UNIVERSITY
OF MIAMI

TroGuard: A System-wide Defense Mechanism Against Web-Based Socially Engineered Trojan Attacks

Rui Han, Saman Zonouz, Mihai Christodorescu

August 9th, 2013

Introduction

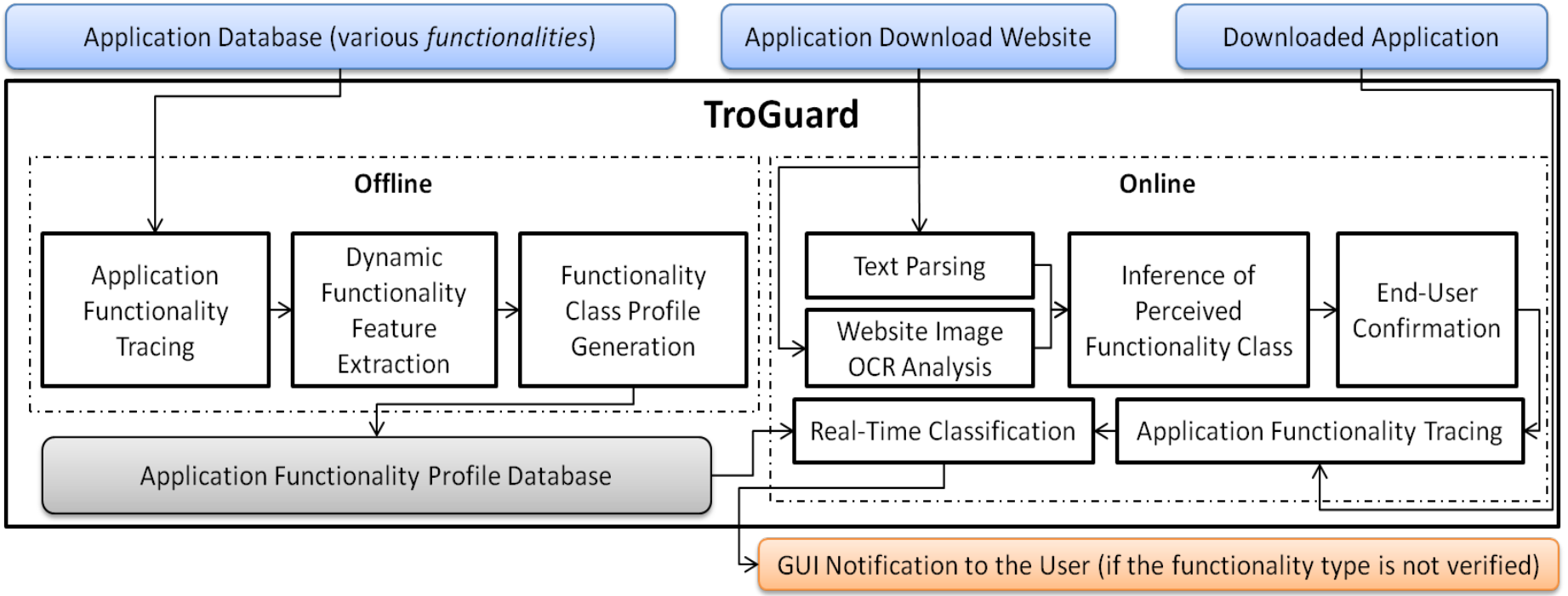
- Download exploit vulnerability of web browser becomes harder. (Memory-page protection, address randomization, etc.)
- Socially-engineered download attacks are prevalent.
- For instance: fake anti-virus, fake games, fake video codecs, fully functional pirated software, etc..
- Malicious activity: Upload all sensitive local info to server, open a socket to connect to botnet, etc.

Exist Techniques

- Dynamically updated blacklists (Anti-virus software, Google SafeBrowsing API)
- Content Agnostic Malware Protection (CAMP)
- Combination of blacklist and mapping between file system and inferred user-consent (BLADE)

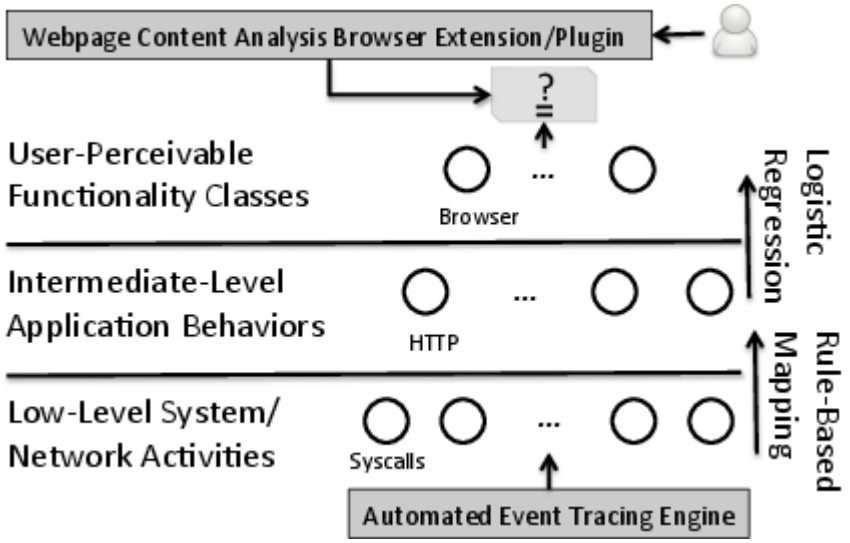
The TroGuard Architecture

- Comparing the user's expectation of application functionality with actual functionality at runtime



Challenges

- Have software classes meaningful to users
- Have software exhibit some unique usage of system operations.
- Runtime monitor and profiler
- Web page analysis
- System integration



Contributions

- A new approach in Trojan detection
- An end-to-end system to identify mismatches of user-perceived and actual software categories.
- A comprehensive evaluation over a large data set (100 different application profiles)

Threat Model

- Attacker has full control of download Website
- No exploits involved
- Web browser and OS are trusted base
- Rely on the willing cooperation of the user

TroGuard Design

- Functionality classes

Table 1: Functionality classes in TROGUARD and matching software categories of three software-download web sites.

TROGUARD Functionality Class	Softpedia.com	download.cnet.com	Tucows.com
Graphics Editor	Artistic Software	Graphic Design Software	Design tools
Game	Games	Games	Games
Browser	Internet	Browsers	Internet
Instant Messenger (IM)	Communications	Communications	
Media Player	Multimedia	MP3 and Audio Software	Audio and Video
Audio Editor			
Video Editor		Video Software	
Office	Office	Productivity Software	Business
Integrated Dev. Environment (IDE)	Programming	Developer Tools	Dev. and Web Authoring
Calculator	Utilities	Utilities and OS	Home and education

Website Analysis

- performs website content analysis to show the recommendation window
- make use of the text and its related pictures in the current web page.
- OCR engine (Tesseract) to extract the text in the image file
- Single word description
- Multiple words description

Feature Extraction

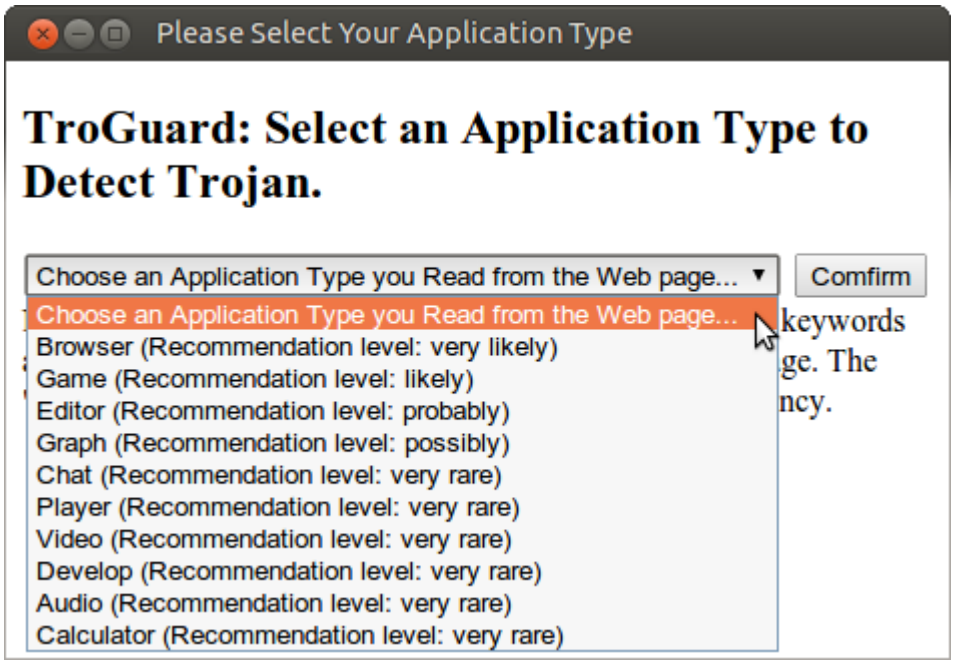
- Kernel space features
- User space features

Four Classes:

- File System Attributes
- Network Attributes
- Resource Usage Attribute
- User Interactivity Attribute

TroGuard Implementation

- Browser Extension for Website Analysis
- Chrome Extension
  - Background page
  - Content scripts
  - Popup.html



Profiler

- System-call tracing tool
- User-space component for collecting high-level information
  - User interactivity
  - Resource consumption
  - IP addresses
- Trace Parser
  - Generate .arff file as data set for Weka

Offline Profile Database Training

- 100 application
- Run with user interaction
- Trace for 60 second
- Parser divide into 10 second data points
- 600 data points

Table 4: Studied Applications

ID	Func. Class	Applications
0	Browser	chrome, firefox , opera, epiphany, midori, chromium, netsurf, arora, xxxterm, rekonq
1	Office	kile, geany, texmaker, calligra-words, soffice.bin, lyx, tea, jed, emacs, vi
2	Games	sol, wesnoth, glchess, neverball, kmahjongg, supertuxkart, hedge-wars, pingus, frozen-bubble, eboard
3	IDE	anjuta, codelite, codeblocks, net-beans, monodevelop, kdevelop, spyder, monkeystudio, drracket, idle
4	IM	skype, kmess, emesene, kopete, pidgin, psi, gajim, empathy, amsn, qutim
5	Graphic-Editor	gimp, pinta, imagej, inkscape, kolourpaint, rawtherapee, mypaint, gpaint, gnome-paint, pencil
6	Media-Player	smplayer, vlc, audacious, quodli-bet, gmusicbrowser, qmmp, abraca, amarok, guayadeque, aqualung
7	Video-Editor	openshot, lives, iriverter, kino, pitivi, videocut, winff, arista-gtk, kdenlive, curlew
8	Audio-Editor	audacity, avidemux, dvbcut, og-gconvert, kwave, wavbreaker, mp3splt-gtk, mhwaveedit, fillmore, soundconverter
9	Calcu-lator	grpn, gcalctool, EdenMath, speed-crunch, kcalc, keurocalc, extcalc, gip, galculator, gnome-genius

Experimental setup

- System 1: Main experiment platform
  - OS: Ubuntu 12.10
  - Processor: Intel Core i7 3.6 GHz
  - Memory: 16 GB RAM
- System 2: for application trace collection
  - OS: Ubuntu 12.10 virtual machine on system 1
  - Processor: 4 Cores (PAE/NX enabled)
  - Memory: 4 GB RAM

TroGuard Evaluation

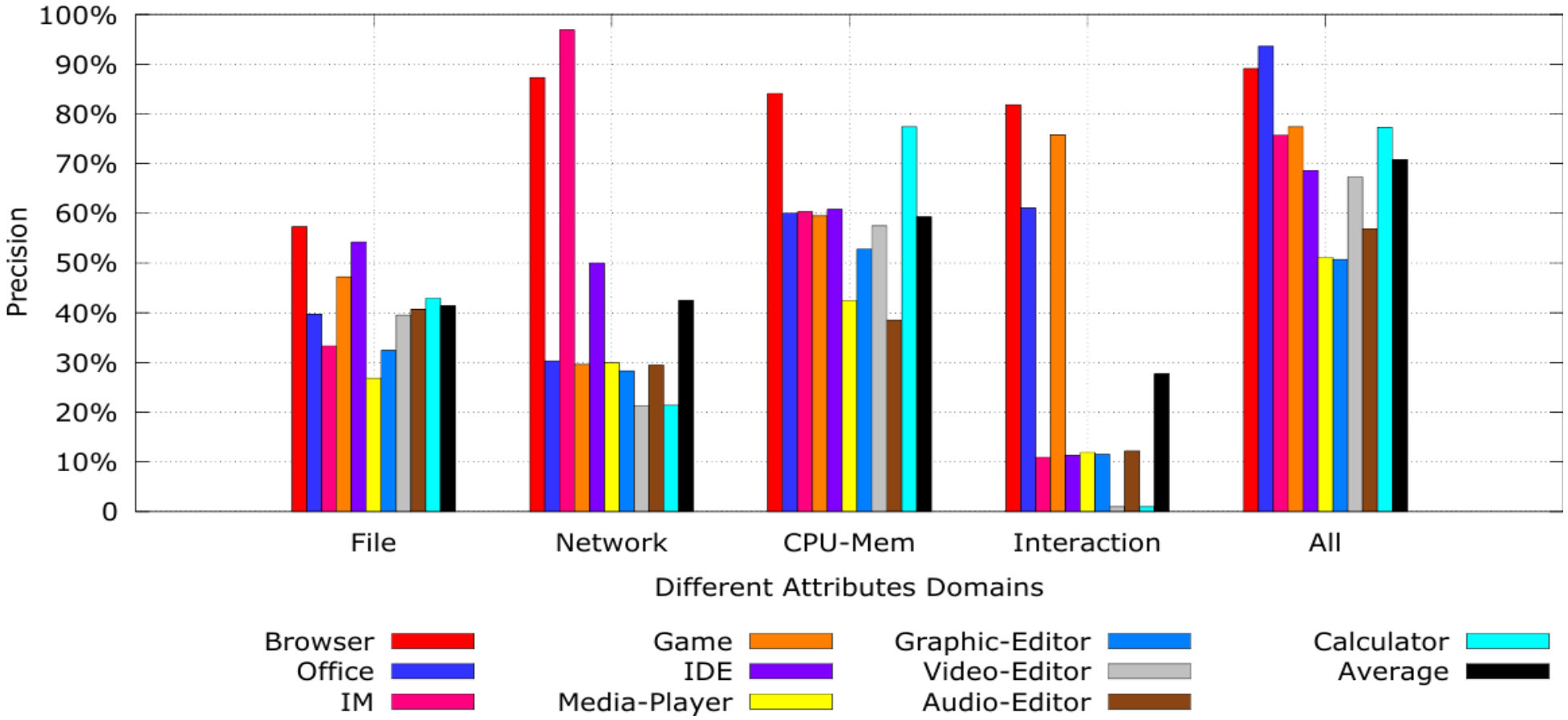
- Accuracy

$$\text{Recall} = \frac{\text{number of documents retrieved that are relevant}}{\text{total number of documents that are relevant}}$$

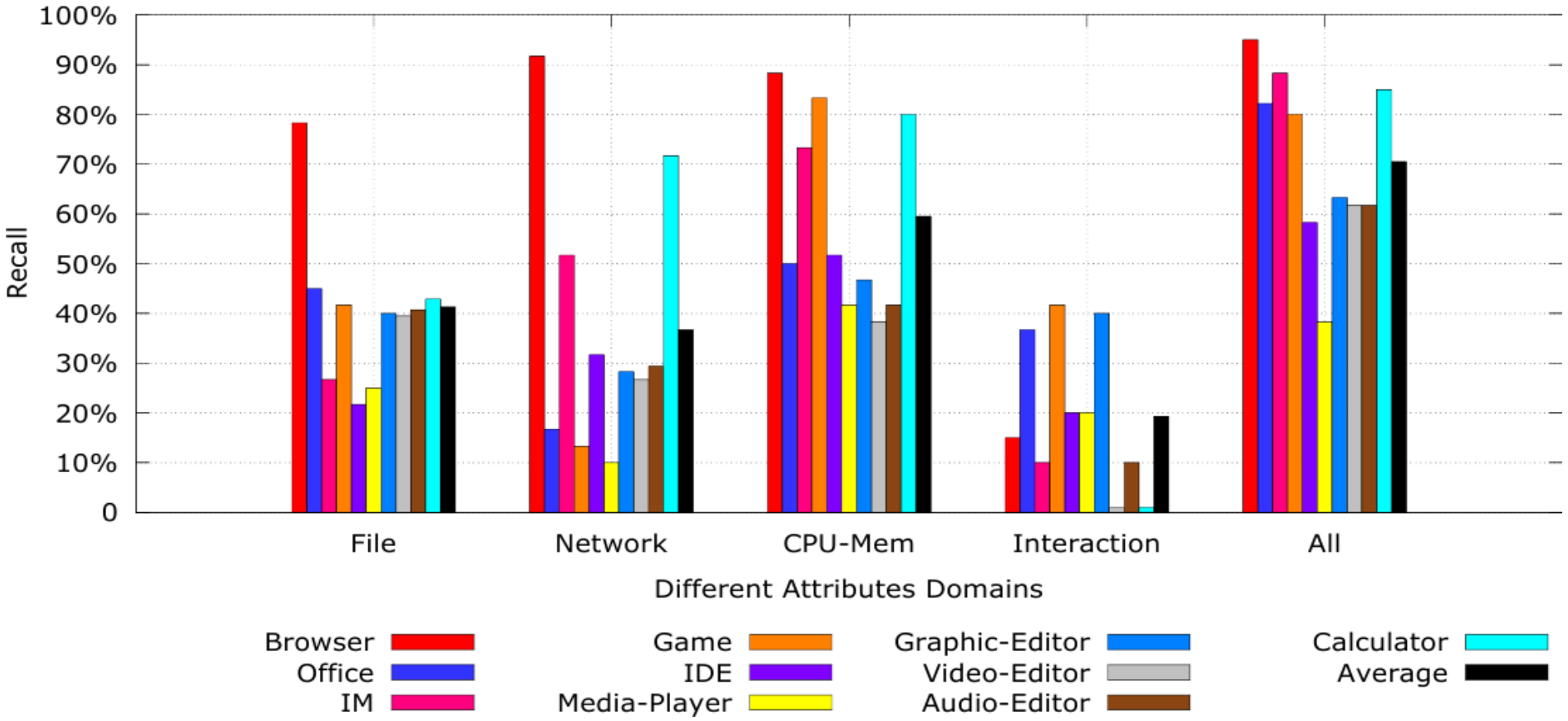
$$\text{Precision} = \frac{\text{number of documents retrieved that are relevant}}{\text{total number of documents that are retrieved}}$$

- Confusion Matrix

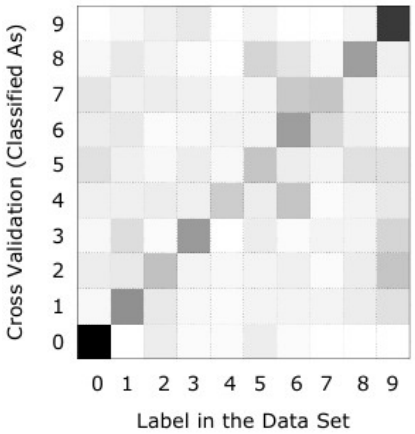
Precision


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
Precision
File
Network
CPU-Mem
Interaction
All
Different Attributes Domains
Browser
Office
IM
Game
IDE
Media-Player
Graphic-Editor
Video-Editor
Audio-Editor
Calculator
Average

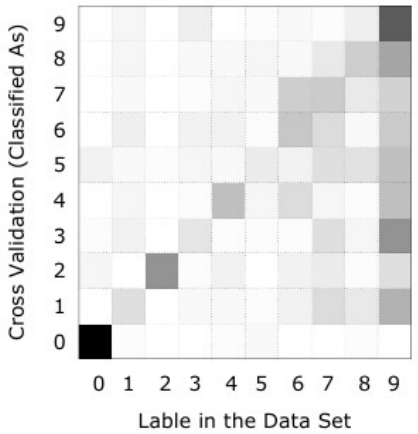
Recall


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
Recall
File
Network
CPU-Mem
Interaction
All
Different Attributes Domains
Browser
Office
IM
Game
IDE
Media-Player
Graphic-Editor
Video-Editor
Audio-Editor
Calculator
Average

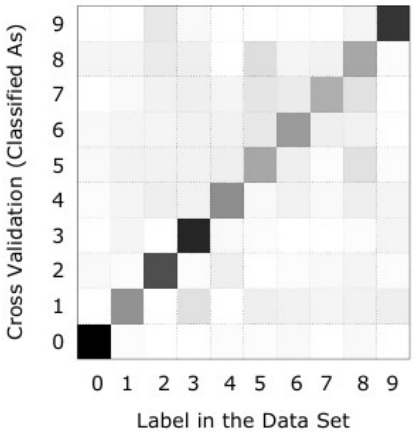
Confusion Matrices



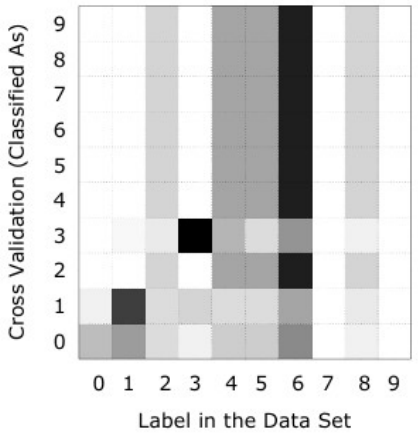
(a) File-system activity only.



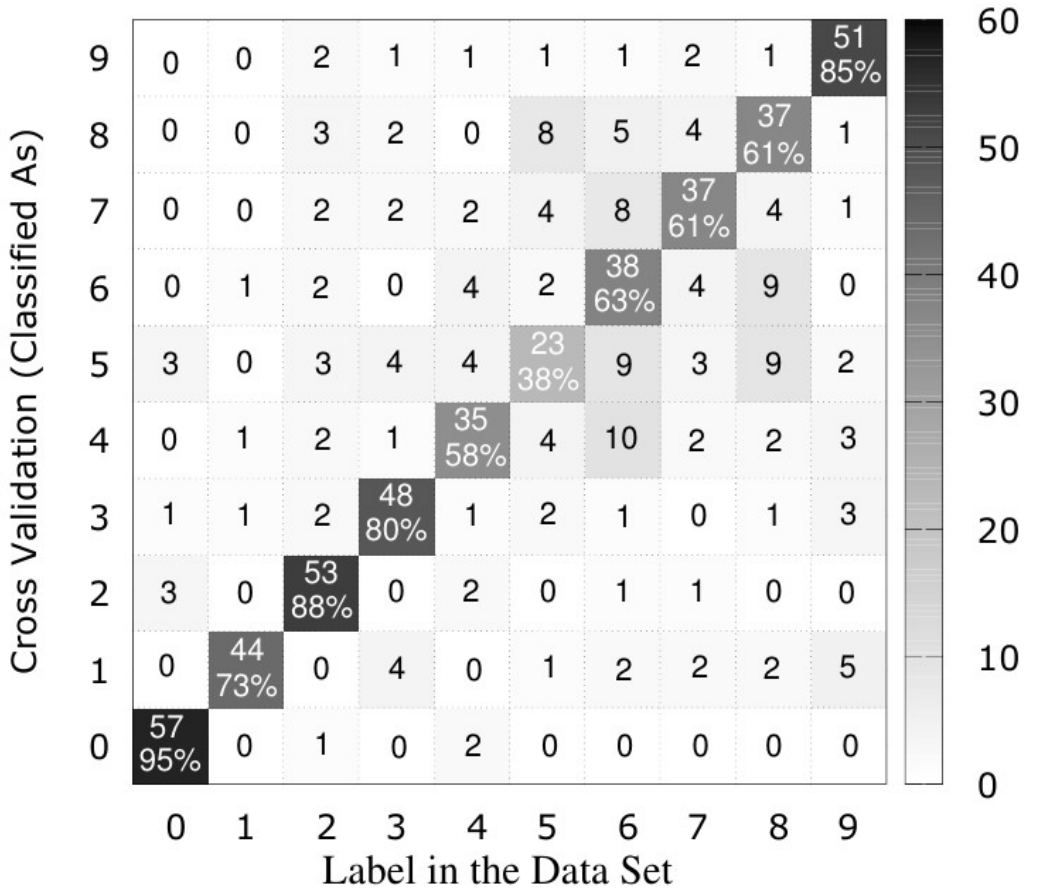
(b) Network activity only.



(c) Resource usage only.



(d) User interactivity only.



(e) All activity attributes.

Website Analysis Accuracy

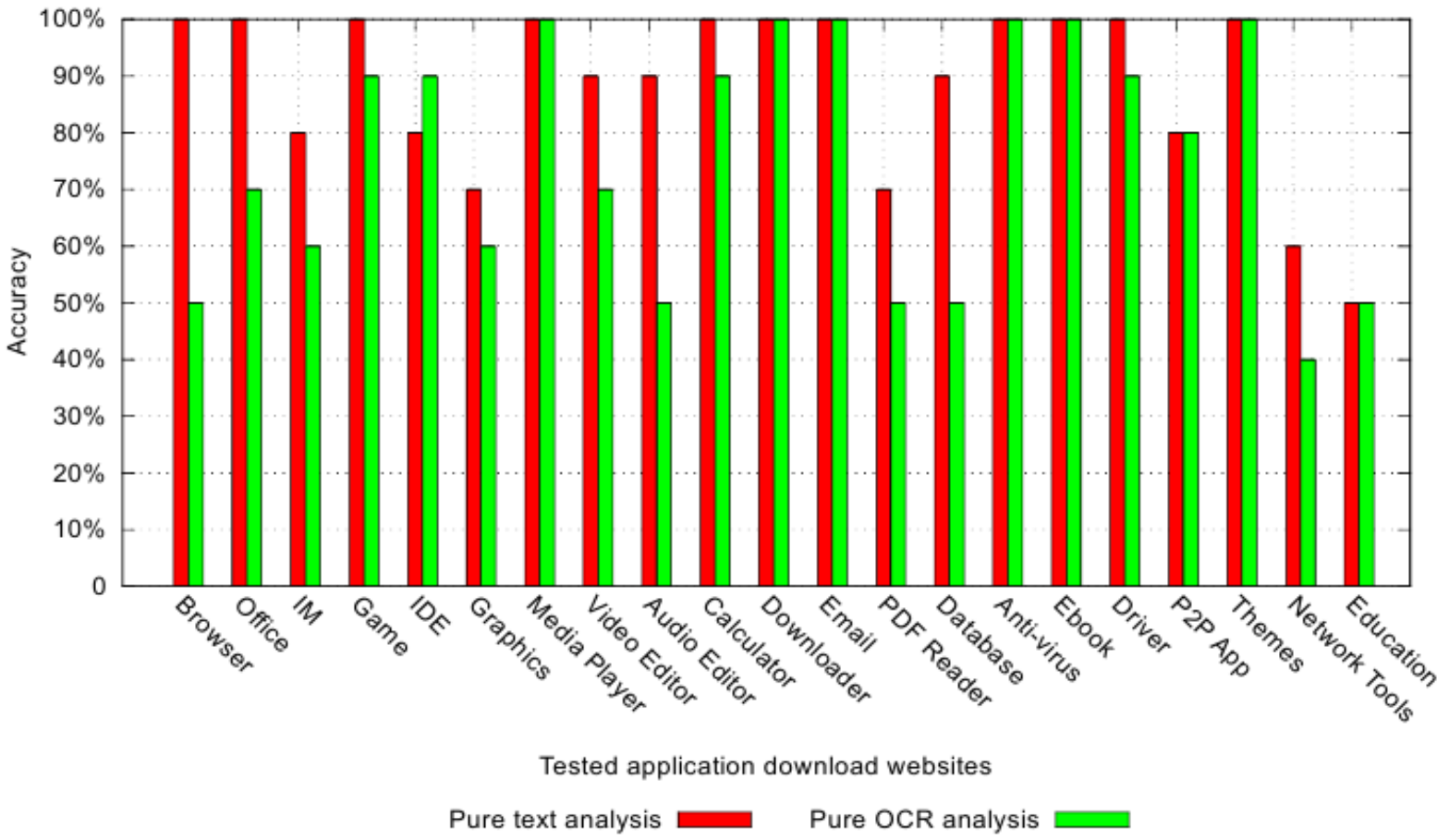


Figure 6: Website Analysis Accuracy Evaluations

Performance

- Classification Performance Evaluation

Table 2: Training times (seconds)

	File	Network	CPU-Mem	Inter.	All
#Attr.	44	20	8	9	81
Time	0.49	0.14	0.19	0.4	0.82

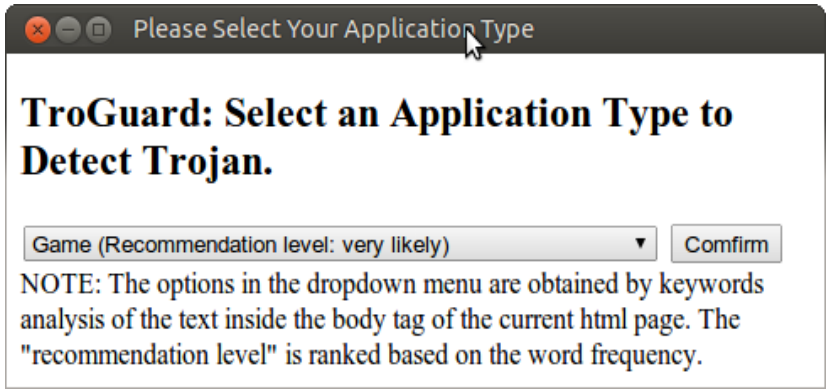
- Website Analysis Performance

Web Page source	CNET	Tucows	Softpedia	Download3k	Soft32	Soft82	Download3000	Average
Pure Text Analysis	0.606	0.209	0.337	0.334	0.308	0.262	0.591	0.378
Pure OCR Analysis	69.871	35.715	25.427	34.532	29.361	45.556	31.784	38.892

Table 3: Website analysis times (seconds)

Case Study

- Craft a trojan (Freesweep + metasploit payload)
- Detect against the build application database
- Results



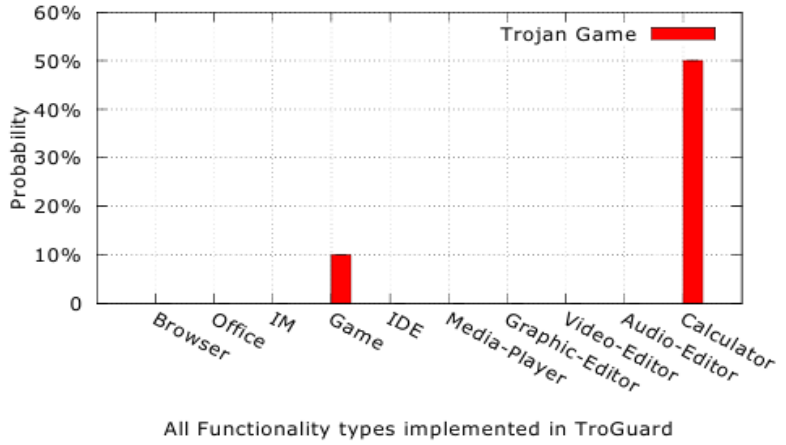


Figure 9: Final Trojan Detection Result

Future Work

- Expand the sample to 500 or more
- Static analysis
- Binary analysis?
- Automatic generate SELinux Policy
- Debian package system(directory policy)

4N6 Group

- Founded by Dr. Saman Zonouz in August 2011
- Research Interest
  - Computer Security and Privacy
  - Intrusion Response and Recovery Systems
  - Automated Intrusion Forensics Analysis
  - Intrusion Detection and Root-Cause Analysis
  - Trustworthy Cyber-Physical Power-Grid Critical Infrastructures
- 1 Postdoc Researcher, 7 PhD. Students, 3 Ms. Students,
- Sponsors: NSF, DOE(ARPA-E), ONR, Fortinet.

Thank you very much.
Questions?