MD5 Chosen-Prefix Collisions on GPUs

Marc Bevand
m.bevand@gmail.com
marc.bevand@rapid7.com

Agenda

- MD5 on GPUs
- Dec 2008: rogue CA certificate on PS3 cluster
- MD5 birthday search
- Results & performance

MD5 on GPUs

- MD5 is optimized for 32-bit architectures
- 32-bit integer & logical instructions
- GPGPU tech makes it possible to run arbitrary code
- GPUs are massively parallel chips with lots of ALUs

MD5 on GPUs (cont'd)

- Let me repeat: ”massively parallel”
- As in hundreds of instructions per clock
- Why isn't everybody doing GPGPU ?! Lack of awareness

Why ATI GPUs (cont'd)

- ATI R700 GPU family (Radeon HD 4000 series):
  - Up to 800 Stream Processing Units per ASIC
  - Clocked up to 850 Mhz
  - Dual-GPU video cards
- Best perf/W and perf/$ (May 2009): HD 4850 X2
  - 2nd fastest video card in the world
  - 1 trillion 32-bit instructions/sec (2 TFLOPS)
  - TDP 230W, Price US$250
- Can't wait to see next-gen R800

Why not Nvidia

- Top-of-the-line member of the Nvidia GT200 GPU family: GTX 295
  - 596 billion 32-bit instructions/sec
  - TDP 290W, Price US$500
- Raw perf/W and perf/$ respectively roughly 2 times and 4 times worse than HD 4850 X2
- However Nvidia CUDA SDK is more mature
- Next-gen GT300 will be better ?

Rogue CA

- When: Dec 2008, paper published in Mar 2009
- Where: 25th Chaos Communication Congress (25C3)
- Who: 7 researchers (Sotirov, Stevens, Applebaum, Lenstra, Molnar, Osvik, Weger)
- What: implemented an MD5 chosen-prefix collision attack on a cluster of 215 PlayStation 3s to create a rogue CA

Rogue CA (cont'd)

- Simplified explanation:
  - Create cert ”A” and rogue CA cert ”B” with same MD5 hash
  - Get a CA to sign a cert signing request that end up producing cert A
  - Steal A's signature and apply it to B
- How to generate A and B with same MD5 hash:
  - ”Birthdaying” stage ← most computing intensive part
  - ”Near collision” stage

MD5 ”Birthdaying”

- We have 2 ”chosen-prefix” bitstrings (certs)
- When processed through MD5, lead to 2 different MD5 states (8 32-bit variables):
  - A, B, C, D
  - A', B', C', D'
- Goal of birthdaying is to append a small number of bits to find a state such as the 8 variables satisfy some conditions (see Mar 2009 paper)

MD5 ”Birthdaying” (cont'd)

- Technique to find these conditions: deterministic pseudo-random walk in search space using Pollard-Rho method
- Same concept as a rainbow table chain ”walking” through the search space except we are looking for collisions !
- Basically this search consists of running the MD5 compression function over and over
- [TODO: schema]

MD5 CAL IL Implementation

- Therefore to optimize the attack, a fast MD5 implementation had to be developed
- Hand-coded one in CAL IL (Compute Abstract Layer Intermediary Language) – a pseudo-assembly language for ATI GPUs

MD5 in CAL IL

- ”CAL IL”: looks as bad as it sounds :)

```
mov r9, l101.zzzz
mov r10, l101.wwww

mov r4.x, l100.x
whileloop
  break_logicalz r4.x
  mov r0, r7
  mov r1, r8
  mov r2, r9
  mov r3, r10

  ixor r5, r0, r1
  and  r5, r5, r3
  ixor r5, r5, r1
  iadd r6, cb0[0].zzzz, cb1[0].zzzz
  iadd r5, r2, r5
  iadd r5, r5, r6
  ushr r6, r5, l1.xxxx
  umad r5, r5, l1.yyyy, r6
  iadd r2, r3, r5

  ixor r5, r3, r0
  and  r5, r5, r2
```

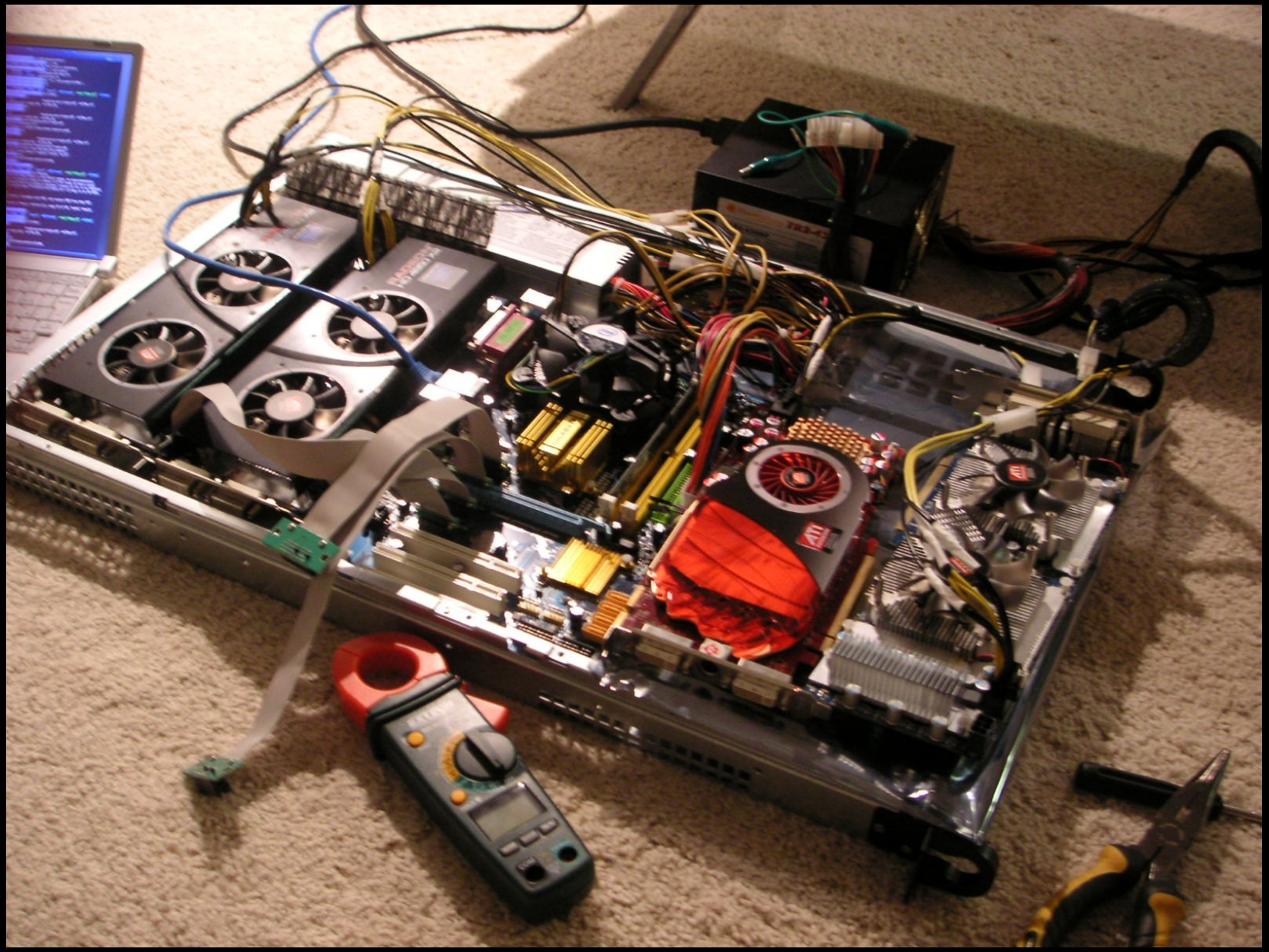
Performance

- 1634 Mhash/sec on HD 4850 X2 (1.6 billion MD5 compression function calls per second) – IOW MD5 processes 105 GByte/s
- Possible future optimization: due to a particularity of the birthday search, the first 14 out of 64 steps of the compression function can be pre-computed – should allow 2090 Mhash/sec

Theoretical GPGPU cracking server

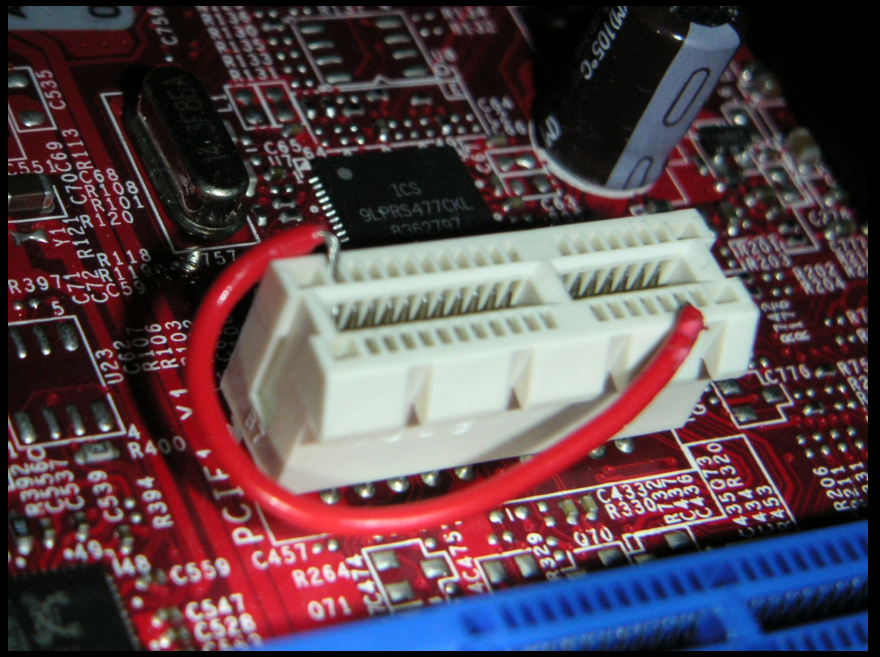
- 4 Radeon HD 4850 X2 in a single machine
- 8 GPUs total
- About US$1500
- Power draw: 950 W from the wall
- Total of 6536 Mhash/s

Here it is



HW Implementation Details

- QEMU/KVM PCI passthrough feature to work around ATI's fglrx.ko driver limitation of 4 GPUs
- Flexible cut-out PCI-Express extenders to down-plug x16 cards on cheap motherboards with x1 slots
- Undocumented secret: short pins A1 & B17 to work around down-plugging compatibility issues



Comparison with PS3 cluster

- 215 PS3s:
  - 28 kW (130 W each)
  - US$86k (US$400 each)
  - 37600 Mhash/s (175 Mhash/s each)
- 6 GPGPU servers:
  - 5.7 kW (950 W each) – 5 times less power
  - US$9k (US$1500 each) – 10 times cheaper
  - 39200 Mhash/s (6536 Mhash/s each) – and a bit faster

Conclusion

- Another blow to MD5 – chosen-prefix collision attack now practical for anybody
- Public CAs have stopped signing with MD5 – what about private/corporate CAs ?
- If a workload can run on GPUs, do it. They are a commodity and so efficient that considering anything else does not make sense.
- Code & tools will be open-sourced on the project page: [TBD]