

# A Global Network for Gathering Network Threat Intelligence

W251 Term Project: Jill Zhang, Todd Young, Leslie Teo

# OUTLINE

- A. Our Objective**
- B. How we collected network attacks - Honey Pots**
- C. How we managed the network - Modern Honey Pot Server**
- D. How we processed the data**
  - a. Streaming via Spark (MongoDB to Spark)**
  - b. Streaming via ELK**
  - c. Streaming via Splunk**
  - d. Historical analysis (MongoDB to Spark)**
- E. Demonstration**
- F. Conclusion**
- G. Challenges and future steps**

# OUR OBJECTIVE

```
Last failed login: Fri Aug 18 23:21:25 CDT 2017 from 58.242.83.34 on ssh:notty
There were 63246 failed login attempts since the last successful login.
Last login: Sun Aug 13 00:42:08 2017 from 132.147.65.14
```

**This project was motivated by the large number of “attacks” or unauthorized login attempts we noticed on our machines**

Where were these attacks coming from? Did it change over time? What else can we find out about these login attempts?

Was this a problem that changes from location to location?

Was this a problem for just SoftLayer?

# COLLECTING LOG DATA

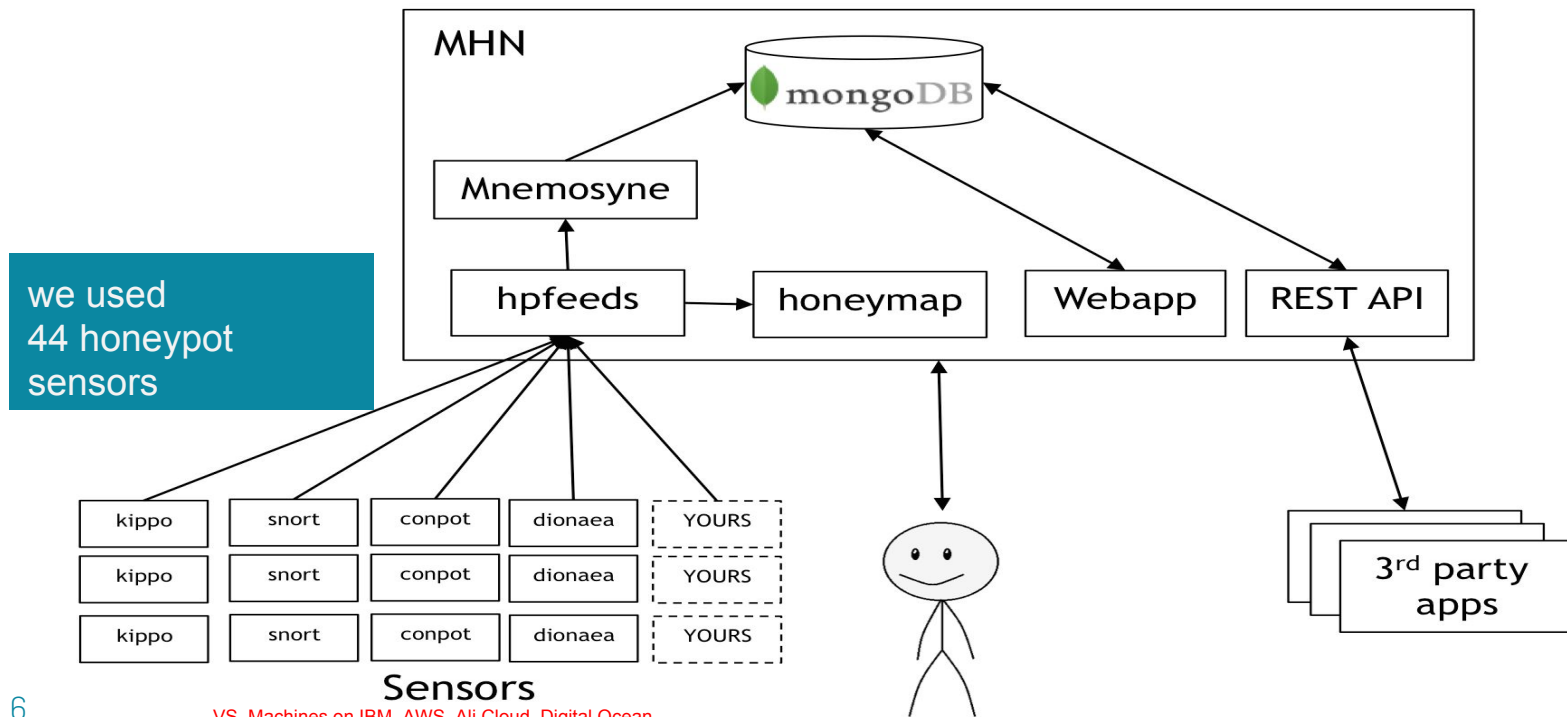
- First, we thought we would collect log data from `/var/log/auth.log`
- Then we found a rich open source solutions called Honey Pots
- Honey pots are set up to deliberately detect, deflect, and study unauthorized access

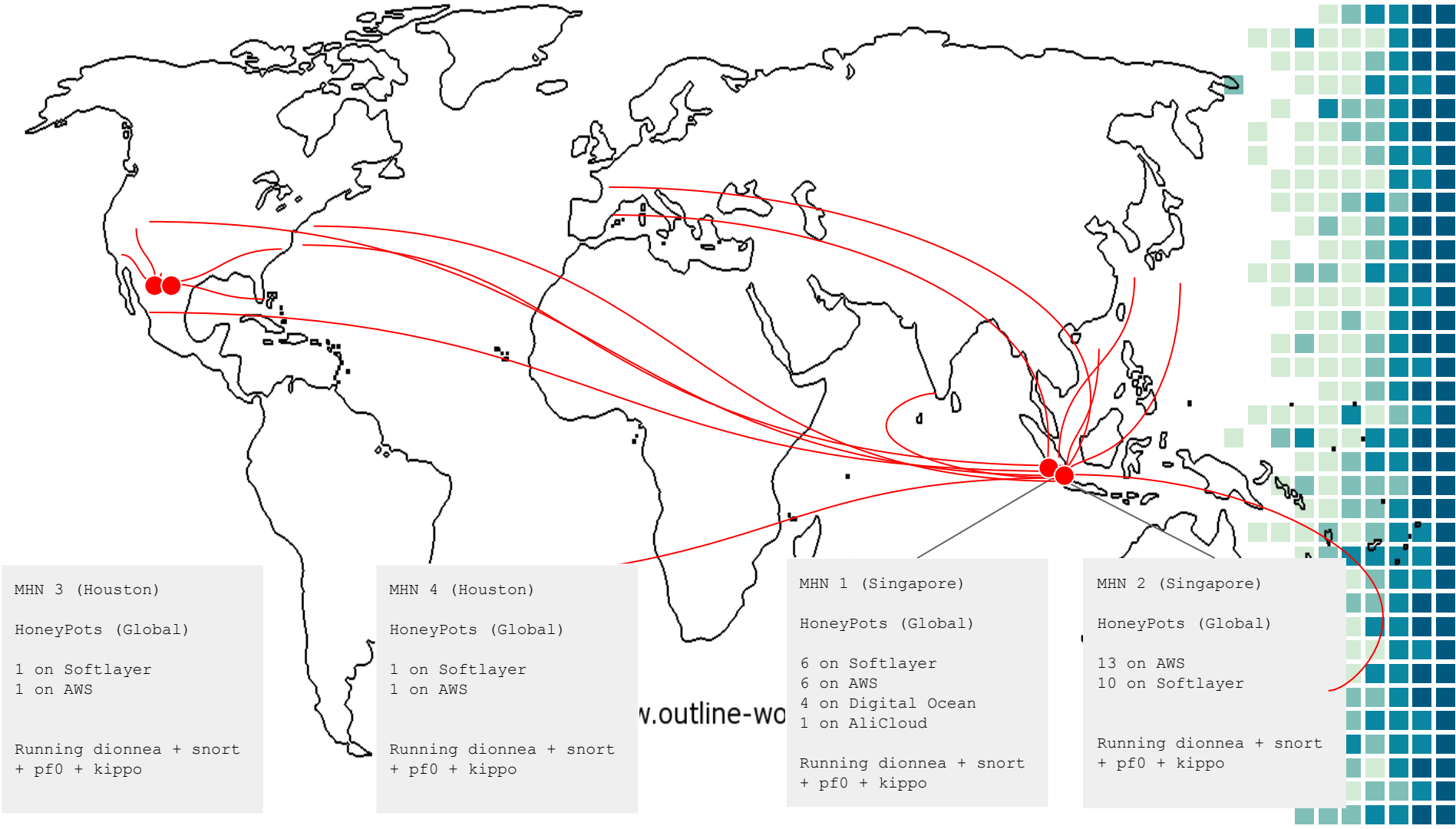


# OUR MODERN HONEY POT SERVER(S)



# HONEY POT SERVER (We created 4)





# TECHNOLOGY USED

## DATA COLLECTION

CLOUD VSs (AWS, IBM, etc)

MNH (open source)

- SENSORS
- HPFEED
- MONGODB
- NGINX

Ansible/Vagrant (to set up servers)

## STREAMING ANALYSIS

MHN NATIVE

ELK

SPLUNK

SPARK STREAMING

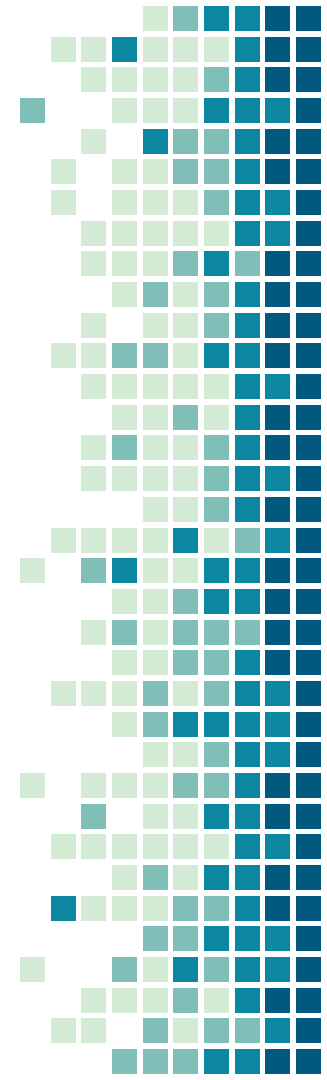
SPARK PYTHON GEOIP

## BATCH ANALYSIS

MONGODB/REST API

SPARK/SCALA

SPARK/PYTHON GEOIP

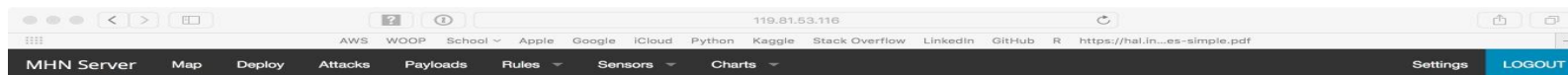




# PROCESSING THE DATA








# MHN Basic



## Attack Stats

Attacks in the last 24 hours: **45,170**

TOP 5 Attacker IPs:

1.  **151.13.11.188 (1,867 attacks)**
2.  **124.158.9.164 (1,853 attacks)**
3.  **121.201.58.37 (1,727 attacks)**
4.  **116.55.242.147 (1,613 attacks)**
5.  **125.77.17.56 (1,515 attacks)**

TOP 5 Attacked ports:

1. **445 (18,635 times)**
2. **5060 (4,348 times)**
3. **22 (3,682 times)**
4. **23 (1,685 times)**
5. **80 (782 times)**

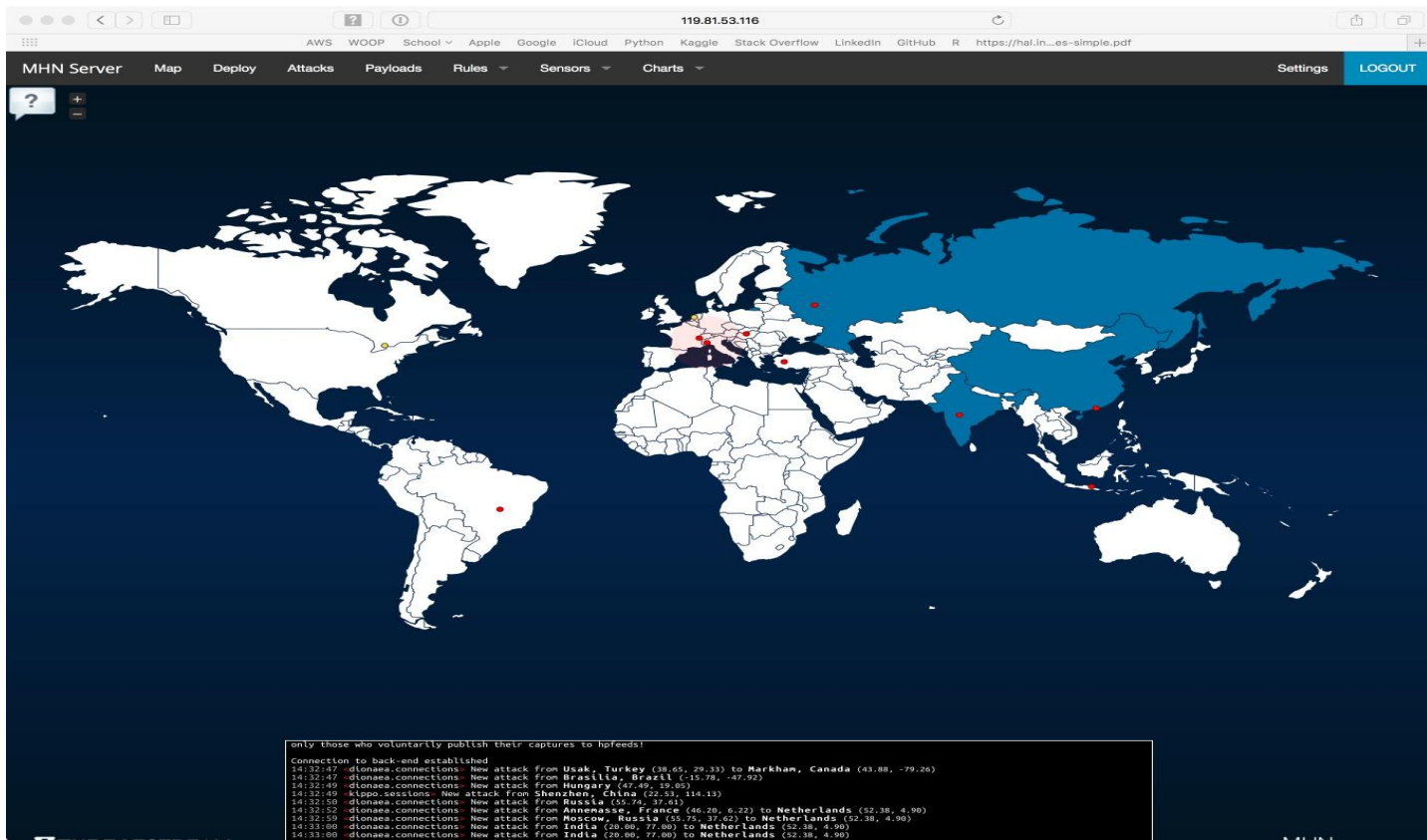
TOP 5 Honey Pots:

1. **dionaea (41,914 attacks)**
2. **kippo (3,257 attacks)**
3. **p0f (4 attacks)**

TOP 5 Sensors:

1. **izJ6c2c42yq1mjdlmdk7fZ (9,214 attacks)**
2. **worker3.ucshang.edu (7,741 attacks)**
3. **worker6.ucshang.edu (5,261 attacks)**
4. **worker5.ucshang.edu (5,216 attacks)**
5. **ip-172-31-9-187 (3,845 attacks)**

# MHN Basic



# Spark Streaming

MongoDB

- Initial the datastream every 10 seconds
- Query the database every 10 seconds before the current time.

Geo Code

- Get the geo locations of the attackers
- Get the destination location of the attacks

SocketStream

- Send the Data to TCP port localhost:6000



Top Countries

Top Cities

Top port

Top Honeypot

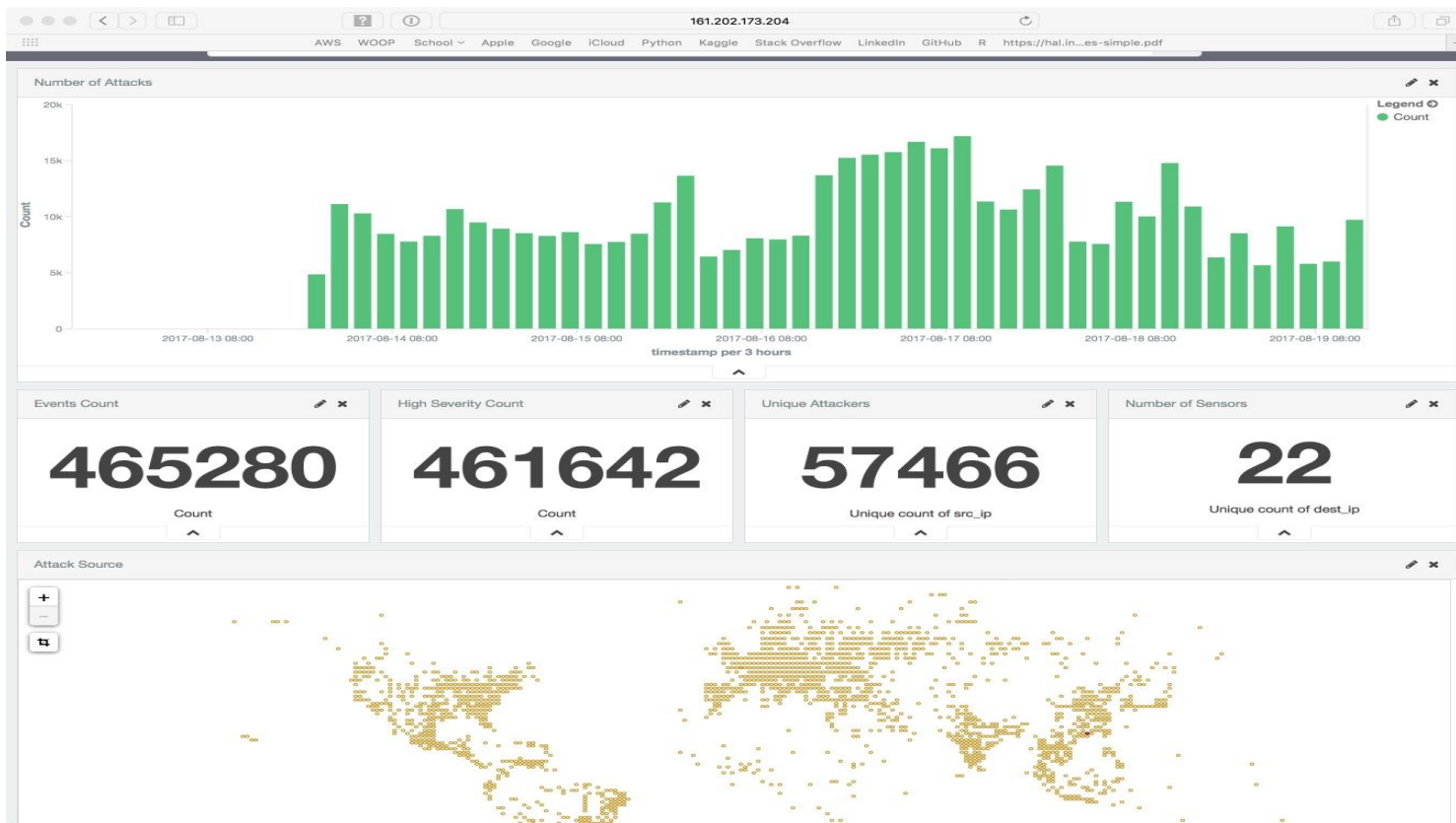
```
Country Info #####  
1 attack come from Russian Federation  
1 attack come from Israel
```

```
City Info #####  
1 attack come from None  
1 attack come from Nizhny
```

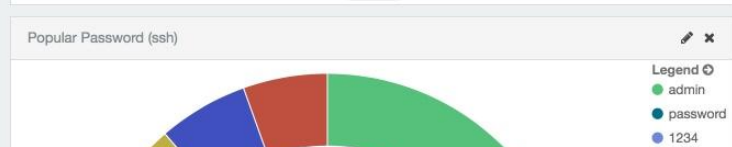
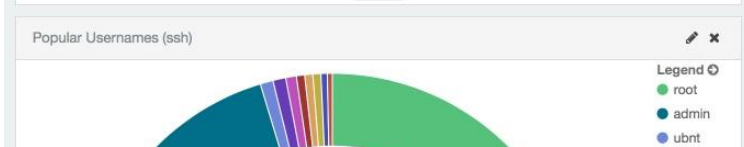
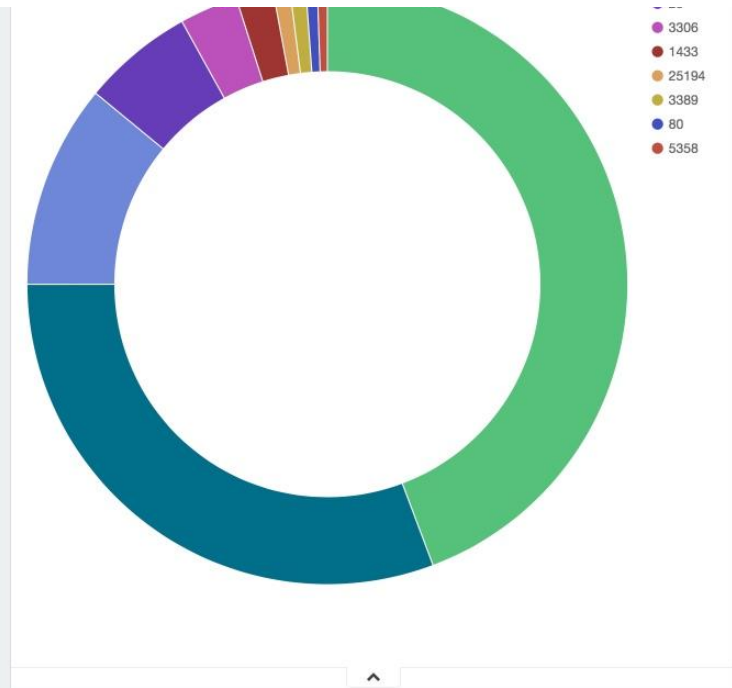
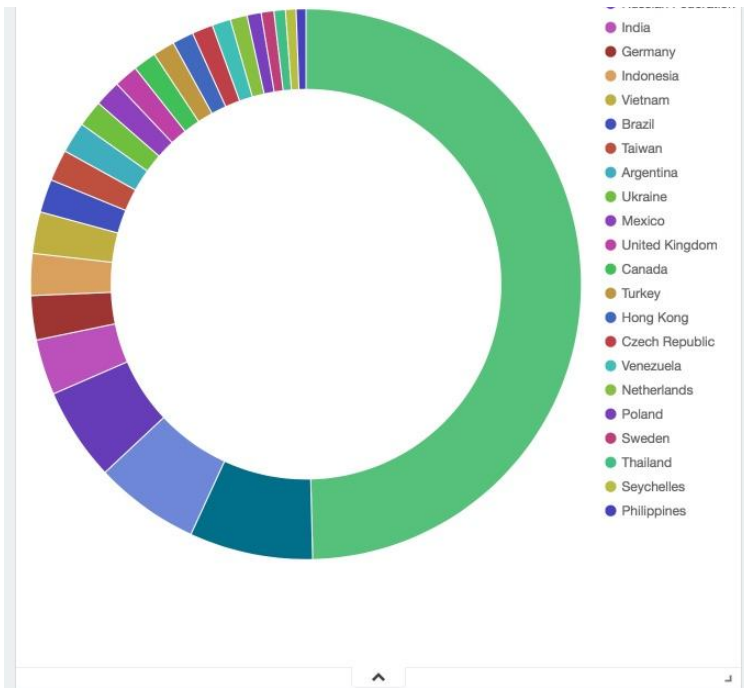
```
port Info #####  
2 attack at 445
```

```
Honeypot info #####  
2 attacks are dionaea
```

# ELK



# ELK



# SPLUNK



# Historical/Complete Analysis

To discover long-term patterns in the attack data, we combined attack data over the MHN servers into a single file and processed the data in batch-mode.

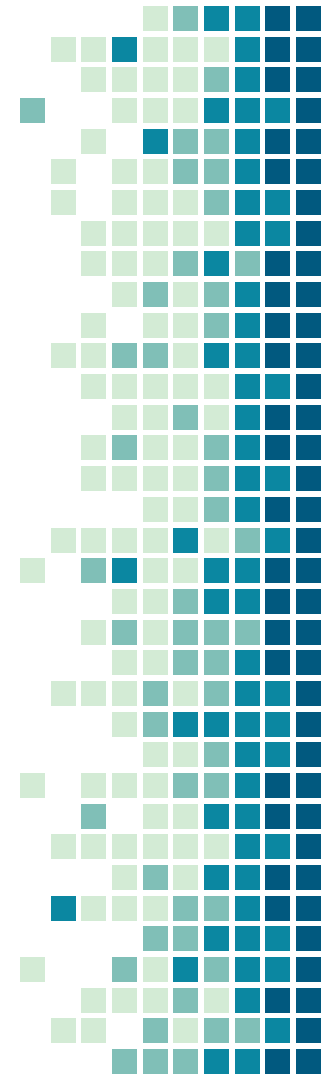
We processed ~2.75 million records this way.

We used an Apache Spark cluster with HDFS to process the records:

- Spark/Scala for json text processing, reformatting the data

- Spark/Python (with GeoIP libraries) for adding city/country information from IP data

- Spark Session for SQL-like summary statistics on data





# Historical Overview

Top countries by number of attacks: China (20%), US (11%), Russia (9.4%)

Top cities by # of attacks w/in China: Shenzhen(13%), Nanjing (10%), Beijing (9.7%)

Top ports attacked: 445 (35%), 5060 (11%), 22 (6.4%)

445 - Microsoft-DS Active Directory (malware: miner-bot, wannacry)

5060 - SIP (steal VoIP services)

22 - SSH (login)

Top attacking IP responsible for 3% of all attacks

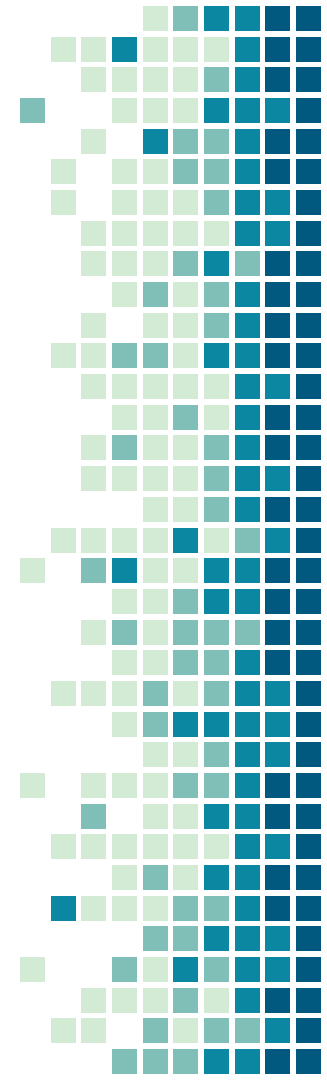


# Demo

**<http://119.81.53.116:3000>  
0 (Map)**

**[161.202.173.204:5601](http://161.202.173.204:5601)  
(Kibana)**

**[119.81.53.116:8000](http://119.81.53.116:8000)  
(Splunk)**



# CONCLUSIONS

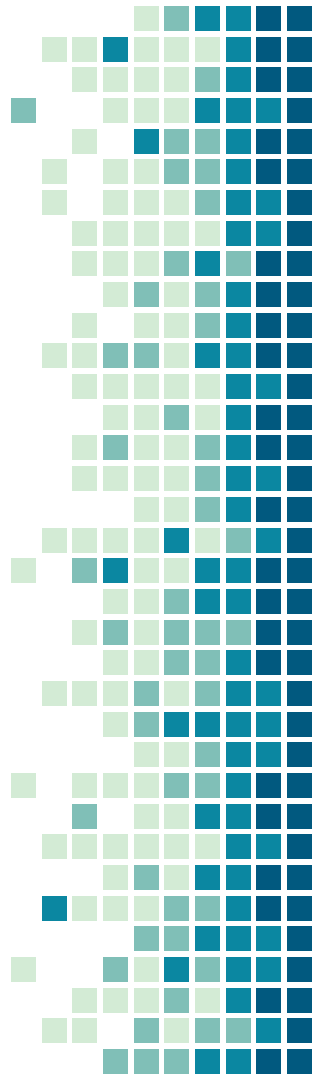


# 2,700,000+

In about ONE month!

160,000+ attacks PER DAY

~3,000-10,000 attacks PER HONEY POT/PER DAY



# THINGS WE LEARNED

## LOTS OF ATTACKS

There is no discernable difference in attacks over time. Or between AWS, IBM, Digital Ocean. Clearly these attacks are automated.

## WHY 445?

Un-updated Windows machines can be infected with malware, including WannaCry and Adylkuzz (Crypto-Miner bot-net)

## MOST ATTACKS FROM CHINA

US, FRANCE, RUSSIA are also up there. Some surprises include Vietnam and Indonesia

## DON'T ALLOW ROOT

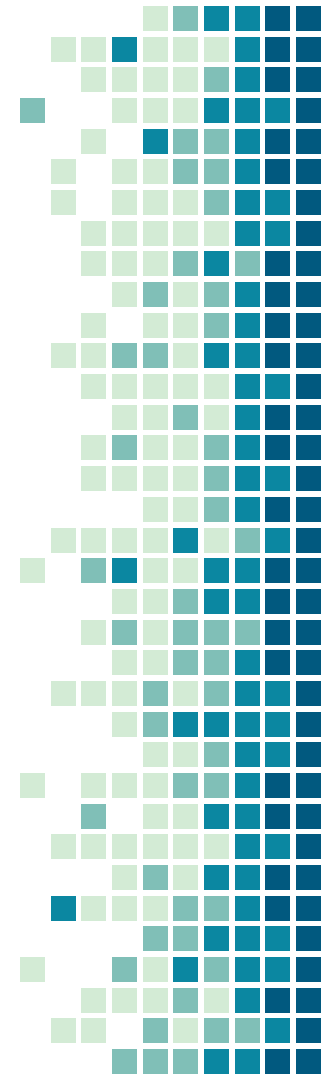
It's a bad idea to allow root logins, especially if password remote login is also allowed. Always change default passwords.

## PORT 445 (NETBIOS)

Other ports most commonly attacked are 22 (ssh), 5060 (SIP - VoIP), SQL related ports (3306, 1403), telnet (23)

## DEFENSE

Honeypots seem like a good way to capture and share threat profiles and sources



# CHALLENGES/FURTHER STEPS

1. It's not clear if source ips are really the source
2. We are reading logs but much more richer honey pots and collection of data on attacks will tell us more about them (e.g. studying actual "breakins")
3. We could build solutions that might be more "elastic"

# THANKS!

Any questions?