Real time machine actionable threat intelligence for OSSEC

Or: Next Gen Active Response

Michael Shinn
mike@atomicorp.com
www.atomicorp.com
Real time machine actionable threat intelligence

What is it?

Threat intelligence that is high quality and therefore a system can take an action without false positives.

How do you do this?

- Make that the requirement :-)
- Collect data from lots of other OSSEC users, not just from your own agents
- Use Machine learning and logic to find bad and good
Examples:

- Known SQL injection attack against a website
- Known bad malware
- Large low and slow brute force attack
- Actions that deviate from “known good”

The magic word is known
But how do you find zero days?

Magic! Machine learning!

And lots of data, from lots and lots of users.
The problem with machine learning is:

It’s only as good as the knowledge it has access to.
What’s machine learning?

Math!

One method, among many but we use:

- Naive Bayes
- Deep Learning
- Unsupervised clustering

Supervised Needs:

- Huge sets of training data
A quick class on deep learning

Deep learning map:

Inputs to Outputs
to find **correlations**

Deep learning is an approximator, it can find a given unknown function:

\[ f(x) = y \text{ between any input } x \text{ and output } y \]

*(assuming they're related, either correlated or causational)*

Or simpler: it finds the right f, or the correct way to transform x into y
A quick class on unsupervised learning

Unsupervised learning finds patterns or structures in data

   Clustering is the most common method

Examples:

- Crime hot spots
- Cell phone tower locations

There, now you’re all experts!
Just kidding, I only have 30 minutes.
Limitations of machine learning

Machine learning can’t do everything

Examples:

- Natural language
- New situations

*Deep learning techniques only work if it’s seen something before, it can’t reason or generalize.*

*At a certain point, more data doesn’t make things better.*
So what’s the solution?

Old fashioned logic to describe the world, **and**

Machine learning, **and**

And unsupervised learning

**Machine learning isn’t the end of programming or design**
Bottom Line

OSSEC now has real time threat intelligence and Machine Learning
Version 0 of OSSEC and machine learning

- Logs collected on the hub from agents
- Machine learning happened on the hub

Problems:

1) the systems “view” of the world is limited to your systems, so it’s ignorant of what good/bad is
2) good corpus to train models is limited to your data
3) more complex machine learning models requires significant amounts of processing power
Version 1: talk to everyone

- The new machine learning model is to collect and share data with every OSSEC user

Advantages:

1) More knowledge - **100,000s of OSSEC users - which can be millions of endpoints!**
2) More power
3) Processing of machine learning can be done with specialized hardware
4) You can detect new novel attacks
5) You can prevent false positives and automate responses
But what about privacy?

Hubs only send conclusions, not logs, pcaps or sensitive data.

Conclusions are IOCs, for example:

- rule id
- hash
- malicious sources (eg IP, CIDRs, domain names)

You control what’s sent.
How Version 1 works:

OSSEC hub sends conclusions to machine learning cluster via JSON via cron.

Machine learning cluster evaluates data from all sources and creates new collections of conclusions.

Hubs download cdbks for local analysis from machine learning cluster.

Hubs use cdbks to ask questions (eg is this file safe, is this source malicious, is this action bad).
Version 2: really real time

- Real time lookups to the machine learning cluster (no local data store, but cached)
- TI integrator performs real time look up of question via DNS protocol

Why DNS?

1) It’s fast and lightweight
2) Everyone allows DNS outbound, and we can cache answers in DNS servers as long as we want (TTY)
3) We can ask almost any question with this protocol
Architecture

Version 1:

Send data: Hubs >(json) TI collector cluster > Machine learning cluster > Rsync clusters
Get data: Hubs <(pull) Rsync clusters

Version 2:

Send data: Hubs >(json) TI collector cluster > Machine learning cluster
Ask question: Hubs >(dns) Machine learning cluster
This system in action

<show TI globe at atomicrbl.com/globe3/>
Future plans

Make rev 2 available in OSSEC
  ● Requires enhancements to OSSEC to do real time lookups (probably in parallel to prevent delays)

Make TI GUI available as PAAS

Multitenant
  ● User defined Tis
  ● User defined machine learning
Assets
Icons from the site
Icons from the site
General icons
General icons