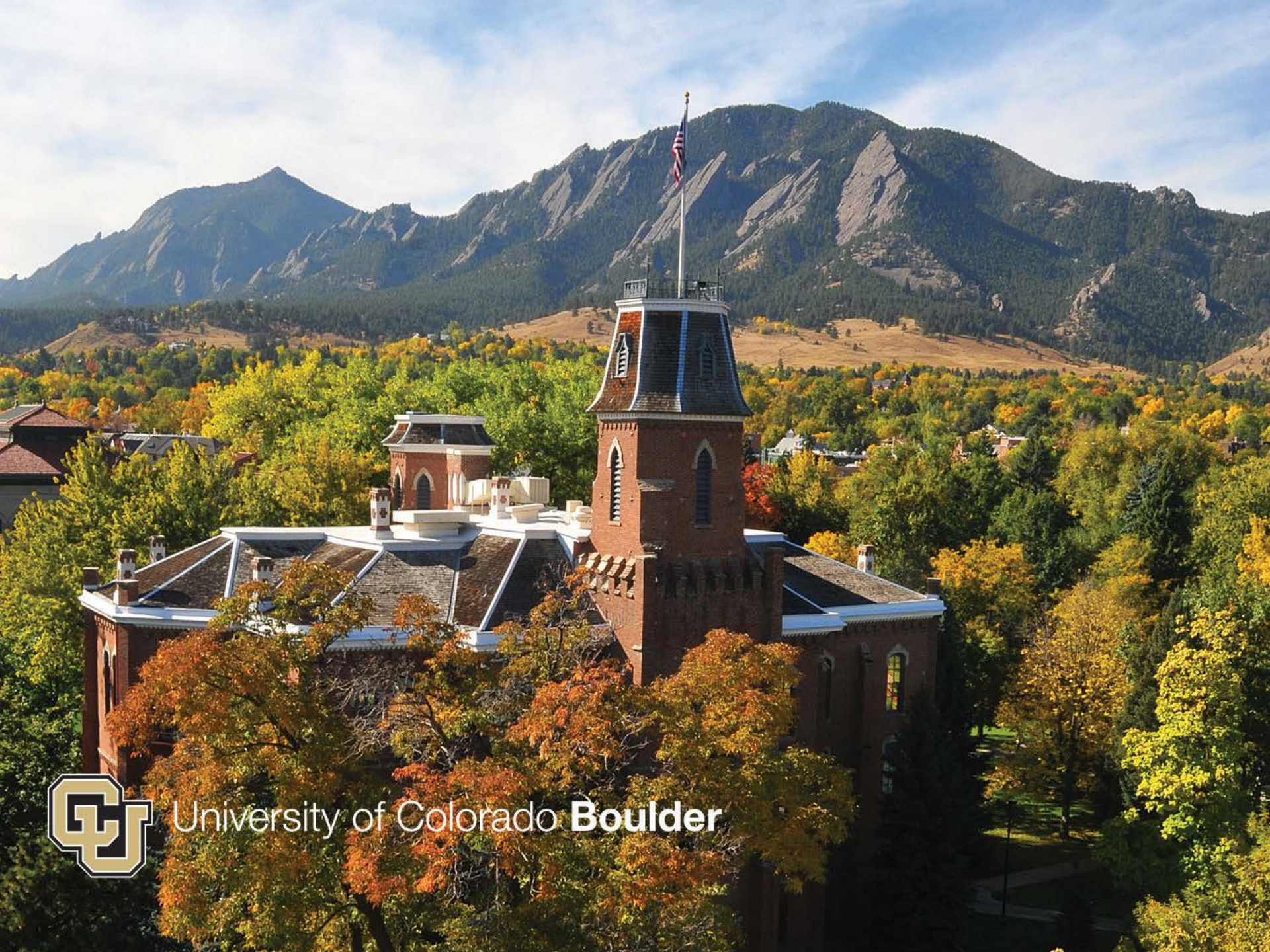

University of Colorado Boulder

Automated Detection and Analysis of IoT Network Traffic Through Distributed Open Source Sensors and Citizen Scientists

Joe McManus

Interdisciplinary Telecommunications Program

University of Colorado

About Me

- Currently:
  - Prof CU Boulder
  - Sr Cyber Researcher CMU/SEI
  - CISO Automox
- Past
  - Director of Security SolidFire
  - Sr Research Manager Webroot

How did this come to be?

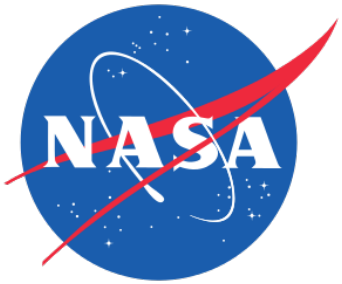
- A security company hired me to create their IoT security product, an agent based security product.
- Does not scale.
  - ESP8266
  - ARM
  - Atmel
  - etc

Question

- **Can we secure the Internet of Things through network based detection leveraging low cost distributed sensing?**

Example of Citizen Scientists

- SETI@home
- Weather Underground
- NASA
- CitizenScience.gov



Sub Question 1

- Using machine learning of network data can distributed analysis and collection of network behavior be used to generate intrusion detection signatures and firewall rules?
- Why not just use a firewall?
- **Home users do not understand NAT/PAT**

Sub Question 2

- Can a low cost sensor network be designed to empower citizen scientists collect data that will be used secure the IoT?

- **NetFlow , DNS, IP Reputation**
- Must be affordable (<$200)

Sub Question 3

- Can effective visualization and machine learning be used to develop tools that will represent a large amount of network data to allow citizen scientists to quickly analyze and respond to the collected data?

- **Visualizations that will encourage the user to explore the data**
- **Do not rely on static signatures**

Problem Space

- 5/11/2017 – Persirai malware infected 100,000 IP cameras made by overseas manufacturers. Actively being used in DDoS campaign.

- Based on a variant of the Mirai BotNet which caused interruptions to Akamai and many other sites by attacking DNS providers.

Mirai as an Example

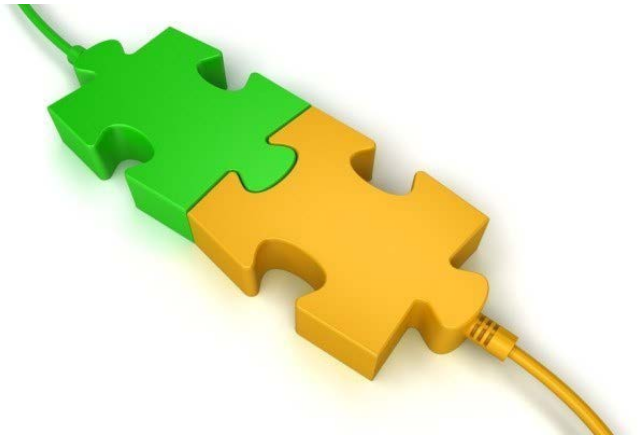
- Why is Mirai a good example ?
  - Even though we use the term Malware, this was not what we'd classify as malicious traffic
  - There was no malicious payload
  - The attack used a list of 50 default usernames and passwords.

Mirai Names and Passwords

USER:	PASS:	USER:	PASS:
root	xc3511	admin1	password
root	vizxv	administrator	1234
root	admin	666666	666666
admin	admin	888888	888888
root	888888	ubnt	ubnt
root	xmhdipc	root	klv1234
root	default	root	Zte521
root	juantech	root	hi3518
root	123456	root	jvbzd
root	54321	root	anko
support	support	root	zlxx.
root	(none)	root	7ujMko0vizxv
admin	password	root	7ujMko0admin
root	root	root	system
root	12345	root	ikwb
user	user	root	dreambox
admin	(none)	root	user
root	pass	root	realtek

Problem Space

- 2/2/2016 Fischer Price patched vulnerability in API of toys that exposed data (audio and video) to attackers.



Problem Space

Director of National Intelligence wa of IoT security threats

He also says that Russia 'will remain a major threat.'

Rob LeFebvre, @roblef
2h ago in Security

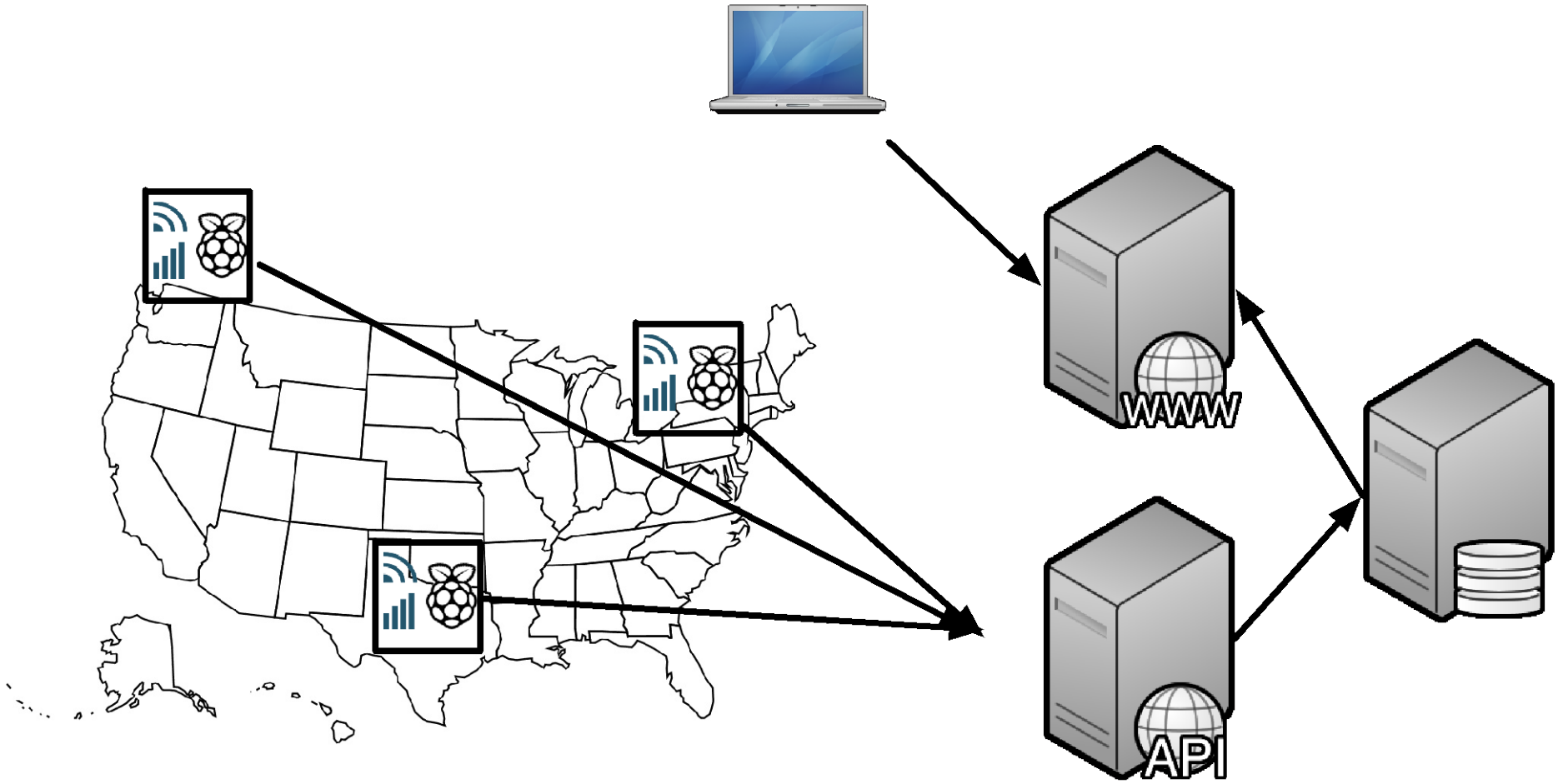
0
Comments

287
Shares

Progress

- Sensor
  - The sensor platform is built
  - Collecting data in 3 installations(have 15 ready to deploy)
- Cloud
  - Cloud server is accepting data
  - Processing and determining type
- Honeypot
  - Collecting flow traffic in a public IoT Honeypot
- Testing
  - Promising with 93% success rate

Sensor Network



Sensor

Sensor

- The sensor is built on the Raspberry Pi 3 with a 7” touchscreen display.
- Data is collected using:
  - NetFlow (SiLK)
  - DNS (Python)
  - IP Reputation (Python)
  - MQTT for data transmission

Visualization

- Traffic flow
- DNS Information
- Security Alerts
- Protections

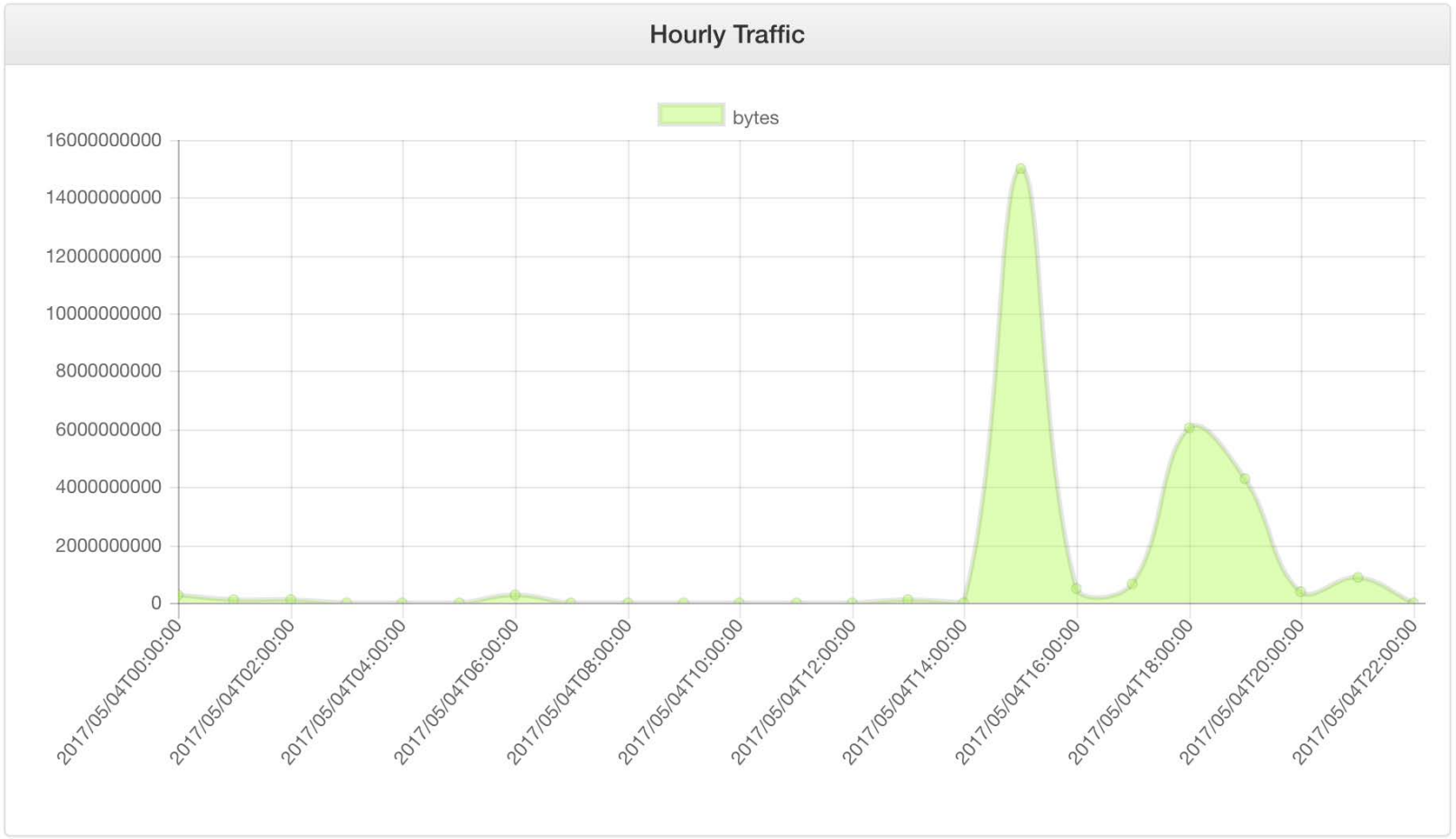
Protection

- As data is analyzed IP or hostname is determined to be bad and MQTT alerts the sensors to block or quarantine traffic.
- Using a software defined network(SDN) certain traffic that is not known bad yet can be rate limited and ACLs are put in place to protect.

DNS


DNS Map
a1851.g.akamai.net.
nest.agkn.com.
ssl.gstatic.com
00792fd90955bc2b54b7-dde9bcd8a491bb35da928cc2123a400b.ssl.cf1.rackcdn.com.
phant.io
a1850.g.akamai.net.
068ed04f23.site.internapcdn.net.
a3r2kfqjg515ls.iot.us-west-2.amazonaws.com.
a1386.g.akamai.net
cdn.netflix.com
0.pool.ntp.org.
cdn.netflix.com.edgesuite.net

Traffic


Hourly Traffic
bytes
16000000000
14000000000
12000000000
10000000000
8000000000
6000000000
4000000000
2000000000
0
2017/05/04T00:00:00
2017/05/04T02:00:00
2017/05/04T04:00:00
2017/05/04T06:00:00
2017/05/04T08:00:00
2017/05/04T10:00:00
2017/05/04T12:00:00
2017/05/04T14:00:00
2017/05/04T16:00:00
2017/05/04T18:00:00
2017/05/04T20:00:00
2017/05/04T22:00:00

Sensor UI improvements

- Allow users to mark traffic as bad/good.
- With enough data push out rules to all.
- For that use black hole or rate limit using SDN.

- Allow users to subscribe to “paranoid”.

- Cloud Environment
  - Linux Servers
  - Restful API
  - Flask
  - Chart Graphing
  - TensorFlow
  - MQTT

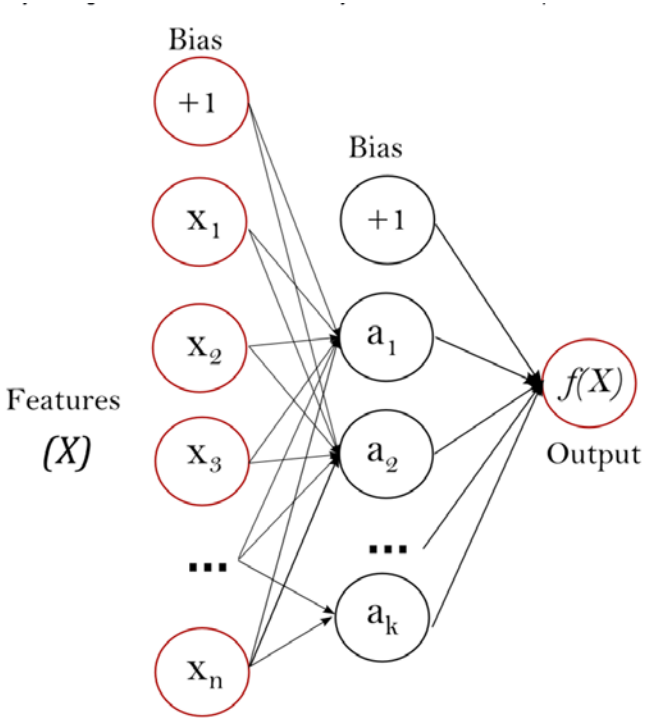
Chart.js

Flask

web development,
one drop at a time

Machine Learning

- Machine learning will be performed using TensorFlow
- Use a supervised neural network
- Multi Layer Perceptron



Test Data

- Training Data was used as follows:
  - 80,000 known malicious traffic flows
  - 10,000 known good traffic flows.
  - 10,000 mixed traffic (5000 bad flows)

Training Data

Malicious Traffic	Good Traffic	Testing Traffic
Reconnaissance: 30,000 Ports scanned using NMAP- TCP, UDP and SYN scans. **Brute Force Password Attack:40,000 login attempts using SSH and Hydra** **SQLi and XSS: W3AF Framework used to test IoT application for two of the top ten OWASP vulnerabilities.**	IoT Device transmitting data in real time. (Web Camera) IoT Device configuration updates. (Thermostat) IoT Device streaming media (Roku Media Player)	NMAP Xmas Scans, non randomized scans. IoT Traffic (Smoke Detector) IoT Traffic (Alarm System) Nikto Web Server Scan Customized Mirai Botnet Code

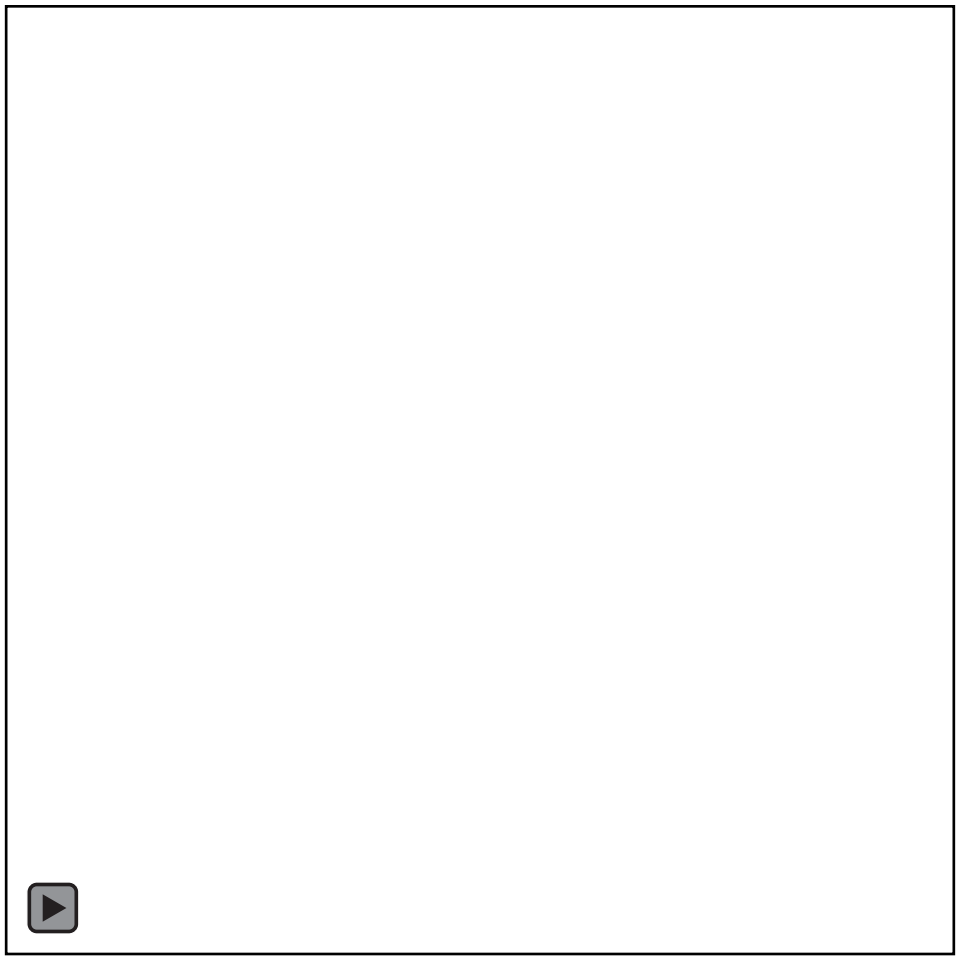
Results

- With the initial training data set loaded the system was able to correctly identify previously unseen attacks with 93% accuracy.

Honeypot

- Honeypot of IoT devices on the public internet
- There is no inbound traffic that is non-malicious
- All of this data is fed to the machine learning environment as known bad

Honeypot



Further Work

- Move cloud to EC2
- Create auto-update feature of sensors
- Distribute sensors to test group
- Increase DNS analysis
- Improve UI

Thank You

- Questions?

- joe.mcmanus@colorado.edu