Traditional AV is Dead?

Real-time Machine-Learning Detection of Modern Malware Downloads

Dr. Marco Balduzzi

Jan. 19 2017, Venice

Who am I?

- M.Sc. In Computer Engineering, Ph.D. in System Security
- On top of things since 2002
- Sr. Research Scientist
  - Web, Malware, Privacy, Cybercrime, IoT, Threats
  - *http://www.madlab.it*

Traditional AV is Dead?

- Signature-based VS Statistical-based
- Signature-based malware detection is inefficient
  - Polymorphism, code obfuscation, packing
  - Analysis is time consuming (static , dynamic)
  - URL blacklists lag behind (DGA botnets)

Traditional AV is Dead?

- Local awareness VS Global awareness
- Local: Looks at one potential malicious file/URL at the time
- Global: Leverages a global situation awareness, e.g. relationships among files and machines

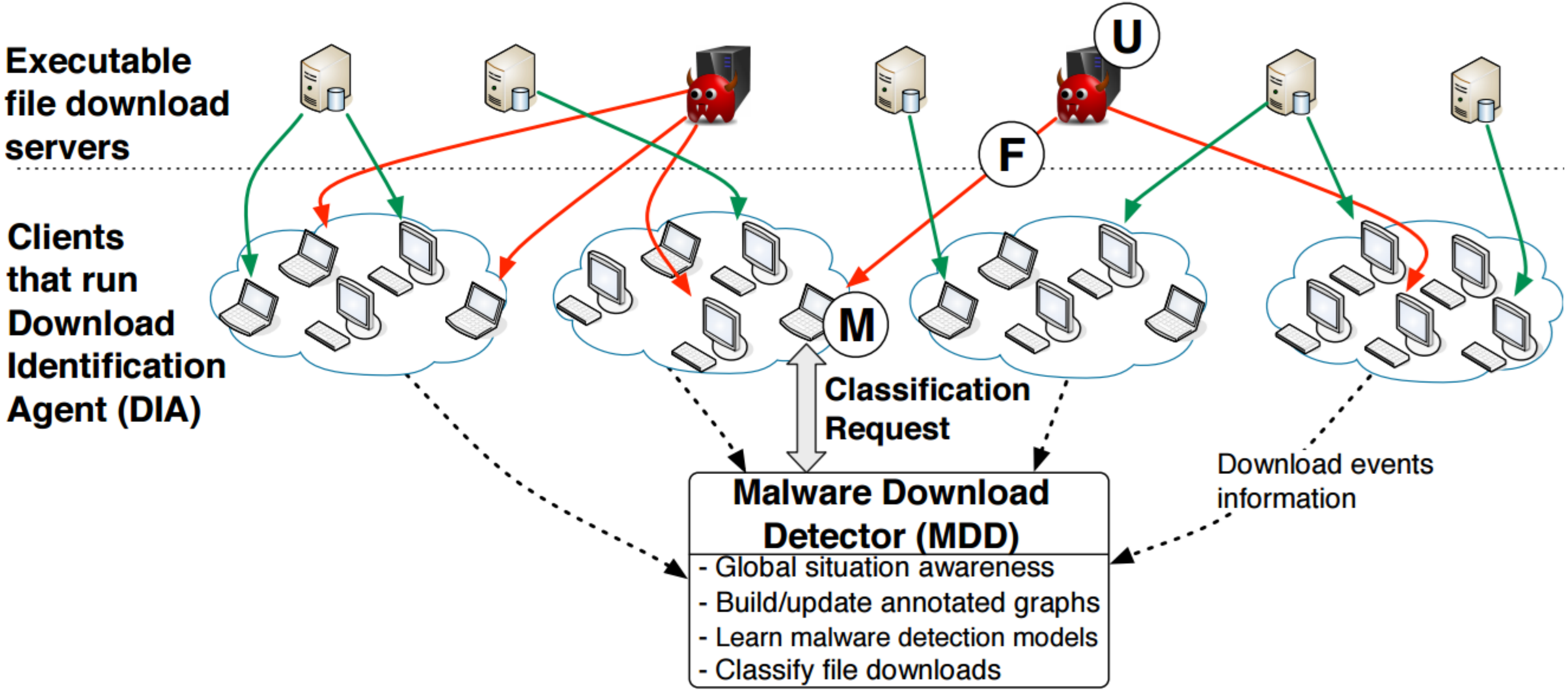
Our X-Gen Approach

- Content agnostic: Files’ or webpages’ content analysis is **not** needed
- Use of relationship patterns
  - 3W: “Who-What-Where” = who downloads what from where
  - Combination of system- and network-level informations
- Statistical-based detection
- Global situation awareness

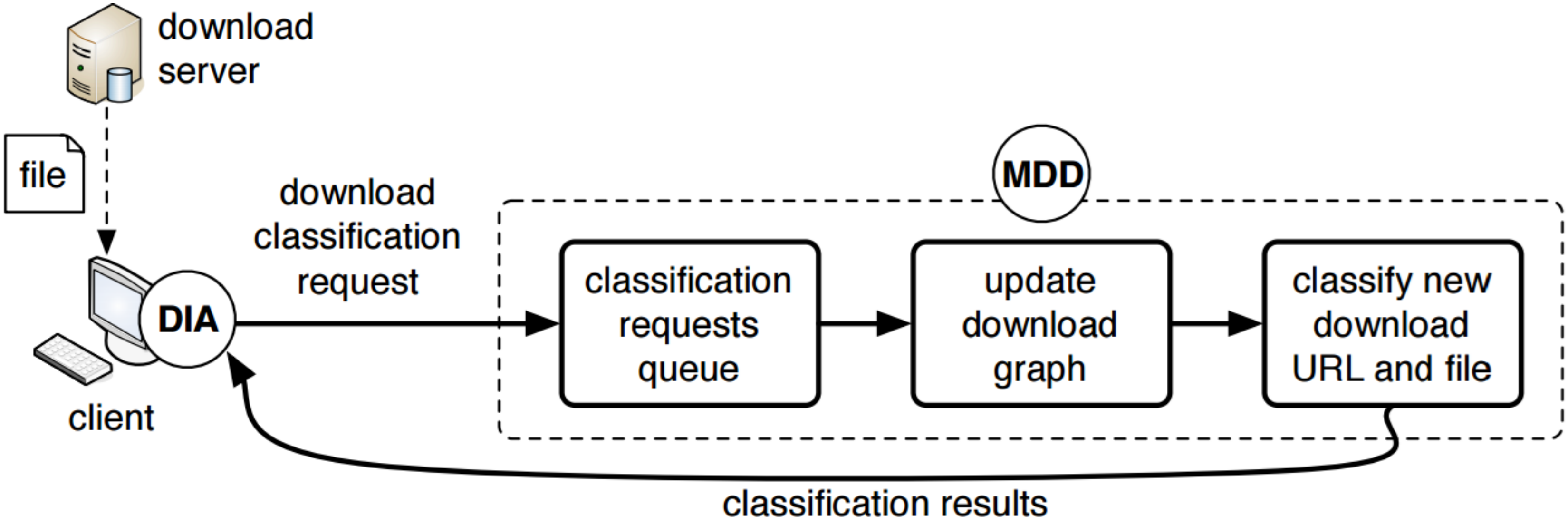
Benefits

- Concurrent detection of malicious download events, i.e. files and URLs
- Complementary approach to existing solutions, e.g. static and dynamic detection
- Efficiency, real-time detection against *unknown* and modern threats

System Overview

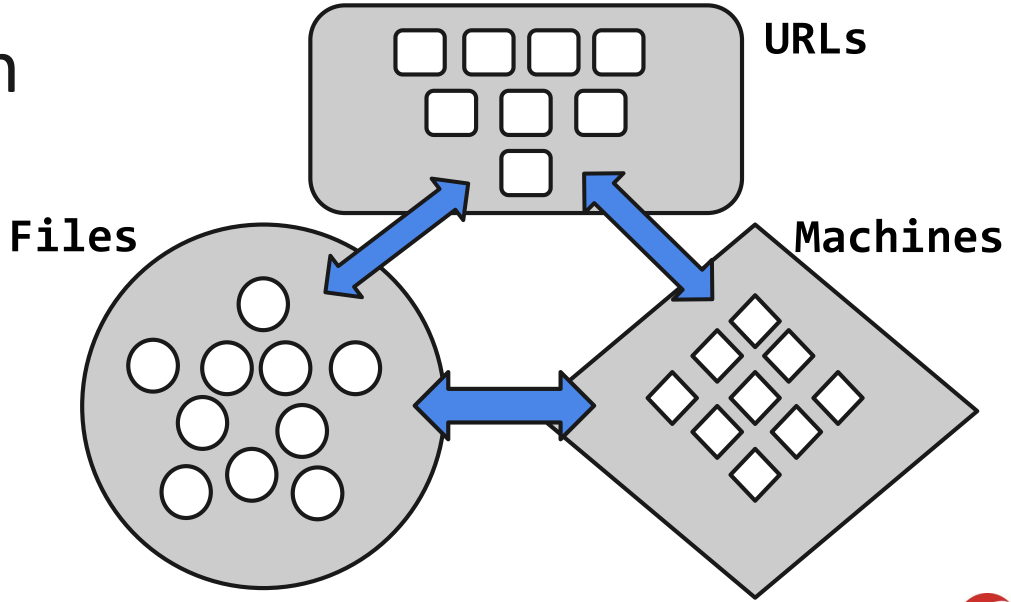


System Overview

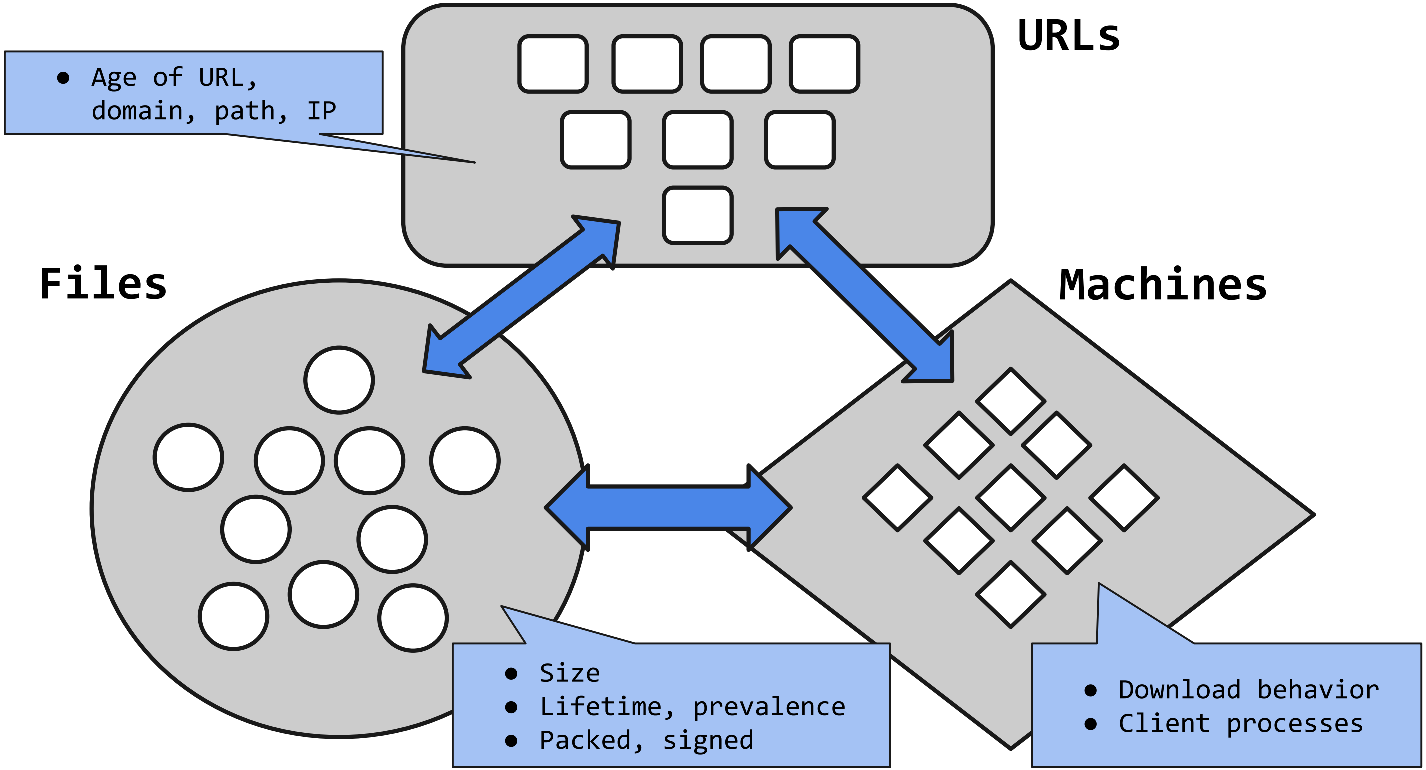


Download Graph

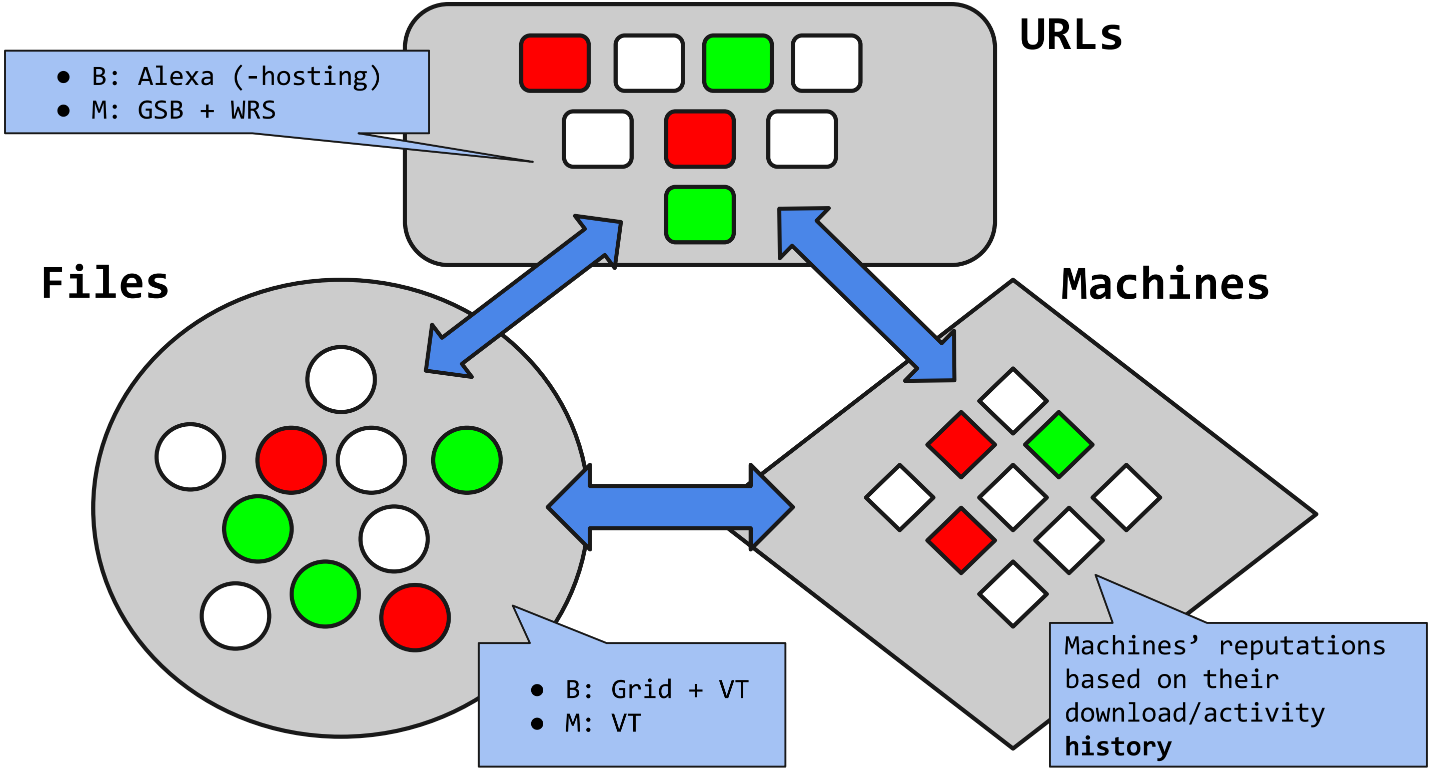
- A representation of a collection of download events
- Global situation



Intrinsic Features

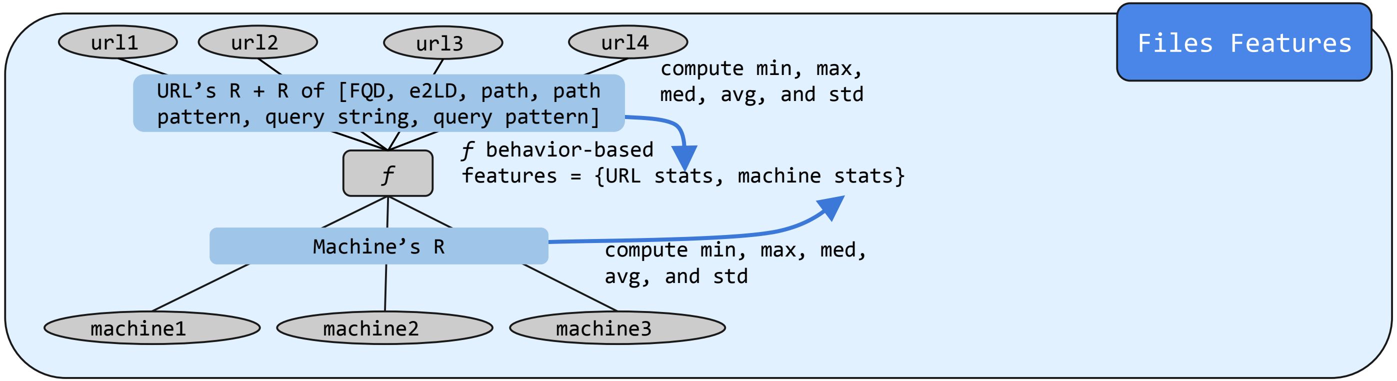


Labels

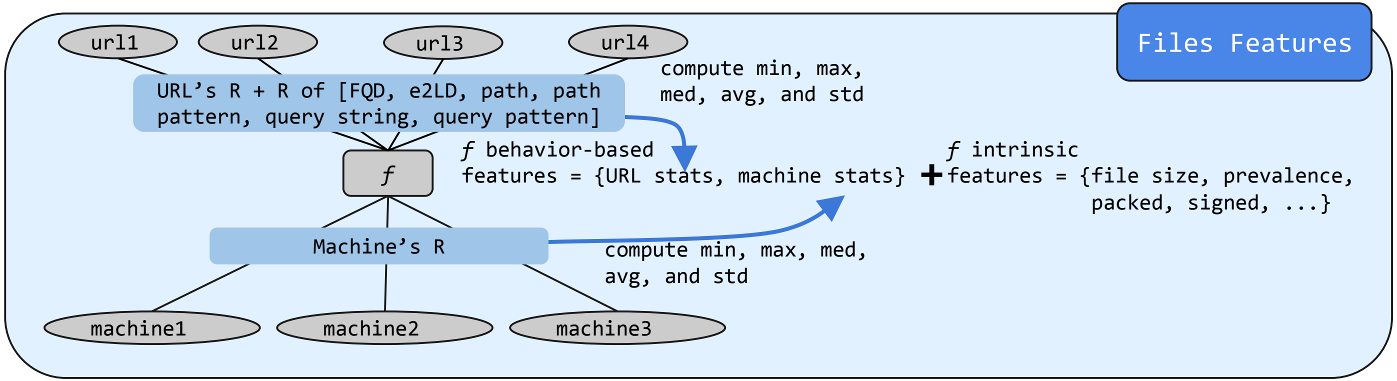


URLs
• B: Alexa (-hosting)
• M: GSB + WRS
Files
Machines
• B: Grid + VT
• M: VT
Machines' reputations based on their download/activity **history**

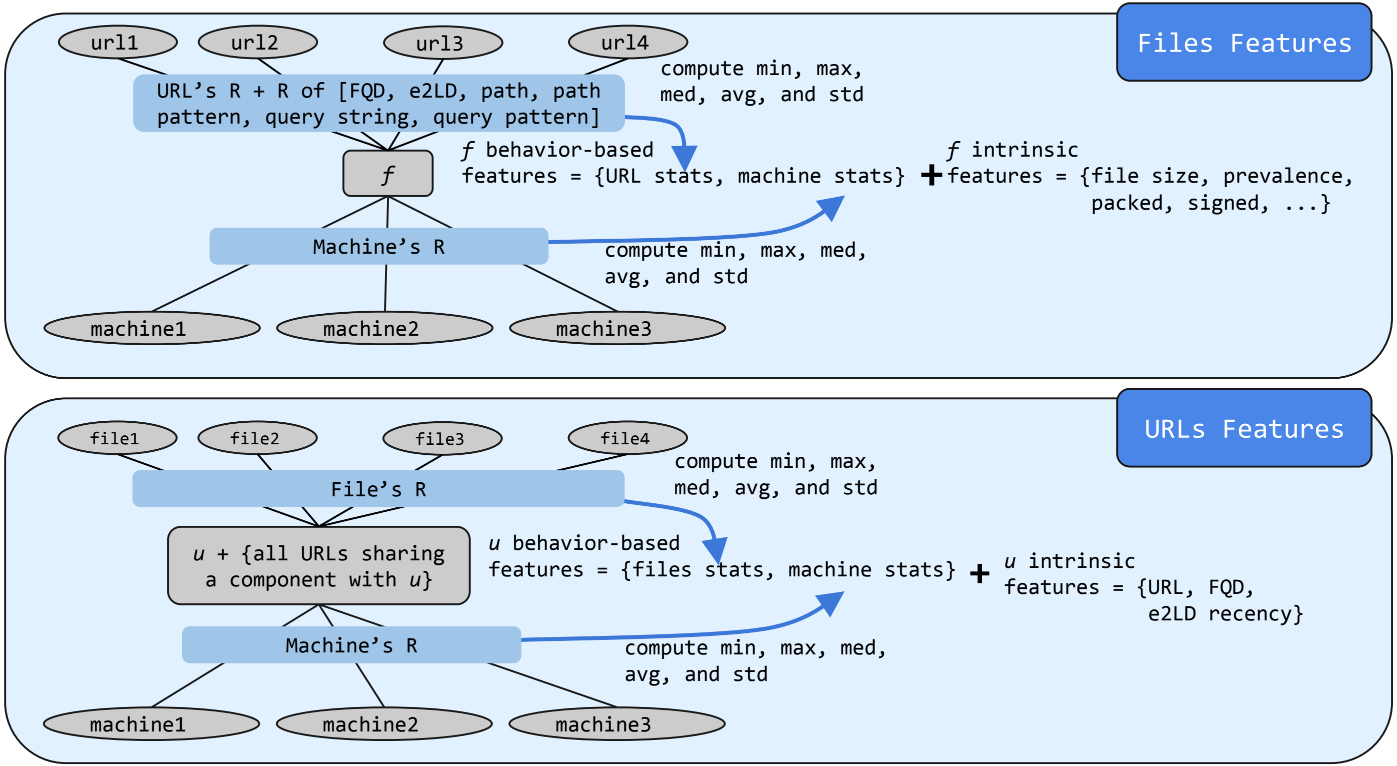
Features Engineering



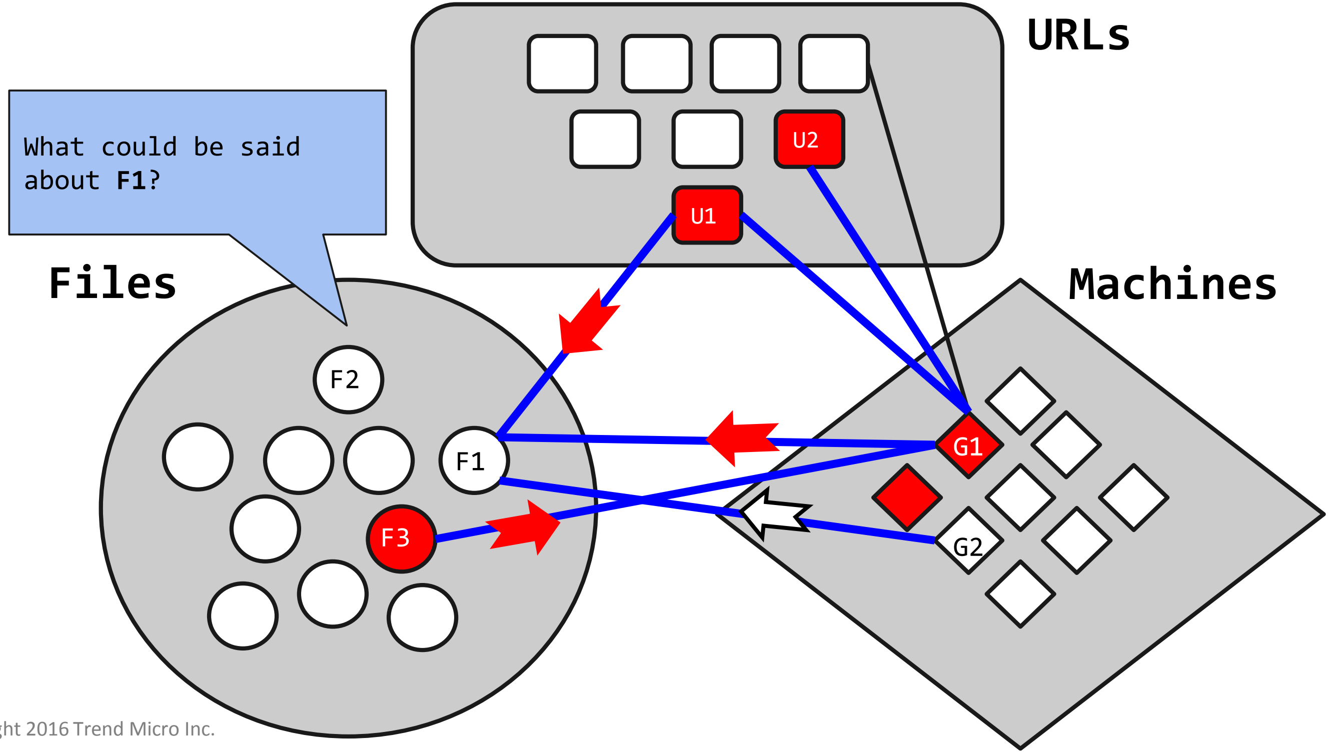
Features Engineering



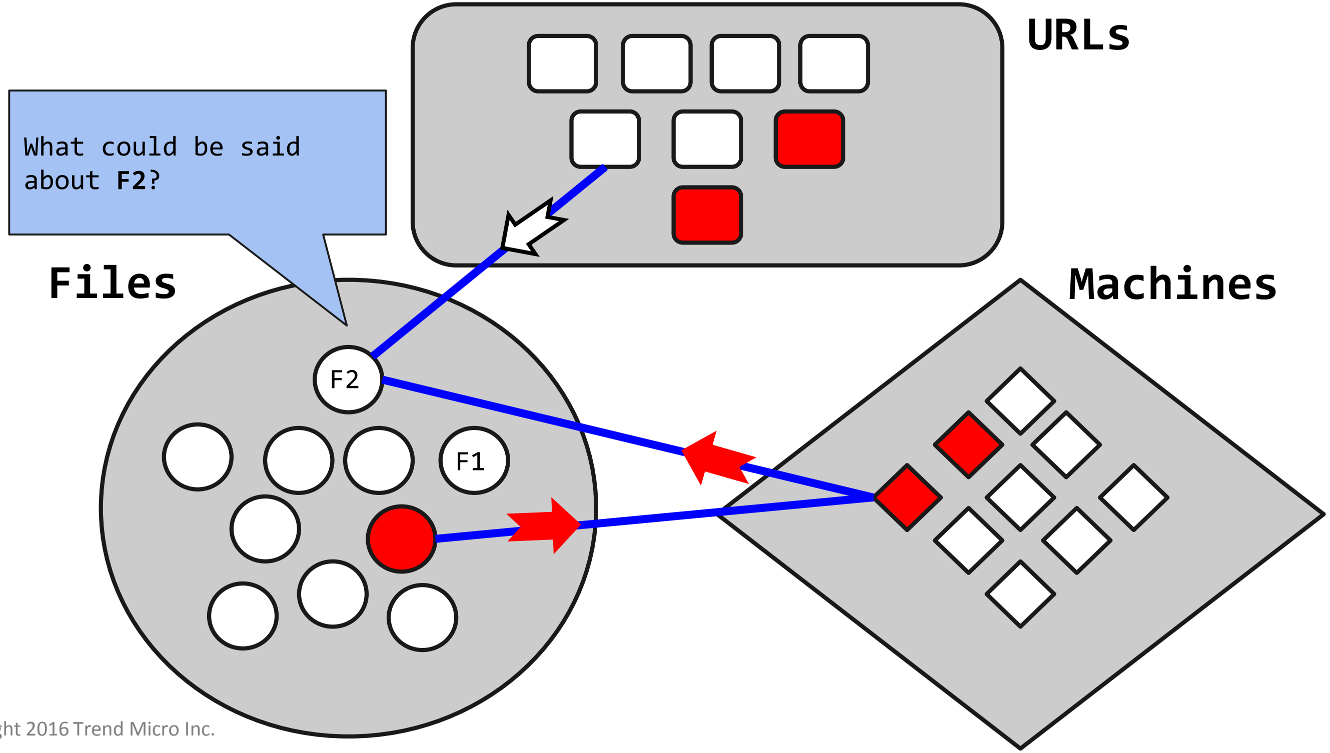
Features Engineering



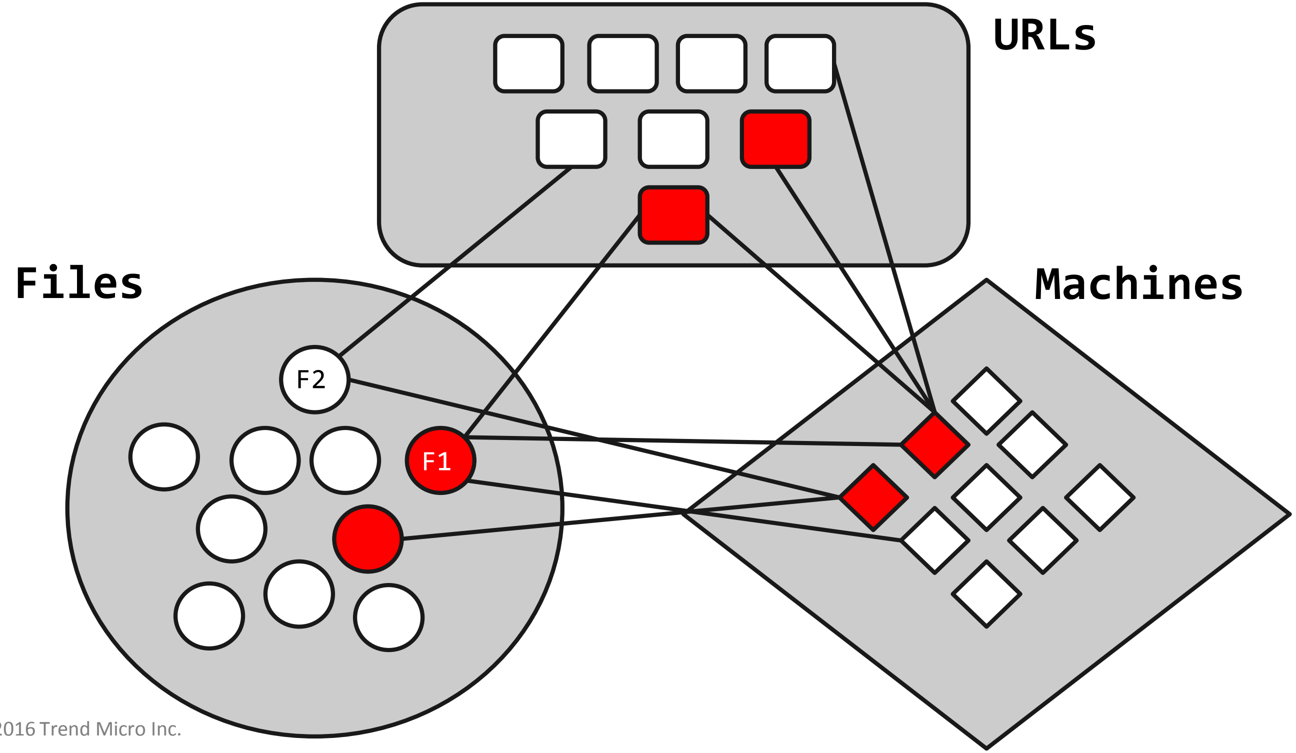
Statistical Classifier, example



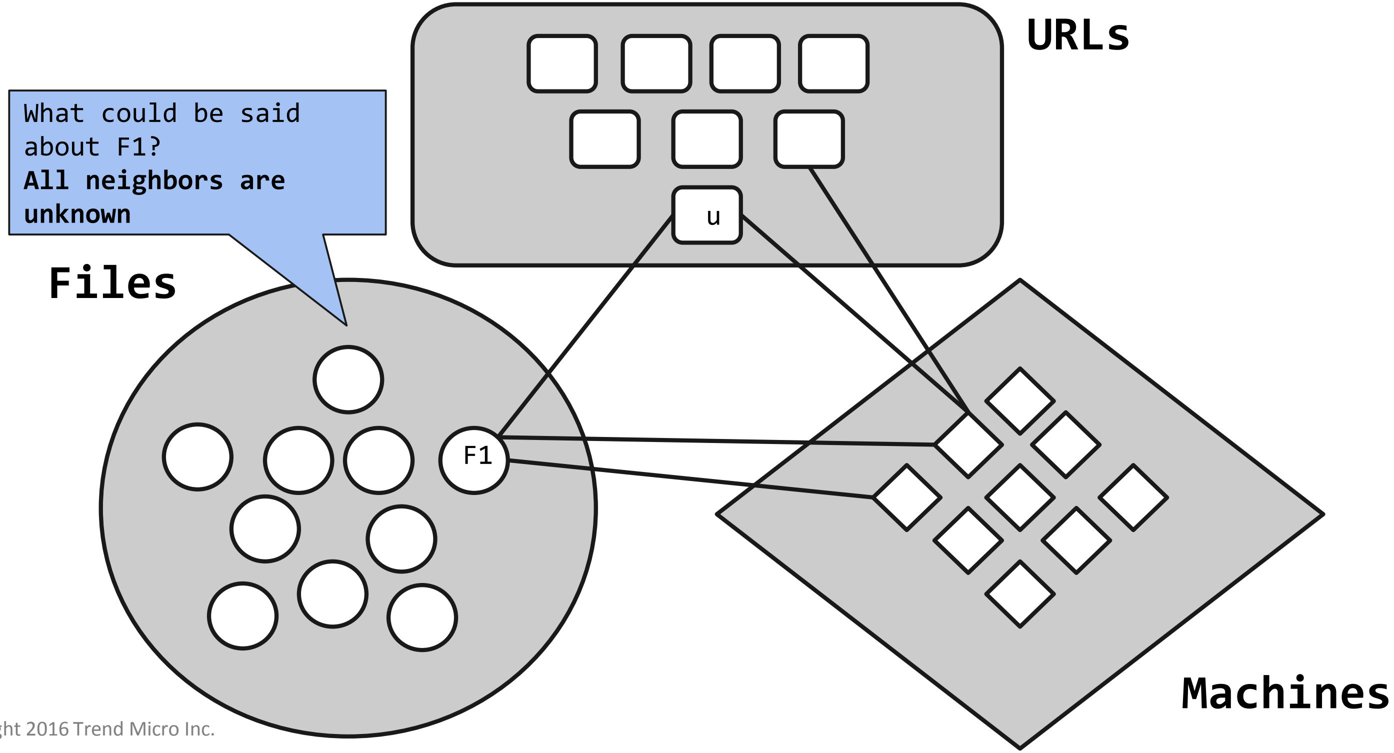
Statistical Classifier, example



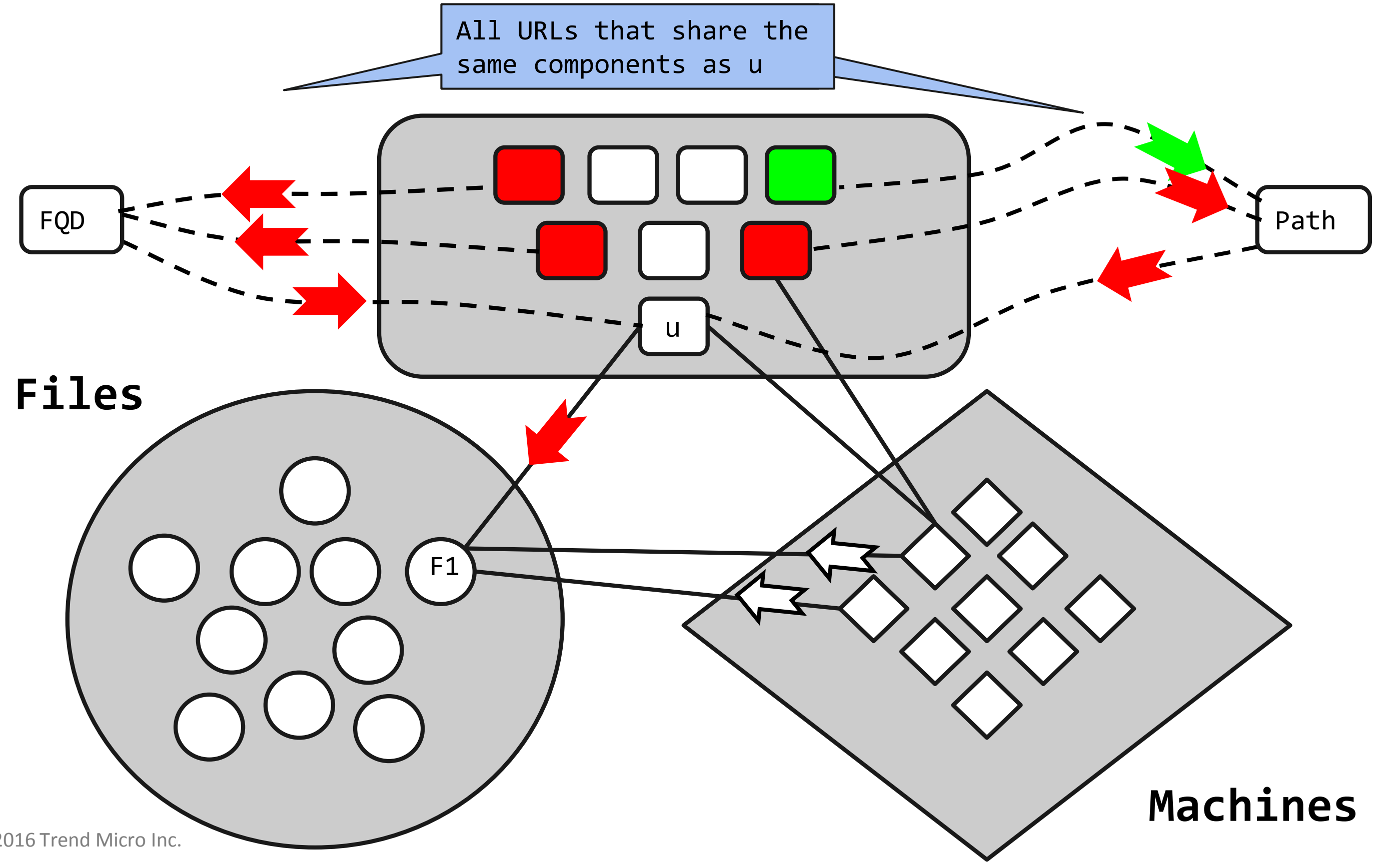
Statistical Classifier, example



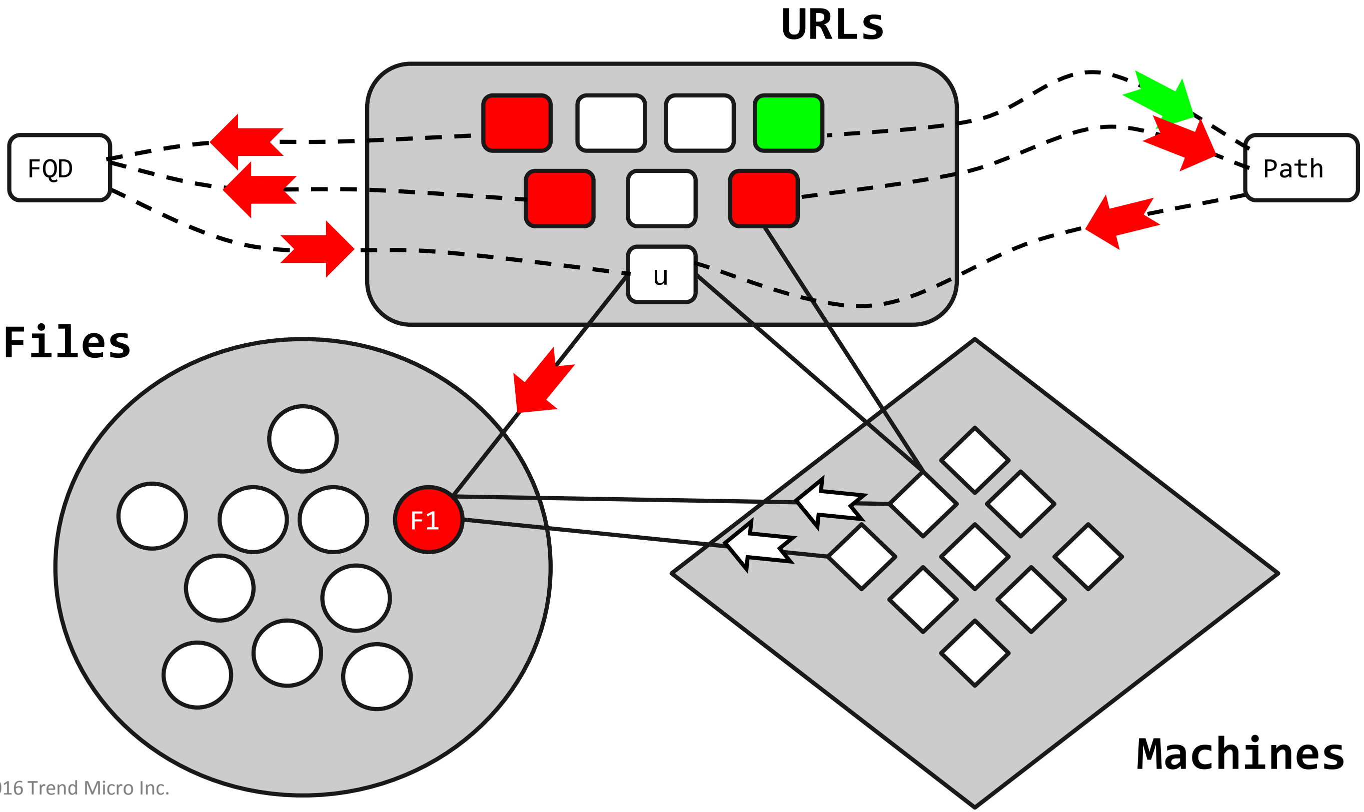
Statistical Classifier, example



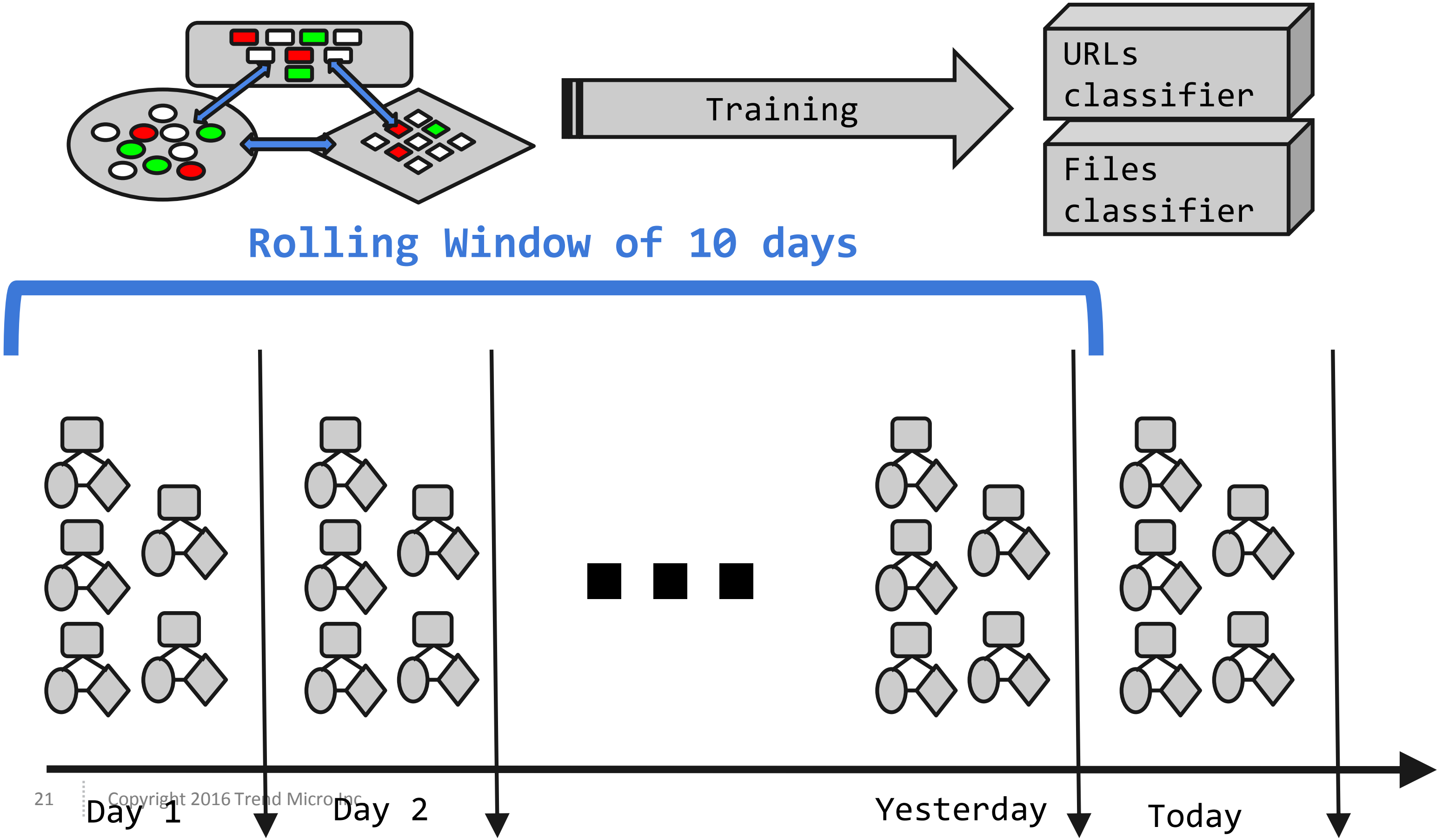
Statistical Classifier, example



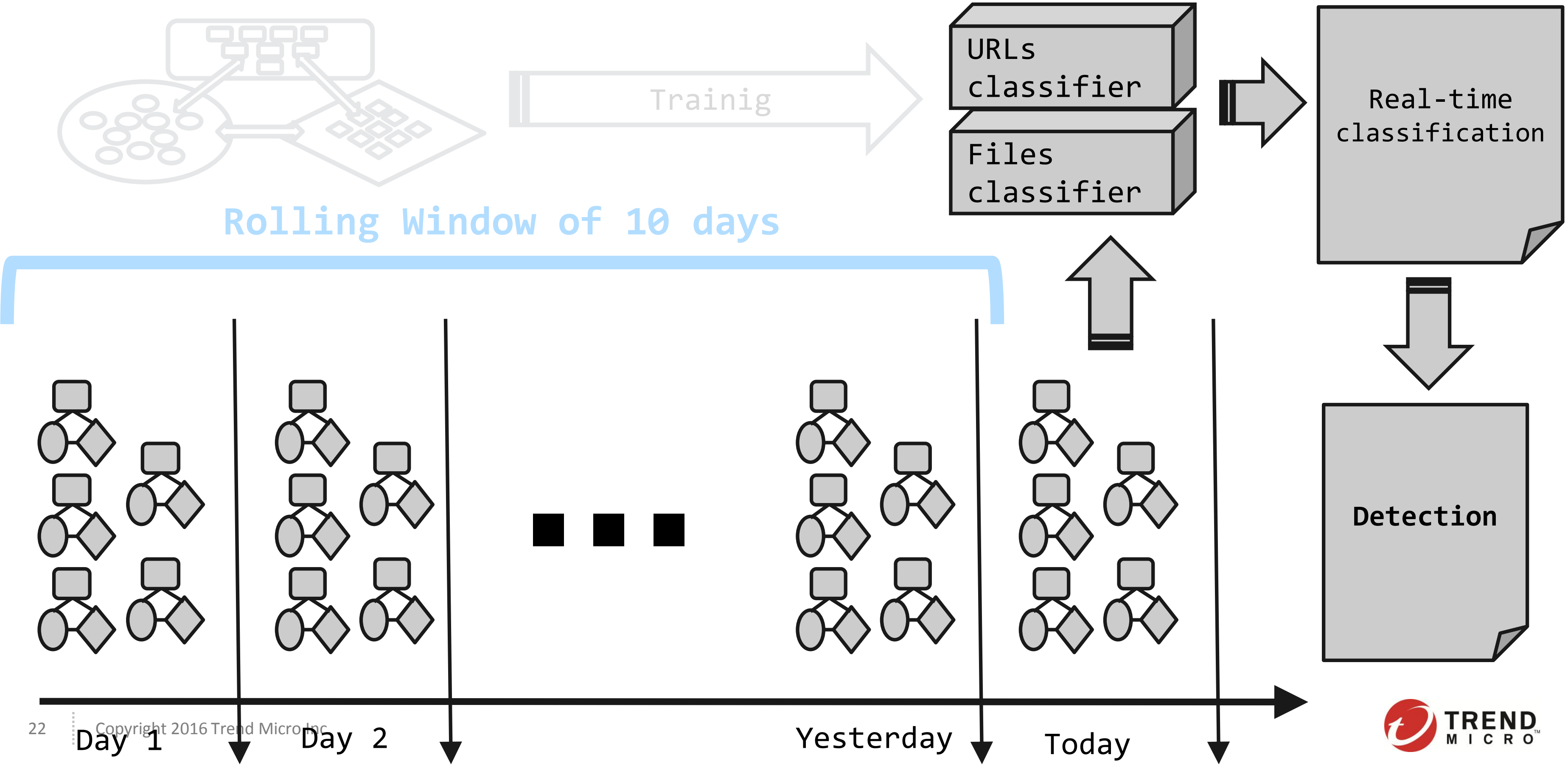
Statistical Classifier, example



Deployment


Training
URLs classifier
Files classifier
Rolling Window of 10 days
Day 1
Day 2
Yesterday
Today

Deployment



Trainig
URLs classifier
Files classifier
Real-time classification
Rolling Window of 10 days
Detection
Day 1
Day 2
Yesterday
Today

Testing Results

- Detection

Month of Test Day	Events		
	test events	TP%	FP%
Feb	4,205	96.2	0.4
Mar	4,581	95.4	0.5
Apr	4,163	97.3	0.5
May	4,004	96.1	0.4
Jun	3,856	94.0	0.5

- Efficiency: requests are served in ~0.16 sec

Early Detection Experiment

- We did classification at t=0.
- We queried Virus Total 6 months after.

- We identified 84% future malware in advance.

Early Detection Experiment

- Droppers and downloaders, e.g. `Win32/InstallCore.MI, TrojanDropper:Win32/Rovnix, Downloader .ATW and MalSign.InstallC.4DB`
- Adware, bots, banking Trojans, key-loggers, e.g.: `Rogue:Win32/FakePAV, Win32:Crypt-QTG, PWS: Win32/Zbot, FakeAV_r.YE, Backdoor.Trojan, and Trojan .FakeAV`

Case Study 1, Wuachos Dropper

- Filename `file_saw.exe`
- Low prevalence
- Invalid signature
- URLs with **no** reputation, BUT path pattern with R of 0.72 (malicious)
  - `/f/1392240240/1255385580/2, /f/1392240120/4165299987/2 -> /H1/I10/I10/I1`
  - 1,445 URLs serving 182 polymorphic malware

Case Study 2, Somoto Adware

- Packed, short lifetime, low prevalence
- 1 graph-connected machines downloaded 1 labeled (known) sample

- Detected a campaign of 695 samples
  - Filename `FreeZipSetup-[\d].exe`
- 616 were unknown to VirusTotal
  - 61 unknown +6 months (10%)

Case Study 3, TTAWinCDM Spyware

- Machine and URL with **no** reputation ☹
- Low lifetime & prevalence & countries
- Mismatch on downloading process
  - Acrobat process + Unauthoritative domain
- Flash 0-day (+2 month)

Conclusions

- Traditional AV is not dead, but tends to become quickly obsolete and inefficient
- Complementary system
  - Content agnostic, statistical based
  - Global awareness
- 90% TP at 0.1% FP
- Detect unknown threats in real-time! ☺

Thanks!

- Questions?

Dr. Marco Balduzzi, @embyte
surname (at) trendmicro.com