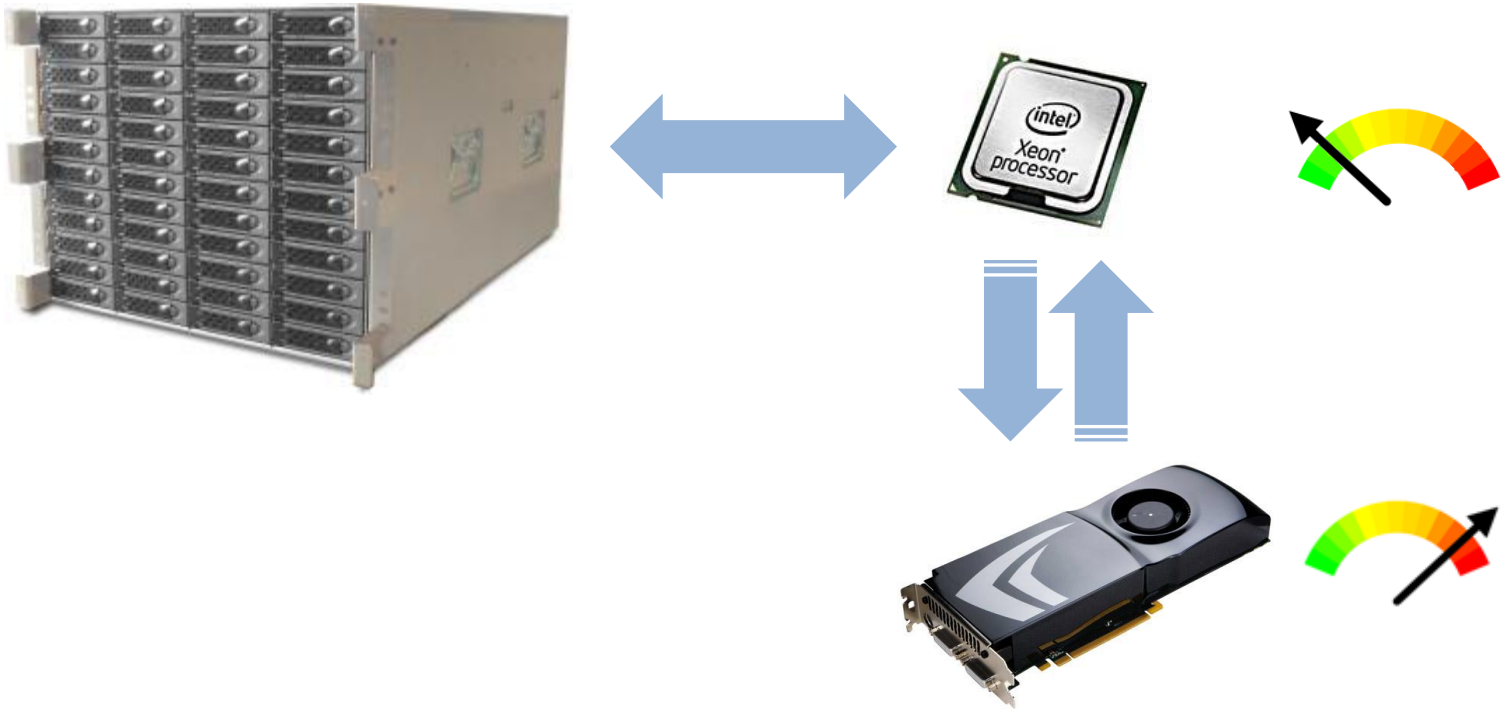
GrAVity: A Massively Parallel Antivirus Engine

Giorgos Vasiliadis and Sotiris Ioannidis
FORTH-ICS, Greece

RAID'10, 15 September 2010

Overview

- Increase the processing throughput of **virus scanning** applications, using the **Graphics Processing Unit (GPU)**



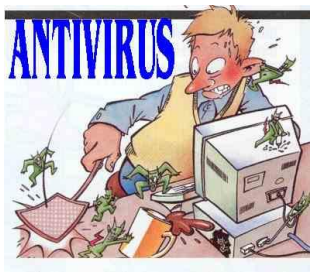
Outline

- Introduction
- Architecture
- Performance evaluation
- Conclusions

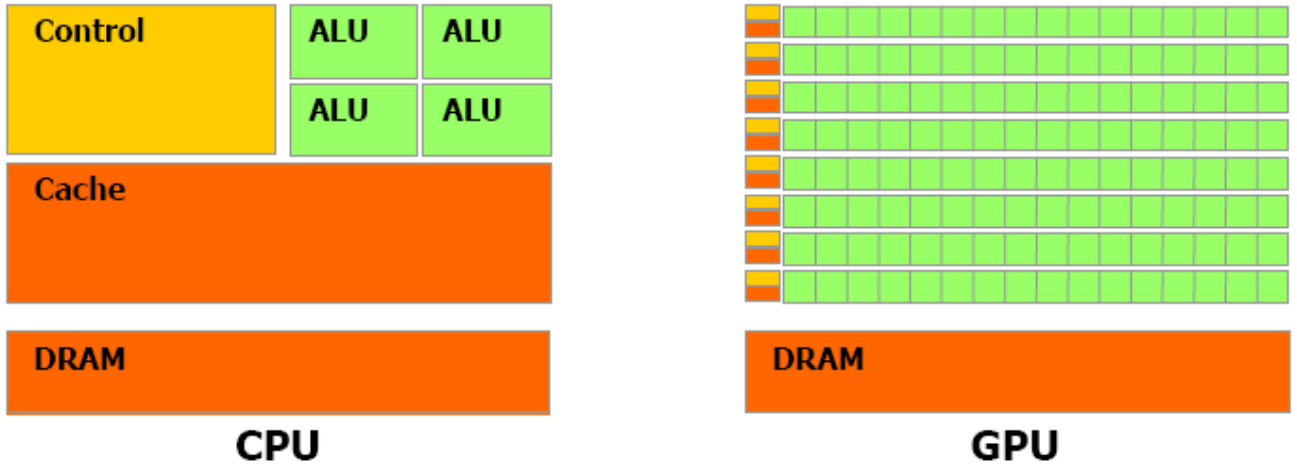
Motivation

- Antivirus software is running on e-mail servers, gateway proxies, user desktops
  - Require significant computational resources
- Graphics cards
  - Easy to program
  - Powerful and ubiquitous

- *Why not use GPUs to speed-up virus scanning operations?*



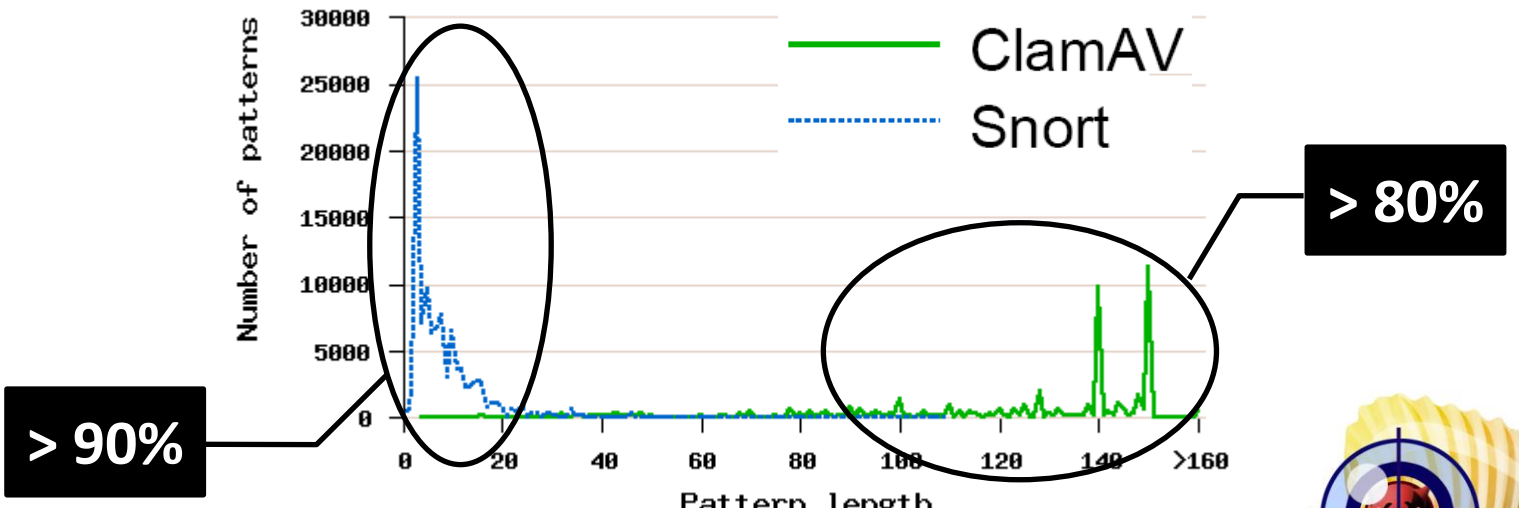
CPU vs GPU



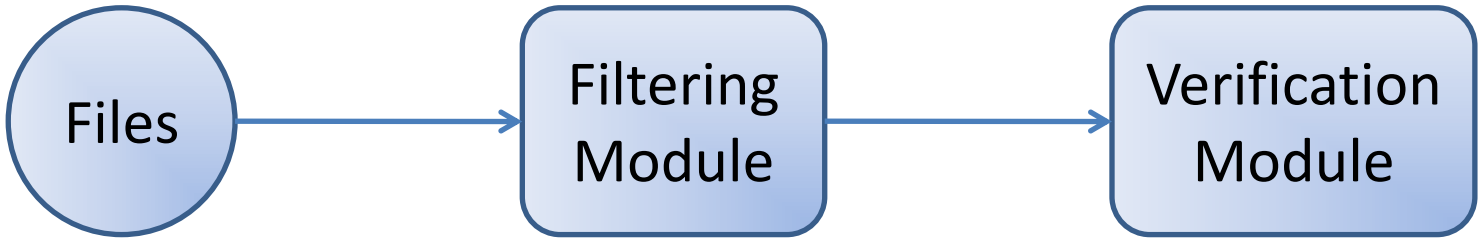
- The GPU is specialized for compute-intensive, highly parallel computation
  - More transistors are devoted to data processing rather than data caching and flow control

Anti-Virus Databases

- Contain thousands of signatures
- ClamAV contains more than 60K signatures, with length varying from 4 to 392 bytes
  - Significant longer than NIDS

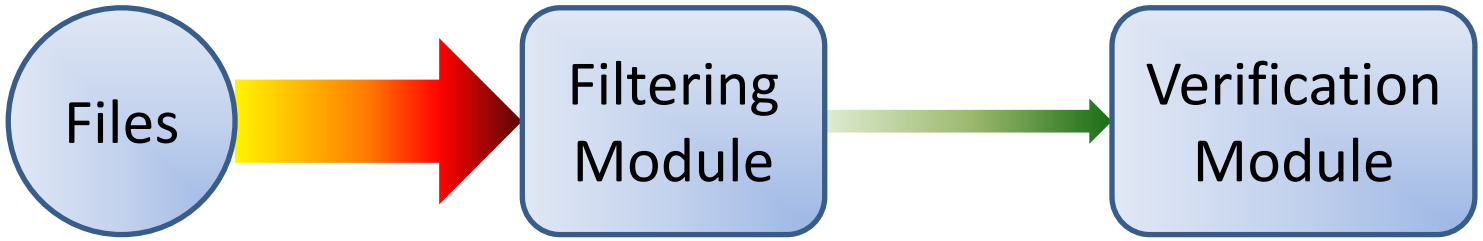


Virus Scanning in ClamAV



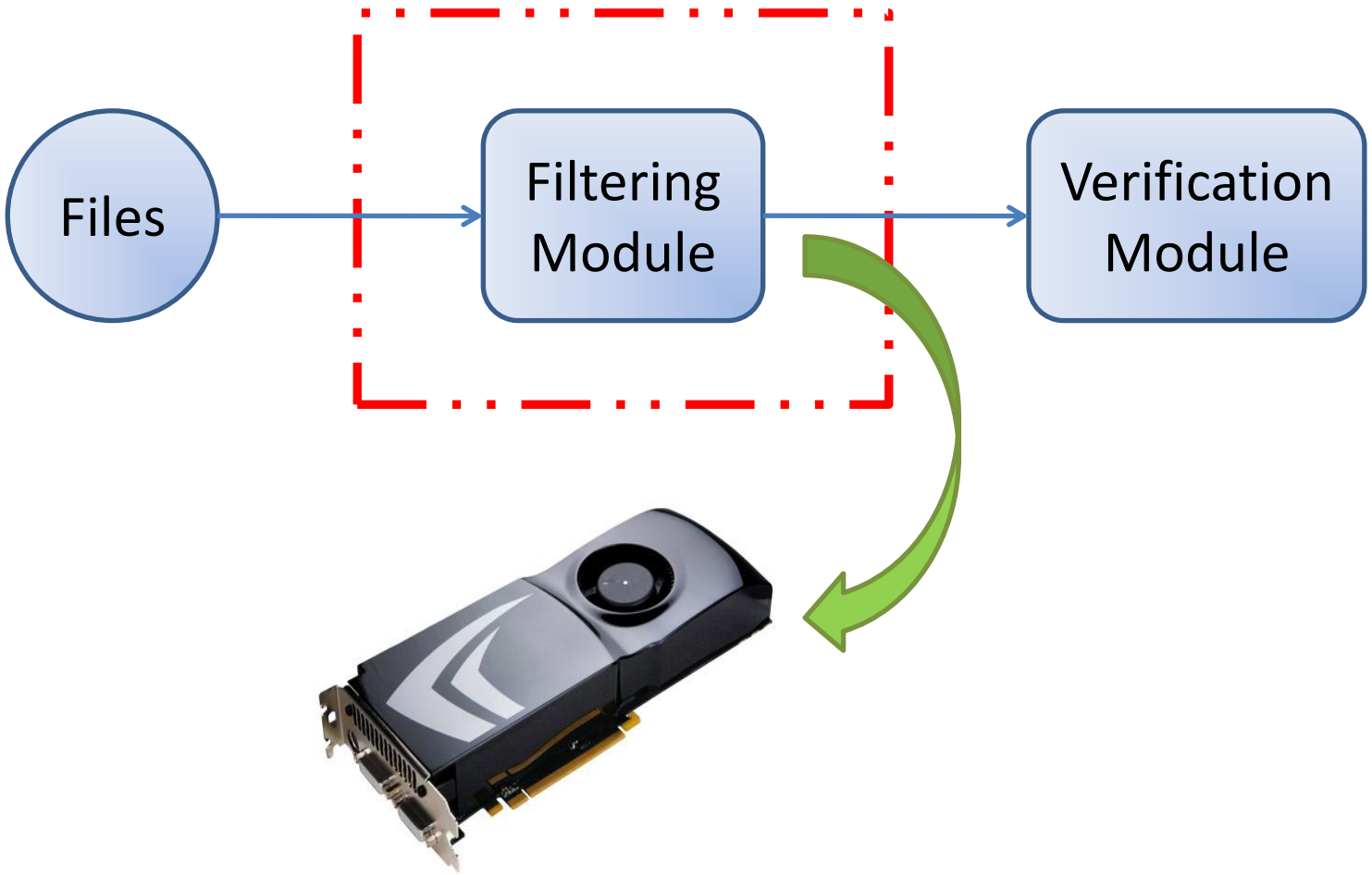
- ClamAV uses a small part from each signature for a *first-pass filtering*
- Every potential match is processed by the verification module

Virus Scanning in ClamAV



- Usually, the majority of data do not contain any virus
  - ⇨Only a small number of file segments pass to the verification module

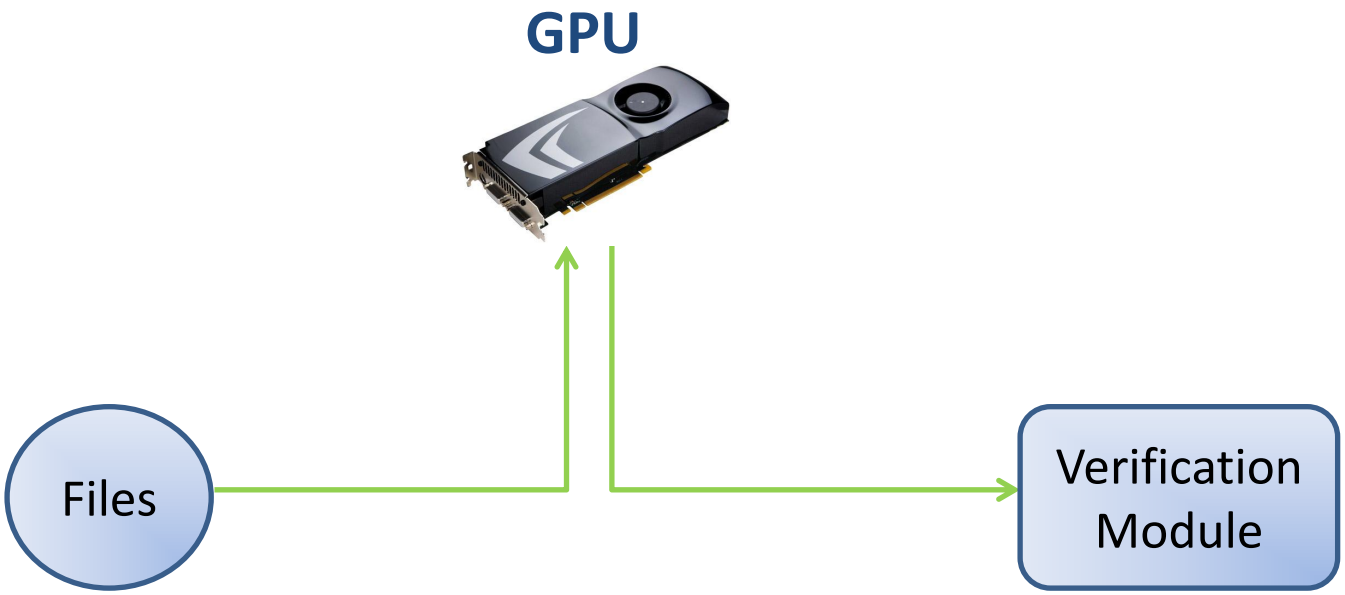
Our Approach: GPU Offloading



GRAVITY DESIGN

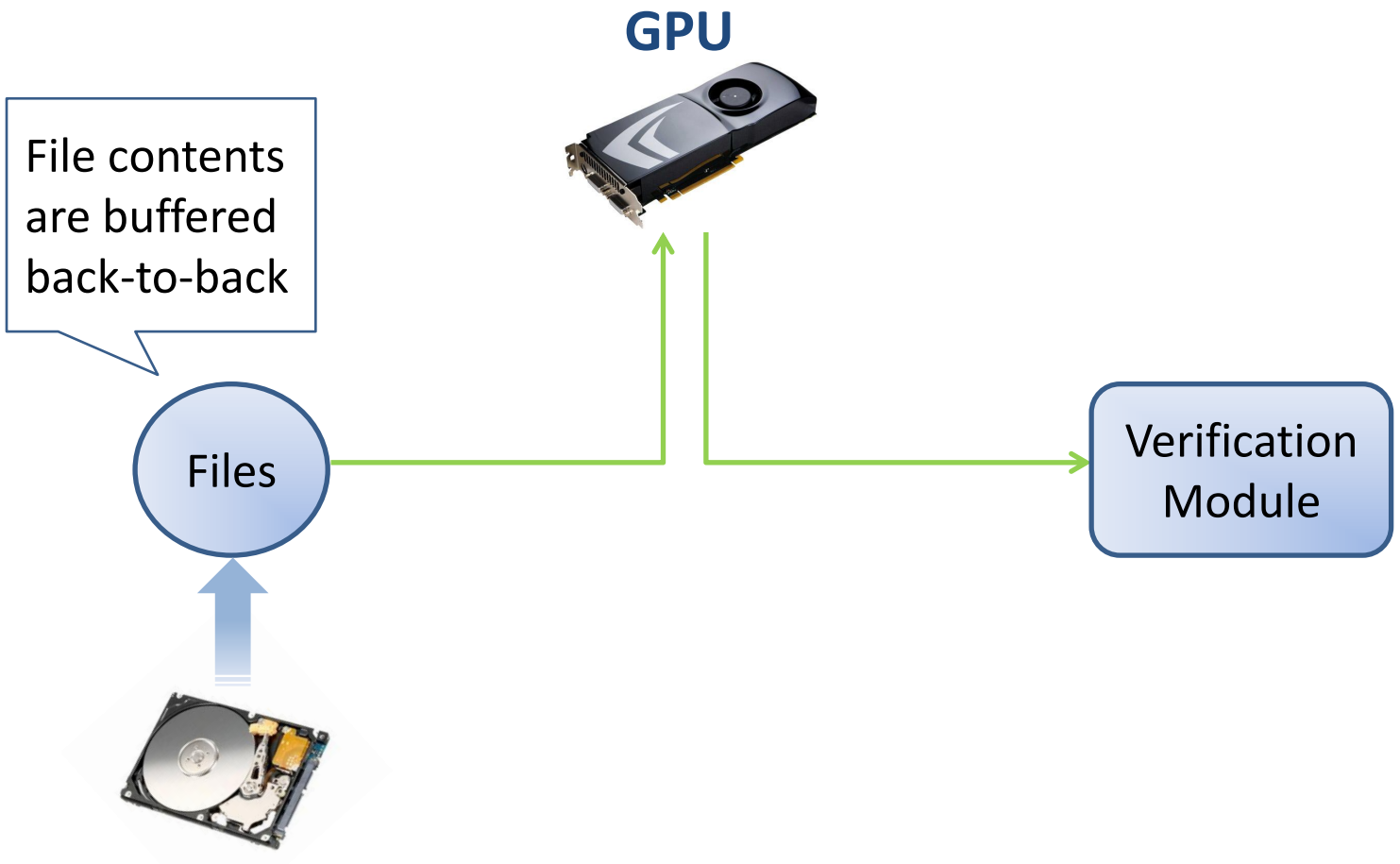
Basic Design

- Three-stage pipeline



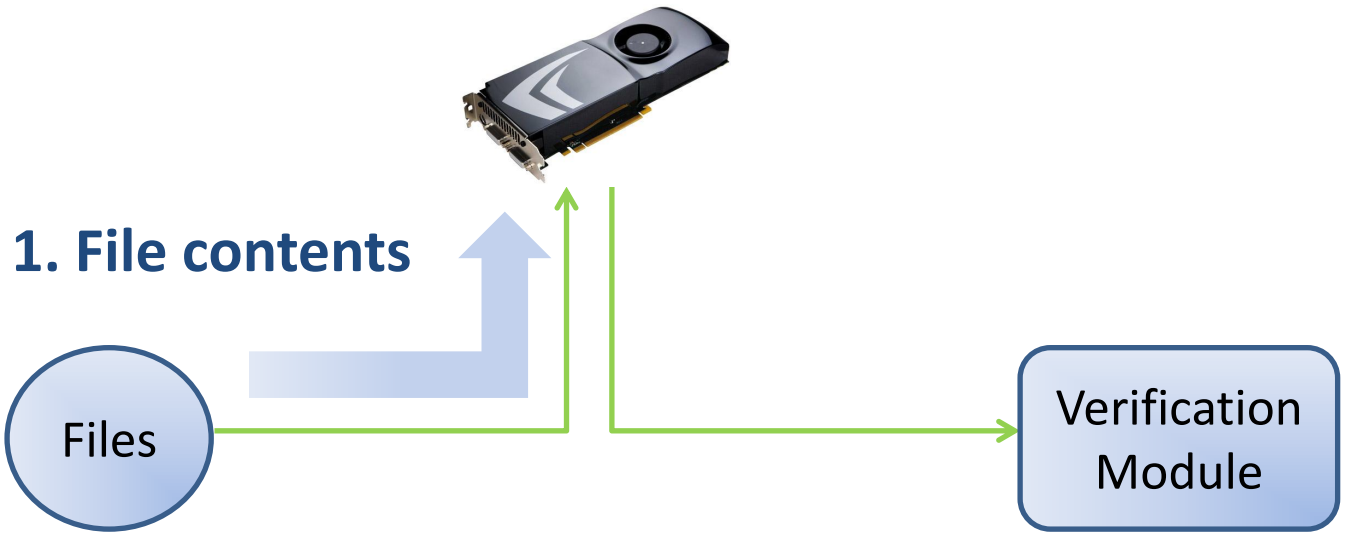
Files Journey (1/5)

- File scanning example



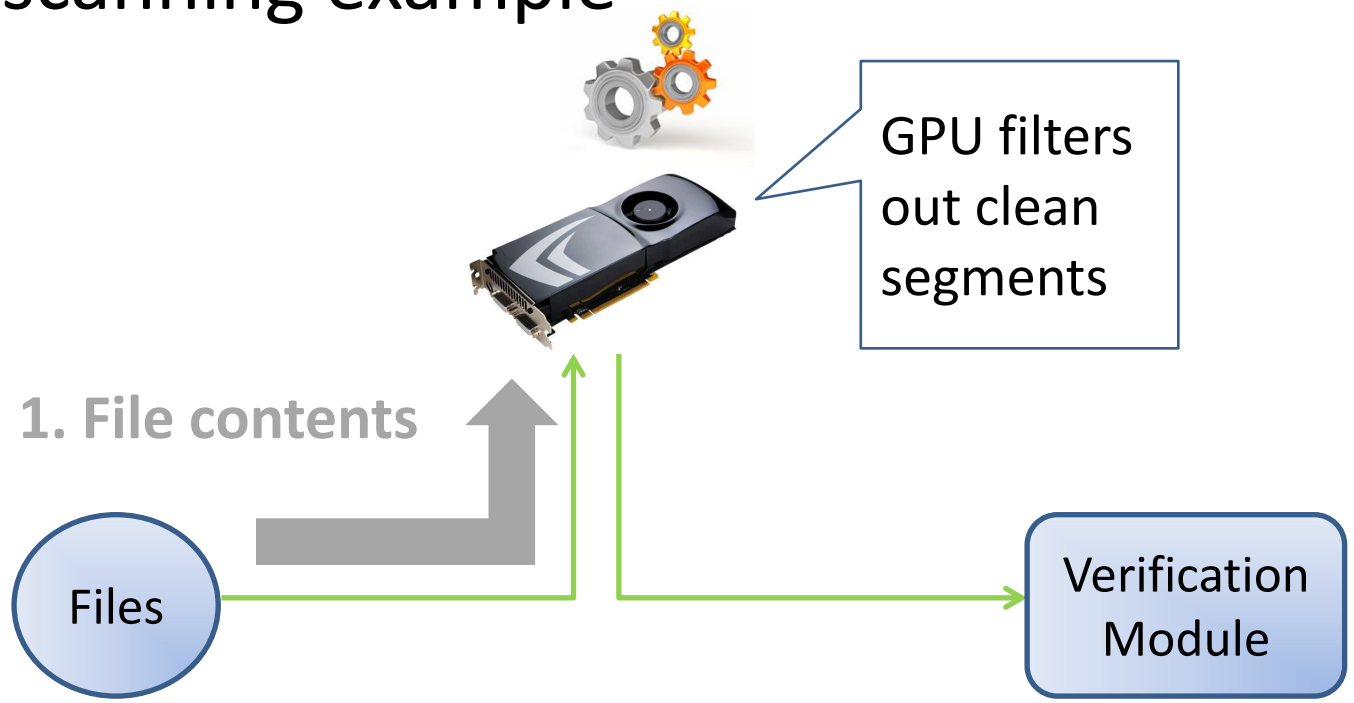
Files Journey (2/5)

- File scanning example



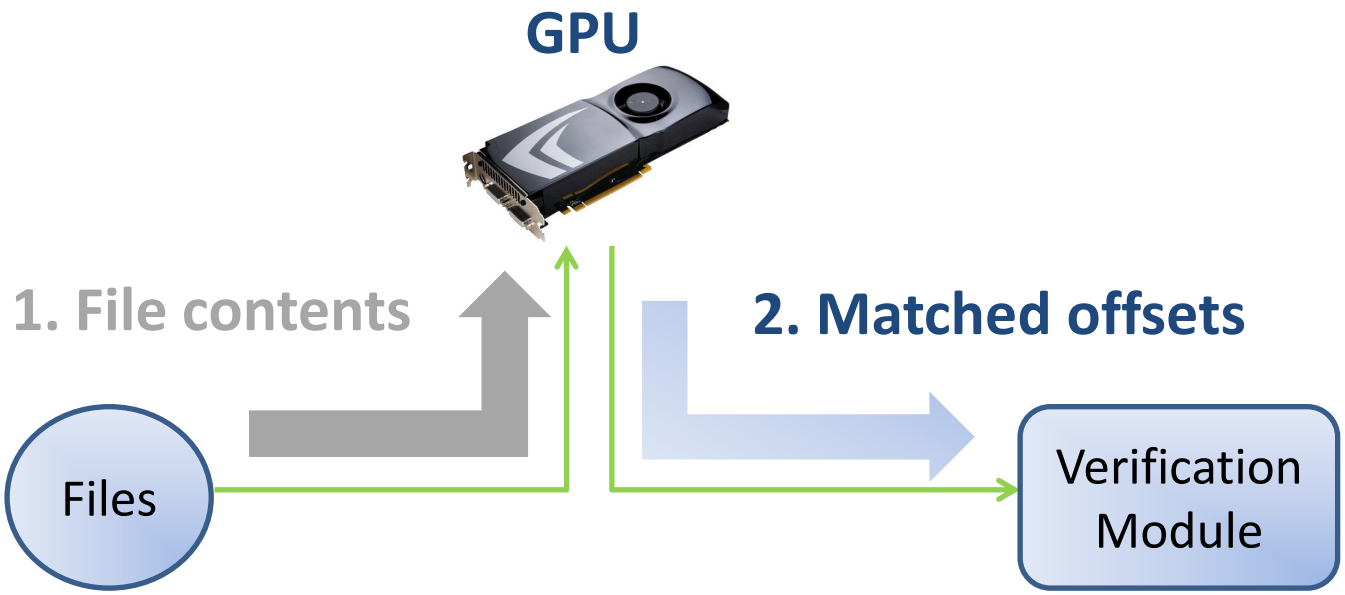
Files Journey (3/5)

- File scanning example



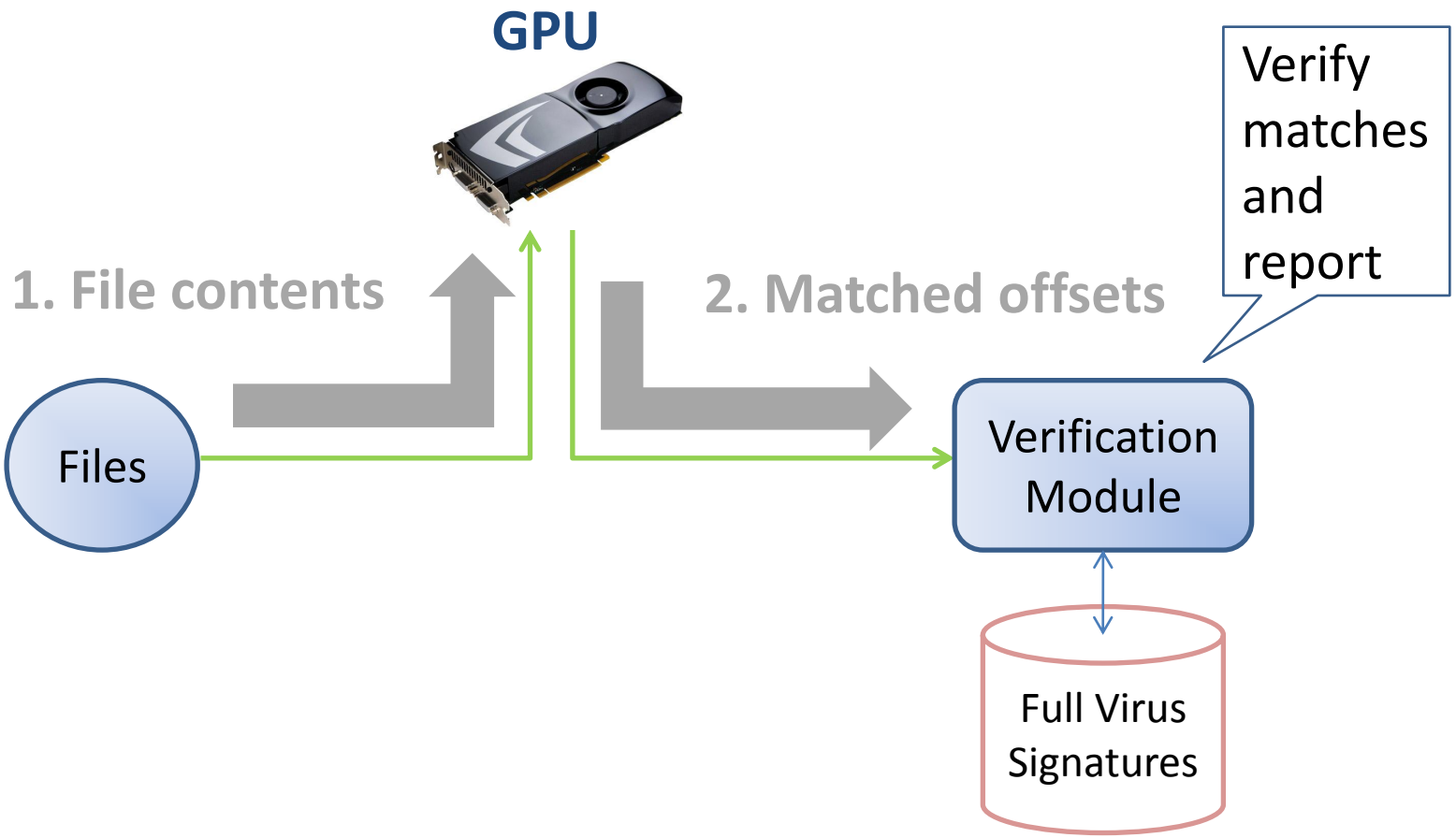
Files Journey (4/5)

- File scanning example



Files Journey (5/5)

- File scanning example



GPU IMPLEMENTATION

Prefix Filtering

- Take the first *n* bytes from each signature
  - e.g.

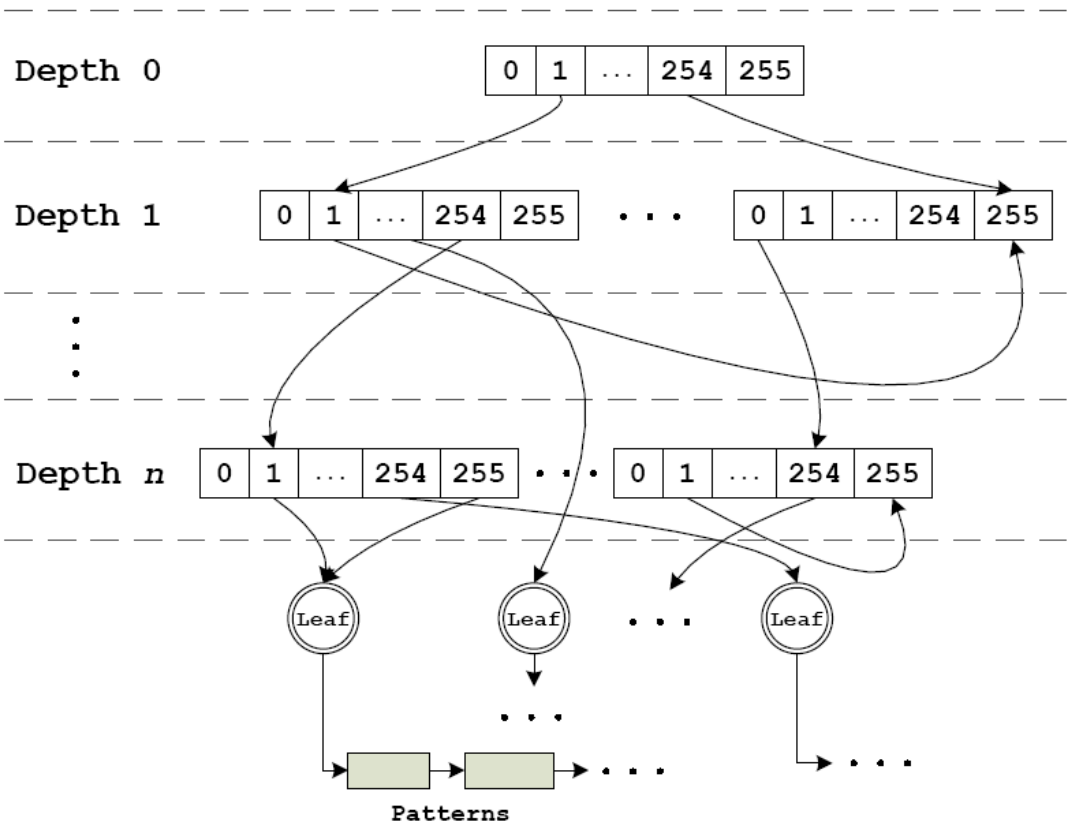
    Worm.SQL.Slammer.A:0:*:

    `4e65742d576f726d2e57696e33322e536c616d6d65725554`

- Compile **all** *n*-bytes sub-signatures into **a single** *Scanning Trie*
- The Scanning Trie can quickly filter clean data segments in linear time.

Scanning Trie

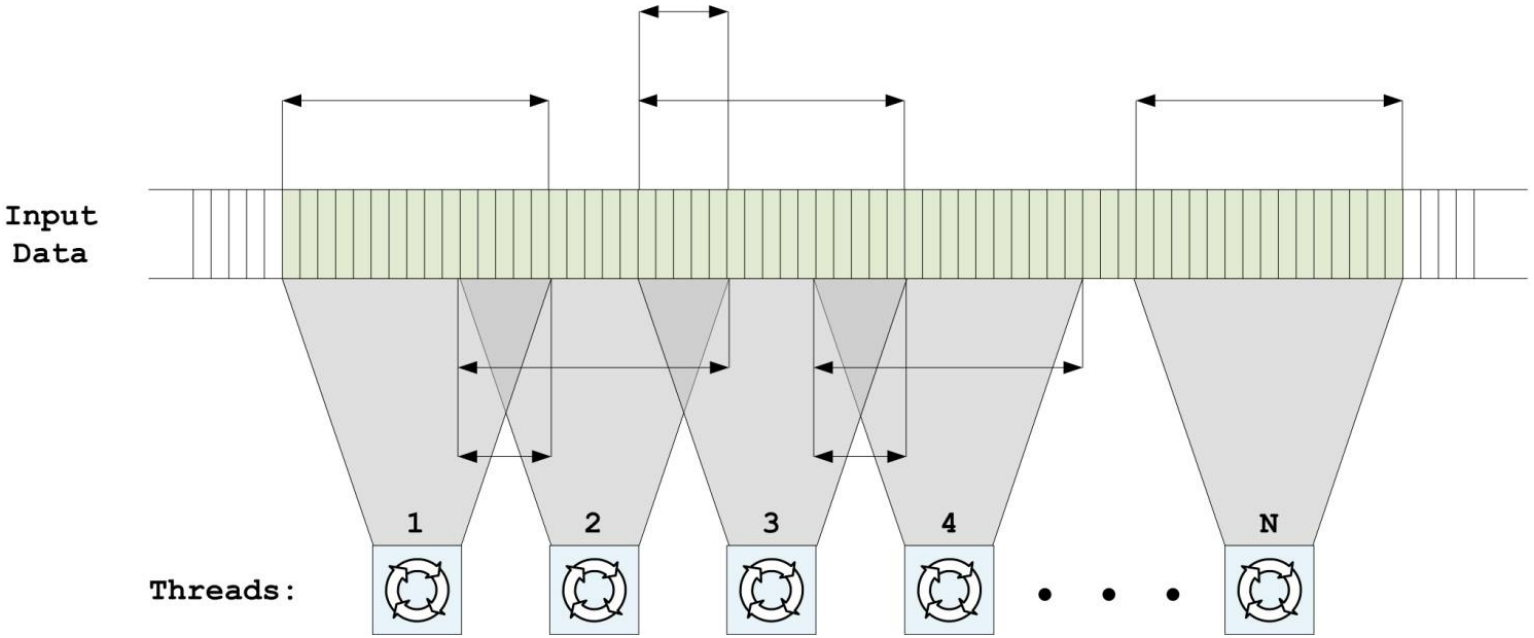
- GrAVity: Variable trie height



4 patterns
(avg) per
14-char prefix

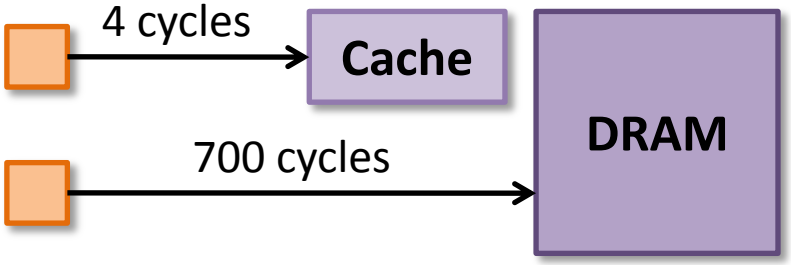
Virus Scanning on the GPU

- Each thread operate on different data
  - May overlap for spanning patterns, but …
  - … no communication/synchronization costs.
  - Highly scalable (million threads can run in parallel)

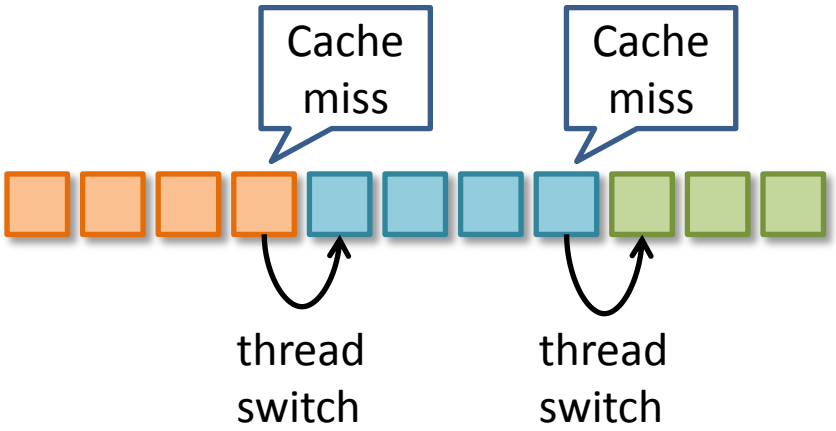


Memory Management Optimizations

- Exploit texture cache, to achieve better reading throughput

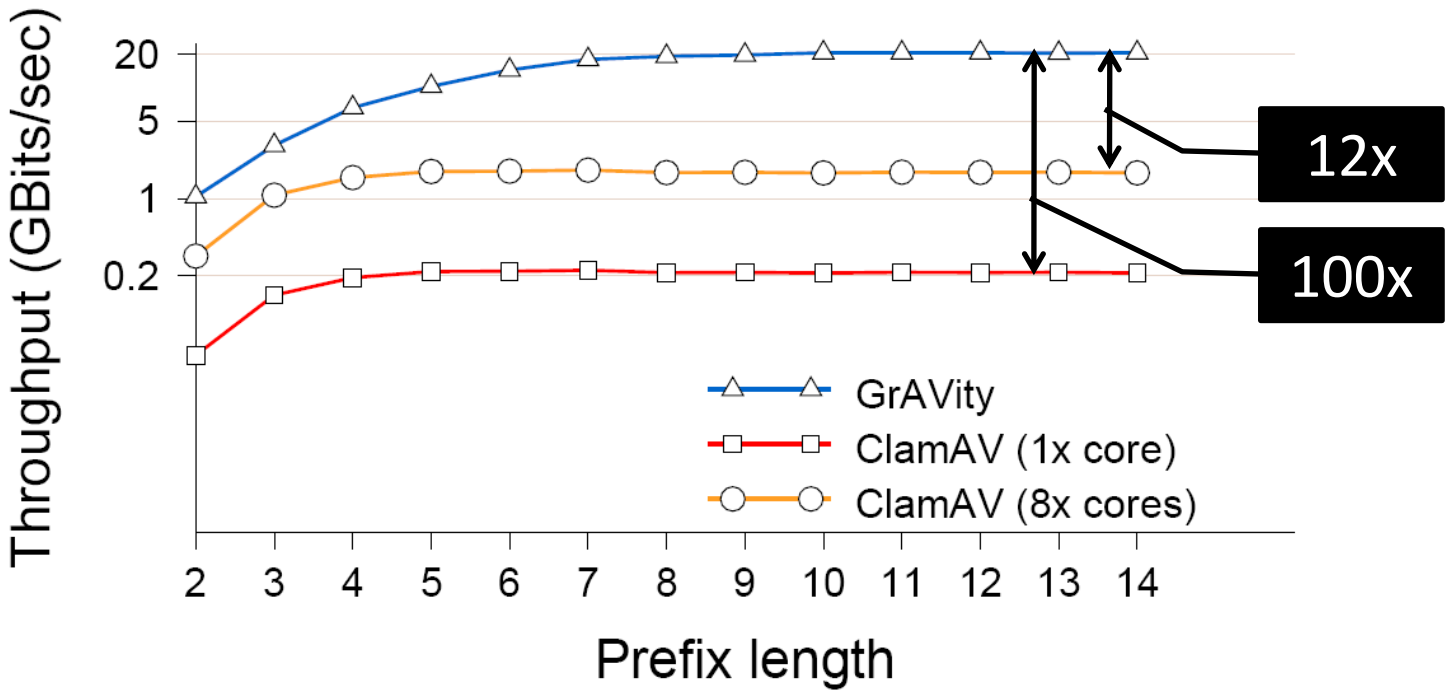


- Cache misses are hidden by running a large number of threads in parallel



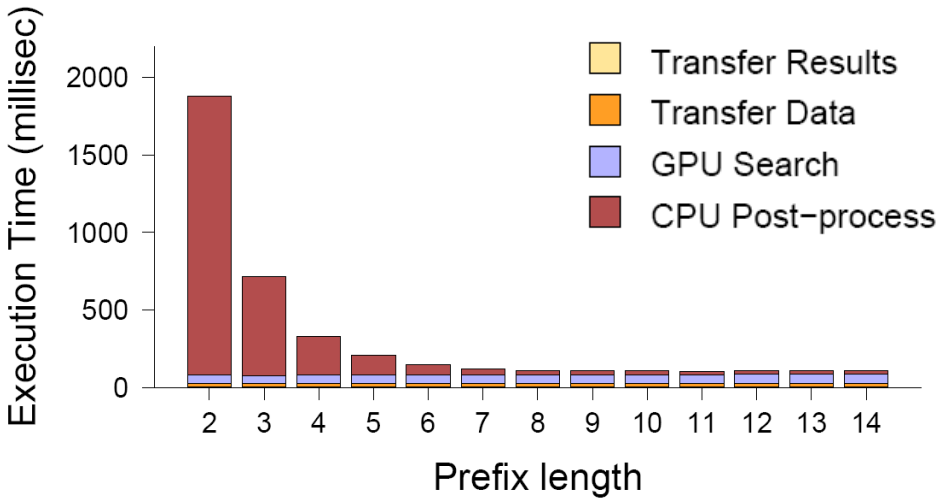
PERFORMANCE EVALUATION

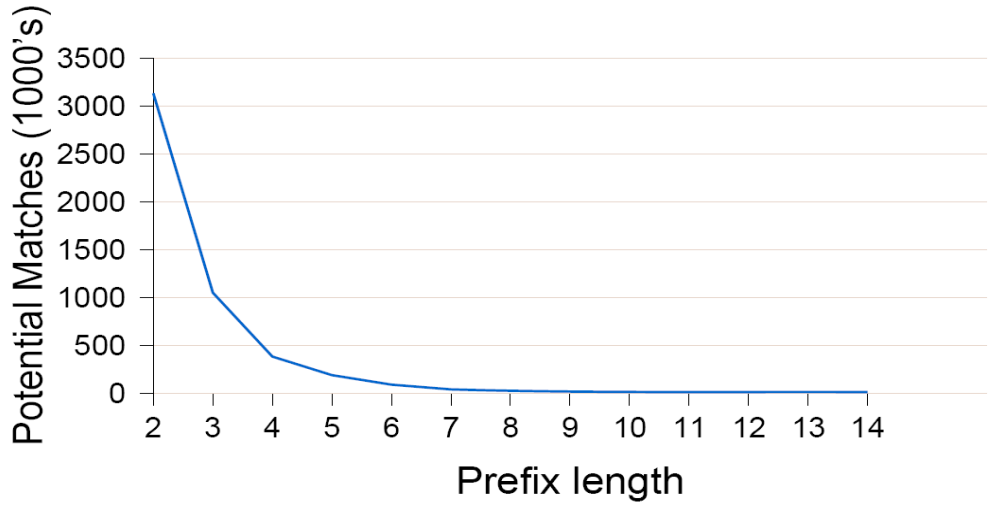
GrAVity vs ClamAV



- Up to 20 Gbps end-to-end performance

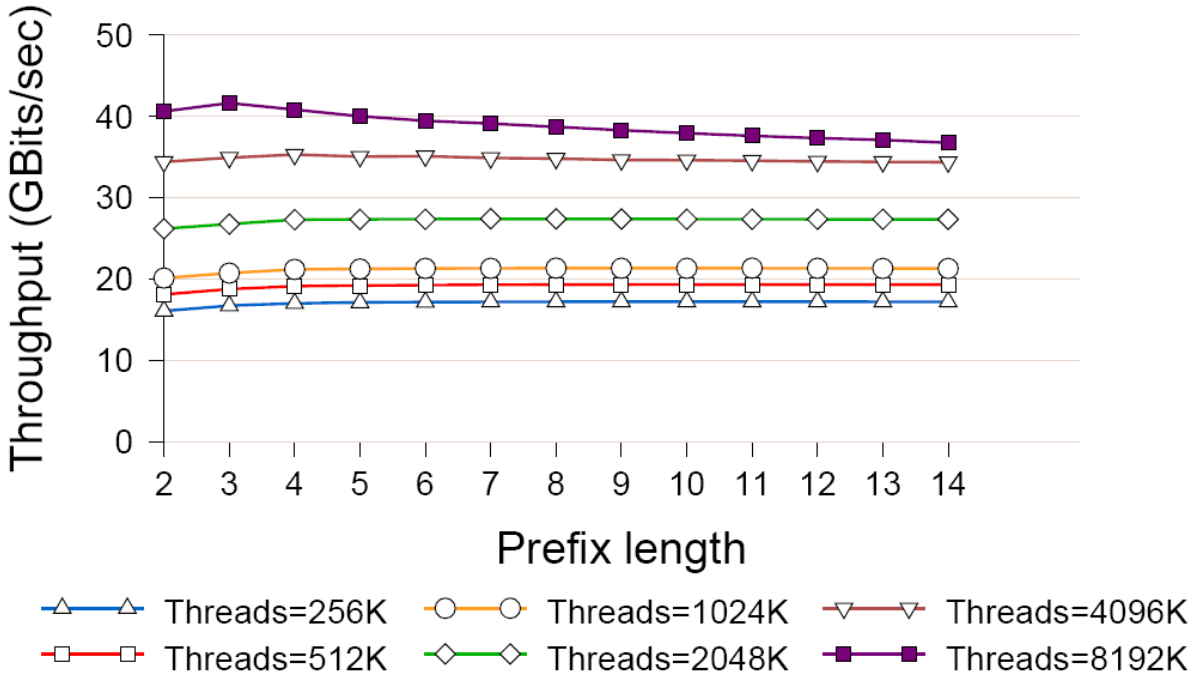
Execution Time Breakdown





- CPU time results in 20% of the total execution time, with a prefix length equal to 14
- Increasing the prefix length, results in less matches

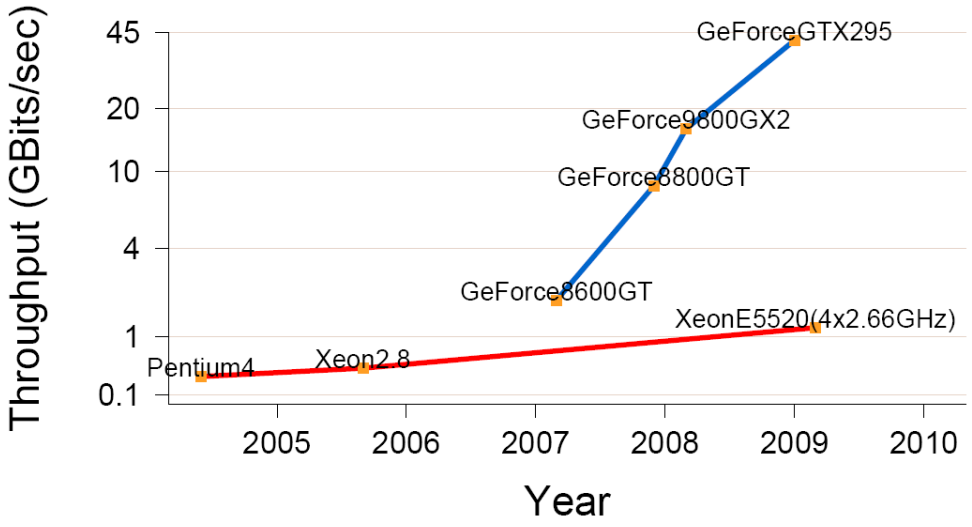
Raw Computational Throughput



- With 8M threads, the GPU achieves 42Gbits/s throughput

Scaling factor

- Fast evolution



Conclusions

- Virus scanning on the GPU is **practical** and **fast!**
- Over 20 Gbit/s throughput
  - Suitable for network-based virus scanning
- Future work includes
  - Adapt memory-efficient algorithms (XFA, D^2FA)
  - Multiple GPUs

GrAVity: A Massively Parallel Antivirus Engine

thank you!

Giorgos Vasiliadis, gvasil@ics.forth.gr
Sotiris Ioannidis, sotiris@ics.forth.gr