Surveillance of / with Small-scale systems

Chamseddine Talhi

École de technologie supérieure (ÉTS), Montreal

Department of Software Engineering and Information Technology

Agenda

- Project Presentation
- Why surveillance of/with small-scale systems?
- Surveillance of Small-scale Systems
- Surveillance with Small-scale Systems
- Project Summary
- Feedback?

Agenda

- Project Presentation
- Why surveillance of/with small-scale systems?
- Surveillance of Small-scale Systems
- Surveillance with Small-scale Systems
- Project Summary
- Feedback?

Project Presentation

- New research thread to Advanced Host-Level
  - 1 year project! 2013/2014

- Team:
  - 1 professor
  - 2 Master students
  - 1 part time research professional
  - Part-time graduate/undergraduate students

- Objectives:
  - Surveillance of small-scale systems
  - Use of massively parallel small-scale systems for the surveillance of other systems

Agenda

- Project Presentation
- Why surveillance of/with small-scale systems?
- Surveillance of Small-scale Systems
- Surveillance with Small-scale Systems
- Project Summary
- Feedback?

Small-scale systems, why?

From Mobile Phones to general-purpose small devices

- « Cabir » 2004 : first mobile phone malware
- « CommWarrior » & « Doomboot » 2005 :
- And …

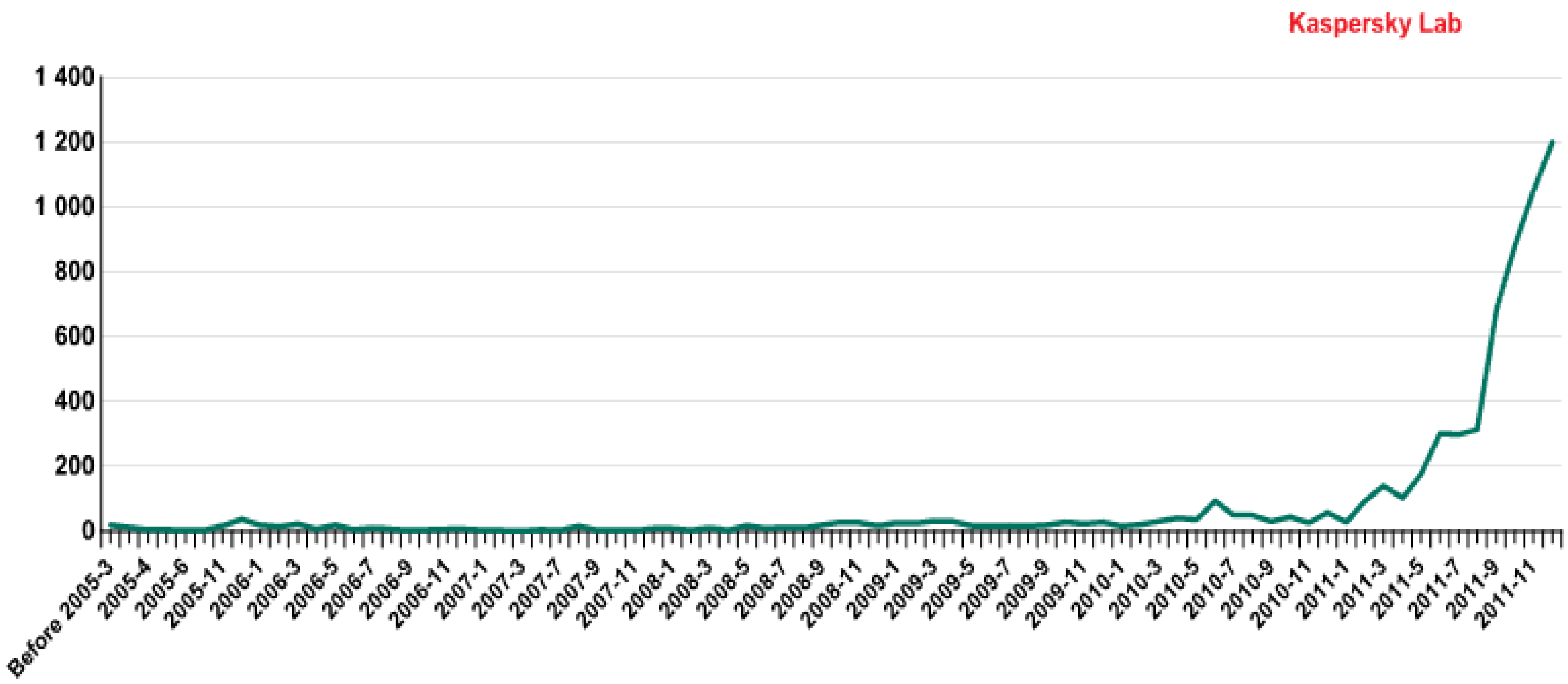
2 years of mobile malware evolution <=>

20 years of Computer malware evolution!!!

More than 1 000 000 variants of malware targeting Android

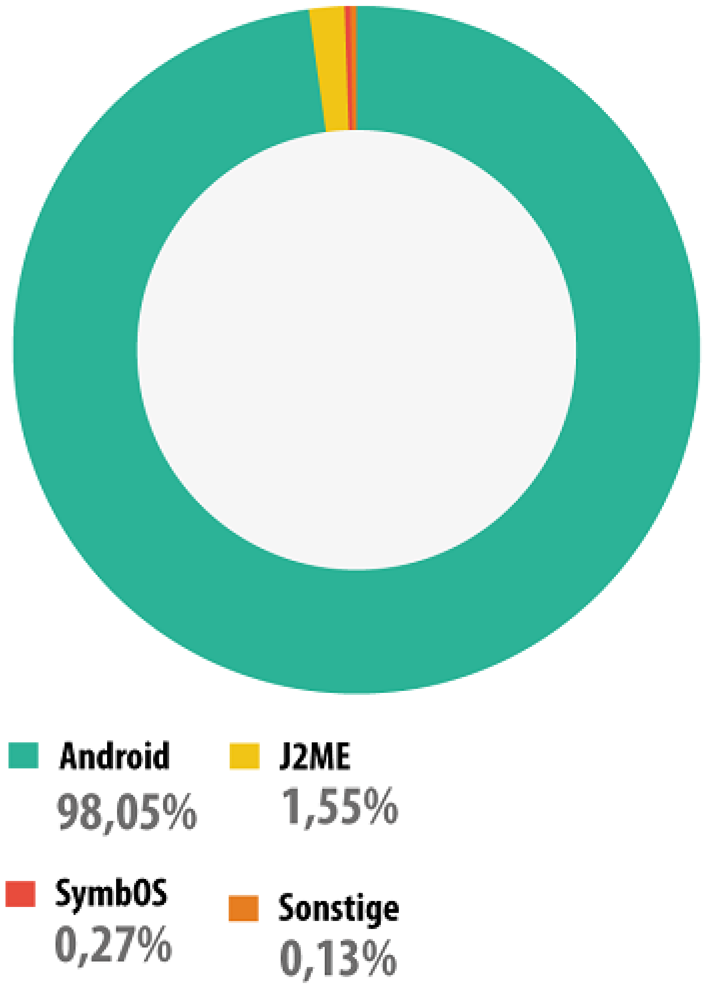
Small-scale systems, why?

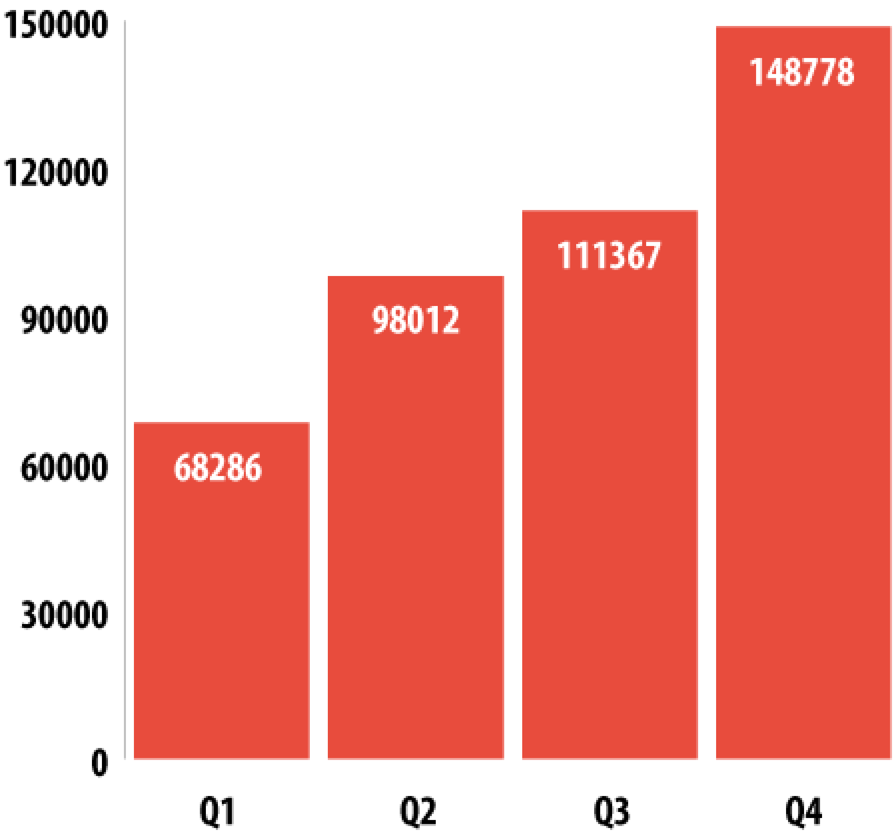
Mobile malwares – Evolution



Small-scale systems, why?

Mobile malwares – 2013 Statistics





Source: http://www.securelist.com/en/analysis/204792318/Kaspersky_Security_Bulletin_2013_Overall_statistics_for_2013#02

Small-scale systems, why?

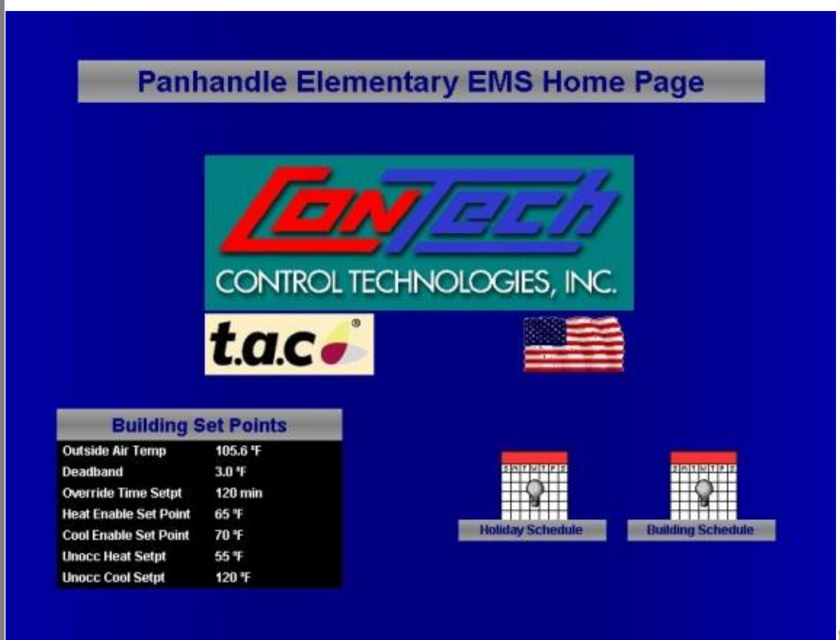
Small-scale systems are not limited to Smartphones!

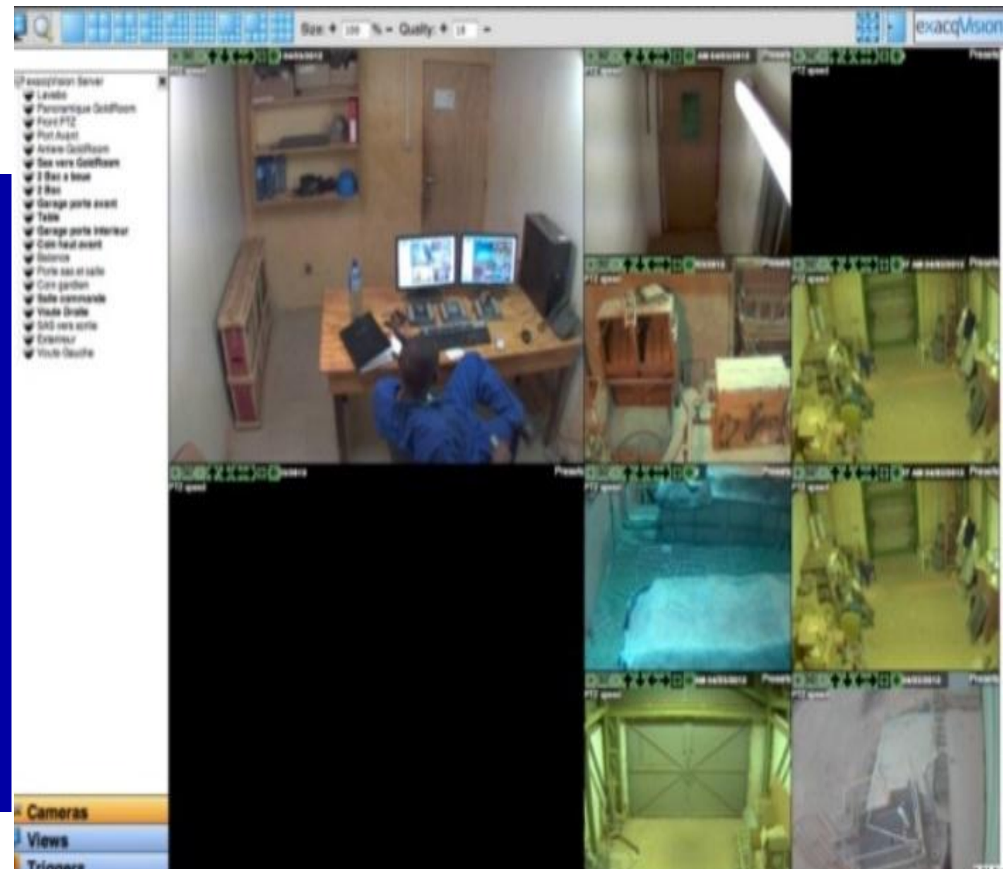
- Linux/Android based devices.
- Shodan : Computer Search Engine

Small-scale systems, why?

Shodan : Computer Search Engine

Privacy? Security?





Small-scale systems, why?

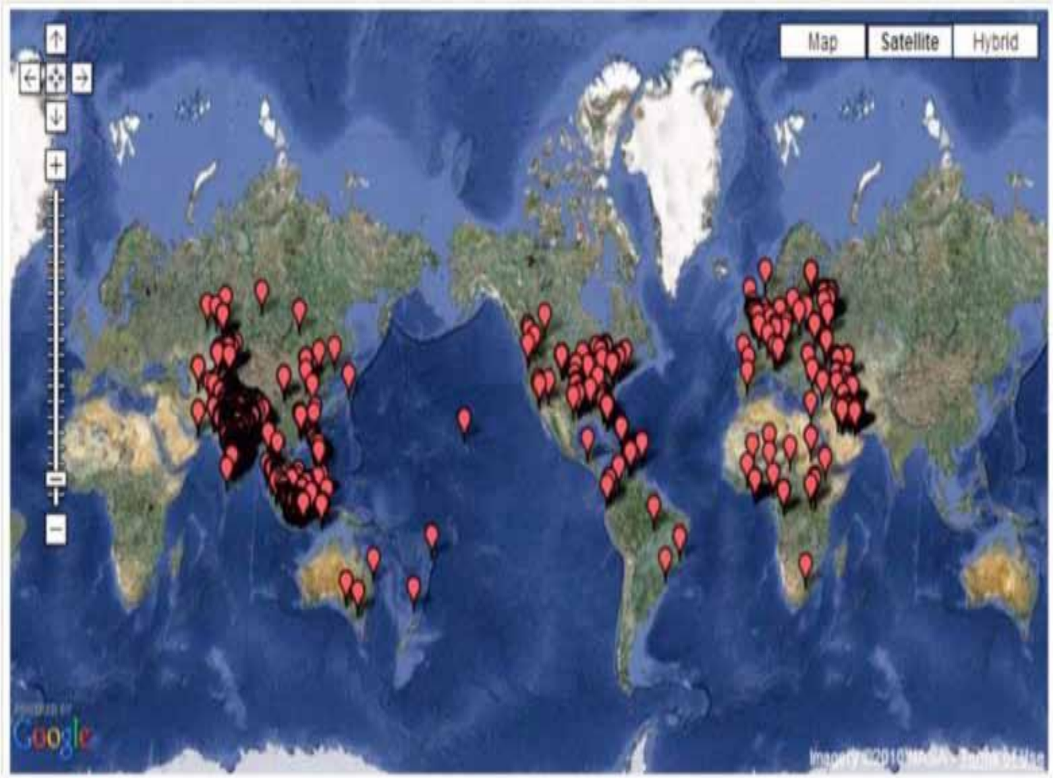
- Malwares in Embedded Systems: next (r)evolution!

Year	Malware /attack	Target	Threats
2009	**psyb0t**	**Linux-based** routers and DSL modems	DDoS
2010	**Chuck Norris Botnet**	**Linux-based** routers, DLS modems	DDoS +DNS Spoofing
	Stuxnet	industrial control systems (ICS)	alter PLCs for supported facilities
2012	**DNSChanger**	computers and routers	DNS spoofing/poisoning
2013	**JUL: GPS attack**	GPS based systems	**total control of system**
	Sept: Linux/Flasher	wireless routers	login credentials captured and transferred to remote web servers.
	Nov 26 : Linux.Darlloz	**Linux-based** computers, industrial control servers, routers, **cameras, set-top boxes.**	generates IP @ randomly, accesses a specific path on the machine with well-known ID and passwords, and sends HTTP POST requests

Small-scale systems, why?

- Stuxnet Malware (2010)!

Country	Infected computers
Iran	58.85%
Indonesia	18.22%
India	8.31%
Azerbaijan	2.57%
United States	1.56%
Pakistan	1.28%
Others	9.2%

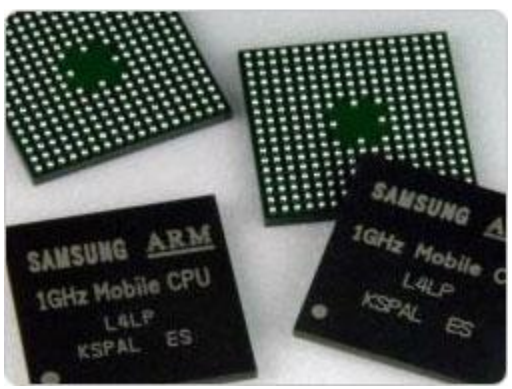


Small-scale systems, why?

- Resource Limitations

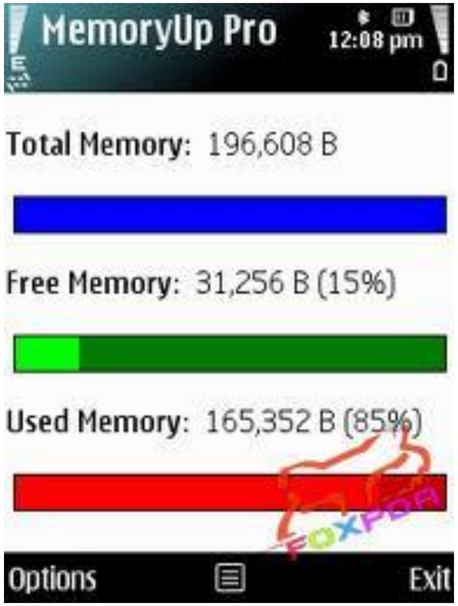
Low power CPUs

- Lightweight processing
- limited multitasking



Battery life

Memory limited to Megabytes



Agenda

- Project Presentation
- Why surveillance of/with small-scale systems?
- Surveillance of Small-scale Systems
- Surveillance with Small-scale Systems
- Project Summary
- Feedback?

Signature vs. Anomaly Detection - Challenges

Signature-based detection

- *Best Multi pattern matching algorithms?*
- *Optimization:* data structures and algorithms, compression, parallel programming, etc.
- Need for Cloud/Server: signatures Database storage, Remote scan.

Anomaly-based detection

- *Machine Learning algorithms:* accuracy (eg. false positives), overhead (eg., memory and power, etc.)
- *Need for remote Cloud/Server:* traces storage and exchange

Signature Detection: Multi Pattern Matching (1)

Empirical study

- Required memory budget : varying numbers of signatures.
- Dataset : Android malwares signatures (MD5 hash).
- Memory budget compared with available memory on a Samsung Galaxy S Vibrant phone.

Small-scale Sys Surveillance

Signature Detection: Multi Pattern Matching (2)

Empirical study

Evolution of Smartphone Memory

year	phone model	memory size (MB)
2002	Blackberry 5810	1
2003	BlackBerry 7210	6
2004	Nokia 6630	10
2005	HTC Universal	64
2006	HTC TyTN 100	64
2007	Iphone	128
2008	HTC dream	192
2009	HTC Magic	288
2010	Samsung Galaxy S	512
2011	Samsung Galaxy S2	1024

Signature Detection: Multi Pattern Matching (3)

Empirical study

512 MB should be enough ... BUT

• Memory reserved by hardware	= **32 MB**
• Android fixed components	= **80 MB**
• Launcher	= **30 MB**
• Live wallpaper	= **20 MB**
• 5 widgets	= **20 MB**
• Android System	= **208 MB**
• **Available memory**	**= 121 MB!**

Small-scale Sys Surveillance

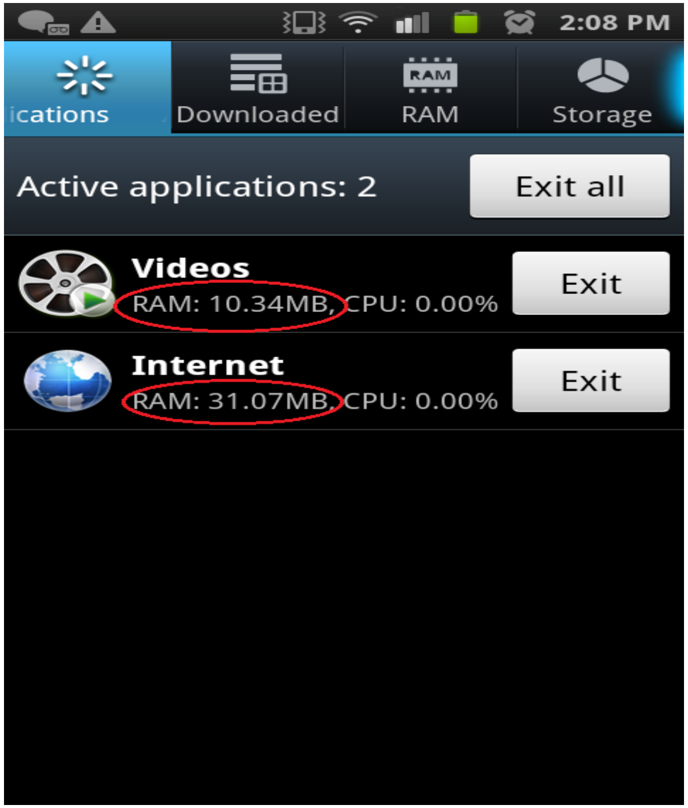
Signature Detection: Multi Pattern Matching (4)

Empirical study

512 MB should enough ... BUT
From 121 MB

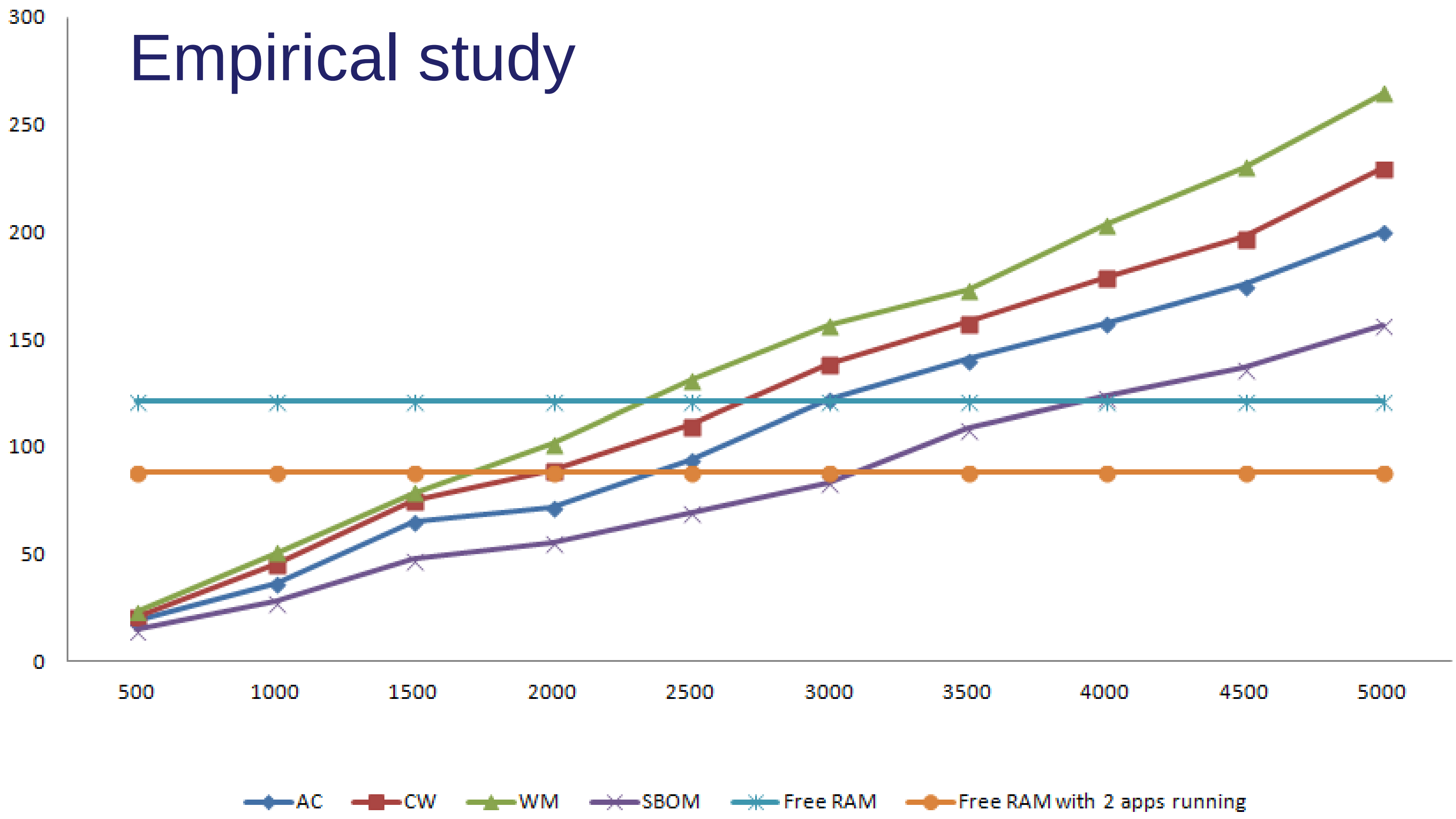
- Video media player consumes 10.34MB
- internet browser consumes 31.07MB

Finally : Available memory is 88MB!!!



Small-scale Sys Surveillance

Signature Detection: Multi Pattern Matching (5)



Signature Detection : Lessons learned

- Fast evolution of signatures database: memory of small-scale systems will never be enough!!!
- A subclass of most important signatures should be maintained
- subclasses of malwares => sub-databases
- Optimize, optimize, …., and optimize

Small-scale Sys Surveillance

Anomaly Detection (1)

Analysis of sys call n-grams

- Look-ahead pairs
- n-gram Trees

Small-scale Sys Surveillance

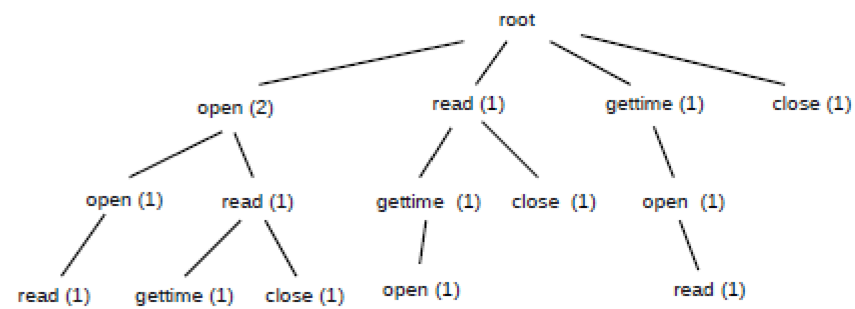
Anomaly Detection : sys call n-gram Analysis (2)

1 2 3

open, open, read, gettime, open, read, close

	Appel système	1 appel après	2 appels après
Appel 1	open	open, read	read, gettime, close
Appel 2	read	gettime, close	open
Appel 3	gettime	open	read

Lookahead pairs

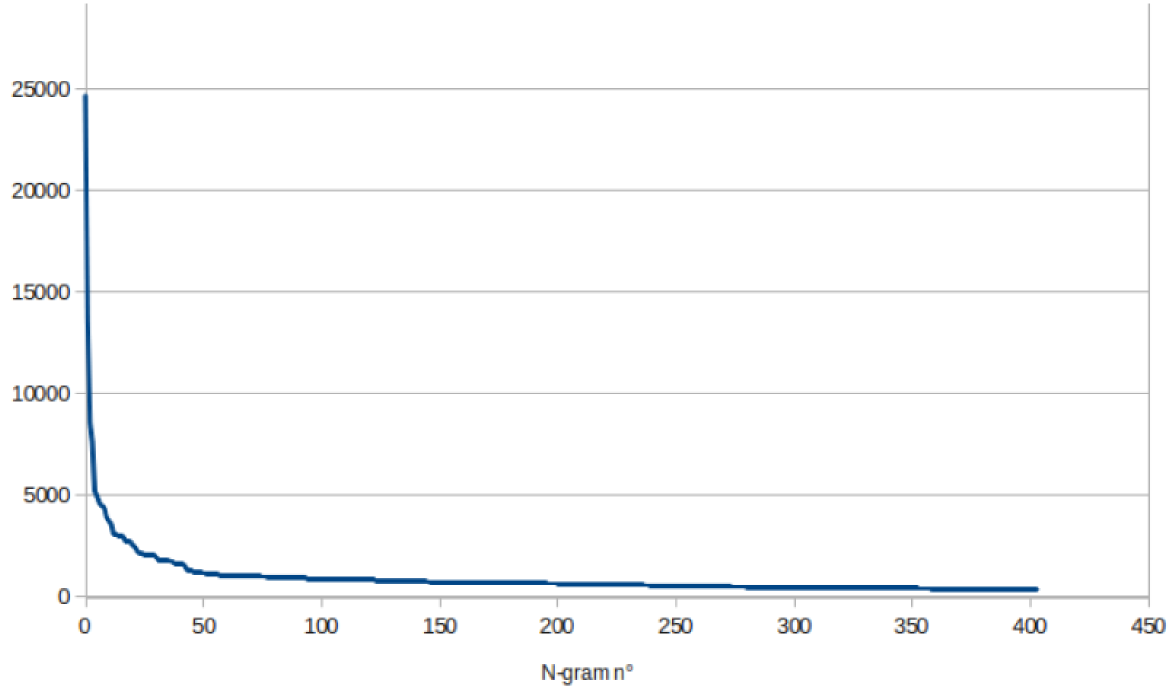


n-gram Trees

Small-scale Sys Surveillance

Anomaly Detection : sys call n-gram Analysis (3)

Possible optimization : Sorted n-gram Tree



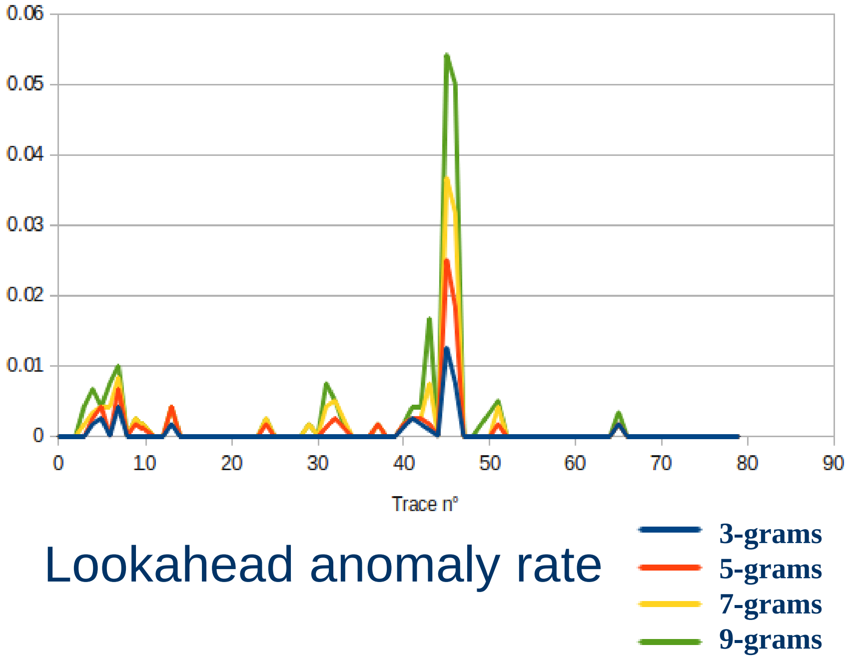
Most frequent n-grams of Angrybirds game

- N-grams sorted according to their frequency inside the normal model
- => Improve analysis time

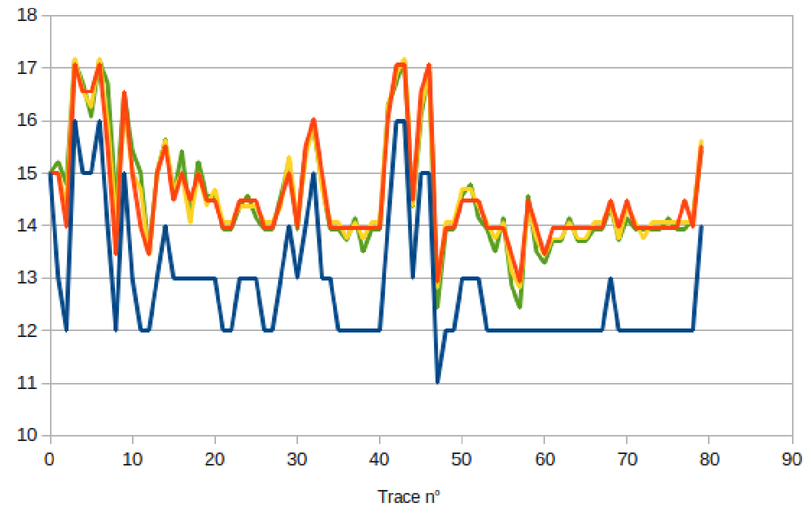
Small-scale Sys Surveillance

Anomaly Detection : sys call n-gram Analysis (4)

Experimental results – injecting 3 function calls in an open source application



Lookahead anomaly rate



n-gram Tree anomaly rate

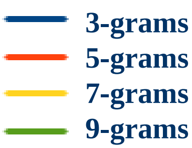
3 new function calls injected in the application : traces 43-48

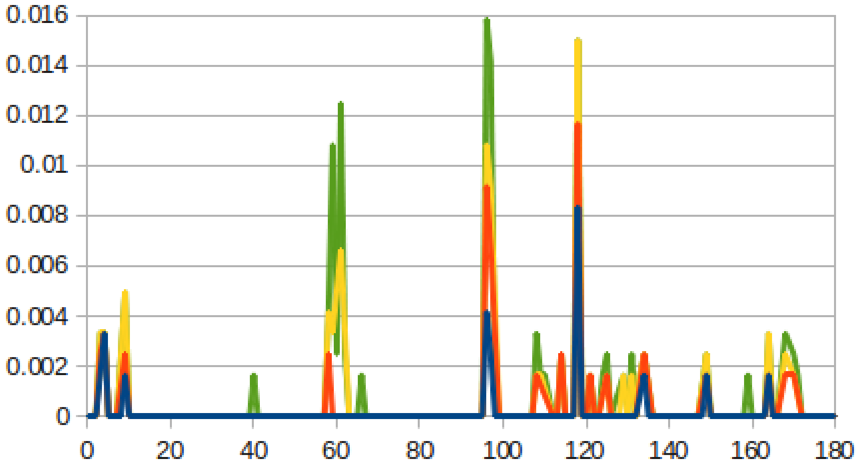
Small-scale Sys Surveillance

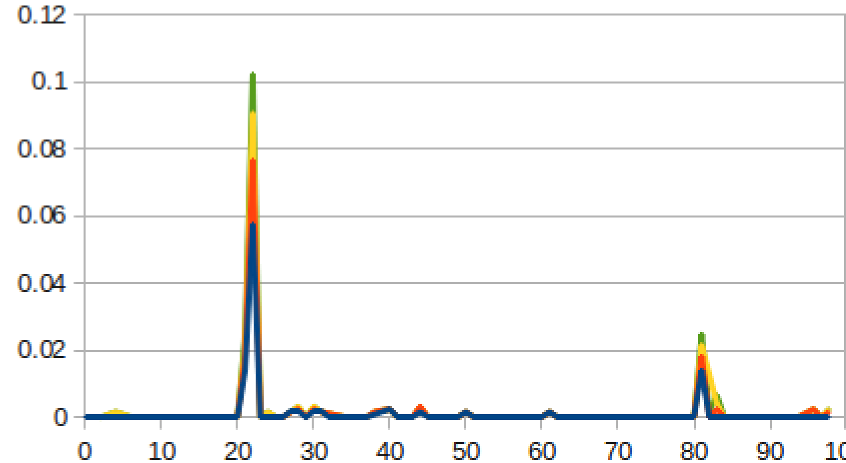
Anomaly Detection : sys call n-gram Analysis (5)

Experimental results – Lookahead model

Angrybird maliciously updated by Droid-KungFu malware







Maliciously updated Angrybird

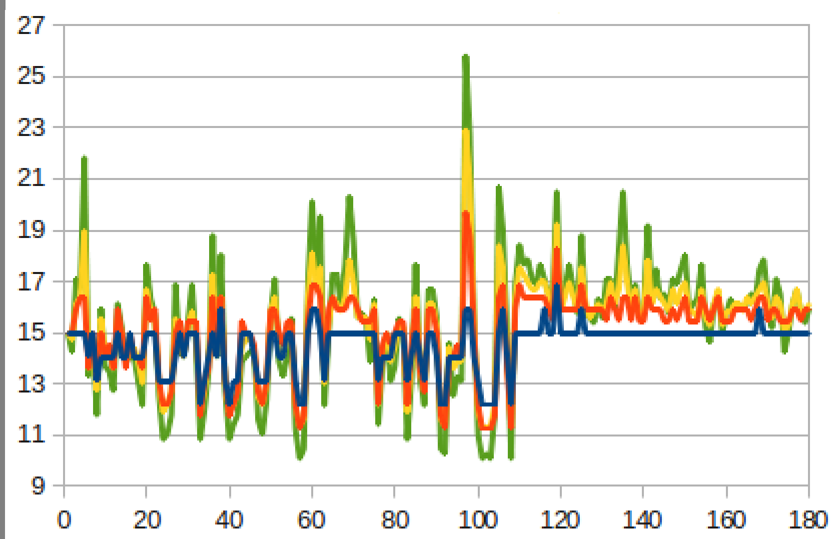
"Safe" update of Angrybird

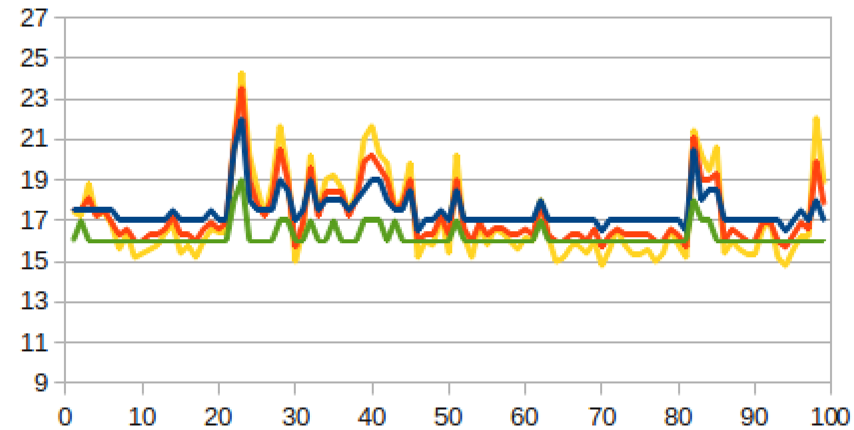
=> Windows >= 5 are good candidates for anomaly detection

Anomaly Detection : sys call n-gram Analysis (6)

Experimental results – n-grams Model

Angrybird maliciously updated by Droid-KungFu malware





Maliciously updated Angrybird

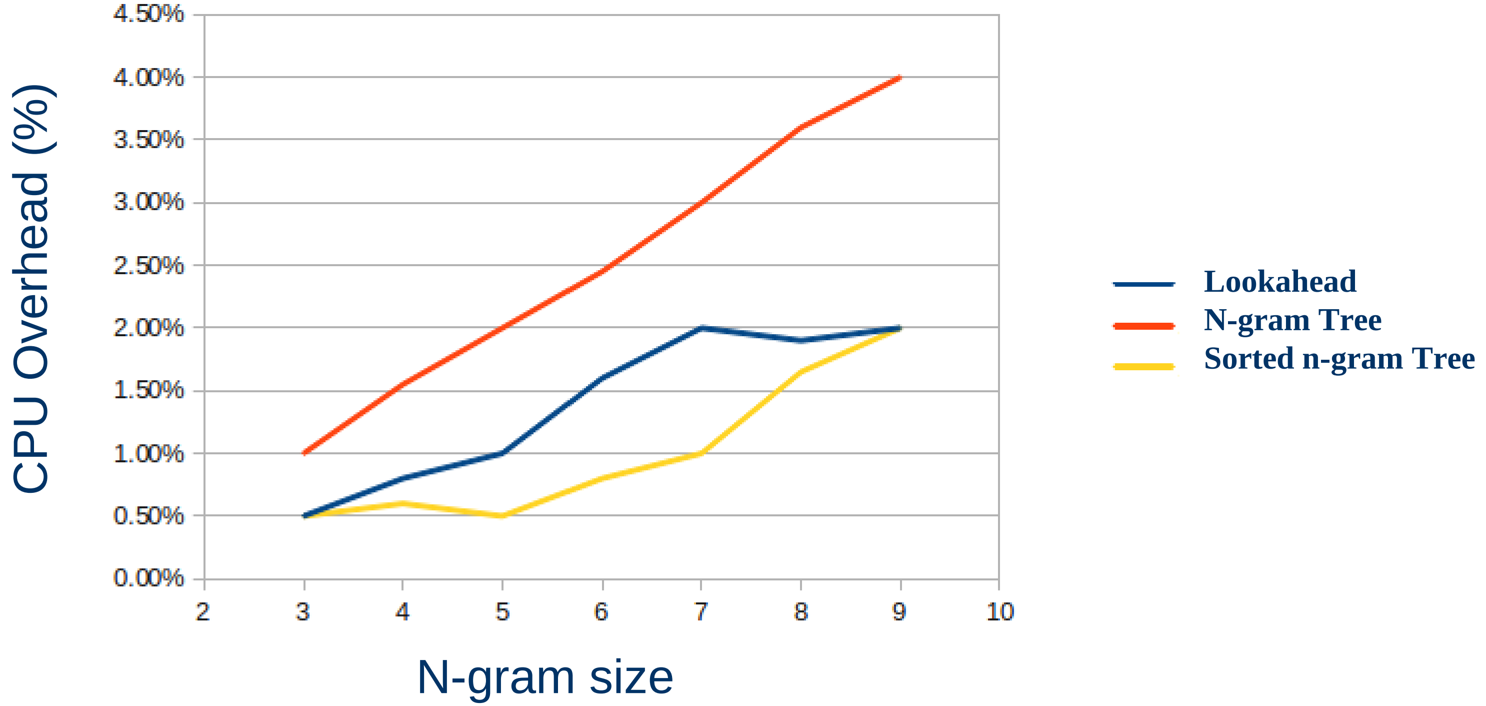
"Safe" update of Angrybird

=> Windows >= 5 are good candidates for anomaly detection

Small-scale Sys Surveillance

Anomaly Detection : sys call n-gram Analysis (7)

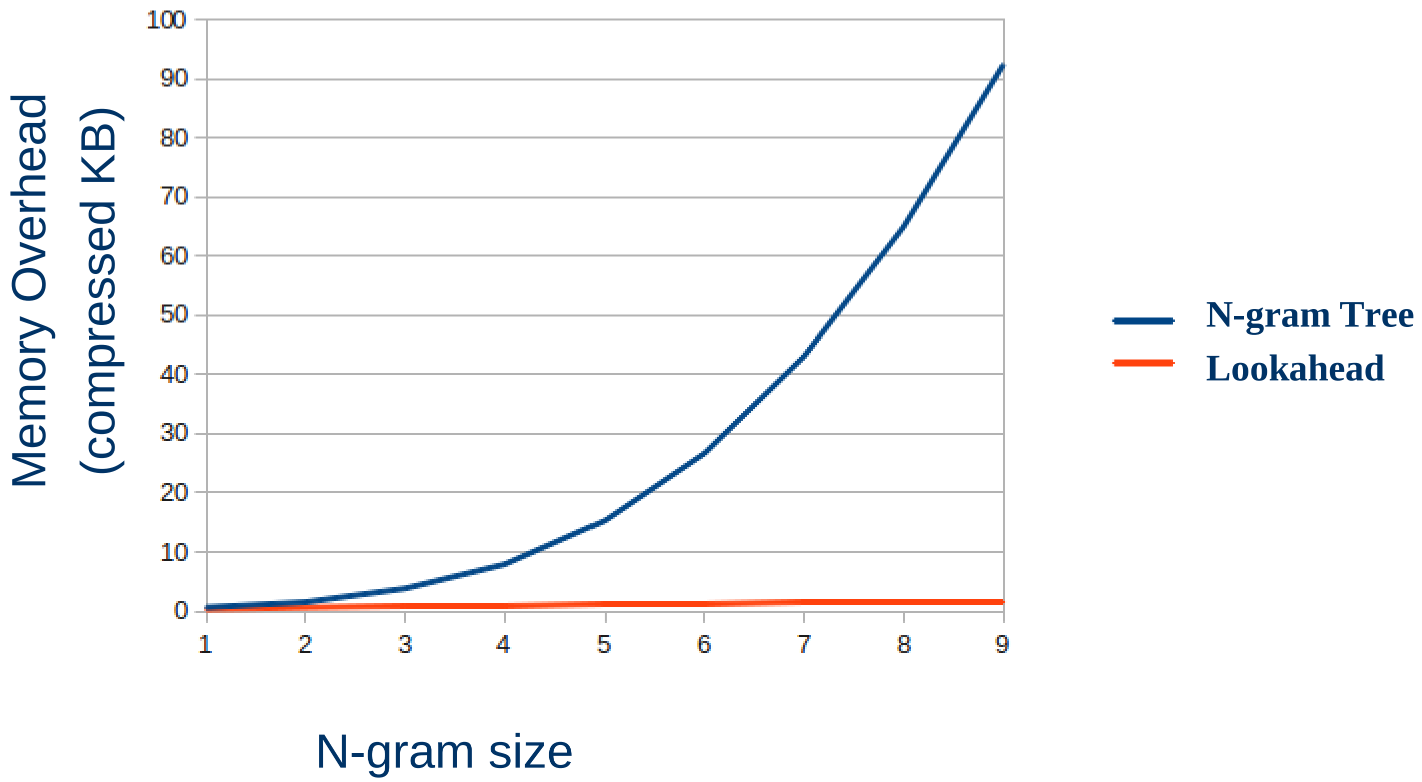
- CPU Overhead



Small-scale Sys Surveillance

Anomaly Detection : sys call n-gram Analysis (8)

- Memory Overhead



Small-scale Sys Surveillance

Signature vs. Anomaly Detection - Challenges

Signature-based detection

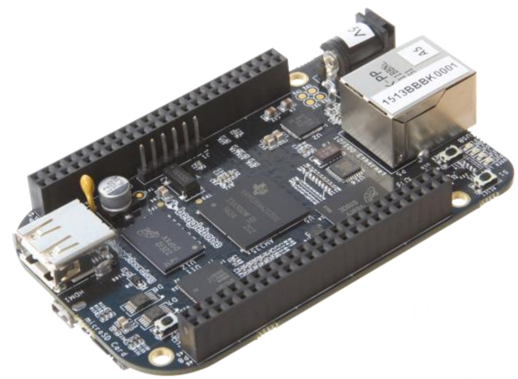
- *Best Multi pattern matching algorithms?*
- *Optimization:* data structures and algorithms, compression, parallel programming, etc.
- *Pragmatic approach:* periodicity, prioritized / specialized signatures, devices collaboration, alert-based, etc.
- Need for Cloud/Server: signatures Database storage, Remote scan.

Anomaly-based detection

- *Machine Learning algorithms:* accuracy (eg. false positives), overhead (eg., memory and power, etc.
- *Adaptive approach*: resource usage of the device, different speeds of the same algorithm, different algorithms, etc.
- *Need for remote Cloud/Server:* traces storage and exchange

Small-scale Sys Surveillance

Evaluation Boards



- PandaBoard, BeagleBoards

- Arndale Board, OMAP5432

Small-scale Sys Surveillance

Evaluation Boards : Use cases

BeagleBone Black:

- Spectrum Analyzer http://www.youtube.com/watch?v=6YhrKMBrJ2g
- Motor Controller http://www.youtube.com/watch?v=34xJIR-mD4A
- Game console http://www.youtube.com/watch?v=U4P_s-7dDRQ
- Web server http://www.youtube.com/watch?v=CDhyVdpXuqQ

Beagleboard-XM:

- Robot Controller http://www.youtube.com/watch?v=FZKtQLj8NLE
- Motor controller http://www.youtube.com/watch?v=bahmjwWKWIo
- Domotic Control System http://www.youtube.com/watch?v=eIAWYCFv0Rw

Pandaboard ES:

- Robot http://www.youtube.com/watch?v=ZWbZBBs9WSs

Small-scale Sys Surveillance

OMAP SOC

	BeagleBone	Overo® FE COM (Gumstix)	Gumstix (DuoVero) Zephyr COM
Manuf.	BeagleBoard.org	Gumstix Inc	Gumstix Inc
CPU	AM335x, 720MHz ARM **Cortex-A8**	OMAP 3530, 600 MHz ARM **Cortex-A8**	OMAP4430, Dual-Core : 1 GHz, **Cotex-A9**
GPU	NEON (SIMD) 2D/3D graphics	OpenGL POWERVR SGX for 2D and 3D graphics acceleration	PowerVR SGX540 ™
Memory	256 MiB DDR2 4GB microSD, Cloud9 IDE on Node.JS	**512 MB RAM** **512 MB NAND** microSD slot	**RAM : 1GB** microSD slot
Features	USB client and Host, **Ethernet**, 2x 46 pin headers, Power consumption 2w	Bluetooth and 802.11b/g, Performance up to 1,400 Dhrystone MIPS, Powered via expansion board (Overo series or custom) connected to dual 70-pin connector	**Ethernet** (10/100 Mbps) **Wifi**, Bluetooth, USB OTG Power: SmartReflex technologies
OS	**Android, Linux**	**Linux** distribution pre-installed. **Android**	**Linux, Android**
Size	76.2 ×76.2 ×16mm	58mm x 17mm x 4.2mm	58mm x 17mm x 4.2mm

Small-scale Sys Surveillance

Military Smartphone/Platforms

	Nautiz X1	Sabre-Tooth	SCORPION H2
SOC	OMAP (TI)	MediaTek	Qualcomm
CPU	OMAP 4430, **dual core**, (1 GHz)	MT6515, **dual-core** (1 GHz)	Snapdragon S3, **dual core**(1.5GHz)
Memory	RAM : 512 MB, flash: 4 GB, MicroSD card slot	RAM : 512 MB MicroSD card slot (32GB)	RAM : 1MB, Flash : 16 GB, expandable to 32GB micro SD
Connectivity	GSM, CDMA, GPS, Bluetooth, 802.11 b/g/n WiFi	Wi-Fi: 802.11 b/g/n, 2G: GSM, Bluetooth	3g/4G compatible, Wi-Fi 802.11 and Bluetooth, GPS
Connectors	E-compass and G-Sensor, Extended battery, Vehicle cradle, 5-megapixel camera, LED flash	2x GSM, Micro SD Card Slot, Micro USB, Gravity and Linear Acceleration Sensor	tactical data radios, extended battery life
features	survive humidity, vibration, drops /extreme temperatures. waterproof and impervious to dust and sand. runs **Android 4.0**	Water Resistant, Shockproof, Dustproof, Battery Standby: 72 Hours, dimensions: 136x75x18mm , weight: 144g Runs **Android 2.3**	run/charge simultaneously via USB port, batteries, or vehicle power. vibration, shock, drop, humidity **Runs Android 4.0**

Agenda

- Project Presentation
- Why surveillance of/with small-scale systems?
- Surveillance of Small-scale Systems
- Surveillance with Small-scale Systems
- Project Summary
- Feedback?

||Small-scale Sys 4 Surveillance

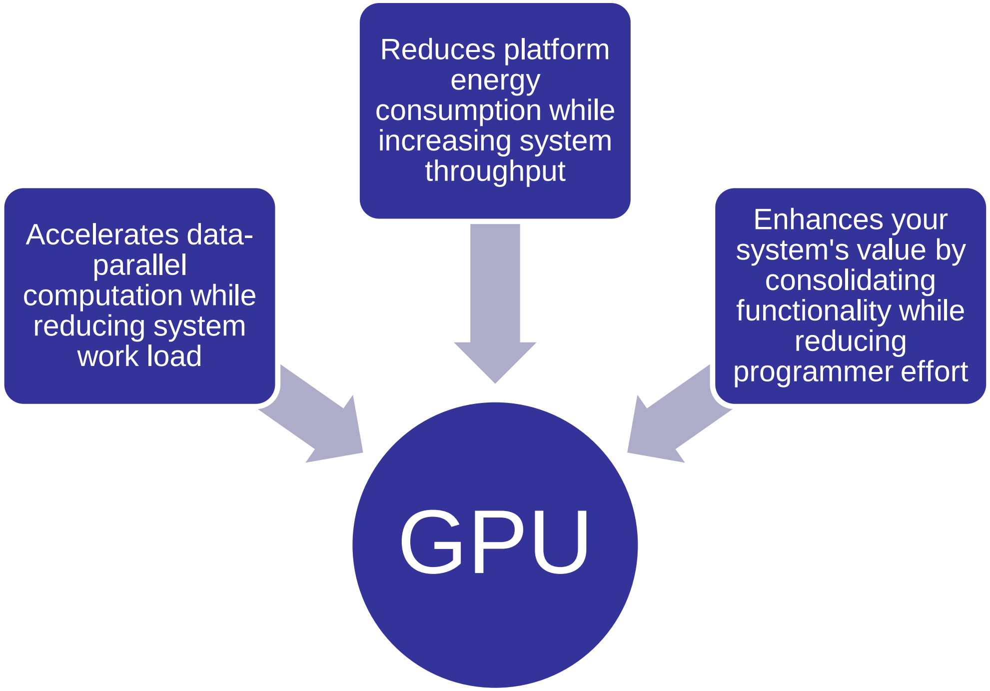
- Massively parallel small-scale embedded systems
- Opportunities for better performance of surveillance techniques

||Small-scale Sys 4 Surveillance

Massively parallel small-scale embedded systems

GPUs : Expanding field for massively parallel computing

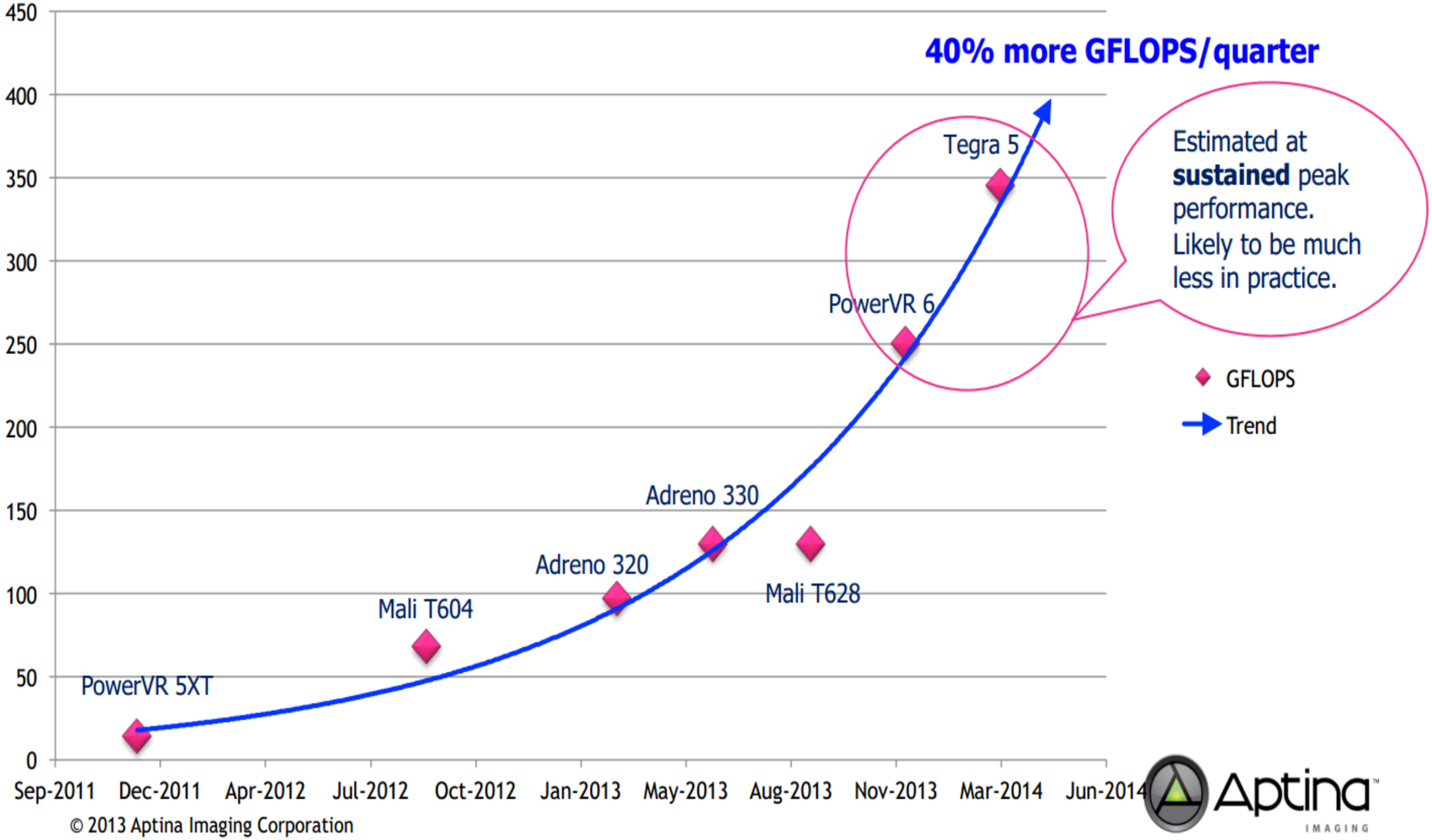
- A Graphics Processing Unit: A co-processor that takes on graphical calculations and transformations so that the main CPU does not have to be burdened by them
- GPUs are the most used platforms for massively parallel programming systems.



Small-scale Sys 4 Surveillance

Massively parallel small-scale embedded systems

Evolution of Embedded GPUs



Mobile Compute driving Imaging use cases

- Requires ***significant*** computing over ***large*** data sets

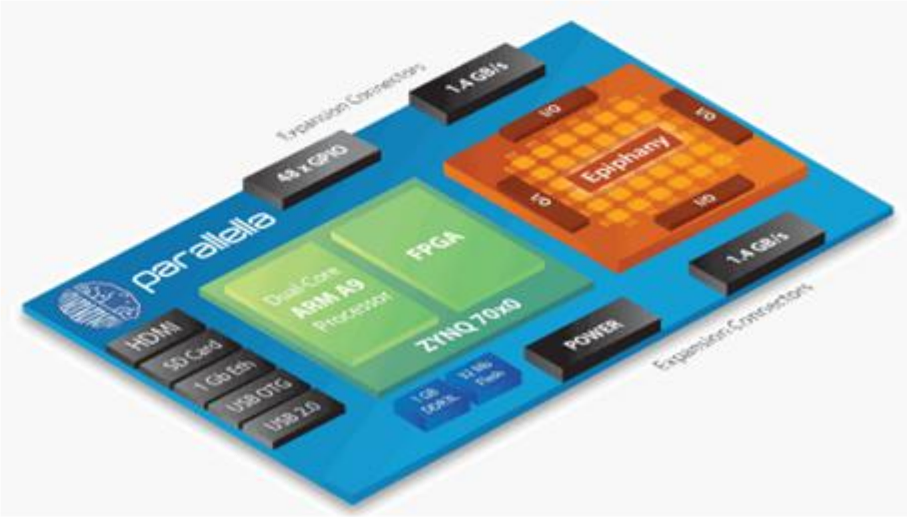
Computational Photography → Face, Body and Gesture Tracking → 3D Scene/Object Reconstruction → Augmented Reality (Time)

Small-scale Sys 4 Surveillance

Massively parallel small-scale embedded systems

Parallella : Super computing for everyone

- Project goal: to democratize access to parallel computing through providing an affordable open hardware platform and open source tools
- The Parallella platform is an open source, energy efficient, high performance, credit-card sized computer based on the Epiphany multicore chips developed by Adapteva.



||Small-scale Sys 4 Surveillance

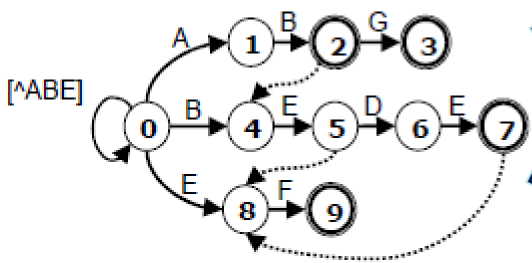
Opportunities for better performance of surveillance

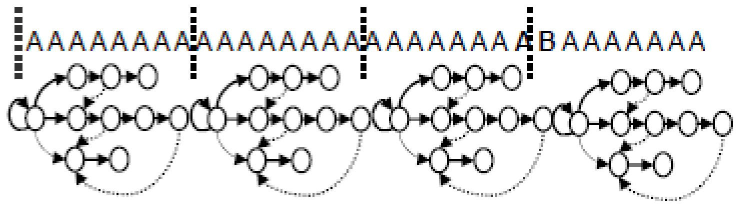
- Accelerating/optimizing surveillance Using Multithreaded Algorithms

  Cheng-Hung Lin; Sheng-Yu Tsai; Chen-Hsiung Liu; Shih-Chieh Chang; Shyu, J.-M., "***Accelerating String Matching Using Multi-Threaded Algorithm on GPU***," Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE , vol., no., pp.1,5, 6-10 Dec. 2010

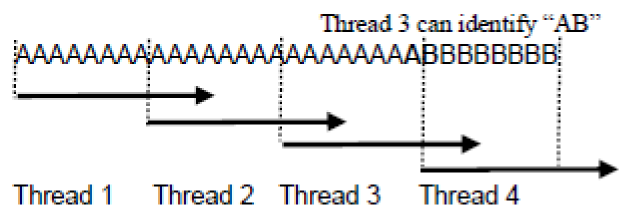
||Small-scale Sys 4 Surveillance

Opportunities for better performance of surveillance





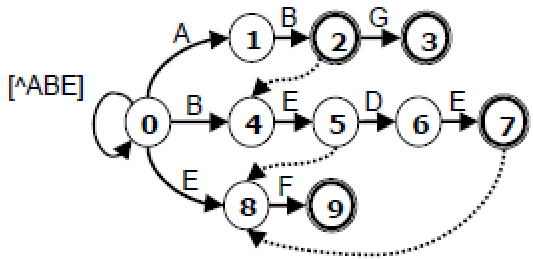
Parallel processing of input stream with
Boundary detection problem!

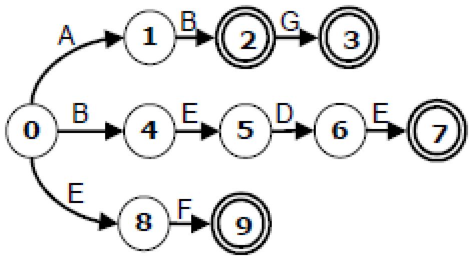


Parallel processing of input stream with
Overlapped segments

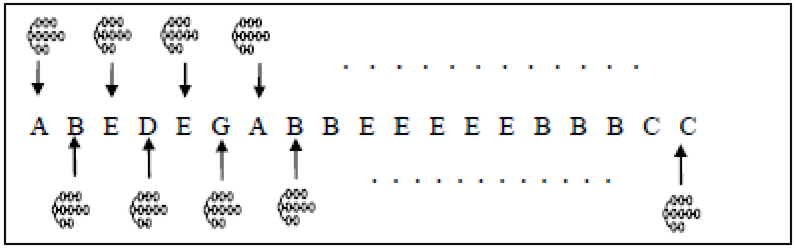
Small-scale Sys 4 Surveillance

Opportunities for better performance of surveillance





AC automaton without failure transitions



||Small-scale Sys 4 Surveillance

Opportunities for better performance of surveillance

Promising improvements !!!

TABLE 1: THROUGHPUT COMPARISON OF THREE APPROACHES

Input streams	CPU_AC *Throughput (KBps)*	Direct_AC *Throughput (KBps)*	PFAC *Throughput (KBps)*
Normal Case	997	6,428	3,963,966
Virus Case	657	4,691	3,656,217
Ratio	1	~6.4	**~4000**

TABLE 2: MEMORY COMPARISON

	Conventional AC			PFAC			
	states	*transitions*	*memory (KB)*	*states*	*transitions*	*memory (KB)*	*Reduction*
Snort rule*	8,285	16,568	143	8,285	8,284	114	21%
Ratio	1	1	1	1	0.5	0.79	

* The Snort rules contain 994 patterns and total 22,776 characters.

||Small-scale Sys 4 Surveillance

Opportunities for better performance of surveillance

	Work	Mechanism
Malware detection	GrAVity: A Massively Parallel Antivirus Engine	Applying a signature filter on GPU
	A Taxonomy and Comparative Evaluation of Algorithms for Parallel Anomaly Detection	Combining different classes of anomaly detection algorithms and address the question of which combination of existing anomaly detection algorithms achieves the best detection accuracy.
	An Efficient Parallel Anomaly Detection Algorithm Based on Hierarchical Clustering	Parallel processing of training and predicting phase Both phases have the same excellent detection performance with serial processing, and it also has better real time performance than serial processing
Pattern Matching	Accelerating String Matching Using Multi-threaded Algorithm on GPU	Proposing a novel algorithm that reduces the complexity of Aho-Corasick Algorithm The new algorithm on GPUs achieves up to 4,000 times speedup compared to the AC algorithm on CPU
	A gpu-based multiple-pattern matching algorithm for network intrusion detection systems	A GPU-based pattern matching algorithm for NIDS has been proposed in this work. The proposed pattern matching algorithm is based on the concept of WM algorithm. The performance of the proposed approach is around twice of that of the MWM algorithm employed in Snort and can be applied on host-based antivirus systems.
	Bit-Parallel Multiple Pattern Matching	Extension of the bit-parallel Wu-Manber algorithm to combine several searches for a pattern into a collection of fixed-length words. Presenting an OpenCL parallelization of a redundant index on massively parallel multicore processors, within a framework of searching for similarities with seed-based heuristics. Some speedups obtained with gpu are more than 60× on cpu.

Agenda

- Project Presentation
- Why surveillance of/with small-scale systems?
- Surveillance of Small-scale Systems
- Surveillance with Small-scale Systems
- Project Summary
- Feedback?

Ongoing Activities

- Signature based detection:
  - Experimenting existing tools :
    - Antimalware for Smartphone
    - Antimalware for embedded systems
  - Optimized pattern matching algorithms

- Anomaly-based detection:
  - Features selection
  - Lightweight and optimized algorithms
  - Adaptive algorithms
  - Experimenting and adapting algorithms developed by collaborators: Concordia University


Time for
Feedback