Mauro Garofalo

Tutor: Prof. Giorgio Ventre

XXIX Cycle - I year presentation

Flow-Based Anomaly Detection System for mobile malware detection

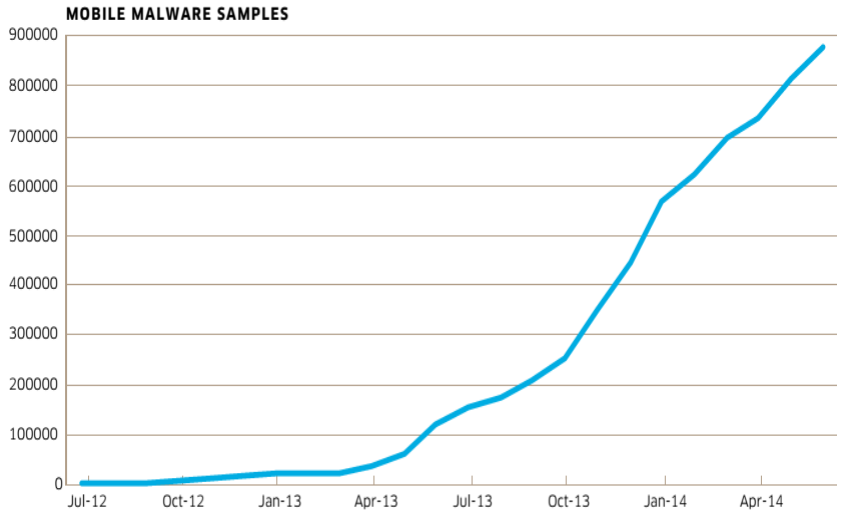
Who am I?

- Graduation: MSc in Computer Engineering
- DIETI Group: COMICS
- Fellowship: MIUR research grant
- Collaboration: DAME

Why so serius about mobile malware?

- Prevalence of mobile devices
- Mobile data traffic grew up to 2.5 EB per month in 2014 [2]
- Work and personal sensitive or confidential information on devices
- User's lack of smart smartphone risk awareness

Norton 2014. Internet Security Threat Report 2014 [4]



KINDSIGHT SECURITY LABS MALWARE REPORT – H1 2014

Also malware exploit more and more network connectivity

Solutions

Common malware detection strategies:

- Host-based approach

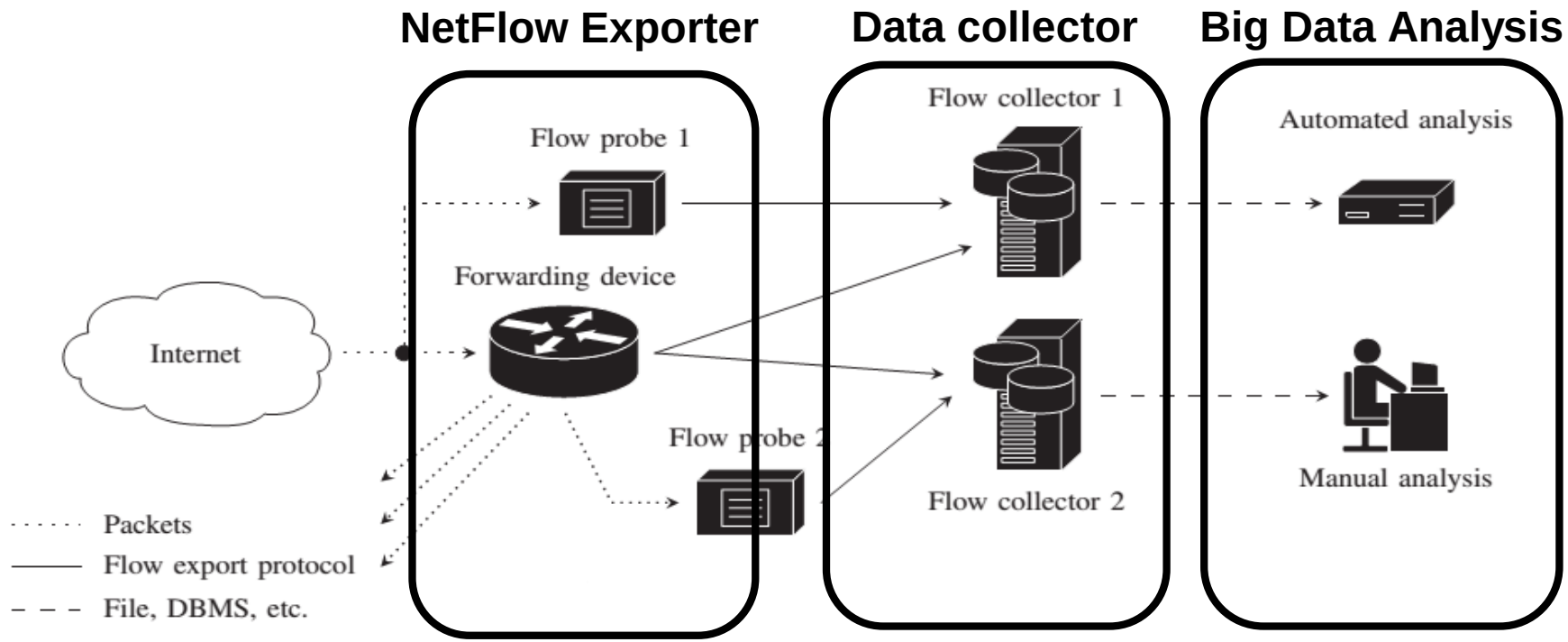
  Analyze the device (running applications, hardware performance in terms of CPU usage, battery consumption, etc.) [4],

- Network-based approach (NIDS)
  - Analyze the network traffic generated by device [5] .
  - Taxonomy: Anomaly Based, Behavior-Based, Protocol-Based and Rule-Based.
  - Approaches:
    - Packet-content based (transport-layer ports, payload, etc.)
    - Flow-based

Flow data contain details like: source and destination IP addresses, port numbers, protocols, etc.

Cisco NetFlow data is generated by network devices like routers and firewalls.

Flow-based Architecture



Our proposal architecture consist of 3 stages:

1. Flow Export, router NetFlow compliant
2. Data collector
3. Data Analysis, using machine learning algorithms

Due the large amount of data involved we can consider this as Big Data

In the first year

- Performance evaluation of machine learning algorithms for parallel computing architectures (GPU) [P2]

Future Work

- Moving to a distributed computing architectures (Hadoop) using “a la Map – Reduce” computing frameworks that enable to use machine learning techniques (Spark)

- The architecture is just defined and its prototype is under construction

Productions

International journal papers

[P1]“DAMEWARE: A Web Cyberinfrastructure for Astrophysical Data Mining”, Brescia M, Cavuoti S, Garofalo M, et al. (2014) Publications Of The Astronomical Society Of The Pacific, Vol. 126, P. 783-797, ISSN: 0004-6280

Conference

[P3] “Acceleration of Machine Learning Models based on GPGPU technology for fast data mining in multidisciplinary physical environments”, Garofalo M, Guarino D, Brescia M, Cavuoti S, Pescapè A, Longo G, Ventre G, Perspectives of GPU computing in Physics and Astrophysics. Università la Sapienza di Roma, Roma 2014

List those in preparation

“A survey on Big Data and Networking”

“An architecture for flow-based anomaly detection based on Big Data analysis tools”

References

- [1] Gartner 2014 http://www.gartner.com/newsroom/id/2791017
- [2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update
- [3] - IDC Q3-2014 http://www.idc.com/prodserv/smartphone-os-market-share.jsp
- [4] Norton 2014. Internet Security Threat Report 2014
- [5] Hien Thi Thu Truong, Eemil Lagerspetz, Petteri Nurmi, Adam J. Oliner, Sasu Tarkoma, N. Asokan, Sourav Bhattacharya, “The Company You Keep: Mobile Malware Infection Rates and Inexpensive Risk Indicators”, 2013, arXiv:1312.3245
- [6] Yajin Zhou; Xuxian Jiang, "Dissecting Android Malware: Characterization and Evolution," Security and Privacy (SP), 2012 IEEE Symposium on , vol., no., pp.95,109, 20-23 May 2012
- [7] Bhuyan, M.H.; Bhattacharyya, D.K.; Kalita, J.K., "Network Anomaly Detection: Methods, Systems and Tools," Communications Surveys & Tutorials, IEEE , vol.16, no.1, pp.303,336, First Quarter 2014
- [8] Zibin Zheng; Jieming Zhu; Lyu, M.R., "Service-Generated Big Data and Big Data-as-a-Service: An Overview," Big Data (BigData Congress), 2013 IEEE International Congress on , vol., no., pp.403,410, June 27 2013-July 2 2013
- [9] Zaslavsky, A.; Perera, C.; Georgakopoulos, D., "Sensing as a Service and Big Data", 2013, arXiv preprint arXiv:1301.0159
- [10] Hofstede, R.; Celeda, P.; Trammell, B.; Drago, I.; Sadre, R.; Sperotto, A.; Pras, A., "Flow Monitoring Explained: From Packet Capture to Data Analysis With NetFlow and IPFIX," Communications Surveys & Tutorials, IEEE , vol.16, no.4, pp.2037,2064, Fourthquarter 2014

Next Year

	Credits year 1								Credits year 2	Credits year 3		
		1	2	3	4	5	6					
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	Estimated	Total	Check
Modules	20		3		3		6	12	18	0	30	30-70
Seminars	7		0,6		2	0,9	2,2	5,7	6	0	11,7	10-30
Research	33	10	6,4	10	5	9,1	1,8	42,3	36	60	138	80-140
	60	10	10	10	10	10	10	60	60	60	180	180

Thanks for your attention.