Breaking Encryptions In The Cloud

GPU-accelerated supercomputing for everyone

Thomas Roth
BlackHat DC 2011

About The Speaker

- Thomas Roth
- Security and software engineering at Lanworks AG

- Blog: http://stacksmashing.net/
- Twitter: @stacksmashing
- E-Mail: input@stacksmashing.net

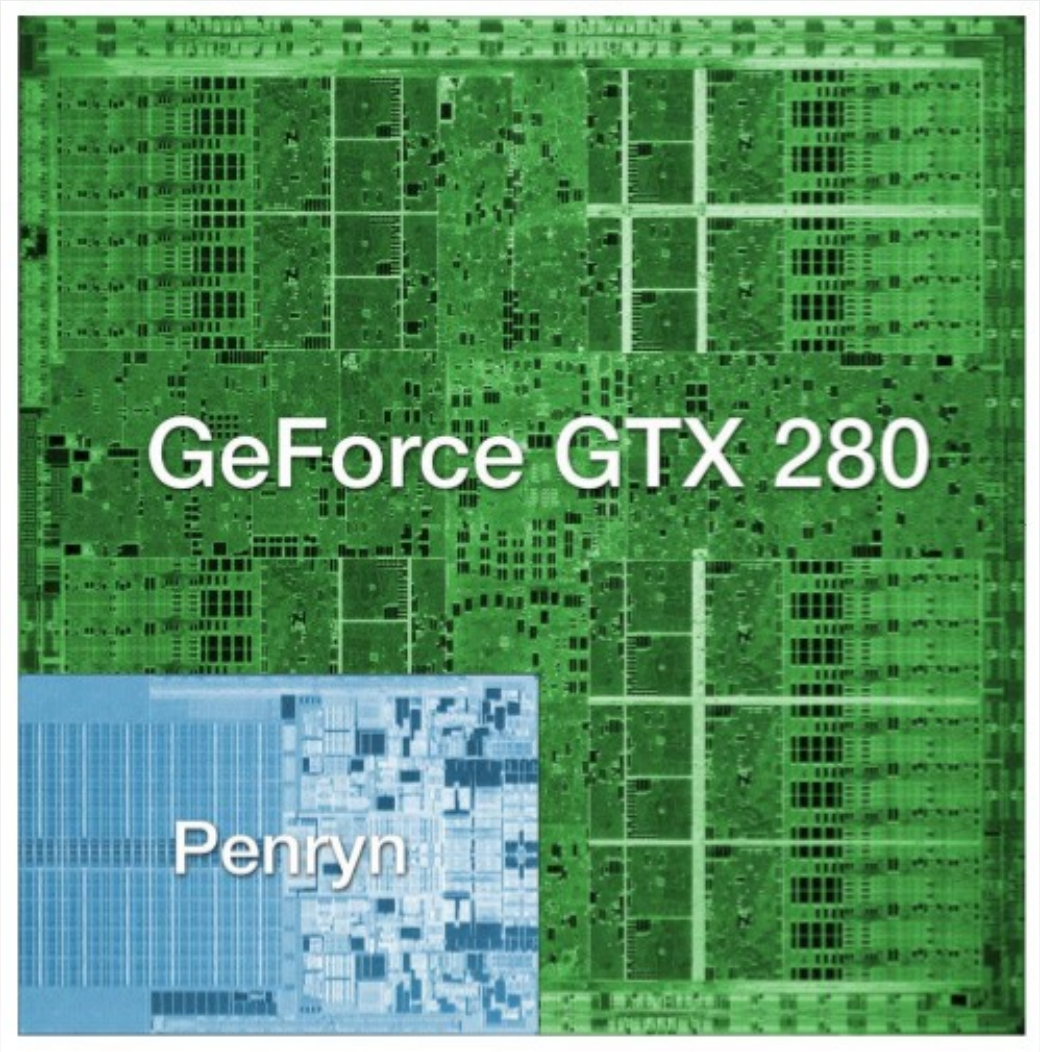
Table Of Contents

- An introduction into GPU computing
- About “the cloud”
- Introducing the “cloud cracking suite”
- Questions and answers

GPU Computing

NVIDIA GTX 480 Graphic Card

GPU Computing: Architecture

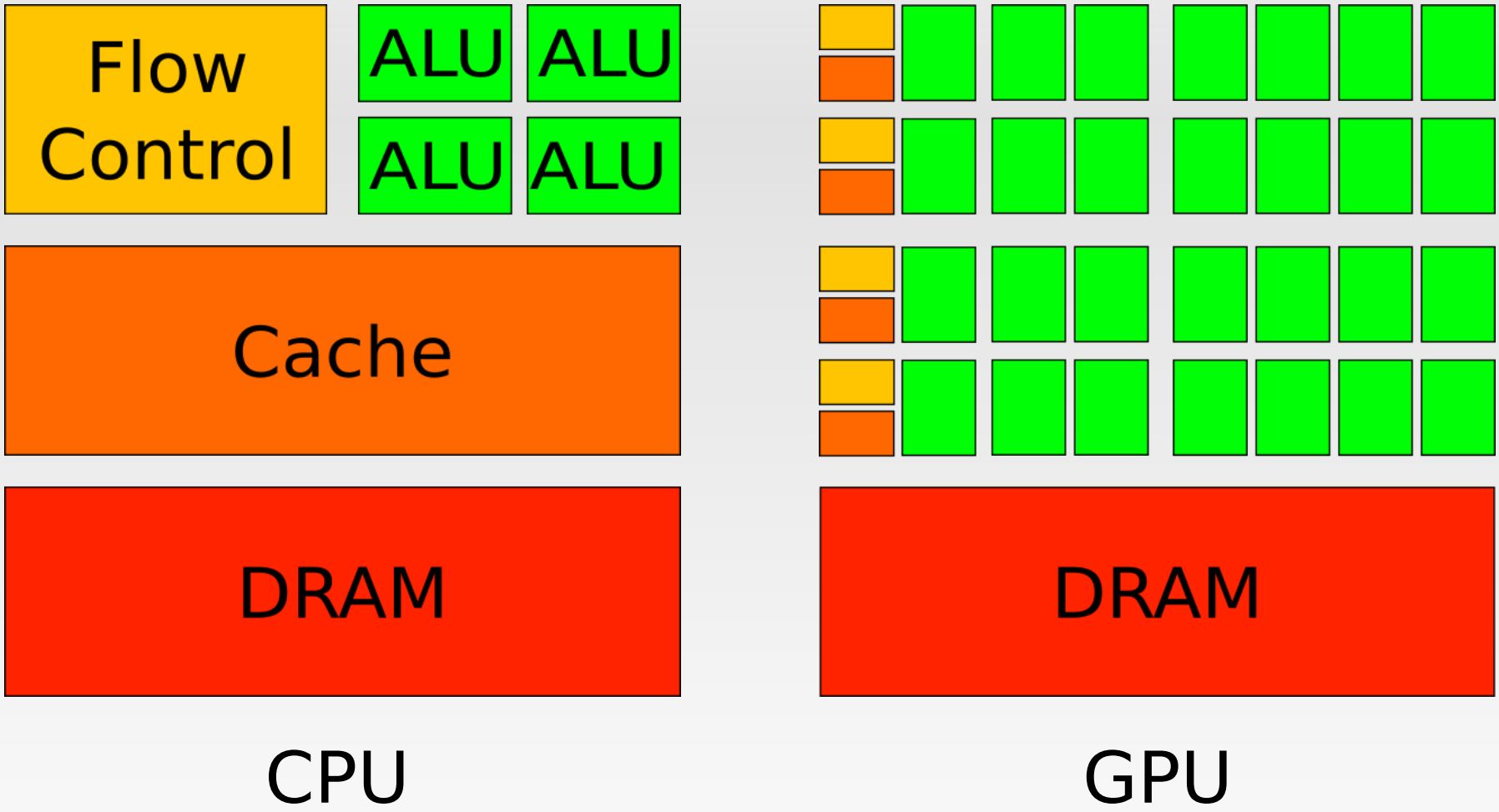


http://www.anandtech.com/show/2549

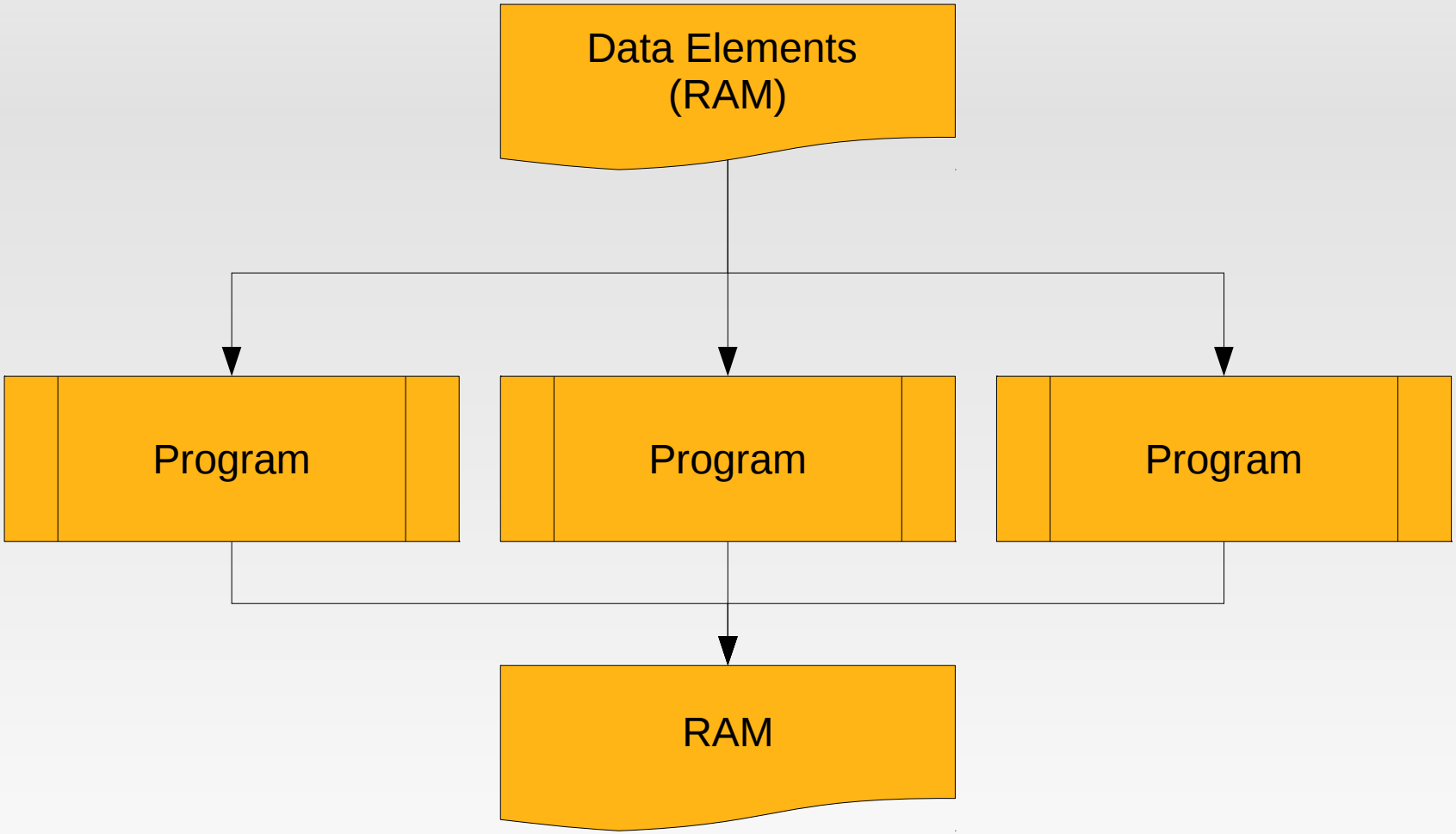
GPU Computing: Architecture

- Modern Graphic Processing Units
  - Highly parallel architecture
    - (> 400 cores)
  - High memory bandwidth
    - (> 170 GB/s)
  - Relatively low power consumption

GPU Computing: Architecture



GPU Computing



GPU Computing

- GPU Computing Frameworks
  - NVIDIA CUDA
  - Khronos OpenCL (Computing Language)
  - Microsoft DirectCompute

GPU Computing: Programming

- NVIDIA “C for CUDA”:
  - “Computer Unified Device Architecture”
  - “nvcc” compiler
  - Separates Host code (CPU) from CUDA code (GPU)
  - Host has to care about Host/GPU memory management

GPU Computing: Programming

- Kernels:
  - Functions that run on GPUs are called kernels
  - Must be callable from *N* threads in any order to ensure scalability for future device generations

GPU Computing: Programming

- Kernels are called from Threads
- Threads are within Blocks
- Blocks are withing Grids
- Several memory spaces:
  - Per-thread local memory
  - Per-block local memory
  - Global memory

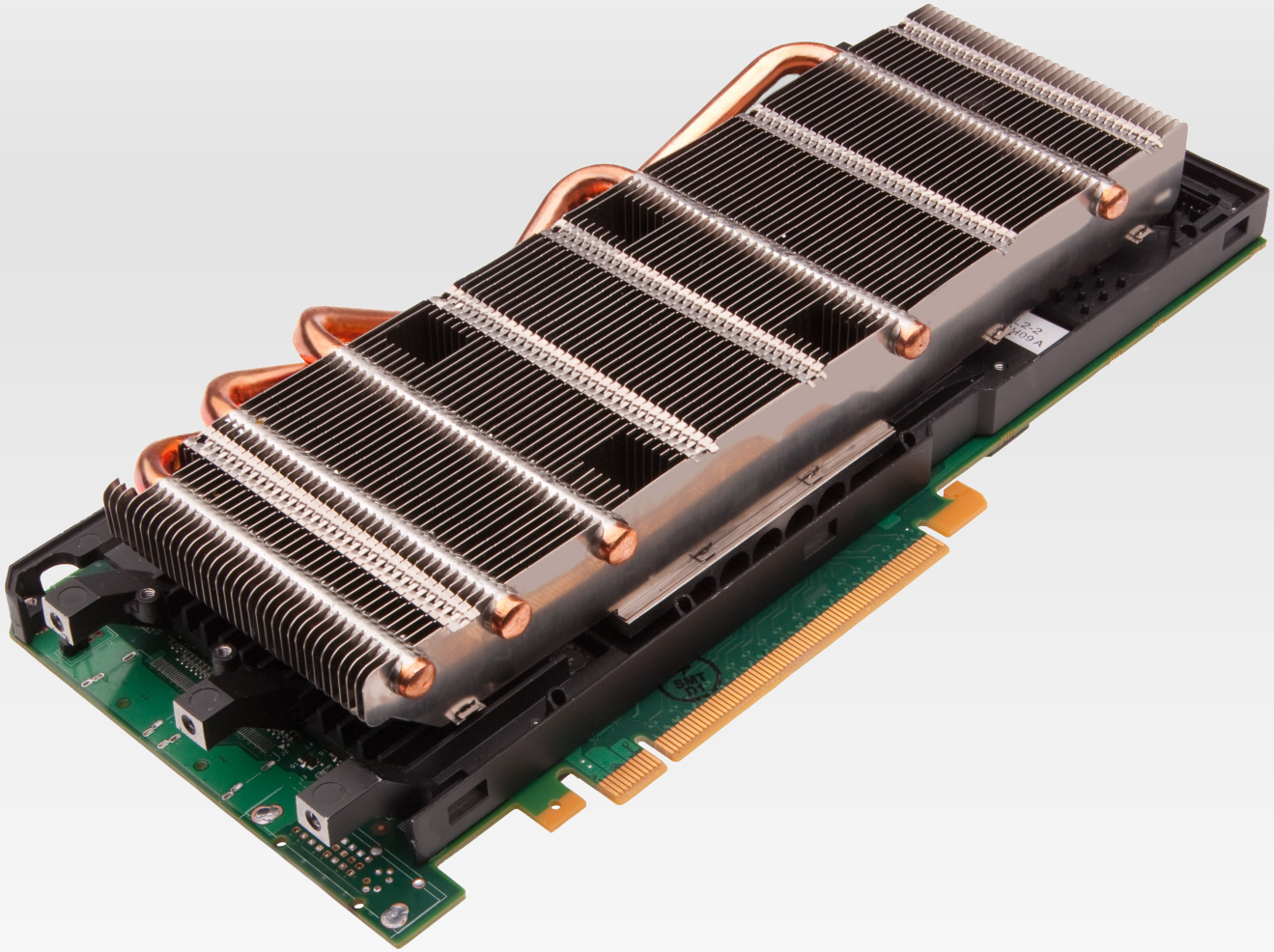
GPU Computing: Programming

- Live demo
  - Comparing CPU and GPU implementations

GPU Computing

- GPU computing in the field
  - NVIDIA Tesla workstations and computing modules
  - 7,168 of them power the worlds fastest super computer (Tianhe-A1) in combination with 14,336 Intel Xeon CPUs

GPU Computing



Computing Module: NVIDIA Tesla “Fermi” M2050

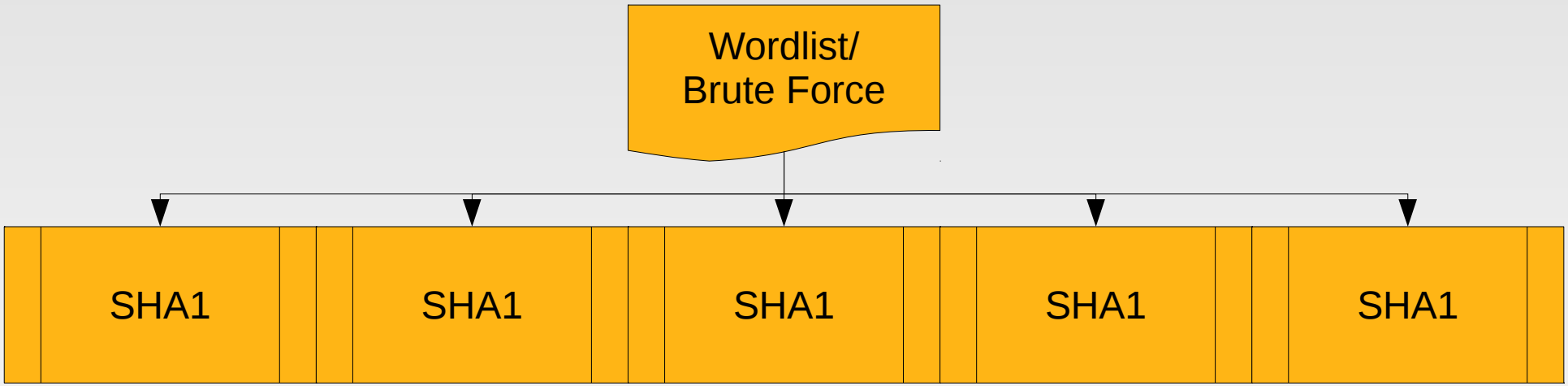
GPU Computing

- The M2050 computing module
  - 448 Cores
  - 3GB GDDR5 RAM
    - 1.55 GHz
    - 148 GB/sec

Double Precision floating point performance (peak)	515 Gflops
Single Precision floating point performance (peak)	1.03 Tflops

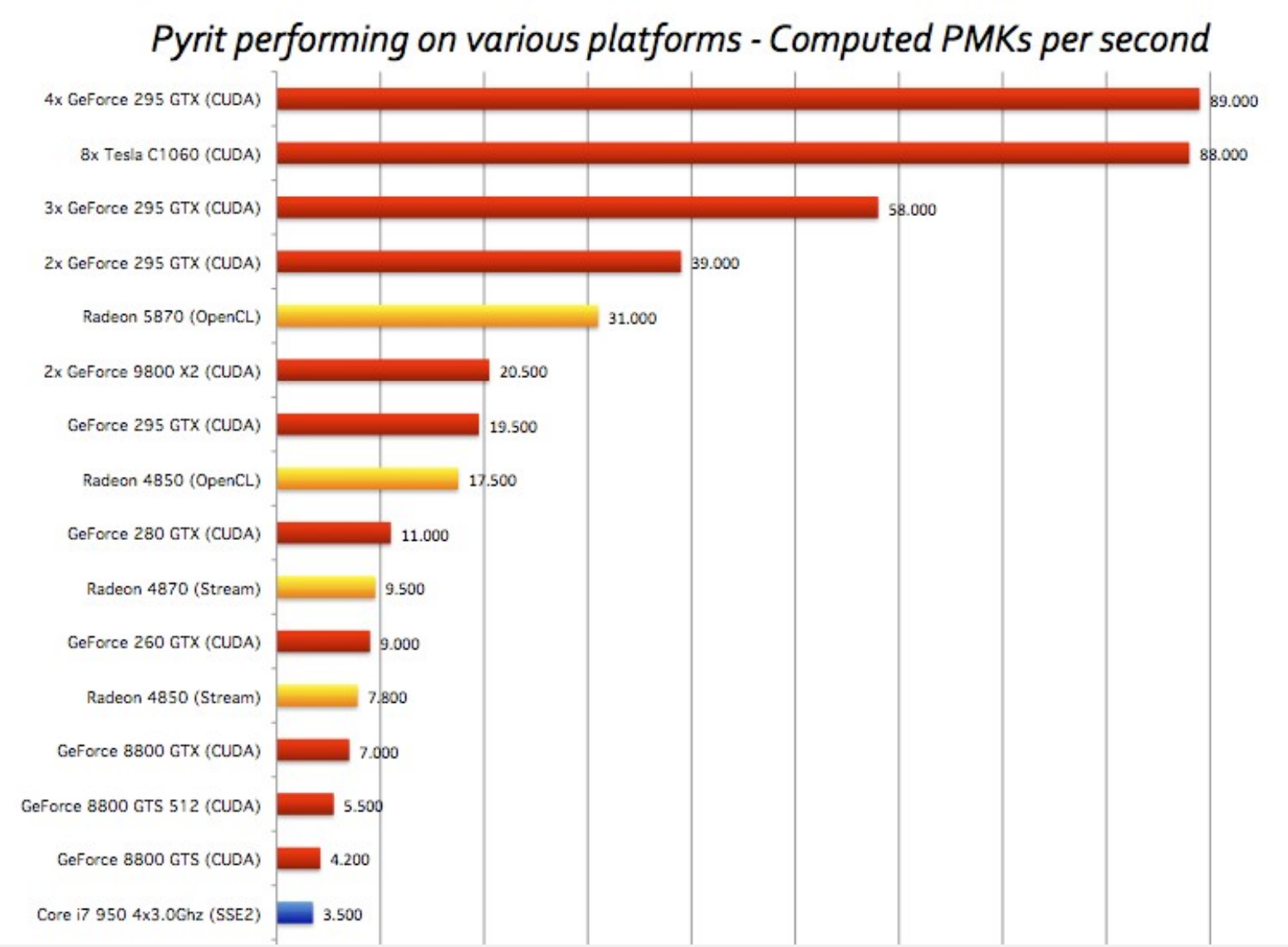
GPU Computing: Breaking encryptions

- Primitive attacks are easy to implement in a distributed manner

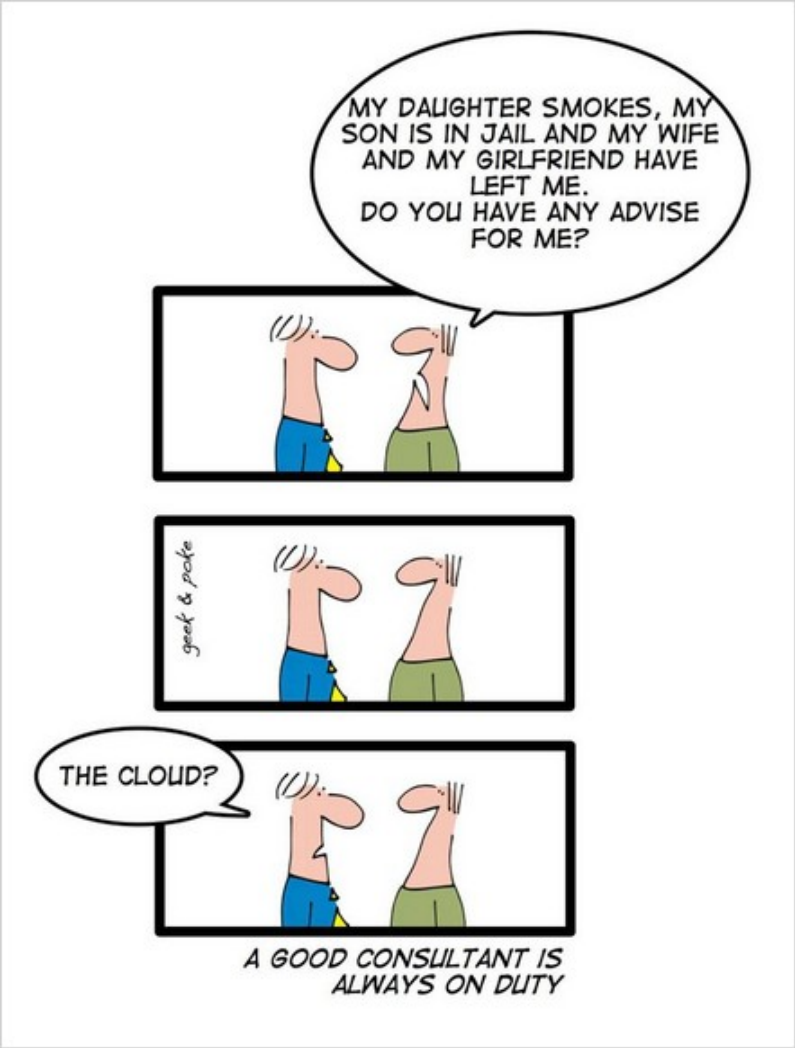


- Exactly what GPUs are made for

GPU Computing: Breaking encryptions



About “the cloud”



- Instances
- Storage
  - Instance Storage
  - EBS
  - S3
- Communication
  - Internal
  - External

About “the cloud”: Instances

- Virtual Machines (Xen)
- Boot from Amazon Machine Images (AMI)
  - Snapshots
  - From VMWare
- Can be started on demand
- Different types
  - (Micro, Small, Large, High-Mem, Cluster Compute...)
- 16K user-data can be supplied.

About “the cloud”: Storage: EBS

- Elastic Block Store
  - 1GB – 1TB
  - Can be mounted as a block device (Unformatted by default)
  - Snapshot creation (Incremental backup)
    - Snapshots are stored in S3
  - Faster than instance store

About “the cloud”: Storage: S3

- Simple Storage Service
  - Object-based
  - Stored in “Buckets”
  - 1B to 5TB
  - REST/SOAP
  - HTTP as download protocol

About “the cloud”: Communication

-
- Internal:
  - IP address via DHCP and internal hostname
    - domU-12-31-35-00-35-F3.z-2.compute-1.internal
- External:
  - Public IP and DNS name
    - ec2-72-44-45-204.z-2.compute-1.amazonaws.com
- Booth are released on termination of the instance.

About “the cloud”: GPU Instances

- Cluster GPU Instances
  - 22GB RAM
  - 2 x Intel Xeon X5570
  - 2 x NVIDIA Tesla “Fermi” M2050
  - $2.10/Hour
  - Spot instances often around $0.70

The “cloud cracking suite”

- Framework for distributed encryption breaking
- Written in Python
- Consists of two parts:
  - ccs-server
  - ccs-client

- http://stacksmashing.net/cloud-cracking-suite/

The “cloud cracking suite”: Server

- Runs on an instance
- Communicates with other instances
- Provides RPC interface
- Preparing the job for the cracking engine
- Controls the cracking engine
- Terminates the instance

The “cloud cracking suite”: Cracking-Engines

- Extensions for new ciphers:
  - Have to provide a Python API
  - Should care about the Hardware
  - Has to report back to the server

The “cloud cracking suite”: Client

- CLI for controlling servers
- Launches instances
- Prepares & uploads data
- Takes care of the initial communication between the nodes
- Used to get the status of the instances

The “cloud cracking suite”: Benchmarks

- Up to 50.000 PMKs/s per instance using the Pyrit cracking-engine at $2.10/h
- 400.000 PMKs/s using 8 instances at $16.80/h
- Easily scales much further

The “cloud cracking suite”

- Live demo:
  - High-speed, GPU accelerated WPA-PSK handshake cracking using CCS and the Amazon cloud.

Questions and answers

- Thanks for listening, hope you enjoyed it.

- If you've any questions left, feel free to contact me:
  - input@stacksmashing.net