Cyber-Security Threats Origins and their Analysis

Maroš Čergeť, Ján Hudec

Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava,
Ilkovicova 2, 842 16 Bratislava, Slovakia
xcerget@stuba.sk, jan.hudec@stuba.sk

Abstract: The number of cyber-attacks on the Internet increases greatly and this trend does not seem to stop any time soon. The spread of malware is fuelled by various factors, with the technology and Internet access becoming relatively affordable globally and forced home-office of various professions, which increases the possible threat exposure time. Threat actors also employ various attack vectors, often utilizing some form of position deception to hide their activity from the authorities. Depending on the attacker's skillset, motivation and available resources, the attack may prevail into successful data breach, theft or data integrity violation. These outcomes may sometimes have tragic consequences. Without access to any kind of private data banks, this work was limited only to publicly available sources alongside with their drawbacks. This paper proposes a tool which is able to accept various sources of data, be it providers of suspicious IP address lists, lists containing IP addresses that are known to be part of any kind of TOR/VPN network, blocklists that contain various data and lastly also geolocation databases as means of gathering intel about IP addresses that are either part of suspicious lists or inside of manual queries. The proposed tool was then tested on publicly available data and the results, originating mainly from generated maps and graphs of various categories, alongside with the actual tool were compared to other cyber-threats origin information services and to other statistics about the situation in the online field.

Keywords: geolocation; threat origin analysis; cyber threats, public sources; public lists; information gathering; suspicious IP addresses

1 Introduction

With the continuing trend of more and more devices being connected to the Internet, the risk of exposure to some kind of cyber threat increases, even though digital skills of Internet users increase, and the number of Cyber security solutions is on the rise. The main factor to consider is that, even if only 1 ‰ of users was subject to successful attack, it would mean that 1 thousand people out of 1 million became victims to some kind of malware, phishing, scam, information leak, digital integrity

of data violation, etc. Reasons, for which threat actors often utilize same or similar attack vectors might come with increasing difficulty that is needed, to overcome security obstacles utilized by modern computers, or simply, because attack vector reusage, with minor personalization can still affect a great number of devices, especially those who still employ inadequate or zero protection.

Geolocation is a concept providing means of adequate geographical place of origin identification. The ability to track a country, in case of a bigger agglomeration also a precise state or a city may help a company selling advertisements, a website that adapts its language to visitor's expected language or to follow certain legislation and law requirements in different states.

However, some people's intentions are not always lawful, and their goal may be to steal, destroy, modify, control or in other way disrupt the device's intended operation, valuable information or even a whole company. Their motivation may be of pure greed, envy, revenge, but it can also be a thing of political and technological warfare. There are certain parallels that can be found even in the work *Art of War* [14], written by Sun-c'. And similarly, as we can observe the win of a trick over brute force in the legend of *Troy* [15], hackers often utilize means of social engineering or combination of various attack vectors, with great impact.

Devices are becoming more affordable and the number of connected devices to the Internet grows each day. This growth was accelerated also due to the fact that during Covid-19 pandemic, many companies moved their employee's daily operation online, greatly increasing the risk of their exposure to certain dangers, which did not go unnoticed by the attackers, further expanding their operations.

Their impact would be highly limited if they could be tracked easily, so they employ various evasive [11] and counter-detection techniques, to secure the highest up-time. In case that their goal is to demand a ransom for an unblocking of victim's computer, which is not always a guarantee as a deciphering part of the malware may not be present intentionally, they have moved the payments into crypto world, where payments appear to be more anonymized.

The aim of this paper is to get an insight into the Cyber security situation, or as it may be, its appearance through information that is publicly available from various sources. Precisely speaking, discovering the most notable countries of origin that appear to be the source of selected attack types, with an attempt to further improve the results with other findings, performing automatic data fusion, resulting in enriched outputs, providing outputs in the selected form, with the help of the designed tool, utilizing proposed algorithm that is visualized in the diagrams. Further goal is to offer an evaluation of generated results, creation and later description of figures and tables that we deem interesting, while also taking one, the publicly available nature of the data and second, the size of country size into account. The expectation and a major challenge of this research is that even if we had all the available data, not just public, it would still render the outputs incomplete, as many (especially ongoing attacks) attack vectors, threat actors and agents are yet to be discovered, if ever.

2 Related Work

There are multiple Cyber Security solutions providers, which make and deploy honeypots, IDS, firewalls, antiviruses, anti-spam, botnet detectors, etc. Gathered data is then used for further analysis and security optimization of their products or adapting solutions for their customers. Their data is not usually publicly available for further use, yet some of these providers publish visualizations with anonymized data showing current/historical situation about ongoing attacks, countries of origin/target with utilized ports, allowing for some result filtration [9].

Among those that we found and deemed most interesting are *Digital Attack Map* [1], which also shows important notice about current large and unusual attacks, and also *Talos* [8], which is a visualization from Cisco. Apart from the general view, it also gives a summarized view on top 10 spam and malware senders, with granularity focused on organizations and countries.

Other works focus mainly on using IP geolocation as means of pure blocklist adaptation or helping law forces in an attempt of investigation to hold certain criminals accountable, when the Internet Service Provider (ISP) can be contacted with the obtained geodata, to get the Network Address Table (NAT) mapping, incident logs to confirm the time accuracy of the occurred events, which can then lead to getting the Media Access Control (MAC) address of the host, their real name, real address [5].

3 Tool Design

Due to the technological limitations such as lack of live data about suspicious activity caught from honeypots, which are generally reserved and kept private by Cyber Security companies protecting certain institutions [7], it is difficult to get a true perspective of the situation, which gets even more complicated when we consider the fact that skilled criminals employ various means of protection, such as: Virtual Private Network (VPN) services [12], hijacked servers, botnets, The Onion Router (TOR) routing [13], cellular data from Subscriber Identity Module (SIM) cards not fixed to their Identity Document (ID) cards, temporarily paid hosting services [2].

We can explore the Internet and search for various sources that do claim to have discovered this information, but since we focus on publicly available (and free) resources, the accuracy of this data is limited. Nonetheless, approach that we propose aims to make use of these, to an extent, unreliable information [4] and provide an insight on the current situation and compare it to other available statistics, then judge the results.

3.1 Focus on Attack Types

It is important to mention that these categories are grouped into suspicious lists represented by their respective instances and tagged accordingly. Further geolocation and information gathering can be performed on the whole list by performing one action. If the list is too big, it is divided into batches that are then geolocated according to the limits of geolocation services presented in the system, which is further elaborated later in this article.

Dridex, QakBot, Emotet

Separate tags in the system. Found blocklist return data about IP addresses connected with attacks utilizing these malware types, almost entirely cryptocurrency ransomware [3], [6], [10].

Botnet

A found list of IP addresses that claims to hold information about IP addresses that were part of botnet. No more information is known.

Spam

A found list of IP addresses that claims to hold information about IP addresses that are suspicious of spam activities targeted on forums. No more information is known.

Mail

A found list of IP addresses that claims to hold information about IP addresses that are suspicious of attack on mail servers. No more information is known.

Resilient

A found list of IP addresses that claims to hold information about IP addresses that are online for at least 5 weeks and with at least 5000 recorded attacks are tied with them. No more information is known.

Brute-force

A found list of IP addresses that claims to hold information about IP addresses that are suspicious of brute force attacks, cracking passwords on websites, etc. No more information is known.

3.2 Geolocation

When it comes to geolocation, it is important to track the number of allowed requests per, e.g.: minute, week, month. In this work a minute interval was chosen globally for all the geolocation databases. Therefore, researcher needs to provide the system with correct information about the valid limits of the added provider. Empirical experiment is also recommended for some thresholds observations. Then also, the way how to request a response is needed. Names of the fields and format

in which the response is received is also required, in order to map them to fields in our database. We have chosen such Application Programming Interface (API)s that cover as much information as possible, as some services are not providing, e.g.: Autonomous system number/name or other provide us with information whether the IP is hosted, uses cellular data, … That will be another aspect of the provided detail alongside with other findings.

3.3 Blocklists

Blocklists generally provide a list of IP addresses that can be used as means of website protection. It is important to update them as, regular owner can regain control after some time, etc. These lists may contain tags as reason of presence in the list or even some other useful fields: country, Autonomous System (AS), etc.

3.4 General Idea of Outputs

The thought behind the output is that the viewer visits a certain page that is of three types. A graph, map, or an individual/fused output.

3.4.1 Individual/Fused Outputs

The term *individual* and *fused* always describes one IP address that underwent geolocation procedure, either by manual request via form, or selecting certain suspicious list (or its part). Visitor is always able to see more information about the sources of this data/findings. Individual means that information is provided only by one geolocation provider and fused means, that data is acquired and joined from all geolocation providers. The data is complemented with other module findings (cover lists, blocklists, suspicious lists).

3.4.2 Graphs

Graphical view, its data is issued for the visitor by his browser fetching specific files, generated by the system every N second (e.g.: 60 seconds, interval that can be increased/shortened) from data that is publicly available and present in the system. Chosen graph categories are *top* and *comparison*. The idea behind TOP is that the graph shows top 5 values in graphical way, where the bar type can be changed dynamically, and top 500 values presented in a table under graphs, in case more detail is required. The idea behind comparison is that we identify certain interesting topics, in which we can specify values up front, and the interesting thing is the difference in metrics that is observed in those values. In the top graph category unique values have to be identified dynamically, therefore, the limit for top 500 values in table is presented.

Top

- *Origin* – countries that appear to be used as threat actor's source most often.

- *Signatures* – biggest occurrence of signatures present in the lists in the system.
- *Ports* – ports found to be used most often in attacks.
- *AS* – autonomous systems out of which the threat actors seem to originate.
- *Tags* – types of suspicious lists and IP addresses used in geolocation.

Comparison

- *Online vs Offline* – number of IP addresses available vs already down [2].
- *Disguise* – number of IP addresses that use (and which) technologies to hide their real position by using another IP address(es) and those that do not.
- *http(s)*– number of threat actors that use SSL and those who do not.
- *IP vs domain* – number of IP addresses which operate under a domain, likely to act as someone more trustworthy, perhaps to act as some other similar domain.

3.4.3 Maps

Maps are visualized on interactive 3D model of Earth, where similar categories are shown with the added visualizations of ransomware, countries of origins, etc. Identified points are placed on the model, where they can be clicked on to get more information about them. This can be later used in documenting the outputs of this academic work and comparing with other publicly available data:

- *Specific origins for Dridex/Qakbot/etc.* – notable often occurring world points.
- *Specific origins for suspicious lists* – most occurring locations of each type.
- *Other categories chosen from Top* – places of origins for some categories.

3.5 Presented Algorithms

Every algorithm presented in this part is abstract. The exact implementation varies in detail and is different in a way that, e.g.: data about VPN/TOR/etc. are being looked for in *cover lists.* Other lists follow respectively in general, depending on the created graph/map files topic of interest. When performing suspicious list geolocation, it is a case when the actions of which contained IP addresses are suspected of are clear. Nonetheless, geolocation has to be performed regularly and exploration for other findings is performed as well during analysis. Another case is when the list input is manual. Then, no such information is available beforehand and exploration for potential match is performed for each list respectively alongside the regular geolocation. An optimization is employed, where no deep analysis is performed until the last batch of IP addresses is processed.

3.5.1 Geolocation Algorithm

Algorithm in Fig. 1 shows how various APIs and their limits are utilized, how lists are split into smaller batches that are then requested, processed and saved.

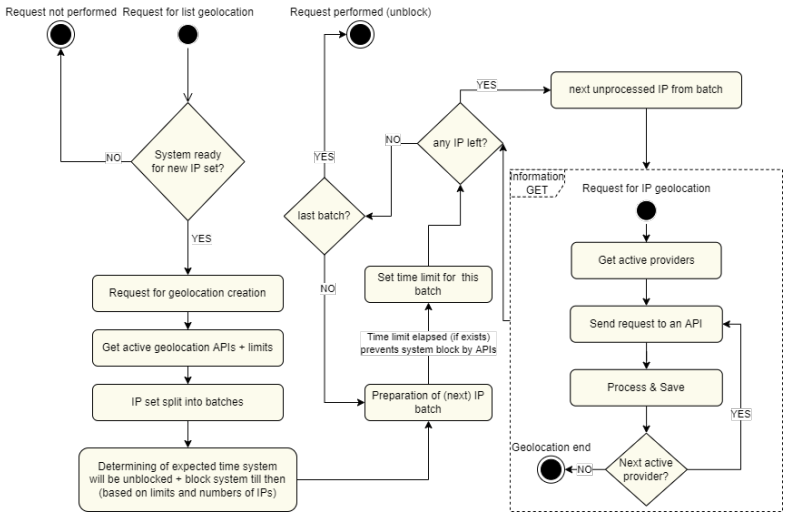

Figure 1
Algorithm showing geolocation requests for APIs

3.5.2 Files for Graphs/Maps Orchestrator Algorithm

This algorithm in Fig. 2 shows that the whole idea is to define event listeners and actions, which generate files for graphs and maps. Then, events are being fired regularly, based on the configuration.

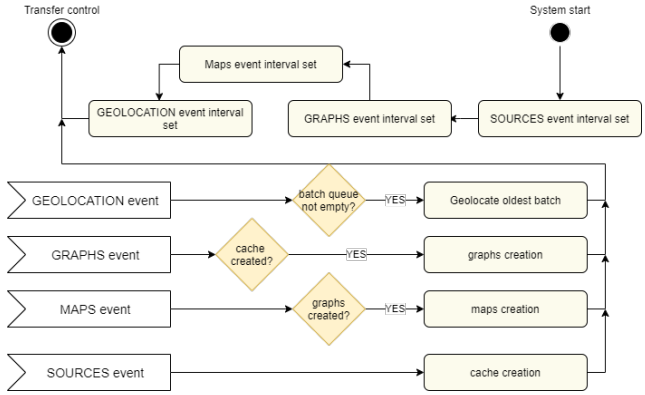

Figure 2
Algorithm that shows regular updates to files for graphs and maps

3.5.3 Findings Locator

This abstract algorithm in Fig. 3 is responsible for matching information from sources/lists present in the system, resulting in returned data, which is then put into file(s) that are utilized by graph/table/map engines.

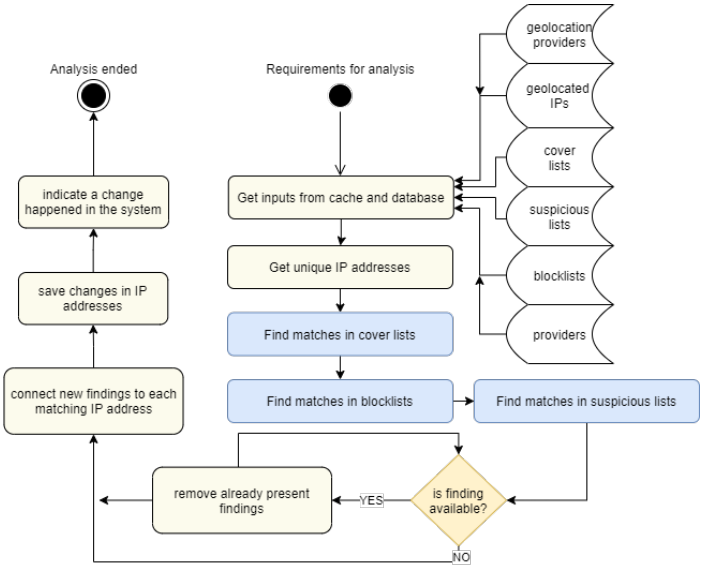

Figure 3
Algorithm showing process of looking for matches in various lists for unique IPs

4 Results

In this part, notable parts of the system outputs are presented, in a way that attempts to provide a meaningful view into the current state of cyber threats in the world, as observed based on acquired and utilized lists, sources and data.

4.1 Used Data

Every source used for analysis, has the nature of being published free of charge. The nature of those sources is, therefore, relatively unreliable. The true potential of the research tool can be, therefore, achieved only by having multiple sources of data that could prevent a great bias. However, as there is no such way on how the trustworthiness of used data can be tested separately, the only possible way is therefore to perform a test and then compare and evaluate gathered information.

4.1.1 Geolocation Services

Chosen attributes of requests are covered by two geolocation services, its APIs contribute to the database. The first service is able to provide 45 IP addresses per minute, the second service has no such limitation. Empirically it was proven that only 3 IP addresses per minute are valid. The system therefore chose 3 IP addresses

per geolocation event. The issue of lengthy lists for analysis is mitigated by implementation of batch pre-insertion control whether the IP address was already geolocated and, except for IPv6 addresses, every IP address is grouped into /24 CIDR subnets, with expectations of observing similar/same geolocation position, thus saving requests:

- IP-API https://ip-api.com/
- IPWHOIS https://ipwhois.io/

4.1.2 Suspicious IP Address Lists

It was not an easy task to choose which publicly available data should be used as a basis for this experiment, as there are some factors that influence these decisions greatly. It needs to be regularly updated, contain sufficient amount of data, while also keep the number of false positives to a minimum. We have chosen to follow a path, where we trust a reputable service [16] that first receives information from Fraud/Abuse specialist, whose servers are often attacked, as a source of IP addresses that are suspicious of performing attacks. We believe that since its data is pre-filtered with reputable whitelists, this data met our requirements.

The system was enriched with lists that are internally marked with tags and all IP addresses in them were subject to geolocation and further analysis. The respective date at which the data was recent is circa 3rd May of 2022 [16]:

- VOIP, SIP, SIP server attacks: https://lists.blocklist.de/lists/sip.txt
- Brute force logins: https://lists.blocklist.de/lists/bruteforcelogin.txt
- Mail, Postfix service attacks: https://lists.blocklist.de/lists/mail.txt
- REG-Bots, IRC-Bots, spam: https://lists.blocklist.de/lists/bots.txt
- Threats responsible for more than 5000 attacks, still online after at least 2 months of activity: https://lists.blocklist.de/lists/strongips.txt
- Attacks on FTP service: https://lists.blocklist.de/lists/ftp.txt

4.1.3 Blocklists

Following sources were added to the system, according to empirical response from these blocklist providers. Respective lists were obtained, out of which interesting attributes such as URL, availability status, IP address, user port, threat reason, were acquired and transferred to the database:

- Feodo Tracker – project of abuse.ch organization that aims to share IP addresses of botnet C&C servers that are responsible for Dridex, Emotet, QakBot, Trickbot, etc. malware family types.
- URLhaus – abuse.ch project that shares malicious URL addresses throughout which a malware of respective family type is delivered.

4.1.4 Cover Lists

There is a solid assumption that skilled attackers utilize one way of position deception as a protective measure in hiding their identity, or another. The identified techniques go as VPN, TOR routing, using Hosting services and cellular connectivity. One of the used geolocation databases provides our system with reasonably sufficient data about the IP addresses nature when it comes to hosting, cellular connectivity, but comes with only joined information whether proxy/VPN/TOR was used without any kind of distinction in between them. Therefore, providers and lists that could enrich our results are of VPN servers and TOR exit nodes IP addresses. The utilized lists in this testing are as follows:

- ProtonVPN – supposedly a list of servers, their IP addresses, of the service ProtonVPN. It is available from GitHub repository, where it continues to be updated in the regular manner and is processed as plaintext https://github.com/X4BNet/lists_vpn/blob/main/ipv4.txt
- NordVPN – supposedly a list of servers, their IP addresses, of the NordVPN service. It is available from GitHub gist file, where it continues to be updated regularly and is processed as plaintext https://gist.github.com/JamoCA/eedaf4f7cce1cb0aeb5c1039af35f0b7
- Tor-IP-Addresses – a list of continuously updated exit nodes of the TOR network available in the GitHub repository https://github.com/SecOps-Institute/Tor-IP-Addresses/blob/master/tor-exit-nodes.lst

4.2 Presented Outputs and Evaluation

The outputs of the system were chosen and put into this part of the work as a way of grouped presentation with the goal of important points summarization that can be further evaluated and compared. In the Fig. 4, the example of interactive 3D map output is presented to portray the way the visualization with clickable Points of Interests and Tabs separating context, containing the view for other attack types, works. Hence, it also serves as a distribution visualization of discovered threat actors. Other map outputs are shortened and put into tables in this work.

Fig. 4 shows that among used data, spam attacks were detected to originate predominantly in the USA followed by Russia and this list goes on with Indonesia, Germany, Ukraine and United Kingdom. The visualization utilizes the WebGL Earth service, where on-map points are clickable and provide more details.

Fig. 5 shows top 5 autonomous systems under which IP addresses suspicious of malicious activity, based on the utilized data, belong. Russian ISPs placed first and fifth, while other places are occupied by USA hosting services.

Description	Country	Count
bots	US - United States	8216
bots	RU - Russia	5134
bots	ID - Indonesia	1024
bots	DE - Germany	530
bots	UA - Ukraine	520
bots	GB - United Kingdom	512

Figure 4
Map visualization that shows distribution of 4th tab, IP addresses acting as bots

Figure 5
Graph showing Top 5 discovered autonomous system origins

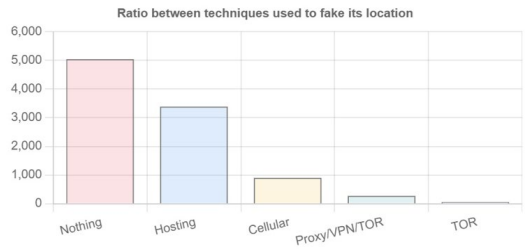

Figure 6
Graph showing Top 5 discovered position deception techniques in use

When it comes to position deception techniques utilization, in the Fig. 6, results from the discovered data can be interpreted in a way that majority of suspicious IP addresses are not using any kind of position camouflage. However, it does not mean, that the threat actors are not using any kind of protection, only that none was detected or they itself were misused for an attack in the form of agent. When it comes to detected deception techniques, the most popular seems to be Hosting, followed by cellular mobile, most likely being just a burner SIM card with pre-paid data plan, and then joined Proxy/VPN/TOR with no way of distinguishing in between them. TOR exit nodes are closing the graph. Specific VPN servers were also linked to some IP addresses, but the quantity was not big enough to overcome TOR exit nodes, for them to be shown in the Fig. 6. The number of IPv4 addresses found to be malicious dominate the graph comparison in the Fig. 7, where threats with IPv6 addresses almost do not exist, while there exists a certain amount of domains that serve as malicious data provider. It is important to mention that this is caused by the small ratio of IPv6 data in suspicious lists.

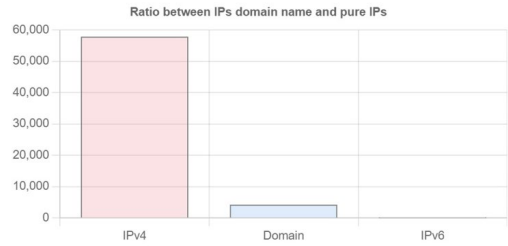

Figure 7

Graph showing the dominance of IPv4 addresses in the included blocklists

Fig. 8 shows unique threat origins (including addresses that use position deception). It can be seen that USA and Russia mostly continue to be seen as the predominant countries used as threat origin, with valid question that resonates:

"Whether this order is a coincidence, or if the situation comes from the fact that some countries have problems with regulation of cyber threats due to their size or possibly due to strategical political motivations."

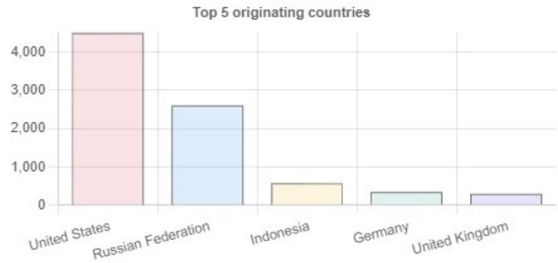

Figure 8

Graph showing Top 5 discovered unique origins of threats

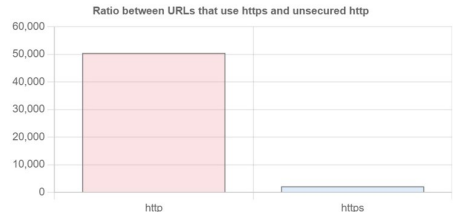

Figure 9

Graph showing the usage dominance of http over more trustworthy https protocol

In Fig. 9 the comparison of *http vs https* usage protocols is shown. It is necessary to mention that for the informed user, this alone might be suspicious. Only 4.12% of threats use https for their malicious activities and attempt to hide the traffic sent.

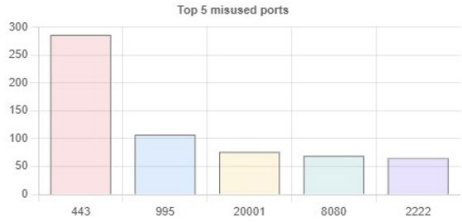

Figure 10

Graph showing Top 5 ports discovered being misused by the attackers the most

From Fig. 10, the prevalence of port 443's misuse is evident. Rest of the ports displayed in the graph continue with a lower, but to themselves similar quantities, trend of which continues even with ports that are not shown in Top 5.

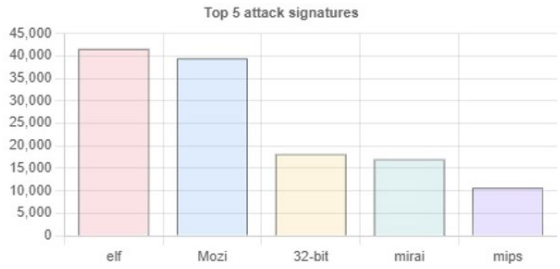

Figure 11

Graph showing Top 5 threats tags present in the system blocklists

Fig. 11 shows the quantity of threats that are present in the system. It may be surprising to see threats aimed at Linux operating systems (elf) in the first place, which contradicts the popular belief of virus free experience. Dominance of botnet malware families (Mozi and mirai) is also present in the graph alongside Linux threats. Less common, but still very potent threats include malware that targets IoT devices specifically (mirai and mips). Among others, there is a great number of threats consisting of shell codes built for 32-bit operating systems.

In Fig. 12, comparison of current threat activity status is shown. The fact that a threat is offline, does not necessarily mean that the URL or IP address is no longer reachable, it may mean that the rightful owner has regained access, in other words, an address no longer serves as threat agent. The majority of threats present in the system are offline. Therefore, if any kind of publicly available (e.g., via REST API) blocklist would be provided by this tool, only IP addresses of threats that are online should be considered 88.38% of threats present are offline.

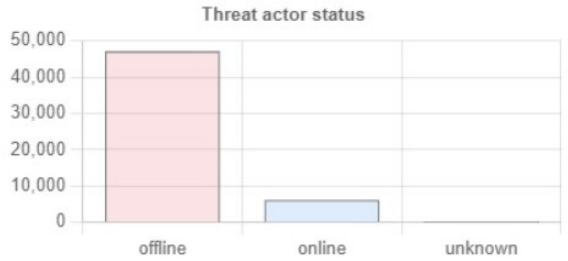

Figure 12

Graph showing that the majority of threats in the system are already offline

Another continuation of figures, showing data from maps could be presented, but it is decided, that presentation in the form of tables may be of more comparison value. The following Table 1 shows the trend of top country order occurrence, listing the possible means of position deception, linked in the data fusion. Based on detected usage, USA Hosting services were found to be the most attractive to threat actors, possibly due to their availability and simplicity of use. Popularity of this technique was also found in Singapore, Germany, China and India. Cellular services were most often misused in USA and Russia, while TOR exit nodes, seemingly random, were found to be used most often in Luxembourg. For some data, the system was not able to clearly identify which technique was used, placing it into joined category Proxy/VPN/TOR, which also showed that the most prevalent origin is in USA and China. One of the major factors contributing to the repeating appearance of certain countries in tables is that not only many providers with massive networks operate there, are as well many companies offering commercial solutions to customers from around the world.

Table 1

Table showing top 10 of grouped position deception techniques identified to be used by threat agents, with the exception of Nothing detected, which was omitted from this graph

Order	Deception technique	Country
1.	Hosting	US
2.	Cellular	US
3.	Cellular	RU
4.	Hosting	SG
5.	Hosting	DE

6.	Proxy/VPN/TOR	US
7.	Proxy/VPN/TOR	CN
8.	TOR	LU
9.	Hosting	CN
10.	Hosting	IN

Table 2 contains information regarding order of botnet malware family type occurrence in a certain country. Majority of the findings belong to QakBot family type with countries like USA, United Arab Emirates, Mexico, India, Brazil and Pakistan being the supposed C&C server botnet origin. Emotet malware family is represented by USA and France while Dridex by USA and TrickBot by Columbia. Further listing of each category would be possible, but this is the chosen view which shows these threats ordered in quantifiable manner as seen in experiment.

Table 2
Table showing top 10 botnet malware families alongside with countries of origin

Order	Botnet family	Country
1.	QakBot	US
2.	QakBot	AE
3.	QakBot	MX
4.	Emotet	US
5.	QakBot	IN
6.	QakBot	BR
7.	Emotet	FR
8.	QakBot	PK
9.	Dridex	US
10.	TrickBot	CO

Table 3
Table showing top 10 threats marked by suspicious lists as ordered by their quantity alongside the supposed country of origin

Order	Threat type	Country
1.	Bots	US
2.	Mail	US
3.	Bots	RU
4.	Mail	RU
5.	Bots, Mail	ID
6.	Bots	DE
7.	Bots	UA
8.	Bots	GB
9.	Mail	DE
10.	Strong	CN

Table 3 shows that majority of threats are aimed towards forums performing malicious spam operations, while being followed by attacks aimed at Mail servers that are both sharing the same origin similarities. USA, Russia, Indonesia, Germany, Great Britain and Ukraine were found to be origins of these threats, with resilient IP addresses of China closing the top 10 table.

Conclusions and Future Research Directions

With the use of the designed and implemented software tool, experimental research was performed, with outcomes showing comparable level of accuracy to published data, confirming the leading positions of USA, Russia and China as countries with the most prevalent probability of being a threat actors' true origin or at least origin of the misused threat agents. Despite this, it cannot be confidently expressed whether this is a true state of the situation, as many position evasion techniques are utilized and even though they were, in lots of cases, identified, the true geographical location could not be determined. Other types of information that the tool provided are believed to be mostly accurate, the only issue being that the input data does not contain information about all types of attacks, but for the sake of keeping false positives to a minimum this set was chosen. The provided insight was created with the help of data fusion, from data that was all publicly available, parsed and filtered.

An important notice is that the generated results presented in this work serve mainly as a snapshot of the situation in May 2022. Ongoing monitoring would probably show changing trends in the most prevalent countries of threats' origins, but we believe that countries like USA, Russia and China would remain at the top. New types of threats could emerge, and they would be observed, as information from blocklists is being updated regularly. When it comes to commercial companies providing free informational visualizations online, the data is as well ever changing. But with limited amounts of mostly anonymized detail and often shown spikes of malware senders in certain countries as Brazil, Vietnam, Hungary, etc., it is not easy to do a thorough comparison. The main advantage they have, is that the data is under their control, giving them the ability to perform more advanced research internally [1], [3], [6], [8].

The strength of this tool is also its main issue. Data is freely available on the internet and even though the provider is generally considered trustworthy [16], it is not a guarantee for other providers, where it can be mostly just assumed. If someone would purposefully change the IP addresses in the public lists, or created others that this system would find and utilize, they would be shown as a potential threat, resulting in a false positive, which if combined with indication of online status could result in them being part of a future blocklist. Would the data in the lists was instead supplied with a trustworthy honeypot, suspicious IP addresses would be only analysed for their attributes and potential cover methods, greatly increasing the trust in results.

As one of the most important concluding remarks, that should be mentioned is, that greater accuracy to the true situation can be achieved only if various lists of

suspicious IP addresses are present in the system, as one type of attack might be more popular in one country than in the other one. It would display a biased information in graphs and this needs to be eliminated to a minimum with a variety, which gets complicated with free sources. Further blocklists that focus on manifold threats and lists focusing on single threats would contribute greatly to the overall results. Accuracy of the outputs is also increased, the more information about servers behaving as position deceptors, again, is present in the system. This would be best served through an integration with some other non-public solution.

The implemented data processor could be suitable for evaluation of future results in an ongoing manner, due to the nature of information liquidity in blocklists, etc. which can be regularly updated automatically. This system may enhance the security of websites that want to increase their security and use a joined blocklist of IP addresses that were found to be suspicious in any way, that could be present to an endpoint via API.

It is necessary to take the outcomes of the implemented tool with a pinch of salt, but its ability to provide educational level of information, in a transparent manner of visualizations in graphical way was achieved and its outcomes were utilized.

The aim of this paper, to provide an insight into the Cyber security situation or as it may appear through the publicly available data is deemed as fulfilled and the tool shall stay in use, yet possibly utilizing premium geolocation services that offer more accurate and up-to-date data, with greater number of requests, possibly allowing much greater throughput. As mentioned earlier, setting up this tool to receive data from some honeypots is also a plan. Few remarks came during the research that deserve to be mentioned here, although answers to those questions are yet to be discovered:

- Is it possible to distinguish between attackers of various experience, skills, motivation and utilized resources on attacks? Did they unleash their maximum potential from qualitative and quantitative perspective?
- Is the origin of the IP address the system shows, when no position deception is discovered, the actual origin and in no way covered with different position deception technique? Or are there any other techniques left yet to be discovered?
- Is the IP address figuring as cyber-threat origin a true threat actor or just a bot in a botnet, when no information about presence in the botnet is discovered?
- Can it be assured that the IP address of the threat agent is still under threat actor’s control?
- Was the attack successful? How much of the attacker’s intentions were achieved? What was the target of the attack? What were the consequences and the harm caused?

The actual usage of publicly available data comes with certain risks of false negatives or false positives – as it might take some time until update in acquired blocklist occurs. Furthermore, this tool does not aim to render paid services obsolete, but to provide an alternative to them. Future elaboration of performance, data accuracy and actual potency to request more IP addresses per minute, may show that access to paid geolocation services is required.

Acknowledgement

This work is a partial result of the Operational Programme Integrated Infrastructure for the projects: Research in the SANET network and possibilities of its further use and development (ITMS code: 313011W988), CEVIS Support of excellent labs research activities STU Bratislava (ITMS code: 313021BXZ1) and ACCORD Advancing University Capacity and Competence in Research, Development and Innovation (ITMS code: 313021X329), co-funded by the European Regional Development Fund (ERDF). This research was also supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic, Incentives for Research and Development, Grant No.: 2018/14427:1-26C0.

References

[1] Digital Attack Map, 2020 [online]. 2020-01-01 [visited on 2021-05-16]. Available online: https://www.digitalattackmap.com/

[2] Newman. L., 2017. How an accidental 'kill switch' slowed Friday's massive ransomware at-tack [online]. 2017-05-13 [visited on 2021-03-02]. Available online: https://www.wired.com/2017/05/accidental-kill-switch-slowed-fridays-massive-ransomware-attack/

[3] Kuzmenko, K., 2021. QakBot technical analysis [online]. 2021-09-02 [visited on 2021-12-04]. Available online: https://securelist.com/qakbot-technical- analysis/103931/

[4] Komosny, D., Vozňák, M., Rehman, S., 2017. Location Accuracy of Comercial IP Address Geolocation Databases. Information Technology and Control. Vol. 46. Available from doi: 10.5755/j01.itc.46.3.14451

[5] Hillmann, P., Stiemert, L., Rodosek, G., 2020. Dragoon: Advanced Modelling of IP Geolo-cation use of Latency Measurements. Available from arXiv: 2006.16895 [cs.NI]

[6] Grammatikakis, K., Koufos, I., Kolokotronis, N., 2021. Understanding and Mitigating Banking Trojans: From Zeus to Emotet. In: 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Available from doi: 10.1109/CSR51186.2021.9527960

[7] Wang, H., Gu, J., Wang, S., 2017. An effective intrusion detection framework based on SVM with feature augmentation. In Knowledge-Based Systems. ISSN 0950-7051, 2017, Vol. 136

[8] Talos, 2020 [online]. 2020-01-01 [visited on 2021-05-16]. Available online: https://talosintelligence.com/fullpage_maps/pulse

[9] Zima, S., 2016. Analysis of geolocation databases. MA thesis. FIT VUT Brno

[10] Rudman, L. Irwin, B., 2016. Dridex: Analysis of the traffic and automatic generation of IOCs. In: 2016 Information Security for South Africa (ISSA), Available from doi: 10.1109/ISSA.2016.7802932

[11] Bukac, V., 2010. IDS System Evasion Techniques. MA thesis, Masaryk University, Czech Republic

[12] Berger, T., 2006. Analysis of current VPN technologies. In: First International Conference on Availability, Reliability and Security (ARES'06), Available from doi: 10.1109/ ARES.2006.30

[13] Bee S., 2020. How Does Tor Really Work? [Online]. 2020-09-19 [visited on 2021-05-13] Available online: https://skerritt.blog/how-does-tor-really-work/

[14] Carlsson, A., Gustavsson, R., 2017. The art of war in the cyber world, 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 2017, doi: 10.1109/INFOCOMMST.2017.8246345

[15] Maher, M., 2003, Fall of Troy VII: New Archaeological Interpretations and Considerations, TOTEM: The UWO Journal of Anthropology, TOTEM Vol. 11 2003

[16] Blocklist.de, 2022 [online]. Export all blocked Ips, 2022-05-05 [visited on 2022-05-05]. Available online: http://www.blocklist.de/en/export.html