AUTOMATED MALWARE (MIS)CLASSIFICATION & CHALLENGES

AVAR 2013

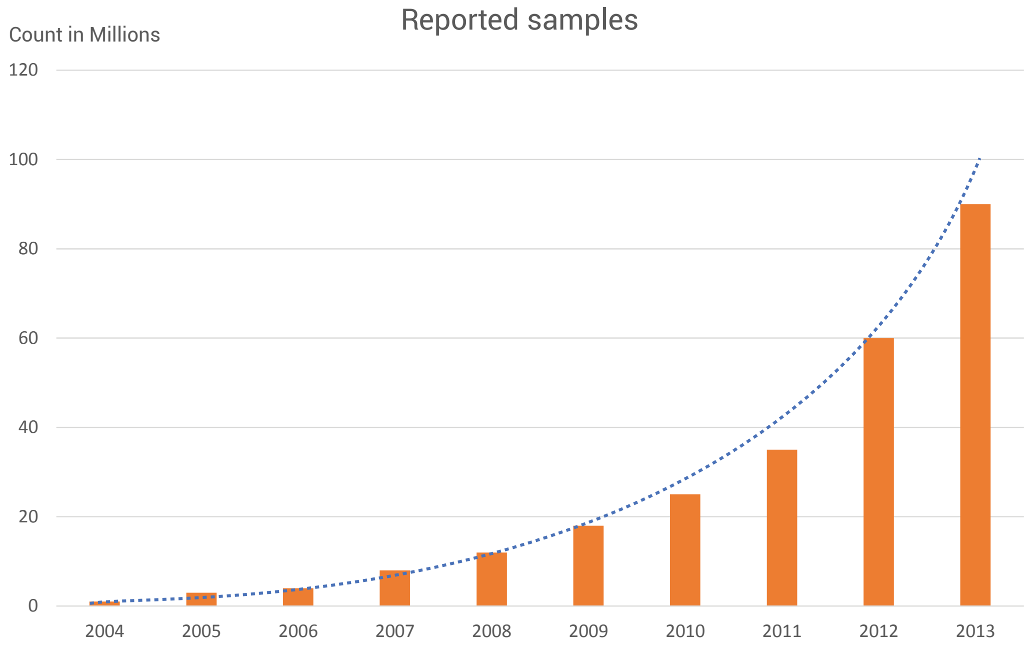
by Rajesh Nikam

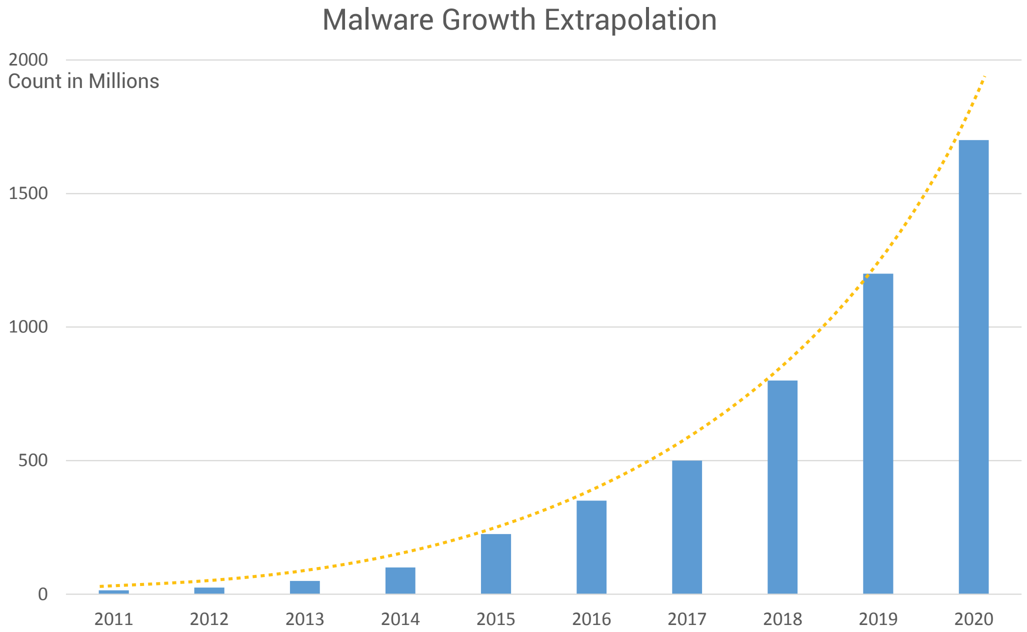
Quick Heal

Security Simplified

CONTENTS

ANTI-MALWARE EVOLUTION





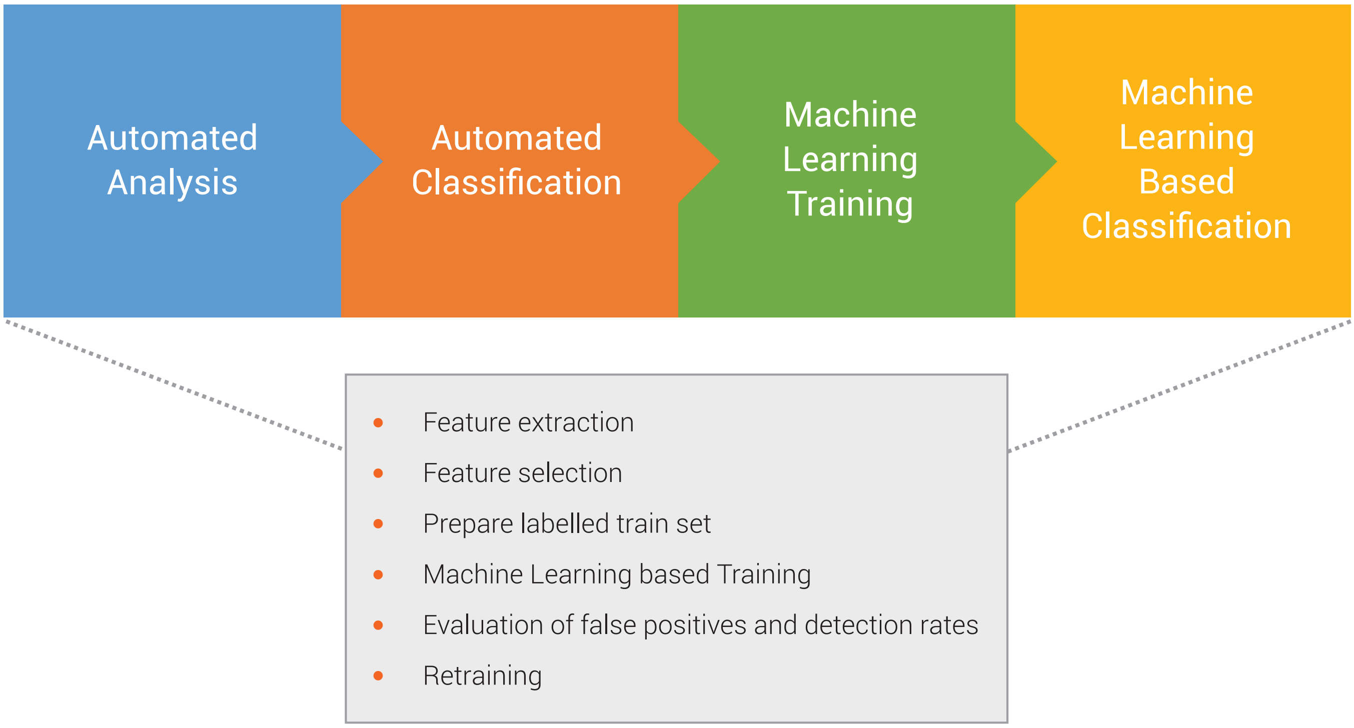
- 8,000,000 samples per month
- 3 minutes per sample for Automated Analysis
- 1,200 samples processed per instance per day
- 6,667 machine days

- 222 machines to complete processing in a month
- 222 * 15 = 3330 machine required for Automated Analysis in Year 2020
- Infeasible to ramp up number of machines with this growth of reported samples

EVOLUTION OF DETECTION TECHNOLOGIES

- CRC on specific parts
- Signature based detections
- Algorithmic detections
- Heuristics based detections
- Support for packers & emulation
- Behavior based detections
- Reputation & Cloud based detections
- Machine Learning based detections

AUTOMATED MALWARE CLASSIFICATION

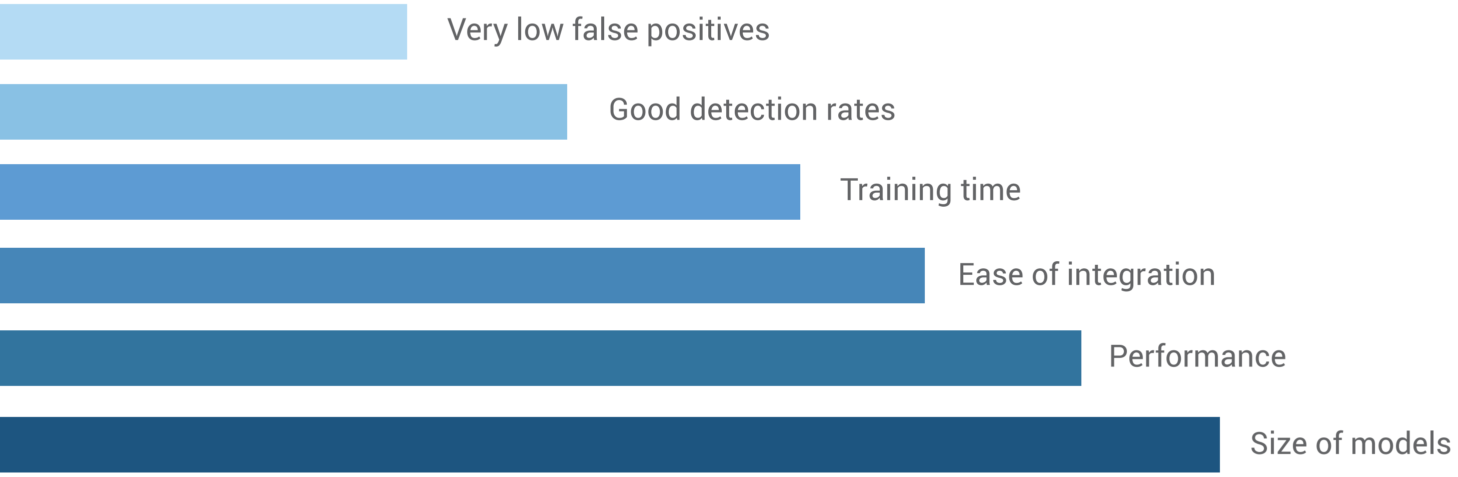

Automated Analysis
Automated Classification
Machine Learning Training
Machine Learning Based Classification

- Feature extraction
- Feature selection
- Prepare labelled train set
- Machine Learning based Training
- Evaluation of false positives and detection rates
- Retraining

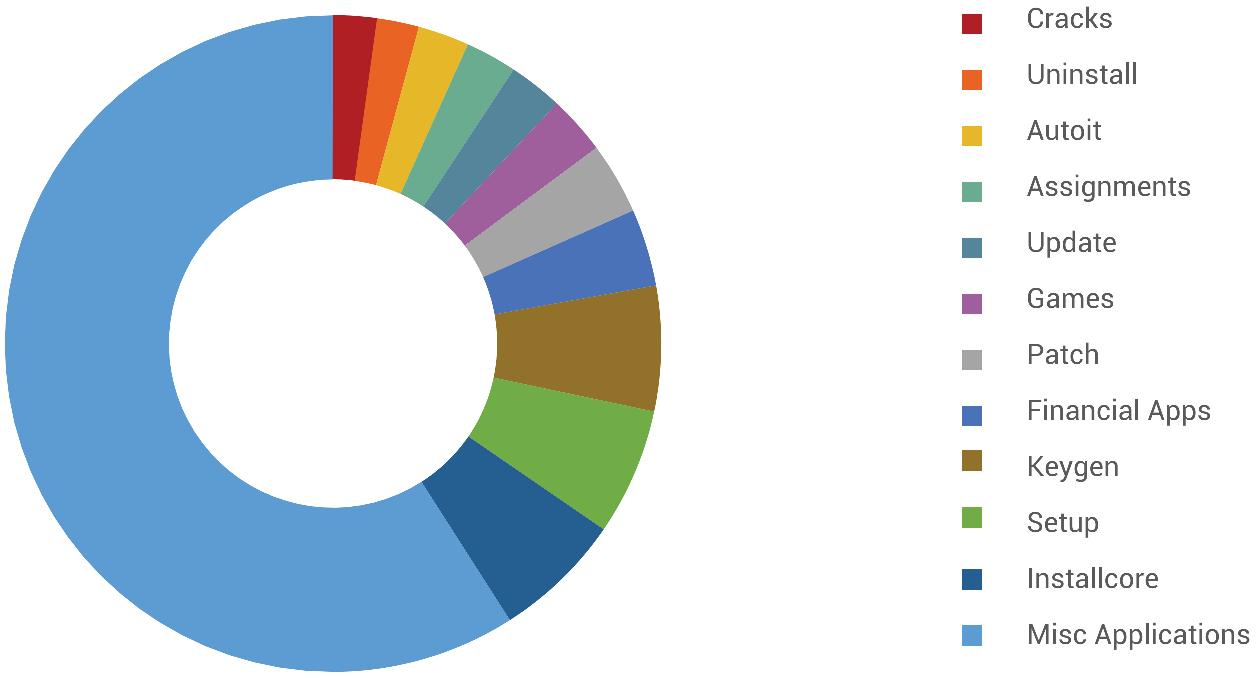
- **Feature extraction**
- **Static features**
  - filetype, compiler, packer, installer identifier
  - n-gram of byte or opcode
  - geometric information of sections
  - anomalies - found in section properties, PE header fields
  - import, export, resource, version information
- **Dynamic features**
  - n-gram of executed instructions
  - api sequence calls
  - identification of anti - { debugging, sandbox, vm, emulation } tricks

AUTOMATED MALWARE CLASSIFICATION

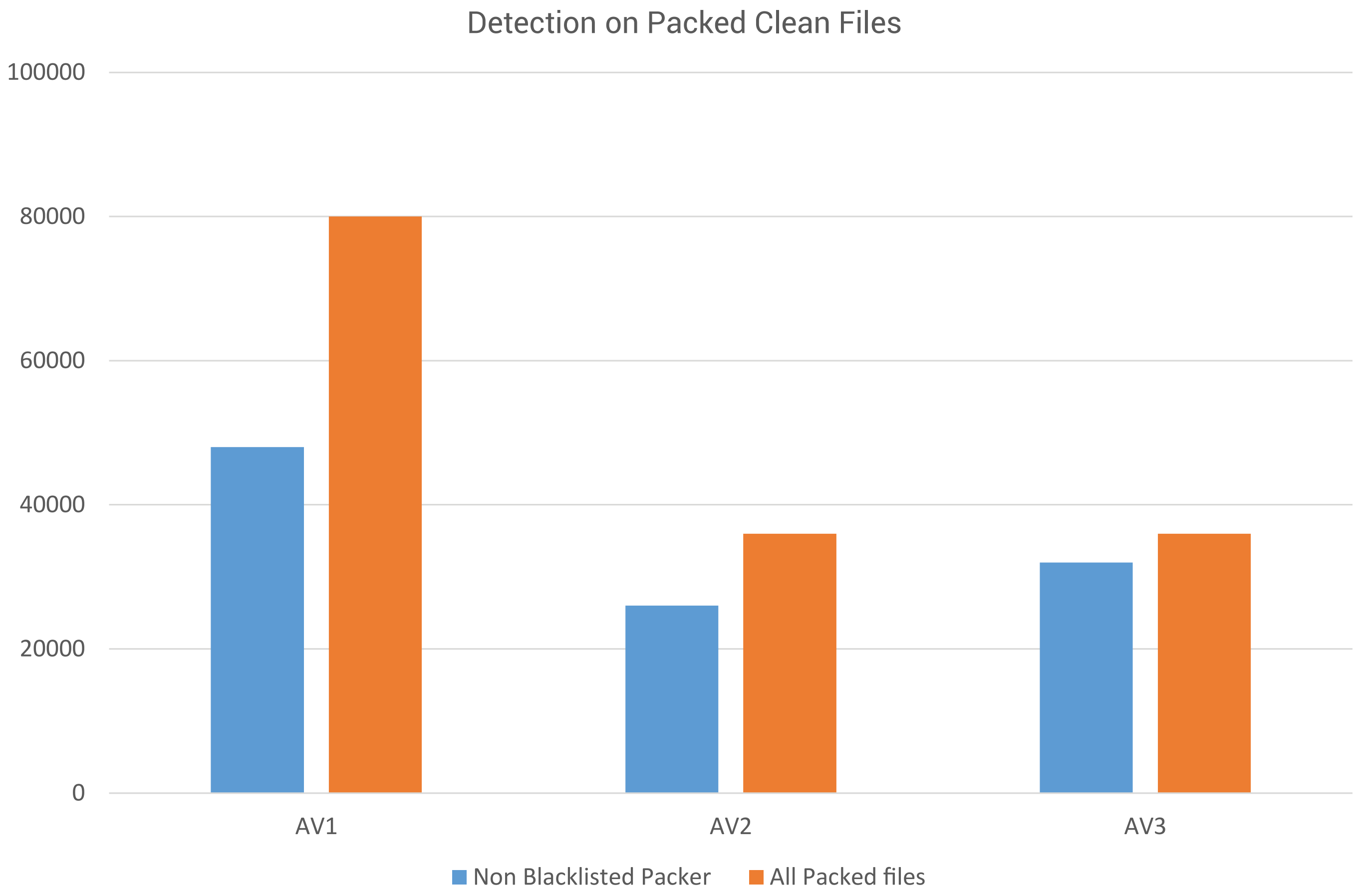
SELECTION OF CLASSIFICATION ALGORITHM



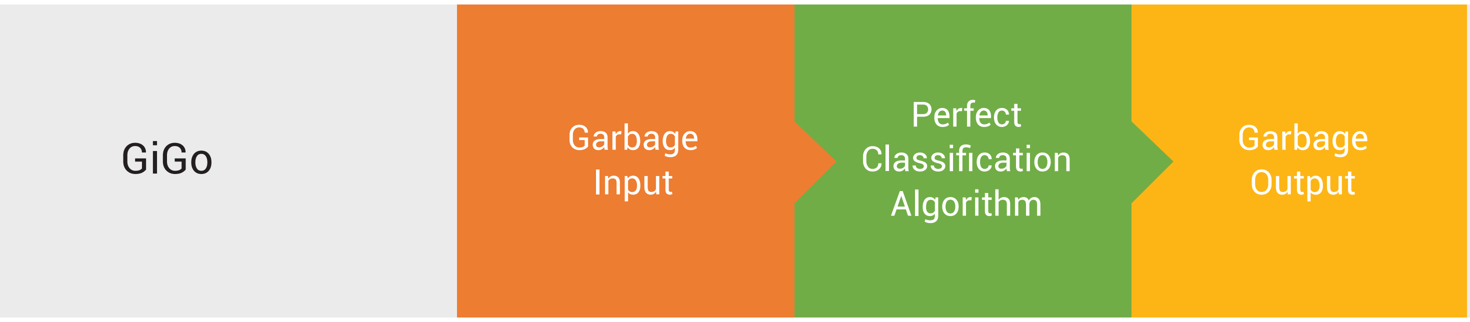
MIS-CLASSIFICATION CASE STUDY 1



MIS-CLASSIFICATION CASE STUDY 2



MIS-CLASSIFICATION CASE STUDY



GiGo

Garbage Input

Perfect Classification Algorithm

Garbage Output

THE ANTIVIRUS UNCERTAINTY PRINCIPLE

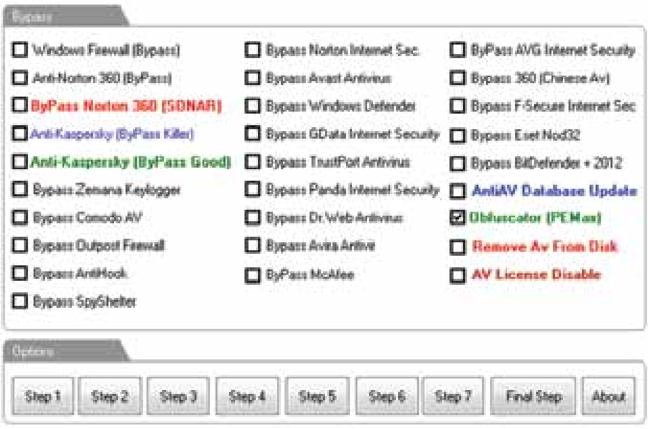
- "The more capable your antivirus detection technologies are in detecting malware, the more frequently false positives will crop up."
- "If you're rarely encountering false positives with your existing antivirus defenses, you're almost certainly missing a whole lot of maliciousness."

- Gunter Ollmann, CTO at IOActive

ATTACKS AGAINST AUTOMATED MALWARE ANALYSIS SYSTEMS

- Multi-component Malware
- Non-executable components like DLL, driver files
- Defeating Entropy analysis
- Delay in execution for specific duration
- Requires user interaction to start functionality
- Payload execution of receipt of instructions from C&C server
- Using Version Information of clean applications
- Using Digital Certificate

- Availability of DIY Tools to use analysis resistance technique]



Bypass

□ Windows Firewall (Bypass)
□ Anti-Norton 360 (ByPass)
□ ByPass Norton 360 (SONAR)
□ Anti-Kaspersky (ByPass Killer)
□ Anti-Kaspersky (ByPass Good)
□ Bypass Zemana Keylogger
□ Bypass Comodo AV
□ Bypass Outpost Firewall
□ Bypass AntiHook
□ Bypass SpyShelter
□ Bypass Norton Internet Sec
□ Bypass Avast Antivirus
□ Bypass Windows Defender
□ Bypass GData Internet Security
□ Bypass TrustPort Antivirus
□ Bypass Panda Internet Security
□ Bypass Dr.Web Antivirus
□ Bypass Avira Antive
□ ByPass McAfee
□ ByPass AVG Internet Security
□ Bypass 360 (Chinese Av)
□ Bypass F-Secure Internet Sec
□ Bypass Eset Nod32
□ Bypass BitDefender + 2012
□ AntiAV Database Update
☑ Obfuscator (PE Max)
□ Remove Av From Disk
□ AV License Disable

Options

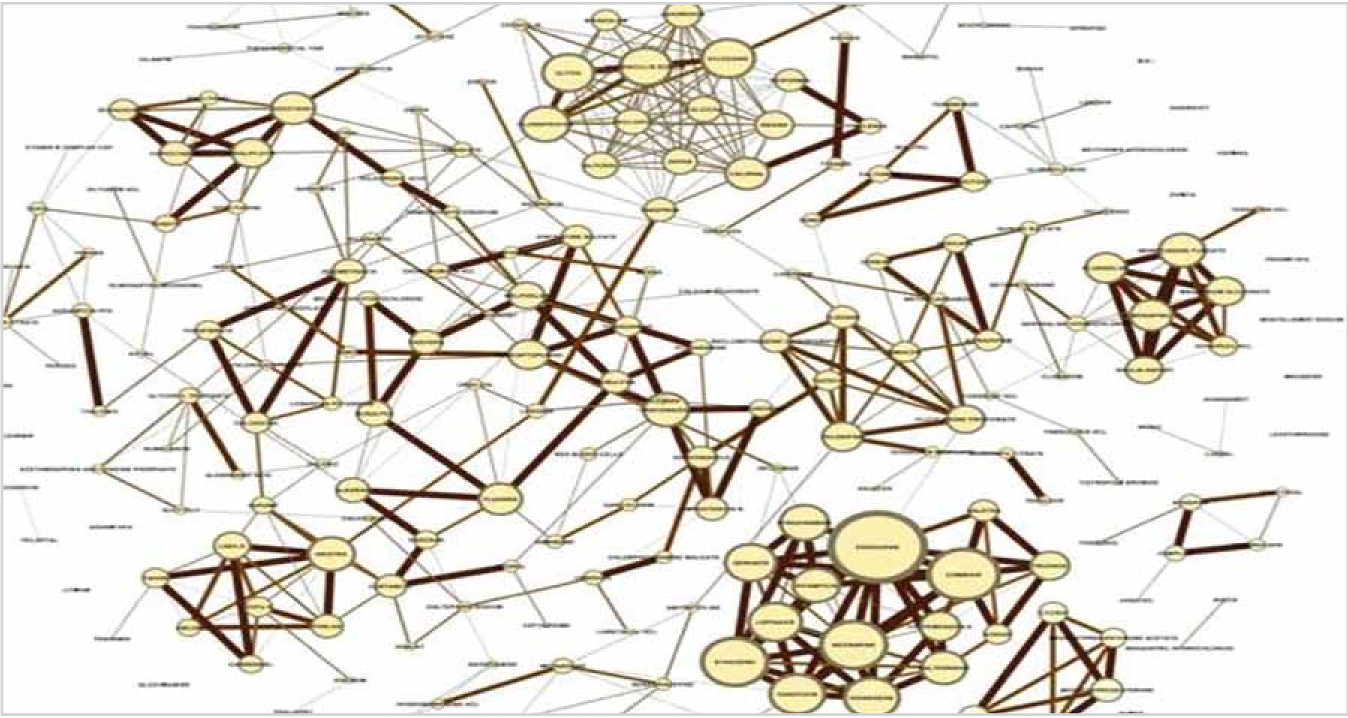
- **Cryptographically unique samples**
  - Downloader component sends unique host identify
    - based on username, computer name, CPU identifier, mac address etc.
  - Unique encryption key is created based on host identity
  - Encrypt payload malware using unique encryption key
  - Creation of unique sample specifically targeted for a victim's machine
  - Could not be correctly decrypted and executed when run in automated analysis environment
  - New generation of analysis reveals environment aware malware

- **Attack on AV vendor automated system**
  - Hundreds of crafted clean files containing signature fragments
  - Other attacks targeting CRC collision weakness
  - Taking advantage of how AV vendors and testers exchange samples
  - AV vendors received thousands of crafted files which poisoned data sources
  - Resulted in false positives on system files
  - Find and fix automation and signature weaknesses

CLUSTERING TO AUGMENT CLASSIFICATION

- Split samples based on file type
- Cluster based on static attribute
- Behavioral analysis & clustering based on dynamic attributes
- Cluster analysis for malicious behavior

VISUALIZATION



CONCLUSION

- 50% YoY growth of reported samples is an alarming situation!
- Find and fix weakness in detection technologies
- Need to re-engineer Automated Systems to be ready for upcoming challenges
- Initiative to share clean samples along with meta information