RSA Conference 2017
San Francisco | February 13 – 17 | Moscone Center

POWER OF OPPORTUNITY

SESSION ID: SPO2-T11

Automated Prevention of Ransomware with Machine Learning and GPOs

Rod Soto
Senior Security Researcher
Splunk, Inc.
@rodsoto

Joseph Zadeh
Senior Data Scientist
Splunk, Inc.
@josephzadeh

$Whoami

- Rod Soto @rodsoto

  Principal Security Researcher at Splunk UBA, former AKAMAI, Prolexic PLXSert. Like to break things, p0wn botnets and play CTFs.

- Joseph Zadeh @JosephZadeh

  Data Scientist at Splunk UBA, building behavioral intrusion detection technologies at scale. Enjoy working on defense projects that combine security, artificial intelligence and distributed systems.

What is Ransomware?

- Ransomware

Current state of Ransomware

ArkansasOnline

32° Little Rock

Search ArkansasOnline

FOLLOW US:

Home News Obituaries Business Entertainment Sports Photos Videos Features Events Classifieds Jobs Homes Autos

Crime Right2Know Traffic Broadway Bridge Archives News Tip Whole Hog Sports Arkansas Life Place an Ad Arkansas Daily Deal

Thursday, December 15, 2016, 9:40 a.m.

TOP PICKS

Mobile Application

ArkansasOnline is just one click away on your smart-phone with our mobile app

Home / News / Arkansas /

$2,440 ransom paid to release Arkansas sheriff's hacked files

By Kenneth Heard

This article was published December 14, 2016 at 5:45 a.m.

8

Comments (3) Font Size

The Carroll County sheriff's office has recovered most of its computer files after overseas hackers breached its system and blocked deputies from accessing information until paying a $2,440 ransom, authorities said.

Chief Deputy Major George Frye said the office's system was "functional" Tuesday and all but a few noncritical files could be accessed after the department's computer network was attacked with "ransomware."

Right2Know

Search mug shots, government salaries, crime maps and more at our Right2Know page.

Latest crime stories

Nearly 4 pounds of marijuana found in Little Rock home, police say

Plea is innocent in 'jihad' threats at Arkansas restaurant

Police arrest North Little Rock man in killing over beer

Juvenile-transfer hearing in Conway couple's

NETWORKWORLD FROM IDG

INSIDER Sign In

Home > Security

PRIVACY AND SECURITY FANATIC

By Ms. Smith | Follow

About

Ms. Smith (not her real name) is a freelance writer and programmer with a special and somewhat personal interest in IT privacy and security issues.

Ransomware attack forces Michigan utility to shut down systems, phone lines, email

BWL, a Michigan municipal utility, was hit with a new variant of ransomware and had to shut down many of its systems

Commonly found Ransomware IOCs

- The modification of the registry keys (Most associated with persistence. I.E execute after reboot).
- Renames and encrypts file extensions of files (Targets User's docs. I.E .doc, xls, ppt, mp3, wallet).
- Modifies Master Boot Record to prevent rebooting, usually encrypting it relocating it and placing a replacement.
- Removal of Volume Snapshot Service files (VSS) or volume shadow files, use for system restoration and backup
- Polymorphic/metamorphic behavior

Enterprises challenged by Ransomware

- Current mitigation technique is… paying…
- Disaster Recovery & Offsite backup.
- Use of Macros/Embedded scripting in Enterprise Document office suites, very difficult and impractical at times to regulate due to business reasons.
- Users are the weakest link, not matter how hardened or strict controls are. It only takes an user action to bypass them. Phishing + Ransomware very effective attack vector.
- New exploitation frameworks/malware using PowerShell to leverage post exploitation.

Ransomware Detection in the New Age

NEW PARADIGM: **DATA DRIVEN INDICATORS**

REAL-TIME & BIG DATA FOUNDATION

BEHAVIOR MODELING

UNSUPERVISED MACHINE LEARNING

ANOMALY DETECTION

THREAT DETECTION

Automation Tools for the Enterprise

- Threat Intelligence Platforms (TIP)
- Threat and Vulnerability Platforms (TVM)
- User and Behaviour Analytics (UEBA)
- Security Incident Response platforms (SIRP)
- Security Operations Automation Platforms (SOAP)

= Automate the ingestion of an unlimited range of contextual & threat data

= Consolidation and normalization (not execution) of of vulnerability assessment results

= Detect and prioritize anomalous/malicious events via machine learning & data-science techniques

= Formalize, enforce and automate incident response playbooks, policies and processes

= provide a selection of connectors, scripts and templates to remediate third-party devices and applications that can be used to fully automate or semi-automate security operations activities.

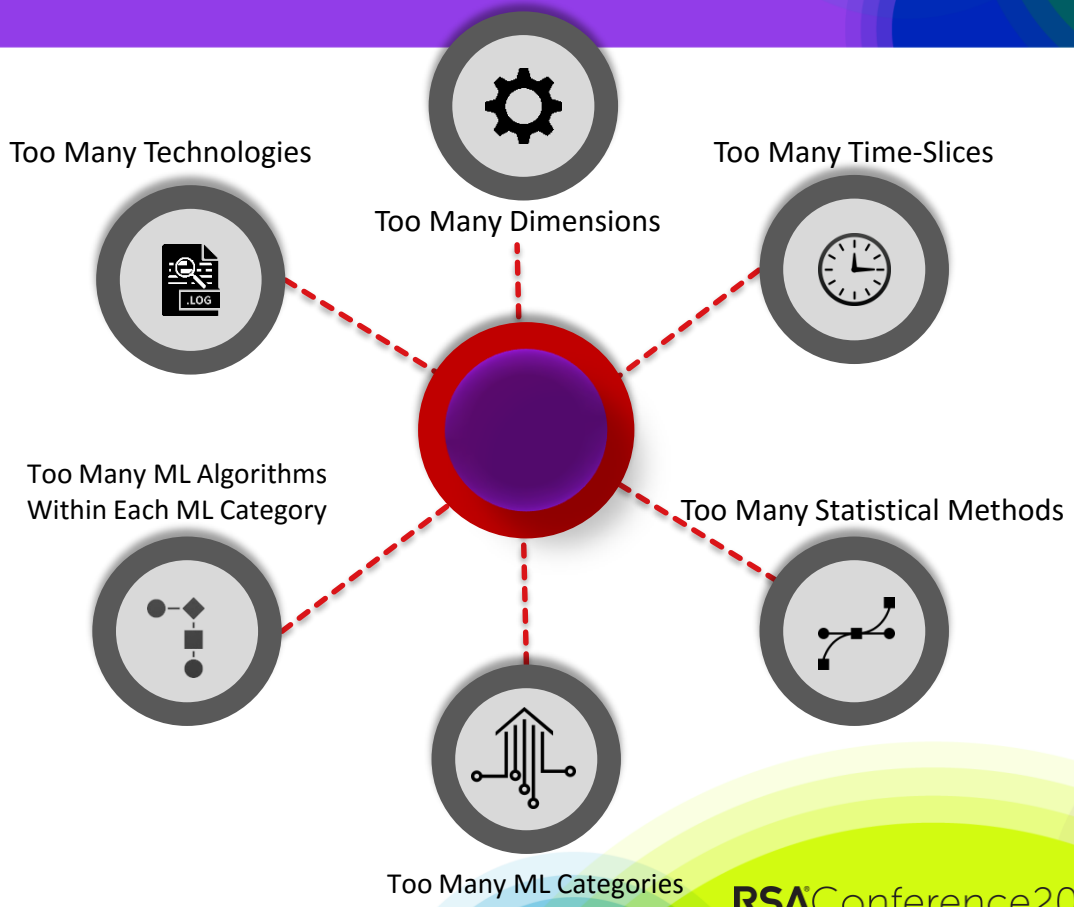
Big Data & Machine Learning

Big Data: Synthesis of technology providing visibility into the analysis of large data sets and the ability to discover patterns, trends, and associations, especially relating to human behavior and interactions.

Machine Learning: Subfield of computer science/statistics. Explores and study construction of algorithms that can learn from and make predictions on Data.

ML At Scale: Multi Faceted Problem

- ML allows us to go beyond of static signature based technologies but can be challenging to deal with for enterprise volumes of user data.
- Combining Traditional Security Tools + Data science creates a scenario where detection of threats based on dynamic and multi contextual indicators is possible (Aktaion is meant to be an example of this).



Guerrilla Machine Learning for Cyber security

- **Fractal Defense**: Reuse logic (and code) across different security use cases. Make behavior based IOC's map to adversary tactics, techniques and procedures for better scalability.
- **Cybersecurity Analytics ROI**: Make security requirements functional by setting realistic benchmarks based on your own data
- **Lambda Architecture**: a generic problem solving system built on immutability and hybrid batch/real-time workflows

Aktaion Detection Workflow

1. Take PCAPs of known (labeled) exploits and known (labeled) benign behavior and convert them to bro format
2. Convert each Bro log to a sequence of micro behaviors (machine learning input)
3. Compare the sequence of micro behaviors to a set of known benign/malicious samples using a Random Forest Classifier (http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/RandomForest.html)
4. Derive a list of indicators from any log predicted as malicious
5. Pass the list of IOCs (JSON) to a GPO generation script (https://github.com/jzadeh/Aktaion/tree/master/python)

Mapping Available Data to a ML Solution

- Ransomware Behaviors:
  - File system Specific
  - Call Back Specific

- Sweet Spot: Exploit Delivery

- Exploit Kits
  - Command and Control behavior can vary widely depending on the post exploit agenda

Training a Random Forest to Detect Exploit Delivery

1. **Initial Redirect From Poisoned Domain:** [29/Apr/2015:16:52:23 -0700] "Nico Rosberg" 192.168.122.177 69.162.78.253 1500 200 TCP_HIT "**GET http://forbes.com/gels-contrariness-domain-punchable/1.html/548828415920276748** HTTP/1.1" "Internet Services" "low risk" "**text/html**" 604 142 "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0)" "http://forbes.com/gels-contrariness-domain-punchable/1.html" "-" "0" "" "-"

2. **Flash Exploit:** [29/Apr/2015:16:52:26 -0700] "Nico Rosberg" 192.168.122.177 69.162.78.253 1500 200 TCP_HIT "**GET http://portcullisesposturen.europartsplus.org/IMvOBBZKDLqAJYIDe02t5hMMNyzBLN_q4kafJkVNqJVTnTmd** HTTP/1.1" "Internet Services" "low risk" "**application/x-shockwave-flash**" 518 821 "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0)" "http:///forbes.com/gels-contrariness-domain-punchable/1.html/548828415920276748" "-" "0" "" "-"

3. **Payload:** [29/Apr/2015:16:52:27 -0700] "Nico Rosberg" 192.168.122.177 69.162.78.253 1500 200 TCP_HIT "**GET http://portcullisesposturen.europartsplus.org/UX7n1YkbNn8FUV6QVtEZLj-p-gLvRKlWEWmz3r7Ug8suRiY_** HTTP/1.1" "Internet Services" "low risk" "**application/octet-stream**" 136 915 "" "" "-" "0" "" "-"

4. **Command and Control:** [29/Apr/2015:16:52:33 -0700] "Nico Rosberg" 192.168.122.177 104.28.28.165 1500 200 TCP_HIT "GET **http://dpckd2ftmf7lelsa.jjeyd2u37an30.com/tsdfewr2.php?U3ViamVj49MCZpc182ND0xJmlwPTIxMy4yMjkuODcuMjgmZXhlX3R5cGU9MQ==** HTTP/1.1" "Internet Services" "low risk" "**text/html; charset=UTF-8**" 566 5 "Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET CLR 3.5.30729)" "" "-" "0" "" "-"

Building a Random Forest

- Random forest trained on labeled malicious and benign samples

Domain Name	TotalCnt	RiskFactor	AGD	SessionTime	RefEntropy	NullUa
europartsplus.org	144	6.05	1	1	0	0
jjeyd2u37an30.com	6192	5.05	0	1	0	0
cdn4s.steelhousemedia.com	107	3	0	0	0	0
log.tagcade.com	111	2	0	1	0	0
go.vidprocess.com	170	2	0	0	0	0
statse.webtrendslive.com	310	2	0	1	0	0
cdn4s.steelhousemedia.com	107	1	0	0	0	0
log.tagcade.com	111	1	0	1	0	0

Learning Machine Libraries : Splunk MLTK, Spark, Weka, scikit-learn

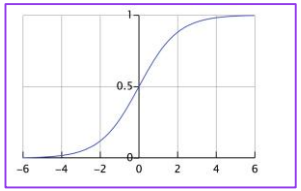
```
Random Forest:
TreeEnsembleModel classifier with 6 trees

 Tree 0:
  If (feature 7 <= 0.0)
   If (feature 11 <= 0.0)
    If (feature 80 <= -51.45518112)
     Predict: 1.0
    Else (feature 80 > -51.45518112)
     If (feature 76 <= 7.0)
      If (feature 60 <= -48.02338409)
       If (feature 81 <= 0.0)
        If (feature 84 <= 2.0)
         Predict: 1.0
        Else (feature 84 > 2.0)
         Predict: 0.0
       Else (feature 81 > 0.0)
        Predict: 0.0
      Else (feature 60 > -48.02338409)
```

Radom Forest Using Splunks Machine Learning Toolkit

- The simple linear model gives us output that separates the **Signal** from the **Noise** (this is not always possible with a model)

receive_time	serial_number	session_id	src_ip	src_port
10/7/15 23:59	sn_1606046662	sid_14787	138.52.78.14	p_57375
10/7/15 23:59	sn_1606046662	sid_1838	205.77.248.110	p_6289
10/7/15 23:59	sn_1606046662	sid_17519	44.165.220.174	p_45700
10/7/15 23:59	sn_1606046662	sid_36258	227.45.212.95	p_33298
10/7/15 23:59	sn_1606046662	sid_48945	40.149.50.140	p_55362
10/7/15 23:59	sn_1606046662	sid_3341	111.234.123.185	p_49284
10/7/15 23:59	sn_1606046662	sid_58015	35.15.188.79	p_50883
10/7/15 23:59	sn_1606046662	sid_21098	243.178.105.215	p_49261
10/7/15 23:59	sn_1606046662	sid_58031	243.178.105.215	p_49262
10/7/15 23:59	sn_1606046662	sid_51761	243.178.105.215	p_49263
10/7/15 23:59	sn_1606046662	sid_56280	243.178.105.215	p_49264
10/7/15 23:59	sn_1606046662	sid_39650	243.178.105.215	p_49265
10/7/15 23:59	sn_1606046662	sid_19759	227.45.212.95	p_58709
10/7/15 23:59	sn_1606046662	sid_23087	212.106.150.184	p_50929
10/7/15 23:59	sn_1606046662	sid_18576	164.51.38.20	p_52000
10/7/15 23:59	sn_1606046662	sid_36199	31.162.206.158	p_43523
10/7/15 23:59	sn_1701003920	sid_331118	243.168.234.169	p_60734
10/7/15 23:59	sn_1701003920	sid_23225	157.228.15.127	p_123
10/7/15 23:59	sn_1701003920	sid_219432	178.32.178.87	p_58570
10/7/15 23:59	sn_1701003920	sid_246545	123.60.19.243	p_57392

Learning Machine:
MLKT Command



Precision	Recall	Accuracy	F1
0.79	0.78	0.78	0.78

Classification Results (Confusion Matrix)

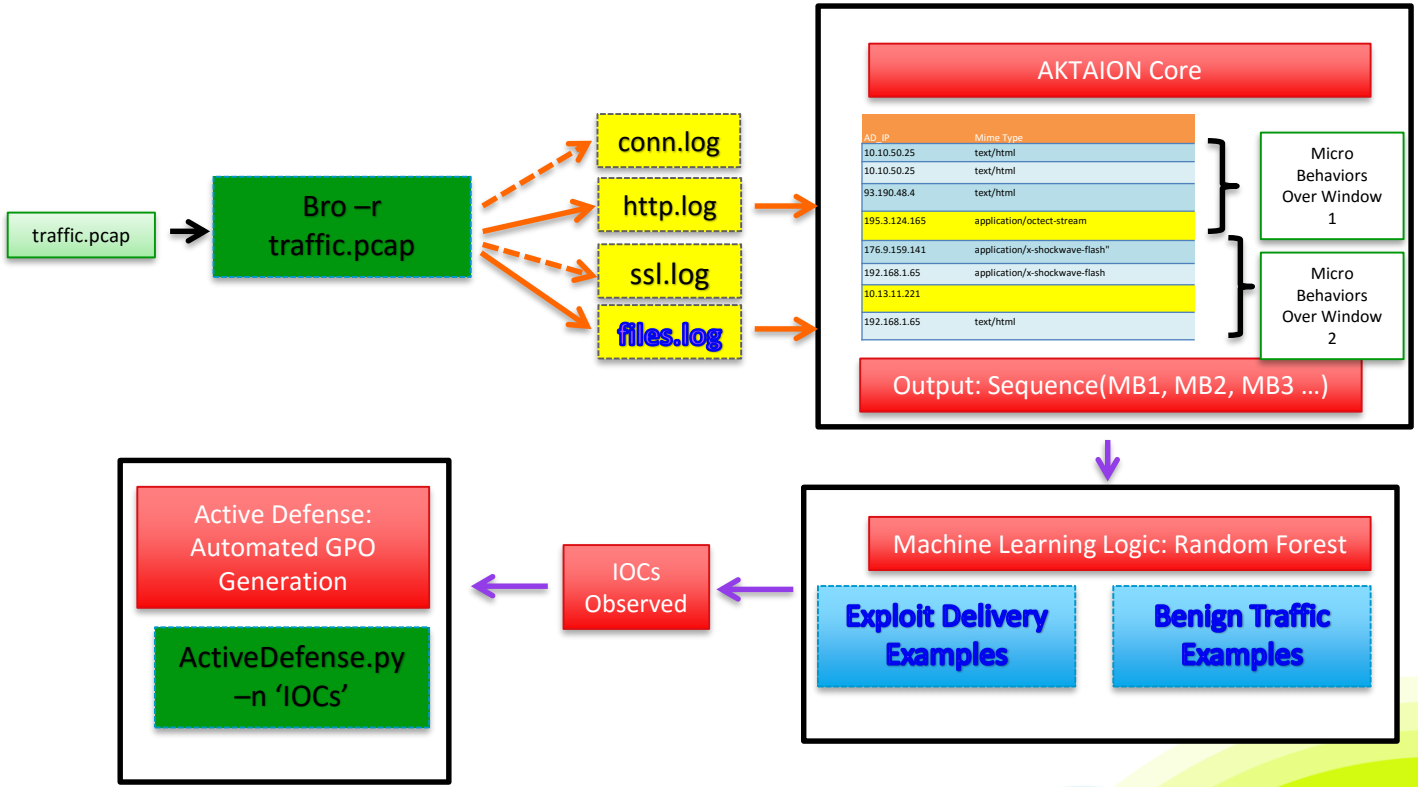
Predicted actual	Predicted no	Predicted yes
no	7671 (77.5%)	2231 (22.5%)
yes	3211 (21%)	12097 (79%)

```
| fit LogisticRegression fit_intercept=true "used_by_malware" from "bytes_received" "bytes_sent" "dest_port" "has_known_vulnerability" "packets_received" "packets_sent" "src_port" into "example_malware"
```

Data Sets Used To Train the Model

- Open Source Examples: github.com/jzadeh/Aktaion/tree/master/data
  - 386 Labeled Exploit chain examples from Contagio (pcap extracts into a generic proxy format). Thanks to the hard work of Contagio and Mila Parkour http://contagiodump.blogspot.com/
  - CRIME Database from DeepEnd Research (DeepEnd Research): www.dropbox.com/sh/7fo4efxhpenexqp/AADHnRKtL6qdzCdRlPmJpS8Aa/CRIME?dl=0
  - Ransomware Samples: small amount of mixed call back/file system level indicators
  - Labeled benign user traffic (days of http user browsing and related activities)
  - Anonymized bluecoat traffic

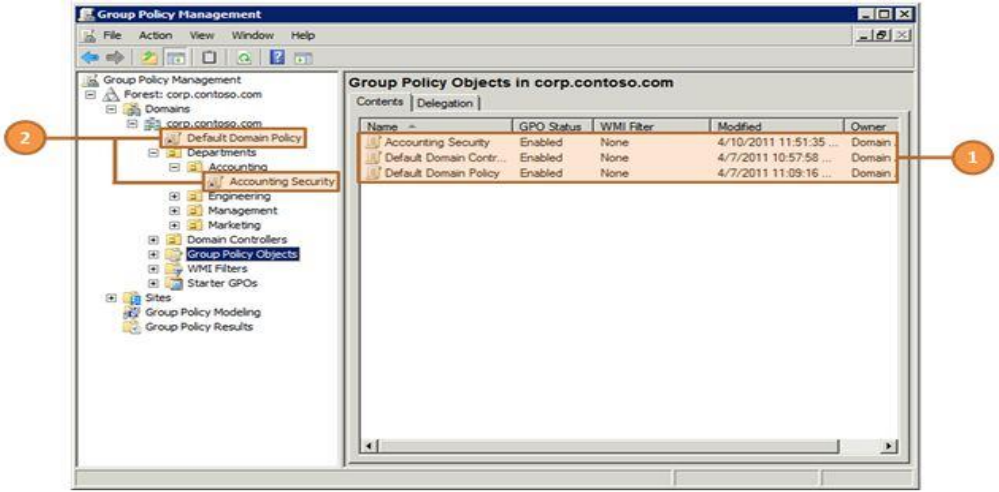
Aktaion Logical Workflow



What is a GPO?

Think of Group Policy as "touch once, configure many."

- Group Policy is simply the easiest way to reach out and configure computer and user settings on networks based on Active Directory Domain Services (AD DS)



Requirements for GPOs (Overview)

- The requirements for using Group Policy and following the instructions that this white paper provides are straightforward:
- The network must be based on AD DS (that is, at least one server must have the AD DS role installed). To learn more about AD DS, see Active Directory Domain Services Overview on TechNet.
- Computers that you want to manage must be joined to the domain, and users that you want to manage must use domain credentials to log on to their computers.
- You must have permission to edit Group Policy in the domain.

Advantages of using GPOs

- With GPOs, administrators can apply settings in granular, distributed and expedited way. (Think permissions, access rights, allowed processes, user/computer profiles)
- Enforce security settings on large scale (I.E password policy, firewall profile)
- Apply and enforce patching and security updates
- Apply security updates in a targeted, prompt and efficient manner.

Security Settings node of a Group Policy object.

- Account Policies (Password Policy, Account Lockout, Kerberos Auth)
- Local Policies (Logons, Filed Read, User Rights Management, Force logoff, halt if unable to audit)
- Event Log (Detailing log of events)
- Restricted Groups (Management of user/group membership)
- Systems Services (Rights given to services, auditing level for systems)
- Registry (Auditing registry keys/sub keys)
- File System (Access/Modification for system files/folders)
- Public Key Policies (Security Certificates)
- Internet Protocol Security Policies on Active Directory (how server responds to a request for IPSec communications)

Machine Learning + GPO = Active Defense.

- By leveraging big data and machine learning we can provide more granular and specific items applicable to Group Policy Objects.
- These ML+BD derived GPOs can be crafted and applied in an automated fashion, speeding up reaction measures.
- These GPOs can be more effective than static based signatures (Think about Malware variants and AV updates)

General Challenges using GPOs

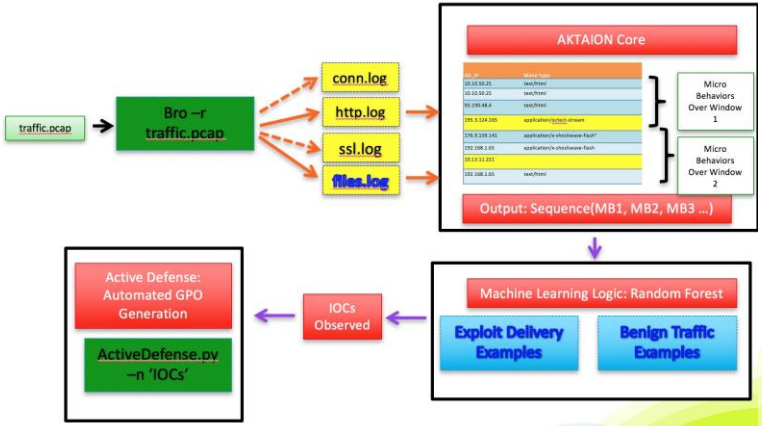
- Scope must be clearly defined. It requires system administrators to organize user, assets, groups.
- There is a level of skill required of administrators in order to apply GPOs efficiently (GPO settings)
- General infrastructure connectivity and redundancy can pose challenges (DNS, Subnets, WAN/LAN, etc)

Machine Learning + GPO = Active Defense.

- By leveraging big data and machine learning we can provide more granular and specific items applicable to Group Policy Objects.
- These ML+BD derived GPOs can be crafted and applied in an automated fashion, speeding up reaction measures.
- These GPOs can be more effective than static based signatures (Think about Malware variants and AV updates)

Proof of Concept

- Ransomware network traffic analyzed using Machine Learning open source tool: (Aktaion - https://github.com/jzadeh/Aktaion)
- This tool analyzes Micro Behaviors present in Ransomware
- Output of tool is input to python script which builds main indicators for GPO generation (Executable name, Domain, IP Address)
- Python scripts executes SSH into an AD host that can push GPO into Windows Domain via powershell.



Proof of Concept

Tool execution

scala-2.11 — java -jar Aktaion-assembly-0.1.jar — 101×53

```
caspidian:scala-2.11 rsoto$ java -jar Aktaion-assembly-0.1.jar
*************************
*************************
** Aktaion Version 0.1 **
*************************
*************************
1: Analyze Bro HTTP Sample
2: Analyze PCAP Sample (Bro must be installed)
3: Demo
Enter Execution Choice [1-3]:
```

Proof of Concept

Tool execution

2016-08-01 04:50:17 INFO WekaUtilities$:101 - Parsed 4 total lines.
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
****
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
****
Exploit detected in window number 1
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
****
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
****
2016-08-01 04:50:17 INFO RandomForestLogic$:183 - IocsExtracted(Set(5.178.71.10, 78.47.139.102, 182.50.130.156),Set(iamthewinnerhere.com, myexternalip.com, graphicstreeme.com),Set(/97.exe?1, /raw, /wp-content/plugins/theme-check/misc.php?34F0103544E2B25192E6AF0913ABE73BC21B0A31B82DC4E8D065CF5E9E55FEA92DB93FE6AEEB312449485E01DC99E4D47932EB53448B09D340AA22EDE68F63A3938F85E00D8EC314F81B2FA6DA02F5F9807B15E9DEFBA2FEA622BBEE35988934E428A133418E0F6B4B11E2918502CB158ABFEAC8D7C77C6542D07AB697F9CDA2EF564892C0B4B680EDB5BB1E6BDB74300CF63F55F4CC39E3E83EE9D8B70685F6D965ED309AF07DDF143D5082AAF0B0D27F422C89DD4F3BFF4CD93A9EBE0A83B5669779E6C050DA4291F89F85727F7EFBFDD96C9149B12C2397F1BA29A7C5CAB5036EB5B02B6ED79379D563C464717B1BE051BA3244EC5F8CE5D5E101F1555486A911A36F546A928CA17CF60FA2FEDEC2F71B2DB6752FC4567112FF797441ECFB6F093FEB8FDF192788AE0FFC9D5662CB88D9F7F8C50576359807C8F8FE4E8AA9965D546DF52000AADC544A03DFFCE596A387D5120254BA0E135ECDB9CB1F1127, /wp-content/plugins/theme-check/misc.php?

Proof of Concept

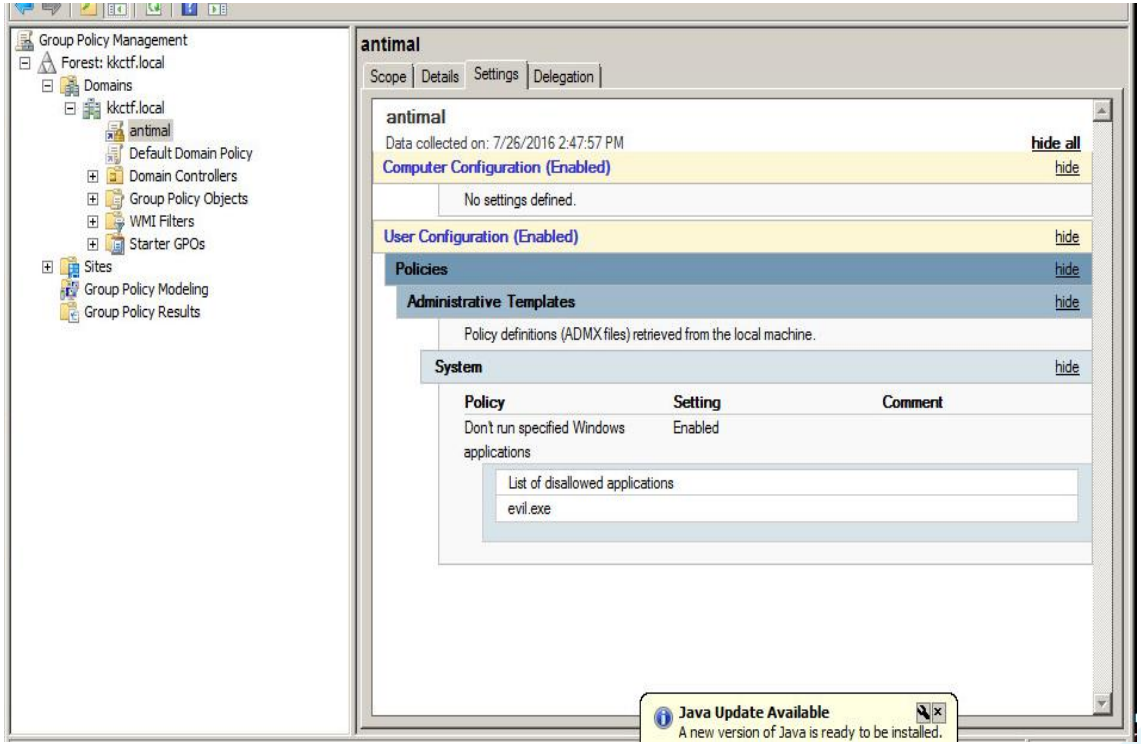
Tool execution – Script pushes GPO into AD

```
research@securityonion14:~/Desktop$ python aktaionAD.py -f event.json
Executing command- C:\Windows\System32\WindowsPowerShell\v1.0\powershell -Inpu
tFormat none -OutputFormat TEXT -command "Import-Module grouppolicy; Set-GPReg
istryValue -Name antimal -Key HKCU\Software\Microsoft\Windows\CurrentVersion\P
olicies\Explorer\DisallowRun -ValueName 1 -Type String -Value evil.exe"
```

```
olicies\Explorer\DisallowRun -ValueName 1 -Type String -Value evil.exe"


DisplayName       : antimal
DomainName        : kkctf.local
Owner             : KKCTF\Domain Admins
Id                : 30b3a3cb-76bf-4345-a3b3-3ac91c21f916
GpoStatus         : AllSettingsEnabled
Description       :
CreationTime      : 7/25/2016 4:54:44 PM
ModificationTime  : 7/29/2016 2:35:22 PM
UserVersion       : AD Version: 15, SysVol Version: 15
ComputerVersion   : AD Version: 0, SysVol Version: 0
WmiFilter         :
```

Proof of Concept

Tool execution – GPO placed at AD



Proof of Concept – Further GPO actions

- Force logoff
- Remove Computer from Domain
- Disable password changes
- Disable access to network shares
- Enforce account lockout
- Prevent further download of payloads from internet
- Apply firewall rules

Conclusions

- Machine Learning + Big Data technologies + GPO can be effectively applied for active defense.
- These tools are available for use without major investment in every enterprise.
- Application of Machine Learning techniques provide enterprises with an alternative to passive, high cost low efficiency signature based technologies.
- Machine learning provides leverage against constant adversarial drift and TTPs