Behavioral Graph-based Detection of Malicious Download Events in Real Time

Babak Rahbarinia*, Marco Balduzzi+, Roberto Perdisci^
*Auburn University Montgomery, +Trend Micro Research, ^University of Georgia
*brahbari@aum.edu, +marco_balduzzi@trendmicro.com, ^perdisci@cs.uga.edu

Introduction

Todays most effective infection vectors:

- Drive-by exploits
- Social engineering attacks
- Second-stage malware drops, etc.

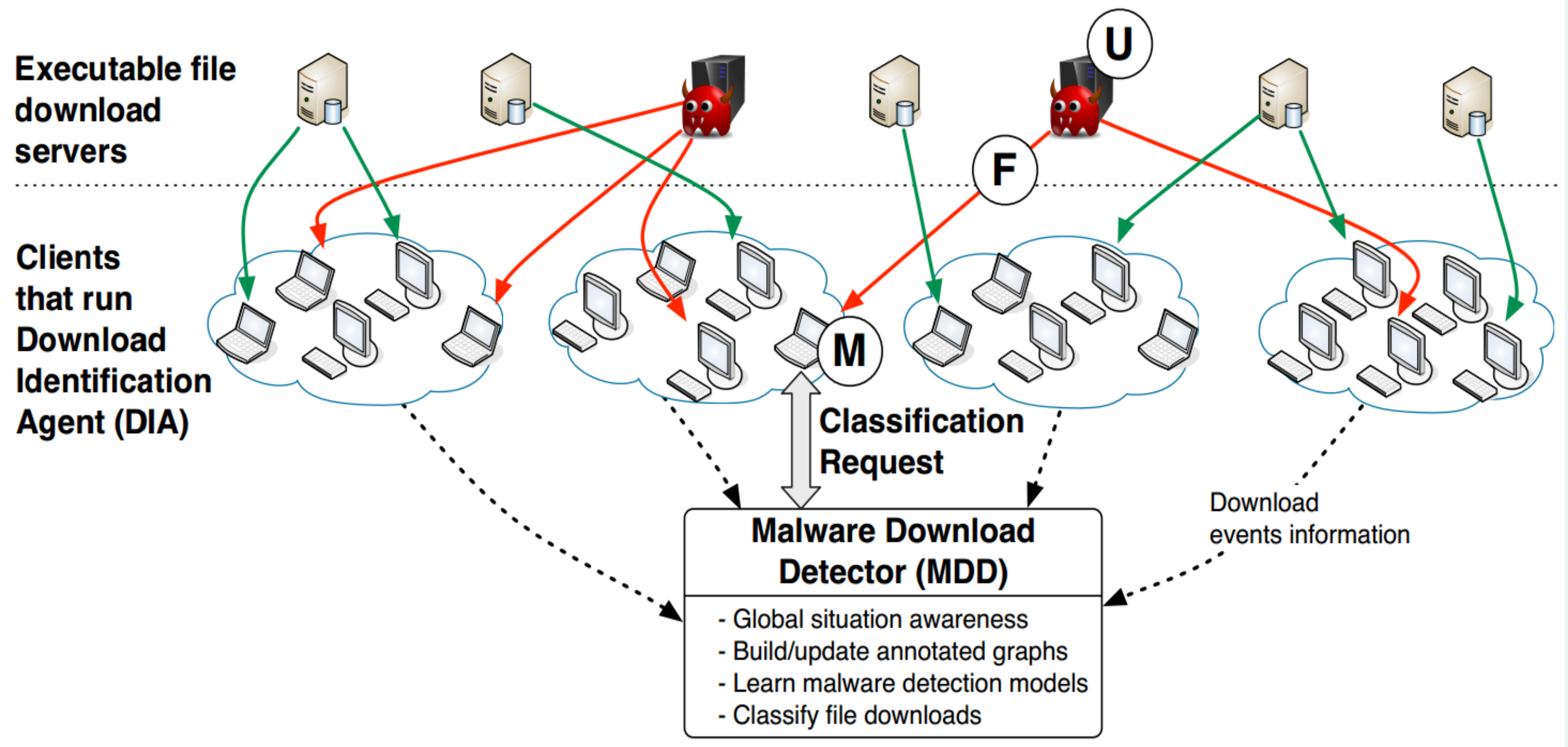
Signature based detection
- Traditional AVs inefficiency (they don't work!)
  - * Polymorphism, code obfuscation, packers, ...
- URL blacklisting
  - * Static, lags behind
  - * Time consuming analysis of individual URLs

Global vs. Local
- Local: looks at one potential malware at a time
- Global: leverages global situational awareness

Behavioral Graph-based Detection

Goals:
- Malware download event detection
  - * Simultaneous detection of files & URLs
- Real time performance

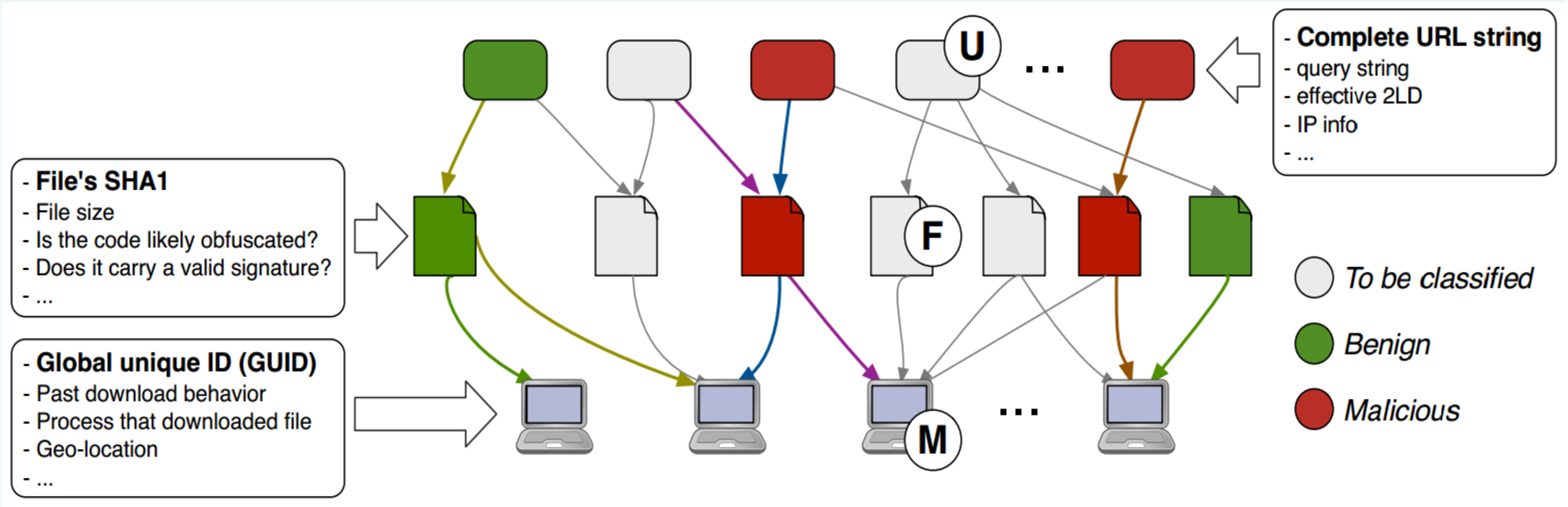
Approach:
- Analyzing behavioral (activity) patterns
  - * Graph inference problem
  - * Graph based learning
- The *"who"*, *"where"*, and *"what"* relationship


Executable file download servers
U
F
M
Clients that run Download Identification Agent (DIA)
Classification Request
Download events information

Malware Download Detector (MDD)
- Global situation awareness
- Build/update annotated graphs
- Learn malware detection models
- Classify file downloads

Annotated, Tripartite Download Graph

Download events: 3-tuples of <files, URLs, machines>
Used to generate a large-scale download graph.

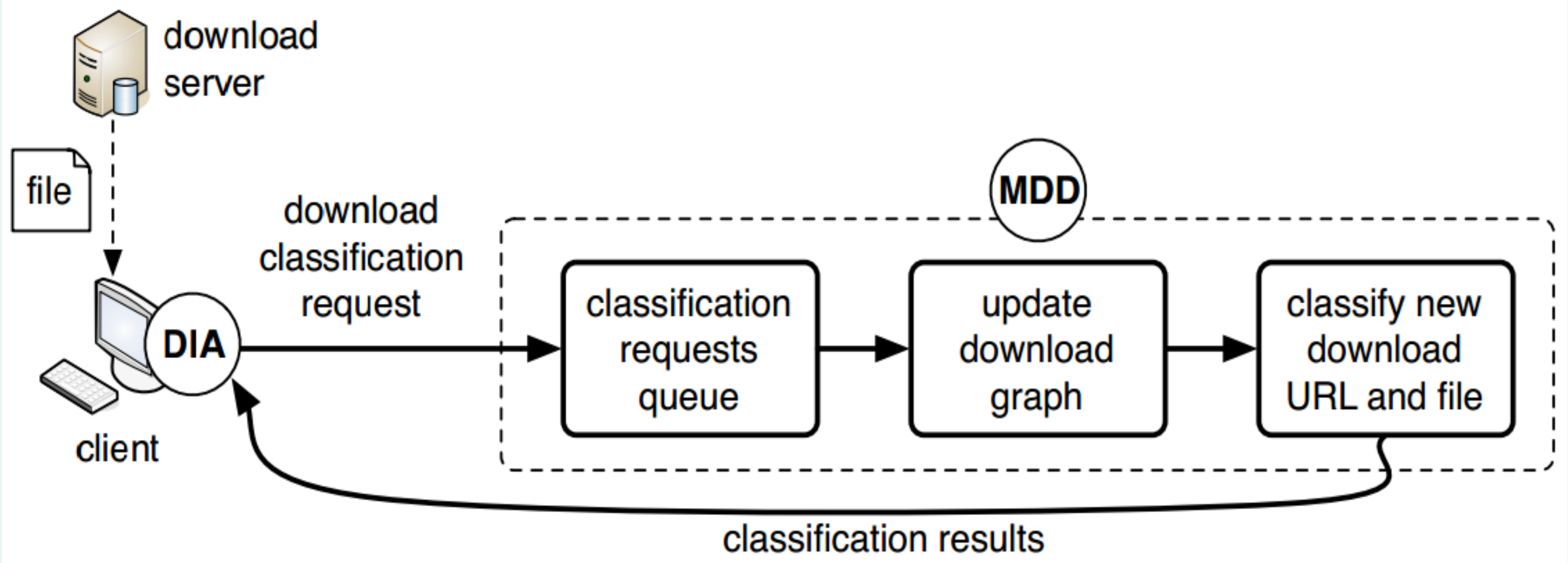


- **Complete URL string**
- query string
- effective 2LD
- IP info
- ...

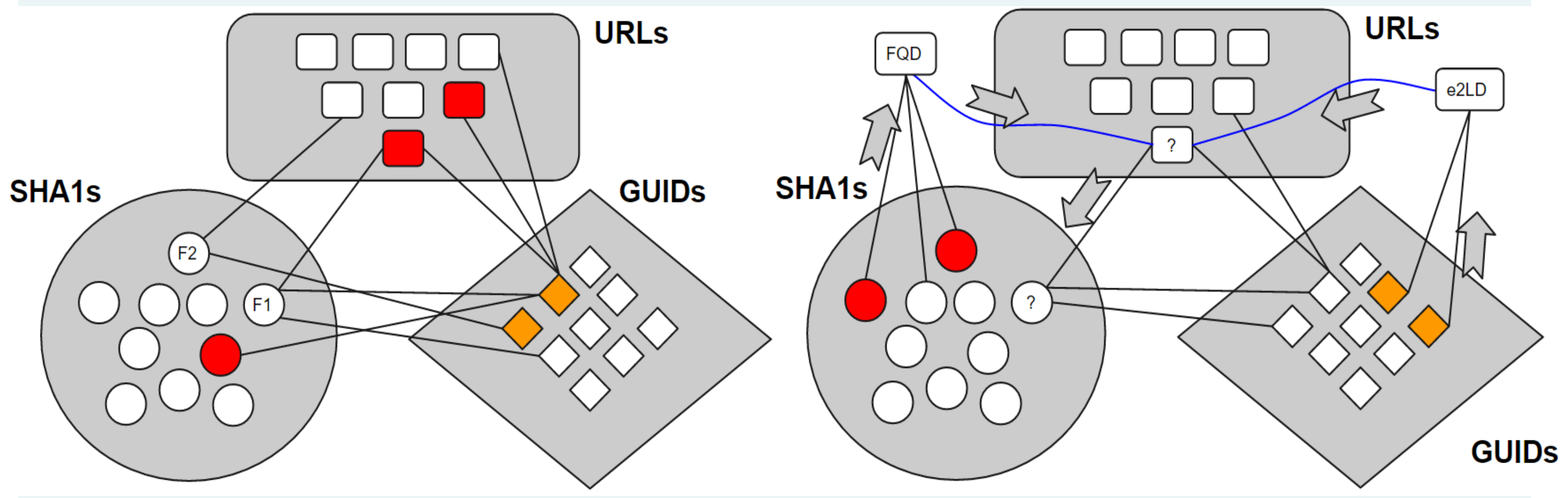
- **File's SHA1**
- File size
- Is the code likely obfuscated?
- Does it carry a valid signature?
- ...

- **Global unique ID (GUID)**
- Past download behavior
- Process that downloaded file
- Geo-location
- ...

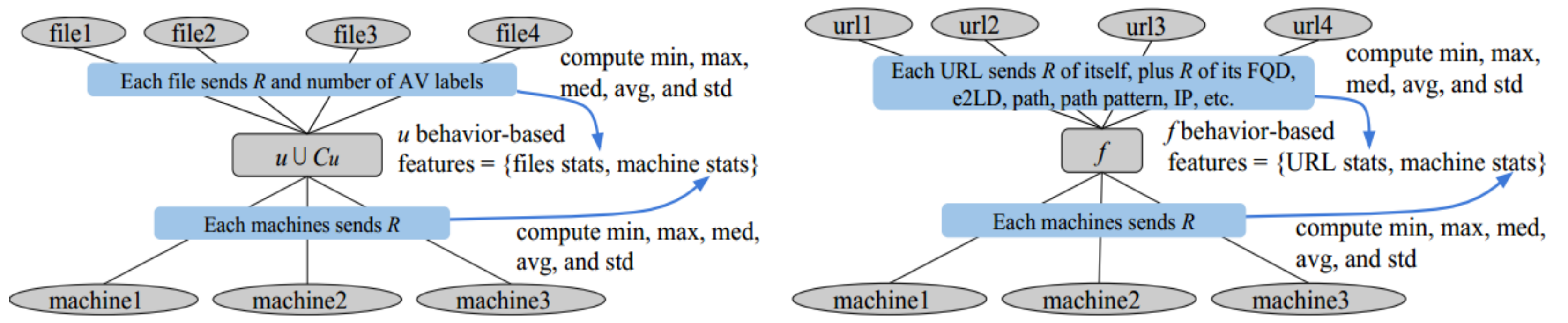
Overview of the System


download server
file
client
DIA
download classification request
MDD
classification requests queue
update download graph
classify new download URL and file
classification results

Intuitions

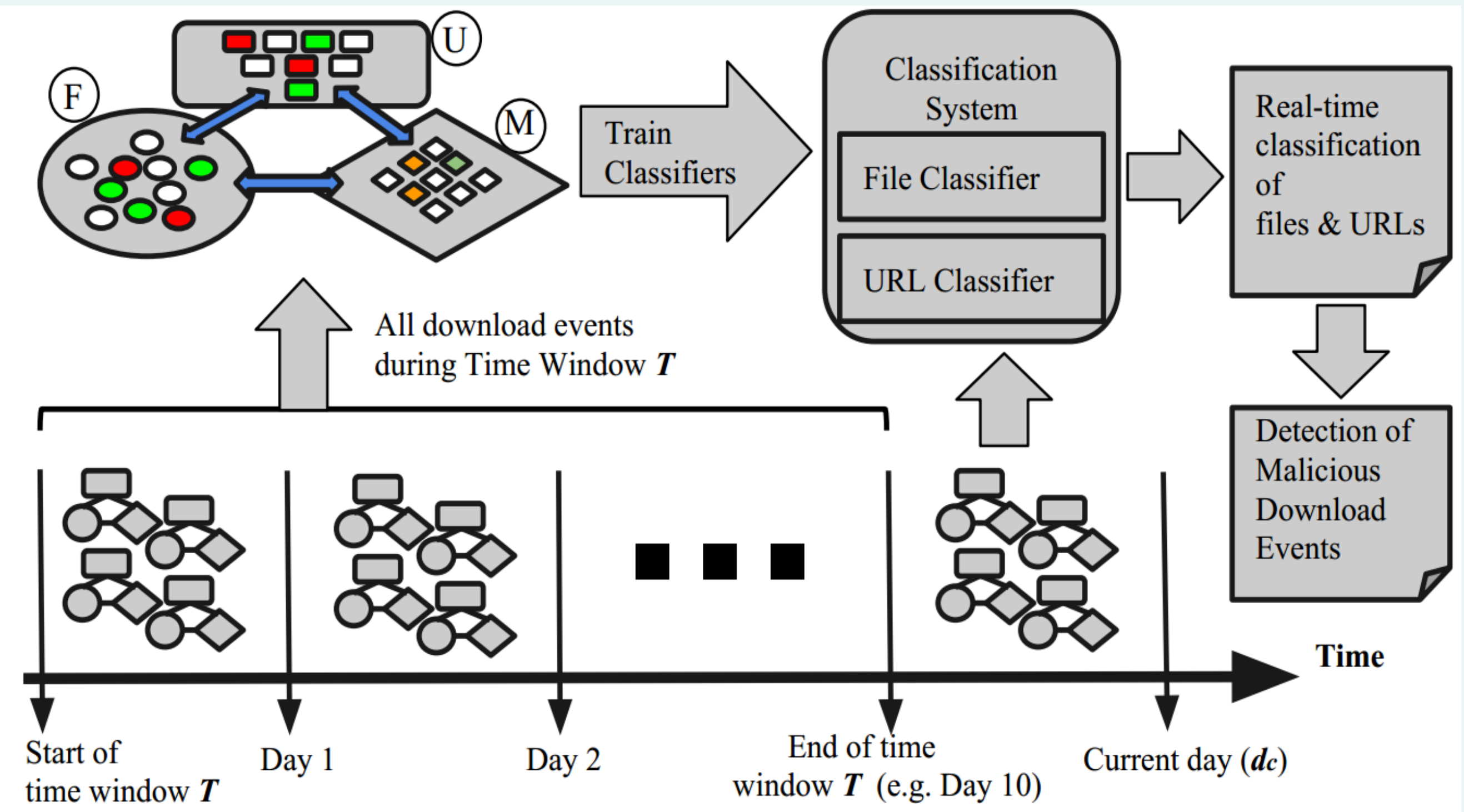


i) Behavioral Features


file1
file2
file3
file4
Each file sends R and number of AV labels
compute min, max, med, avg, and std
u ∪ Cu
u behavior-based features = {files stats, machine stats}
Each machines sends R
compute min, max, med, avg, and std
machine1
machine2
machine3
url1
url2
url3
url4
Each URL sends R of itself, plus R of its FQD, e2LD, path, path pattern, IP, etc.
compute min, max, med, avg, and std
f
f behavior-based features = {URL stats, machine stats}
Each machines sends R
compute min, max, med, avg, and std
machine1
machine2
machine3

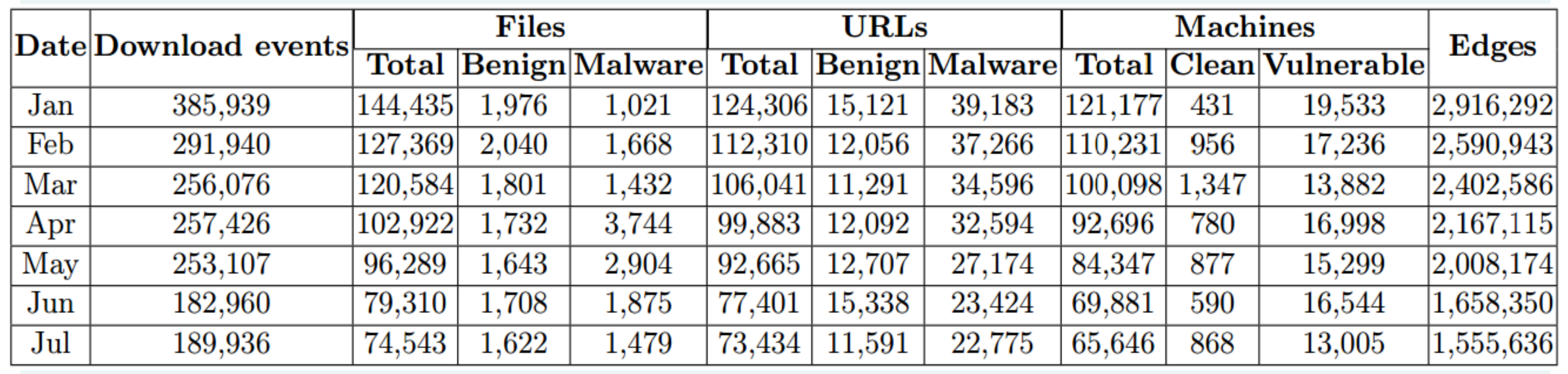
ii) Intrinsic Features
Files (size, lifetime, packed, …), URLs (age, …)

System Operation


F
U
M
Train Classifiers
Classification System
File Classifier
URL Classifier
Real-time classification of files & URLs
All download events during Time Window T
Detection of Malicious Download Events
Time
Start of time window T
Day 1
Day 2
End of time window T (e.g. Day 10)
Current day (dc)

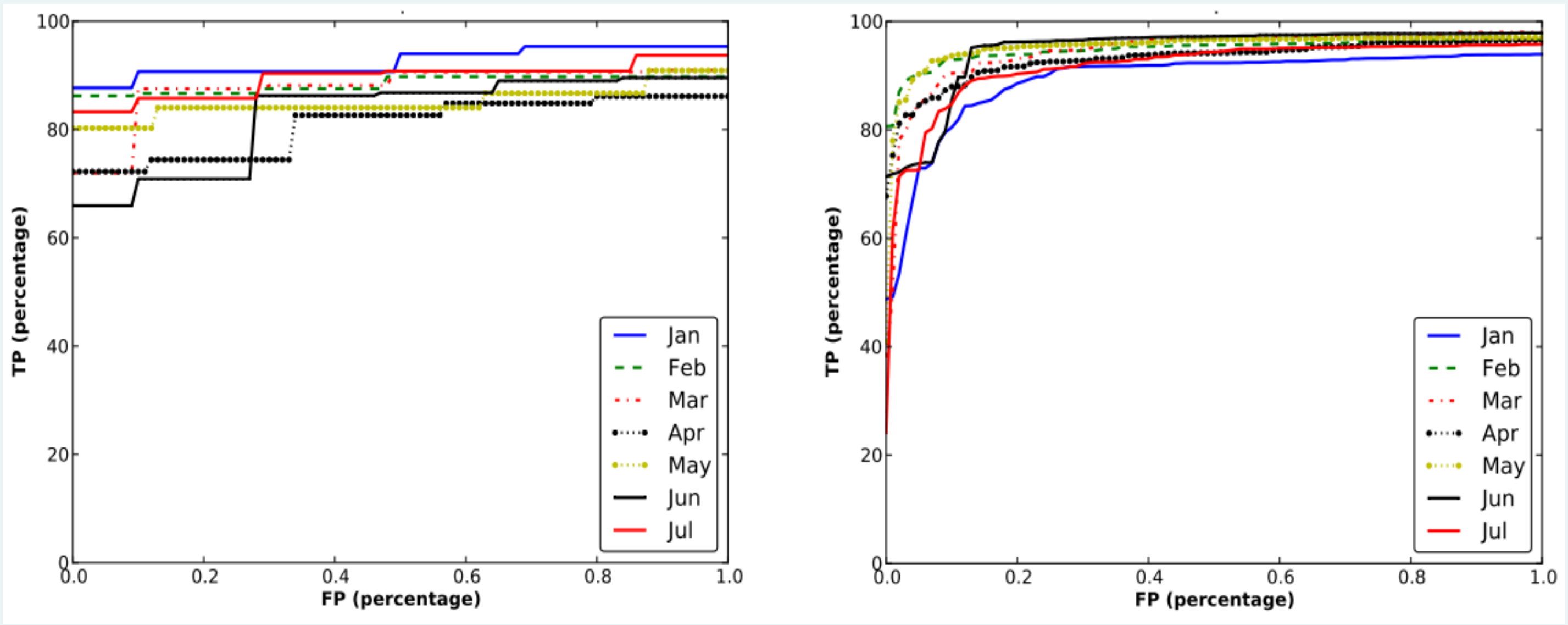
Evaluation Results

The Data



Date	Download events	Files			URLs			Machines			Edges
		Total	Benign	Malware	Total	Benign	Malware	Total	Clean	Vulnerable	
Jan	385,939	144,435	1,976	1,021	124,306	15,121	39,183	121,177	431	19,533	2,916,292
Feb	291,940	127,369	2,040	1,668	112,310	12,056	37,266	110,231	956	17,236	2,590,943
Mar	256,076	120,584	1,801	1,432	106,041	11,291	34,596	100,098	1,347	13,882	2,402,586
Apr	257,426	102,922	1,732	3,744	99,883	12,092	32,594	92,696	780	16,998	2,167,115
May	253,107	96,289	1,643	2,904	92,665	12,707	27,174	84,347	877	15,299	2,008,174
Jun	182,960	79,310	1,708	1,875	77,401	15,338	23,424	69,881	590	16,544	1,658,350
Jul	189,936	74,543	1,622	1,479	73,434	11,591	22,775	65,646	868	13,005	1,555,636

Detection Accuracy



Files

URLs

Contributions

- A system for detection of malware download event
- Real time efficiency
- Combining network- and system-level information
- Real world deployment

*** Partially supported by a grant from the Auburn University at Montgomery Research Grant-in-Aid Program.