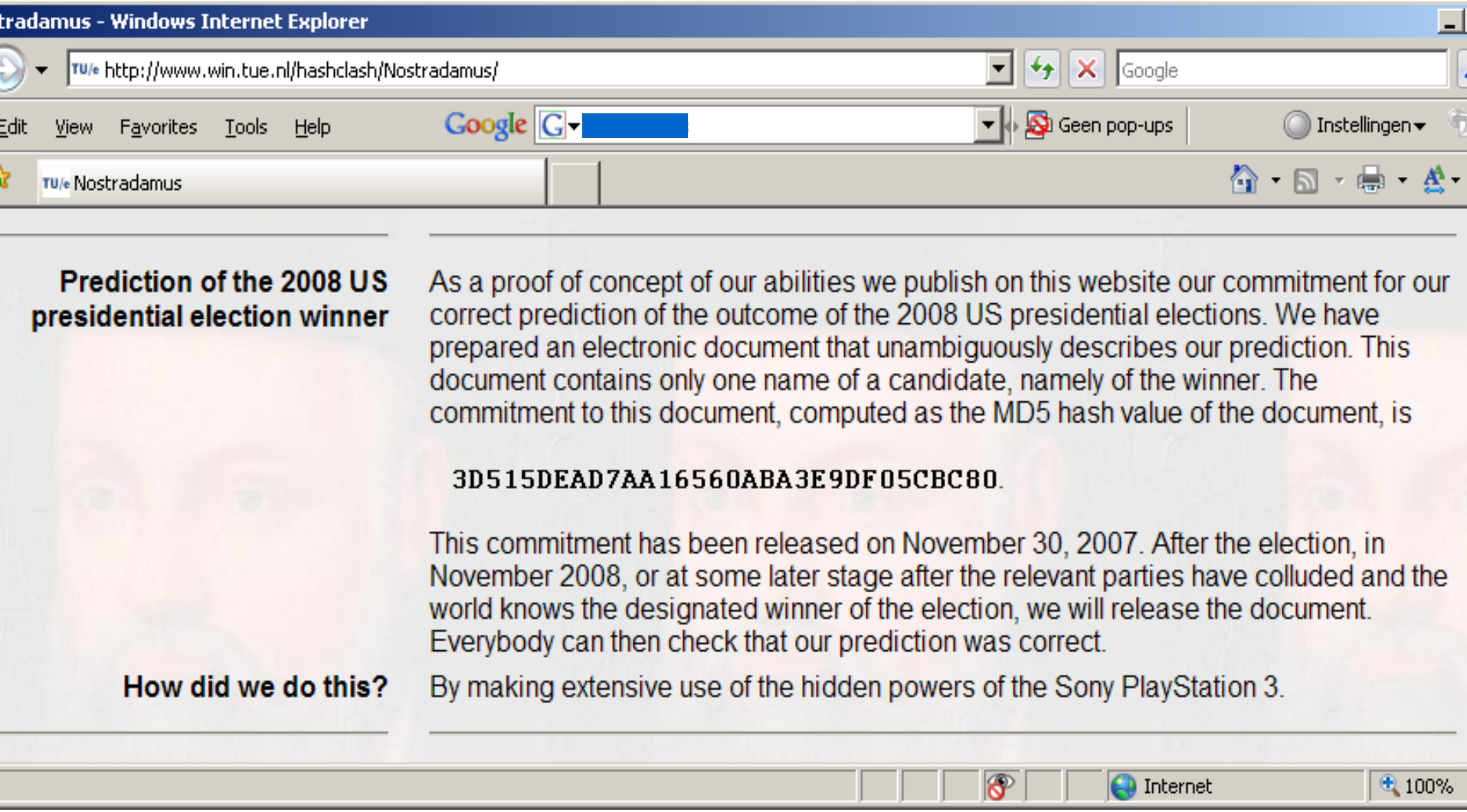
Benne de Weger — TU/e technische universiteit eindhoven

Announcement

We have used a Sony Playstation 3 to correctly predict the outcome of the 2008 US presidential elections. In order not to influence the voters we keep our prediction secret, but commit to it by publishing its cryptographic hash on this website. The document with the correct prediction and matching hash will be revealed after the elections.

Done — Internet — 100%

de “commitment”



het voorspellende document

TU/e technische universiteit eindhoven

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Prediction of the next President of the United States

Marc Stevens , Arjen Lenstra , and Benne de Weger

We predict that the winner of the 2008 election for President of the United States will be:

Barack Obama

commitment (verbintenis)

- op 30 november 2007 publiceerden wij de “MD5-hash” van het voorspellende document
  - commitment: bewijs dat wij dat document op dat moment in bezit hadden
- document zelf bleef geheim
- op 5 november 2008 was de winnaar bekend
  - wij konden het document openbaar maken

- iedereen kan controleren dat de MD5-hash klopt
  - en dat wij dus voorspellende gaven hebben
- wat klopt er hier niet?

wat is een hashfunctie, zoals MD5?

- digitale variant van *vingerafdruk*
- wat is een vingerafdruk?
  - middel om mensen te identificeren
  - uniek
    - zelfs één-eiïge tweelingen hebben verschillende vingerafdrukken
  - makkelijk te maken
    - van een mens een vingerafdruk maken is simpel
  - moeilijk te vervalsen
    - van een vingerafdruk een mens maken is moeilijk…
    - een mens vinden met jouw vingerafdruk is moeilijk…
    - twee mensen vinden met dezelfde vingerafdruk is moeilijk…

wat is een hashfunctie?

- een hashfunctie werkt op een rij bits *m* van willekeurige lengte
- de uitkomst *h*(*m*) is een rijtje bits van vaste lengte
- bijvoorbeeld:

“Nationale Wiskunde Dagen 2009, Noordwijkerhout”

heeft als bit-rij:

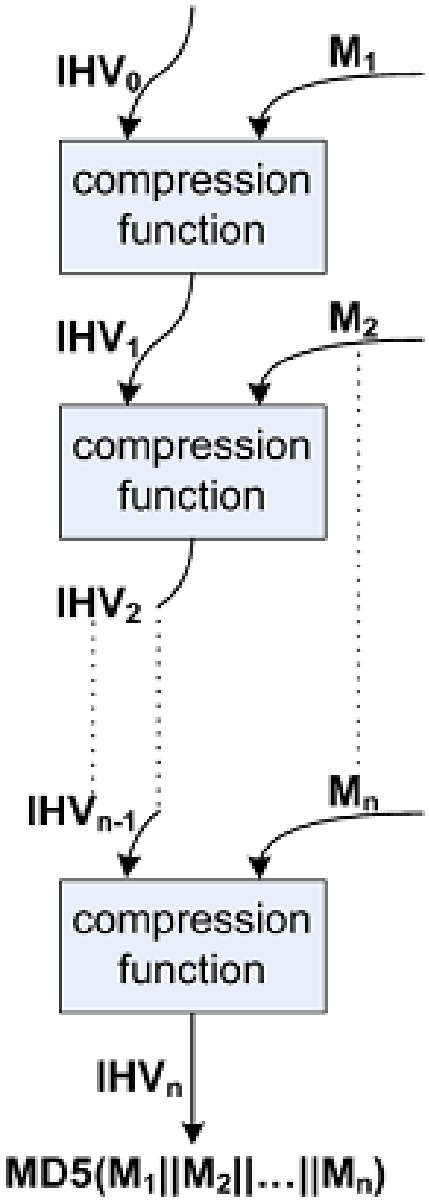
```
01001110 01100001 01110100 01101001 01101111 01101110 01100001
01101100 01100101 00100000 01010111 01101001 01110011 01101011
01110101 01101110 01100100 01100101 00100000 01000100 01100001
01100111 01100101 01101110 00100000 00110010 00110000 00110000
00111001 00101100 00100000 01001110 01101111 01101111 01110010
01100100 01110111 01101001 01101010 01101011 01100101 01110010
01101000 01101111 01110101 01110100
```

en als MD5-hash-waarde (vingerafdruk)

```
00011010 01000111 10000011 01010100 00101110 11101000 11101000
01010110 10010001 01010011 10001011 00100010 01001101 01010111
10110010 11100000
```

hexadecimaal: `5C B2 74 7B 53 B5 5D 5E 9C F3 0A 00 A7 F5 06 03`

hoe ziet zo'n hashfunctie eruit?



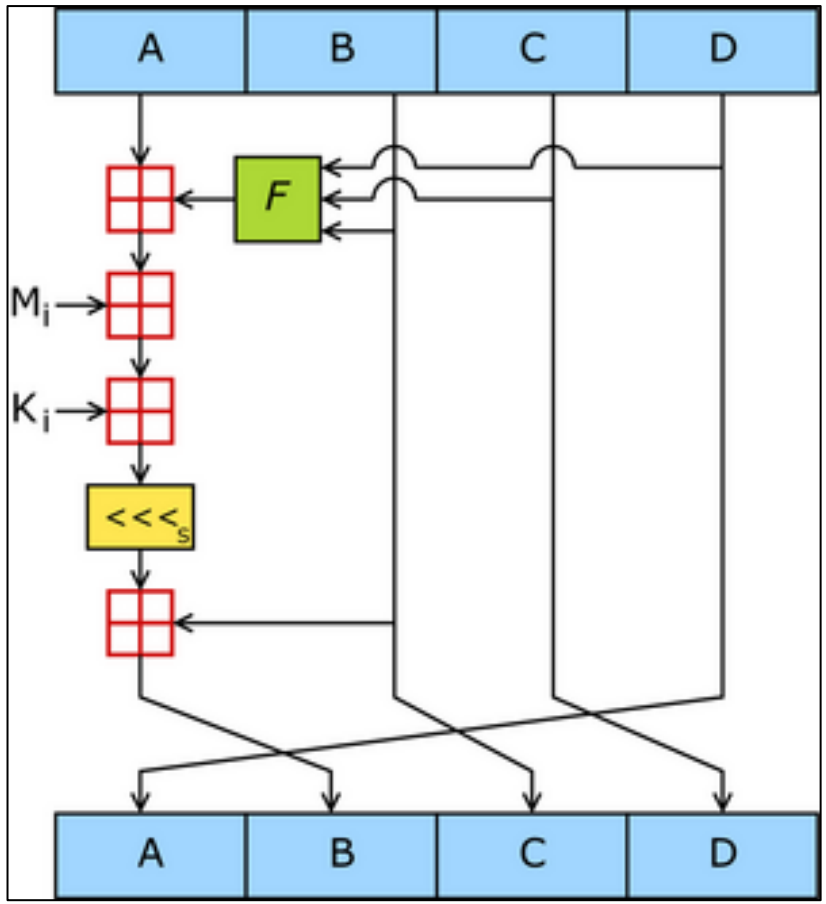
M = blok van 512 bits uit de invoer

IHV = A,B,C,D Intermediate Hash Value

- **IHV_0 is vast**
- **IHV_n is de hash**

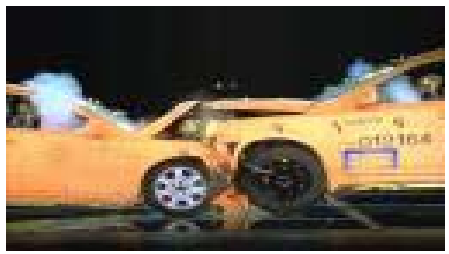
MD5 heeft 128 bits hash

compressiefunctie van MD5: 64 stappen als deze:



eisen voor een goede hashfunctie, lengte *n* bits

- **éénweg-functie:**
  - het is moeilijk om bij een gegeven uitkomst h_0 een invoer m te vinden zodat $h(m) = h_0$
  - het is moeilijk om bij een gegeven invoer m_0 een andere invoer m te vinden zodat $h(m) = h(m_0)$
  - hoe moeilijk? met gokken vind je een oplossing door ongeveer 2^n mogelijkheden af te lopen



- **botsing-bestendig**
  - het is moeilijk om twee verschillende invoeren m_1, m_2 te vinden zodat $h(m_1) = h(m_2)$
  - hoe moeilijk? met gokken vind je een oplossing door ongeveer $2^{n/2}$ mogelijkheden af te lopen

de verjaardagsparadox

- de kans dat in een schoolklas van 25 leerlingen er eentje op dezelfde dag jarig is als jij is ongeveer 7%

- de kans dat in die klas twee leerlingen op dezelfde dag jarig zijn is 57%

- daarom vind je al na $\approx 2^{n/2}$ keer gokken een botsing
- voor MD5 ($n = 64$) is dat $2^{64} \approx 1.84 \times 10^{19}$
  (begint op de grens te komen van wat nu mogelijk is)

- $2^{128} \approx 3.40 \times 10^{38}$ is voorlopig nog wel onmogelijk

toepassingen van hashfuncties

- beschermen van wachtwoorden
- controle op fouten bij downloaden
- om bestanden met dezelfde inhoud makkelijk terug te kunnen vinden
  - politie: doorzoeken van in beslag genomen computers
  - downloaden van films in peer to peer netwerken
- commitments
- digitale handtekeningen
- …

```
-----BEGIN PGP SIGNATURE-----
Version: PGP 6.5.3

iQA/AwUBOqAFfgWv6DmewW5PEQIKcQCg6POZ76PTMfnggkAB4Gc4VbnrjkcAoNPS
UNiTxN+4HXOkNMFkn7QsyncD
=qFzf
-----END PGP SIGNATURE-----
```

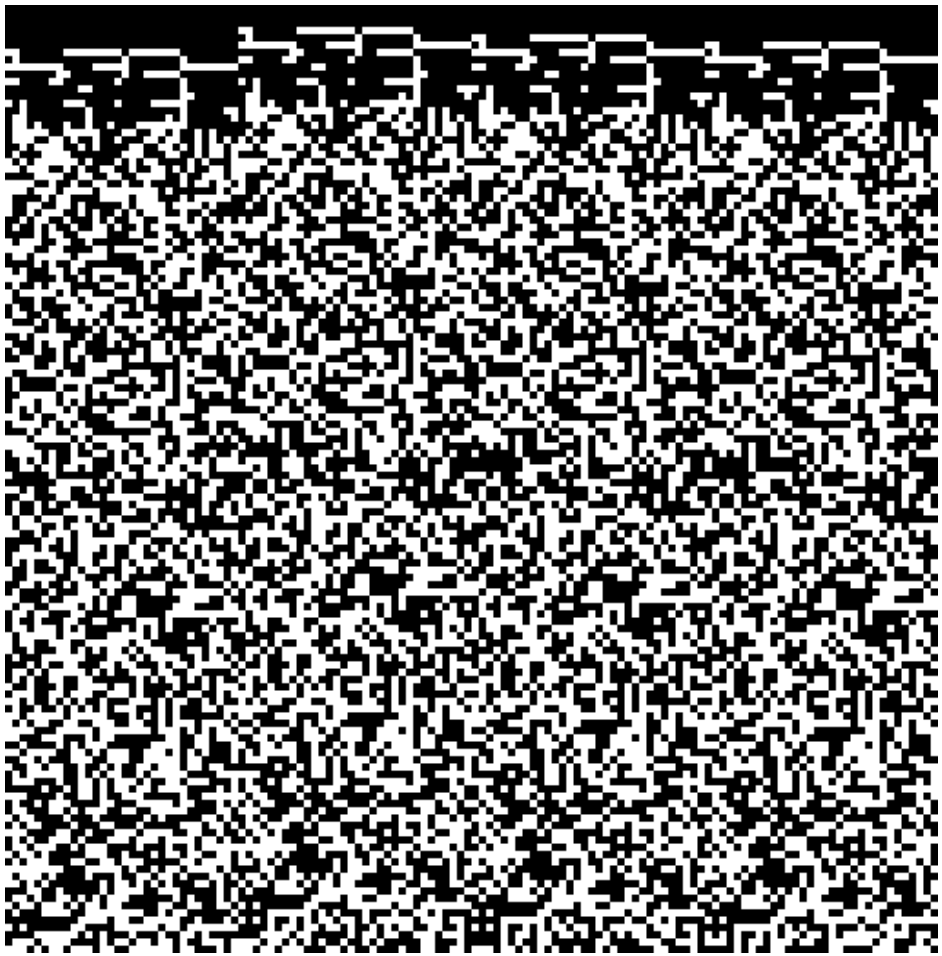
wat klopte er nou niet?

- MD5 is een veelgebruikte hashfunctie, maar…
- …MD5 is niet botsing-bestendig!
  - prof. Xiaoyun Wang (China) vond in 2004 botsingen voor MD5
    - rekentijd: uren op een supercomputer
  - Marc Stevens (Waalwijk) (afstudeerwerk 2007, TU/e):
    - verbeterde methode, helemaal geautomatiseerd
    - rekentijd: 30 seconden op een PC (inmiddels < 1 seconde)
    - ook: nieuw type botsingen waar je veel meer mee kunt

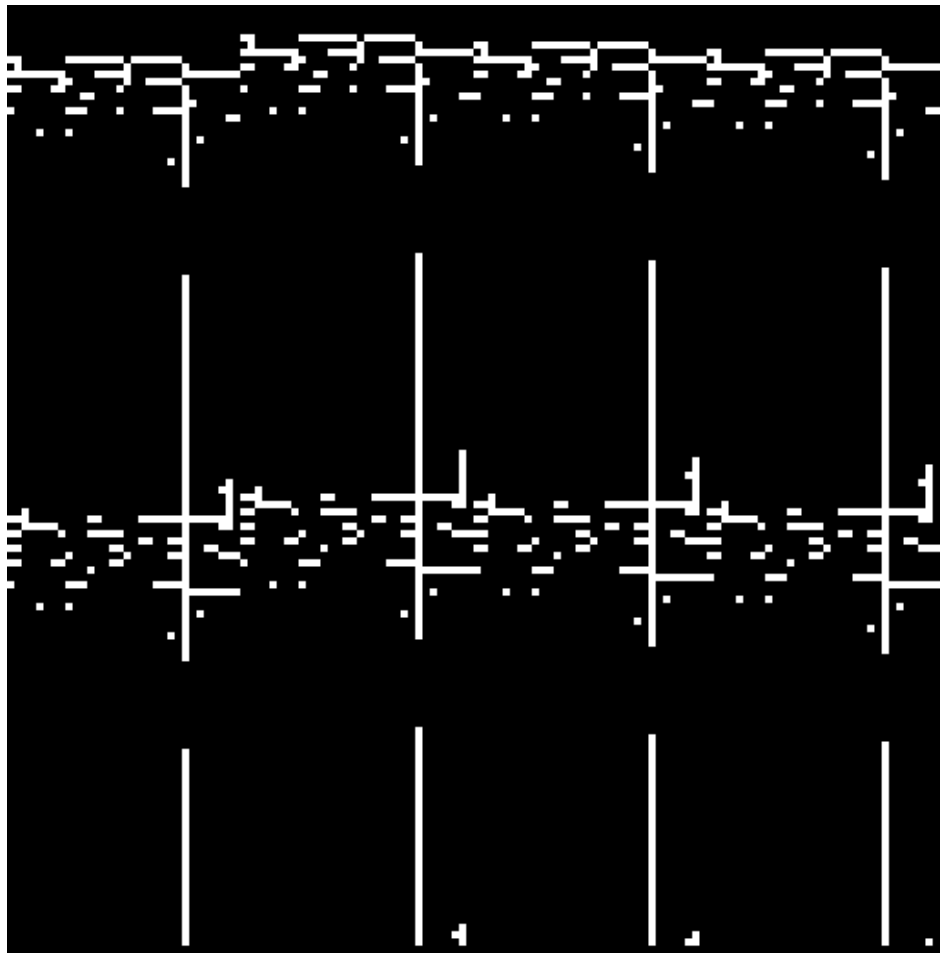
een kijkje in MD5: bit-verschillen in A,B,C,D

zo hoort het

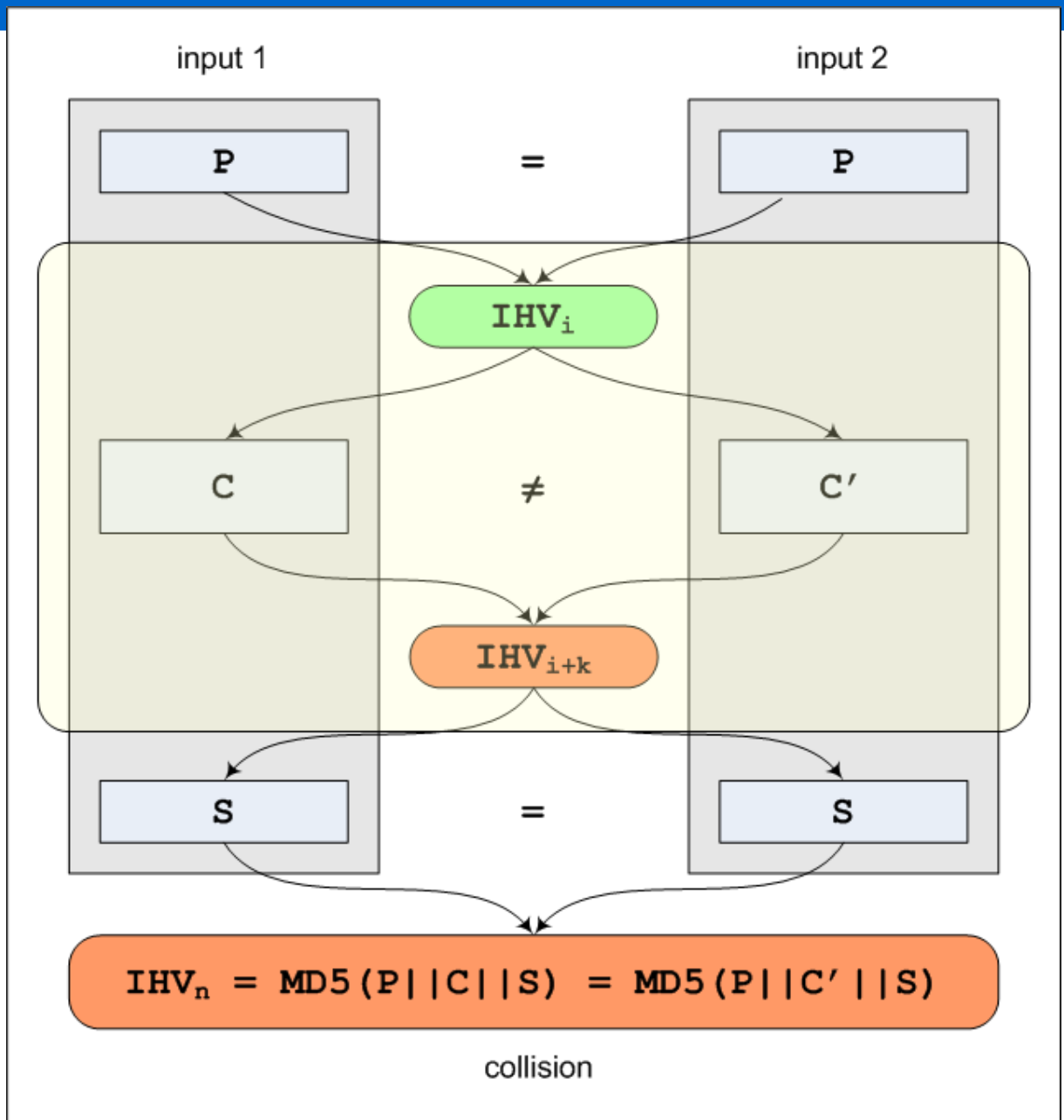
(1 bit verschil in de invoer)



een botsing



botsingen van Wang: identieke IHV



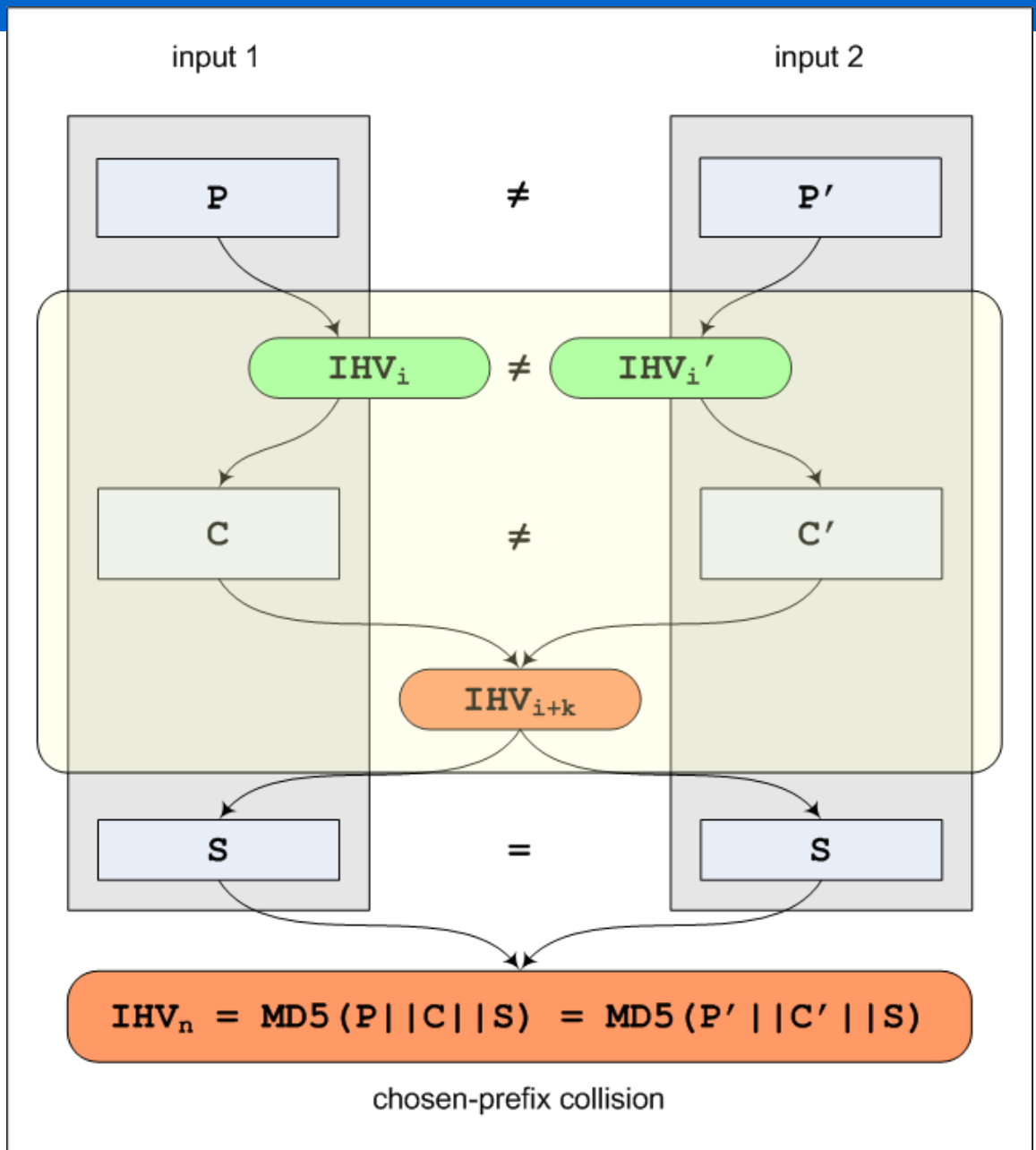
botsing ingebouwd in een document

identieke prefix P

verschillende botsing-blokken C, C’

identieke suffix S

botsingen van Stevens: verschillende IHV



botsing ingebouwd in een document

verschillende prefixen P, P'

verschillende botsing-blokken C, C'

identieke suffix S

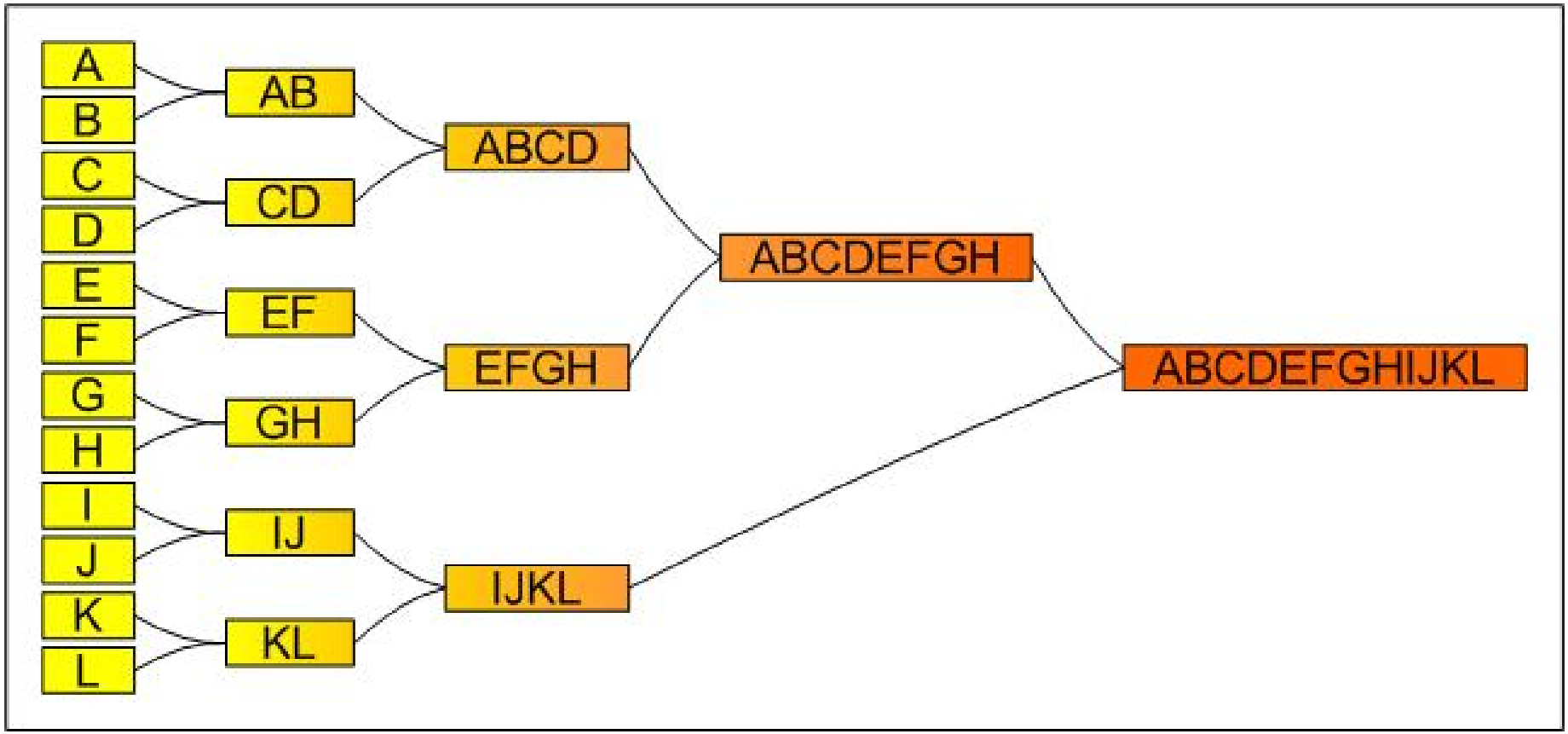
documenten laten botsen

- twee pdf-bestanden, zelf gekozen inhoud
- pdf bestandsformaat staat verborgen code toe
- echte inhoud van de documenten in de prefixen
- bereken de botsing-blokken
- de botsing-blokken in de verborgen code stoppen

- DIT WERKT!

- kan ook met Word, Postscript, …
- je kunt botsing-blokken bijv.
  ook in plaatjes verstoppen

de diamant: meer dan 2 documenten laten botsen



- **12 bestanden met dezelfde hash**

onze voorspelling was nep

us - Windows Internet Explorer

http://www.win.tue.nl/hashclash/Nostradamus/

Google

ew Favorites Tools Help Google Geen pop-ups Instellingen

Nostradamus

Our multi-collision (revealing our secret)

We have prepared twelve different predictions, ten of which are shown in the table below.

A:		John Edwards.pdf	G:		Fred Thompson.pdf
B:		John McCain.pdf	H:		(hidden)
C:		Mitt Romney.pdf	I:		Paris Hilton.pdf
D:		Ralph Nader.pdf	J:		Al Gore.pdf
E:		(hidden)	K:		Jeb Bush.pdf
F:		Barack Obama.pdf	L:		Oprah Winfrey.pdf

All twelve documents we prepared, the ten given above and two hidden ones, have the MD5 hash value

3D515DEAD7AA16560ABA3E9DF05CBC80.

maar we hebben niet gelogen:

we hadden een document met de goede voorspelling

en de goede MD5-hash!

Internet 100%

onze echte voorspelling…

Our real prediction 3D515DEAD7AA16560ABA3E9DF05CBC80.

rekentijd

- botsingen van Wang-type: 2^{16} berekeningen
  < 1 seconde op een gewone PC
- botsingen van Stevens-type: 2^{42} berekeningen
  enkele uren op een PlayStation3
  (of een moderne grafische kaart)

1 PlayStation3
(optimaal geprogrammeerd)
is ongeveer even krachtig
als 40 snelle PCs

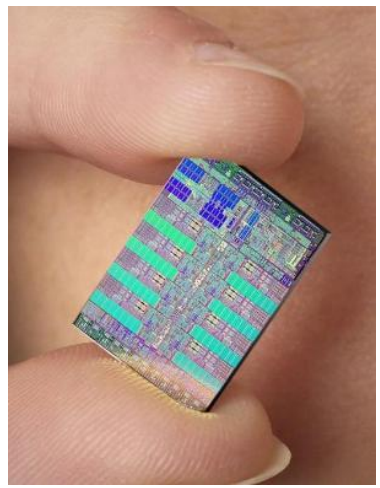
hoe we het tegenwoordig doen…

PlayStation3 Lab van prof. Arjen Lenstra (EPFL, Lausanne, Zwitserland)

cluster van 215 PlayStation3s

waarom PlayStation3s?

- processor van gewone PC:
  - moet veel kunnen, heeft grote control unit en relatief kleine rekenunit
- cell-processor van PlayStation3:
  - kleine instructieset, dus kleine control unit
  - heeft 6-8 sterke rekenunits voeren dezelfde instructie tegelijk uit op verschillende data
- rekenwerk voor botsingen:
  - eenvoudige instructies
  - dezelfde instructies uitvoeren op veel verschillende data
  - ideaal voor de PlayStation3
- de PlayStation3 is open en volledig programmeerbaar



deze week in Newsweek

Attack of the Vigilante Cryptos

The group, unofficially known as MD5 Collision Inc., demonstrated that they'd discovered a way to fabricate rogue certificates copied from the legitimate one they had bought from VeriSign, the world's biggest issuer. A cybercriminal holding a rogue certificate could convince any browser that a fake Web site is authentic.

digitale handtekening

- digitale handtekening onder een document
  - berekend met een geheime sleutel (alleen de eigenaar van de sleutel kan dat)
  - handtekening is te controleren met een publieke sleutel
- handtekening wordt berekend over een hash van het document

- hash-botsing is nu gevaarlijk:
  - twee documenten met dezelfde handtekening

certificaat

- soort digitaal paspoort
- bijvoorbeeld voor websites
  (zoals voor internetbankieren)
  - te zien aan het slotje in de browser
- bevat een digitale handtekening van een CA – Certification Authority – paspoort-uitgever
- je browser controleert zo'n handtekening aan de hand van een lijst vertrouwde paspoort-uitgevers

- er was nog een grote CA die MD5 gebruikte…

een valse CA maken

- die CA een website-certificaat laten ondertekenen (kost $65)
  - daarin hadden we een botsing-blok verstopt
- wij hadden een tweede certificaat, met dezelfde MD5-hash
  - met “CA = TRUE”
- de handtekening van de echte CA konden wij kopiëren van het ene in het andere, valse certificaat
- wij hebben nu zelf een valse paspoort-uitgeverij
  - die wordt door iedere browser automatisch vertrouwd
- wij kunnen in principe iedere website nabouwen en een vertrouwd certificaat ervoor maken
  - valt de bezoeker niet op
- potentieel heel gevaarlijk
  - bijvoorbeeld voor phishing

- genoemde CA gebruikt inmiddels MD5 niet meer

conclusie

- MD5 is niet meer te vertrouwen
  - al sinds 2004 niet meer
  - besef begint nu eindelijk door te dringen
- gelukkig zijn er alternatieven
  - SHA-1: heeft in theorie ook problemen maar is de enige die alle browsers aankunnen
  - SHA-2: voor de nabije toekomst de beste keuze
  - SHA-3: wordt de nieuwe standaard voor hashfuncties
    - competitie van de National Institute of Standards and Technology (USA) is nu bezig
    - winnaar verwacht in 2012