Data Analytics and Visualization In Cybersecurity

Prince Adom, Bruno Gomez, Babafemi Ogundana
Union County College, Union, NJ
Advisor: Dr. Ching-yu (Austin) Huang, chuang@kean.edu
School of Computer Science, Kean University, Union, NJ

Abstract

The advent of computers has seen also seen the rise and growth in intrusion attacks especially on websites and servers. Thankfully, server activity logging has become particularly important in tracking down malicious attackers, detect suspicious activities etc. Server logs keep track of all activities performed by the computer server. In the event of data breaches and unauthorized break-ins, Cybersecurity analysts make use of these files (gigabytes of raw data) to reveal vital information about the attacks as well as hackers.

Background

The purpose of this research is to process server logs (file detailing all activities generated by web servers), gather some useful information about their them and then display this data using tools such as Google Charts, Maps. The research aims to look for and track down all attempted break-ins by hackers, their location, time and date of the attack, etc. The research processed data from three different servers from 06/03/2018 to 07/14//2019, with a total of 668,560 records (execution time of 2 hours)

Methods

- Used PHP to extract (and insert into a database) useful information such as IP, time and date, server name, port number from 152 server logs.
- Used the GeoIP DB API to obtain more information about the IPs such as city, country, latitude and longitude (a new database table)
- visualize the iptable data using Google (Pie, Line and Bar) Charts

Materials

The following tools and materials were in this research:

- **IDEs and tool**s: XAMPP, Putty, Notepad++, FileZilla, Sublime
- **Database**: MariaDB
- **Platform (OS)**: Windows, Linux
- **Frontend**: HTML, CSS, JavaScript
- **Backend**: phpMyAdmin, PHP, MySQL, MySQL Workbench
- **API**: GEOIP DB
- **Data source**: server log files (from 3 different servers)

Summary

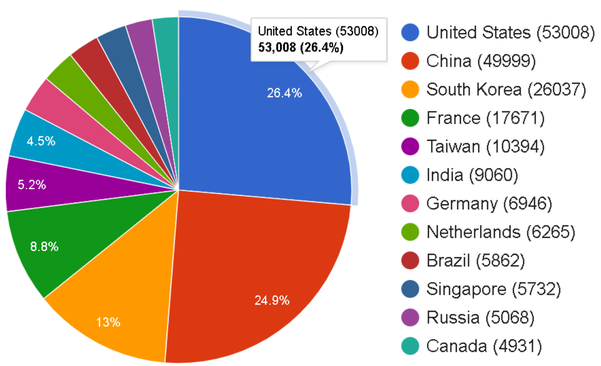
Findings: intruder information (IP address, port number, city and country, longitude and latitude) were obtained

Conclusions: The huge amount of information we obtained from the server logs are in themselves not very useful. Google charts and maps are one way to expand our understanding of this information. Thus, we are able to establish relationships, predict attacks based on trends, and better communicate findings to the world. In a world of massive cyber crimes and threats, data analytics and visualization are the key to ensuring internet security.

Challenges: Inaccurate data may result from intruders using masked IPs, VPNs and Proxy Servers, etc.

Future Work: live and interactive Google Maps and charts means to detect and trace masked IPs and times of the attack by analyzing time zones, etc.

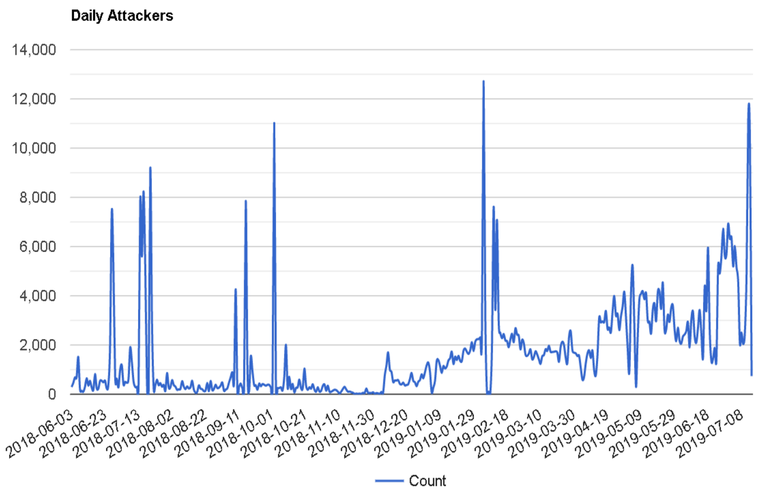
Results



A Google Pie Chart shows the percentage of attackers from the top 12 countries. US, China, and South Korea account for close to 64% of the attacks. Note: we were not able to retrieve the country codes for some of the IPs.

#	IP	count	country
1	40.143.1.60	10853	United States
2	211.50.130.9	10852	South Korea
3	162.243.131.116	10502	United States
4	27.105.92.6	8328	Taiwan
5	117.22.228.210	8097	China
6	198.148.118.199	5386	United States
7	119.10.114.9	4631	China
8	211.234.110.126	3775	South Korea
9	202.125.157.67	3590	Pakistan
10	211.47.191.21	2516	South Korea

The table shows the top 10 unique IPs that have most attempts to attack the servers. The IP 40.143.1.60 had 10,853 attempts. A country could have many different IPs in the database.



A Google Line Graph that tracks the number of attacks spanning from 6/3/2018 to 7/14//2019. Hence, attack frequencies can be narrowed down to specific month. Several specific dates got much more attacks than other dates.

References

- www.php.net: PHP Manual Documentation
- www.w3schools.com: PHP 5.0 Tutorial
- https://developers.google.com/chart/interactive/docs/gallery/linechart: Line Chart Tutorial
- https://developers.google.com/chart/interactive/docs/basic_multiple_charts: Bar & Pie Chart Tutorial

Acknowledgement

We are so grateful to Kean University and Union County College for providing us with the opportunity to take part in the 2019 STEMpact Summer Research. Also, we are thankful to Dr. Ching-yu (Austin) Huang for his incredible mentorship and guidance.