Development of Custom OS Intelligent Application

Data Analysis Report

Department of Information Technology and Management

December 2023

Version *1.0*

Table of Contents

1.1. Overview 3
2. Results 3
2.1 Phishing Dataset 3
2.2 Malware Dataset 3
2.3 Benign Dataset 4
3. Result Applications 5
4. Conclusion 5

1.1. Overview

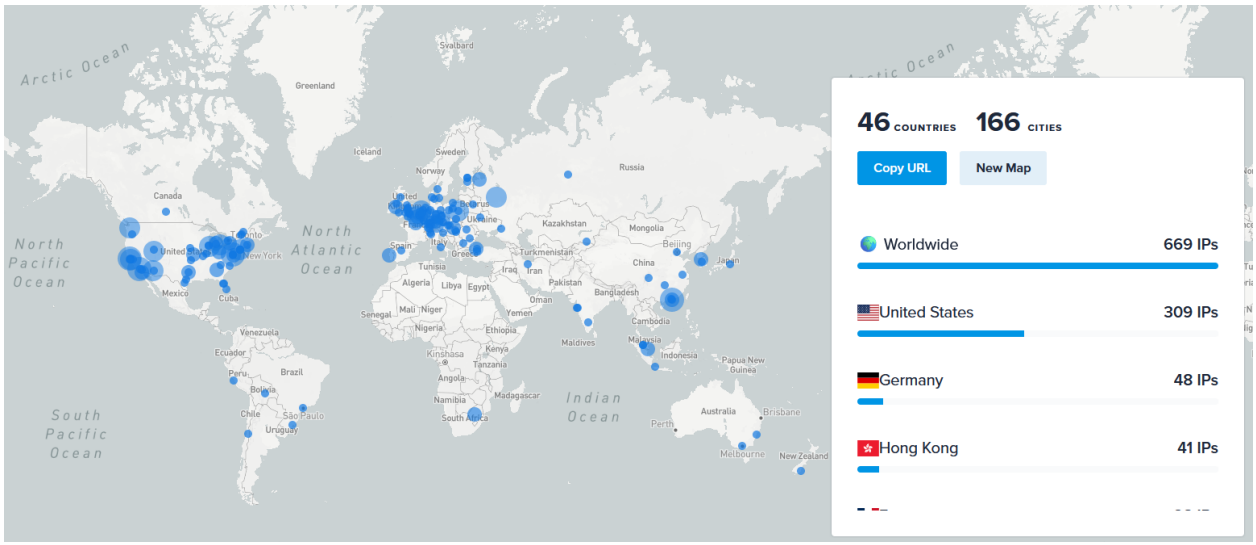
This document provides a comprehensive overview of an open-source intelligent application, presenting the application's outcomes, an in-depth analysis of the results, and their importance, as well as visual representations for a clear illustration.

2. Results

Over 50,00 URLs were scanned through the program, excluding any duplicate URLs. Considering the volume of IP addresses extracted from the data sets, the information has been visualized in a geospatial analysis categorizing IP addresses into three categories: phishing, malware, and benign.

2.1 Phishing Dataset

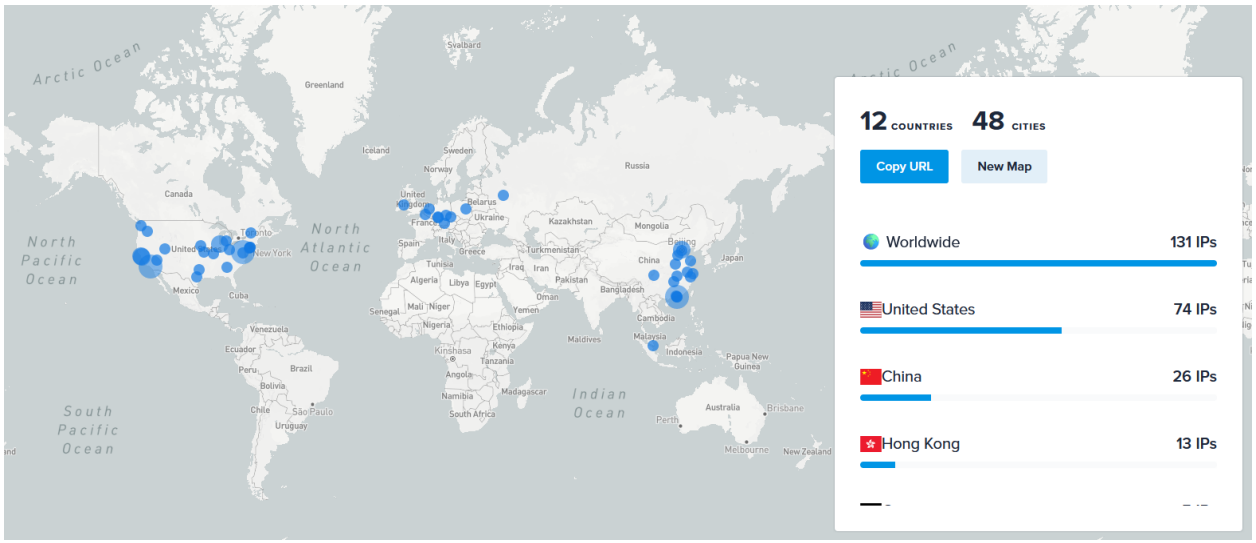
Of approximately 19,000 URLs retrieved from a data set dedicated to phishing, it was revealed that less than 9,000 IP addresses were due to the repetitive amount of data from the URLs. The geospatial analysis of IP addresses related to phishing reveals a significant portion between the United States and Europe, with notable spikes throughout China, Africa, and some parts of South America.



Source: https://ipinfo.io/tools/map/6dea8d14-9b9f-44fd-8699-068bb041637f

2.2 Malware Dataset

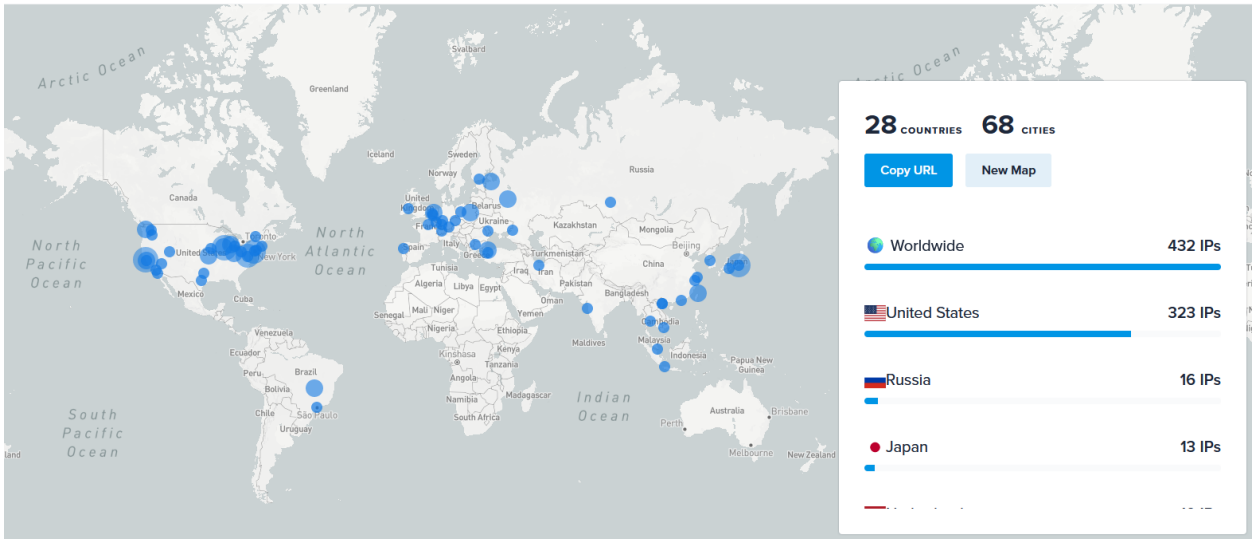
The malware dataset retrieved had a considerably lower number of URLs to scan at 11,000 as a result, only 121 IP addresses were discovered. Similar to the phishing sector, the geospatial analysis of IP addresses connected to Malware reveals a significant portion in the United States. However, in malware specifically, there were more spikes in China and throughout Europe.



Source: https://ipinfo.io/tools/map/0d6cfdb5-3ef1-484c-8437-093210773d6a

2.3 Benign Dataset

The benign data set revealed 253 unique IP addresses. In this data set, a significant portion of IP addresses were considered benign, the majority being located in the United States.

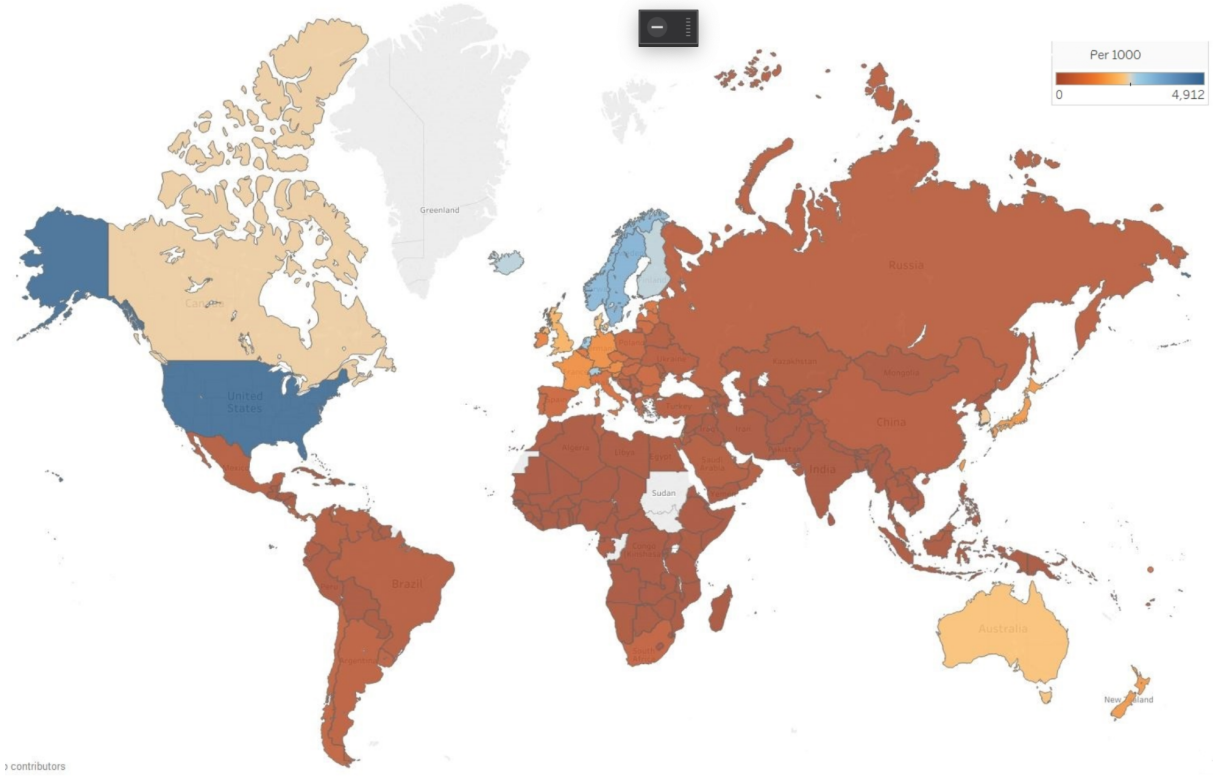


Source: https://ipinfo.io/tools/map/58d7c57b-a012-4e9b-827b-eec5df7df502

3. Result Applications

The analytical and geospatial analysis of this data set provides a distribution of potential threat agents globally. Threat actors can utilize this information to strategically target specific regions for phishing and malware activities including social engineering or infrastructure exploitation, focusing on regions with perceived lower measures of protection against these attacks. The geospatial analysis provides insight into regions with heightened vulnerability for cybersecurity attacks, suggesting a higher likelihood of attack within such.

When considering the large portion of these IP addresses that were concentrated in North America and Europe, it should be noted differences in policy structure. IP addresses are not considered personal data under the U.S. Privacy Act and many state privacy laws, however, they are considered personal data under the EU General Data Protection Regulation and EU case law. Interestingly, the US holds a higher concentration of IP addresses than Europe.



source

4. Conclusion

This research performed an analysis on 50,000 URLs which were geospatially compared to understand trends in phishing, malware, and benign sites. The concentration of malicious IP addresses in these regions provides insight into where specifically cybersecurity strategies can target for additional development in prevention against cyber attacks.