How to attribute blacklisted IPs to RIRs with IP WHOIS data

Posted on March 17, 2021

Analyzing IP addresses is a strategic battlefield in the fight against cybercrime. For instance, there are a number of blacklists and blocklists available, collected with various methodologies and updated dynamically to assist the implementation of IP-based threat risk mitigation measures. Such blacklists are also interesting from a research point of view as they facilitate the study of trends, structure, and dynamics of malicious IPs.

Given a suspicious IP address or netblock, the ownership information is also of paramount importance as it contributes significantly to the knowledge of the infrastructure of potential opponents. This information can be obtained from direct WHOIS lookups. However, WHOIS services normally pose limitations on the amount and frequency of available queries. Alternatively, one can use WhoisXML APIs services. These range from a simple web form for IP WHOIS lookup through a RESTFul API through the possibility to download a comprehensive IPv4 Netblocks WHOIS database along with incremental updates. These facilitate IP WHOIS database queries, highly customized ones without limitations, and also enabling to search in historic data.

In what follows we use an IP WHOIS database set up in MySQL to analyze an actual blocklist of IPs. Our focus is on studying the share of those networks which are administered by APNIC in the blacklist, in comparison to the other RIRs, and to gain an understanding of certain behaviors.

1. Data sources

Our example blocklist will be spamhaus.org's DROP list. Quoting its description:

> The Spamhaus DROP (Don't Route Or Peer) lists are advisory "drop all traffic" lists, consisting of netblocks that are "hijacked" or leased by professional spam or cyber-

> crime operations (used for dissemination of malware, trojan downloaders, botnet controllers)."

A bad enough reputation indeed; such netblocks definitely deserve attention. On 8 February 2021, we downloaded a list which contained 964 IPv4 ranges that time. It is very important to note that our results and conclusions will exactly apply to the given blacklist only and cannot be considered as a holistic picture on the overall malicious IP behavior. Yet we believe that our findings are rather typical.

To get IP WHOIS data we have opted for downloading a full netblocks WHOIS database file dated 2 February 2021, and setting up a MySQL database using the provided support script. This can be fairly efficiently done even on an average laptop and the queries run rather efficiently on a usual MySQL installation. (Most of our research could have been also conducted by using the RESTFul API to carry out the investigation.)

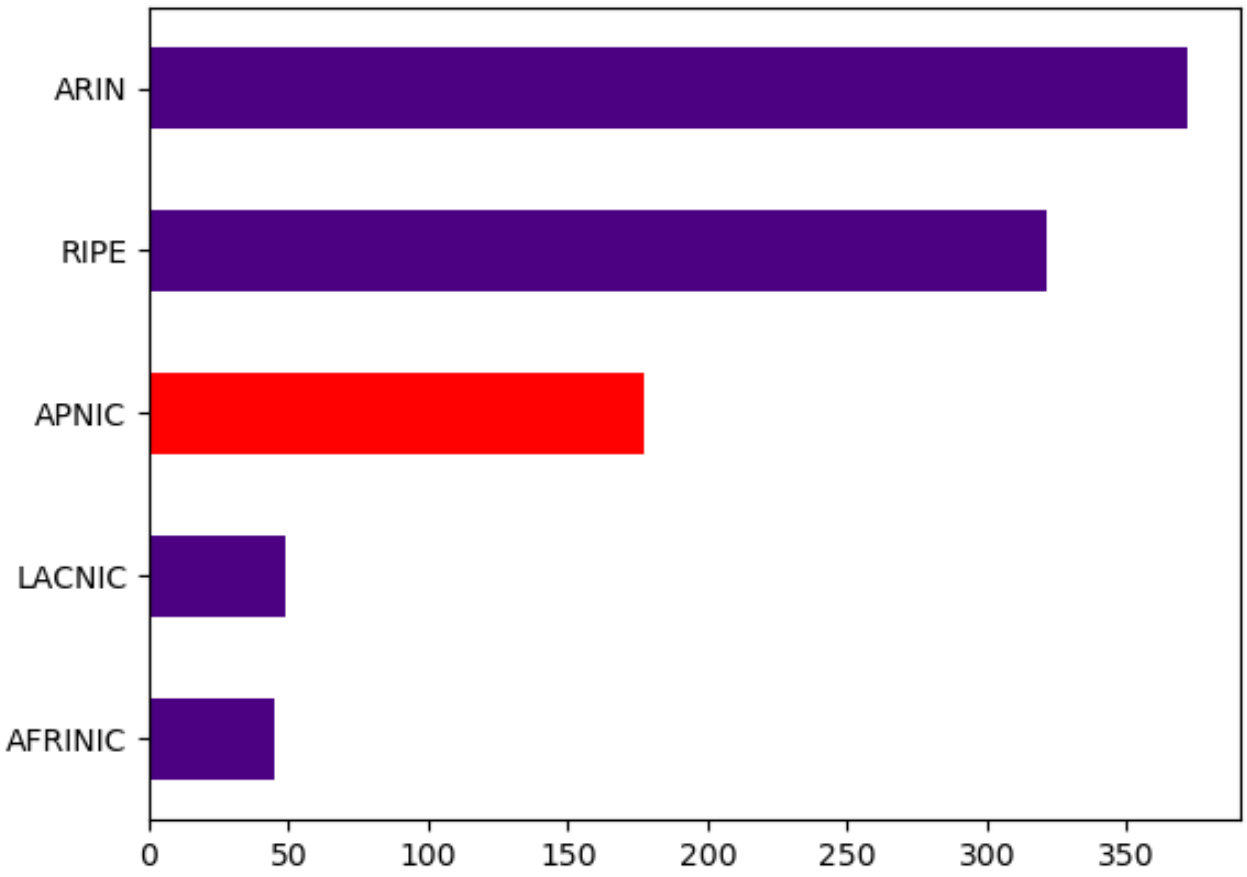
2. Research and findings

The question naturally arises: what can we find out about the netblocks using IP WHOIS data. Obviously, we cannot expect that we will get directly to the cybercriminals so easily, nevertheless we can reveal certain actors who are helping them though probably neither on purpose nor even being aware of it.

For sake of simplicity, we shall only use the first IP address in the netblock in our search. Most of the netblocks in the blacklist are contiguous and have a single IP WHOIS record. Netblocks form a hierarchical structure and thus it is reasonable to search for the smallest netblock containing a given IP as it will coincide to the block in the blacklist.

Based on this, we have collected the WHOIS records of the netblocks in our database using simple MySQL queries. On the basis of this information, we can answer a number of questions.

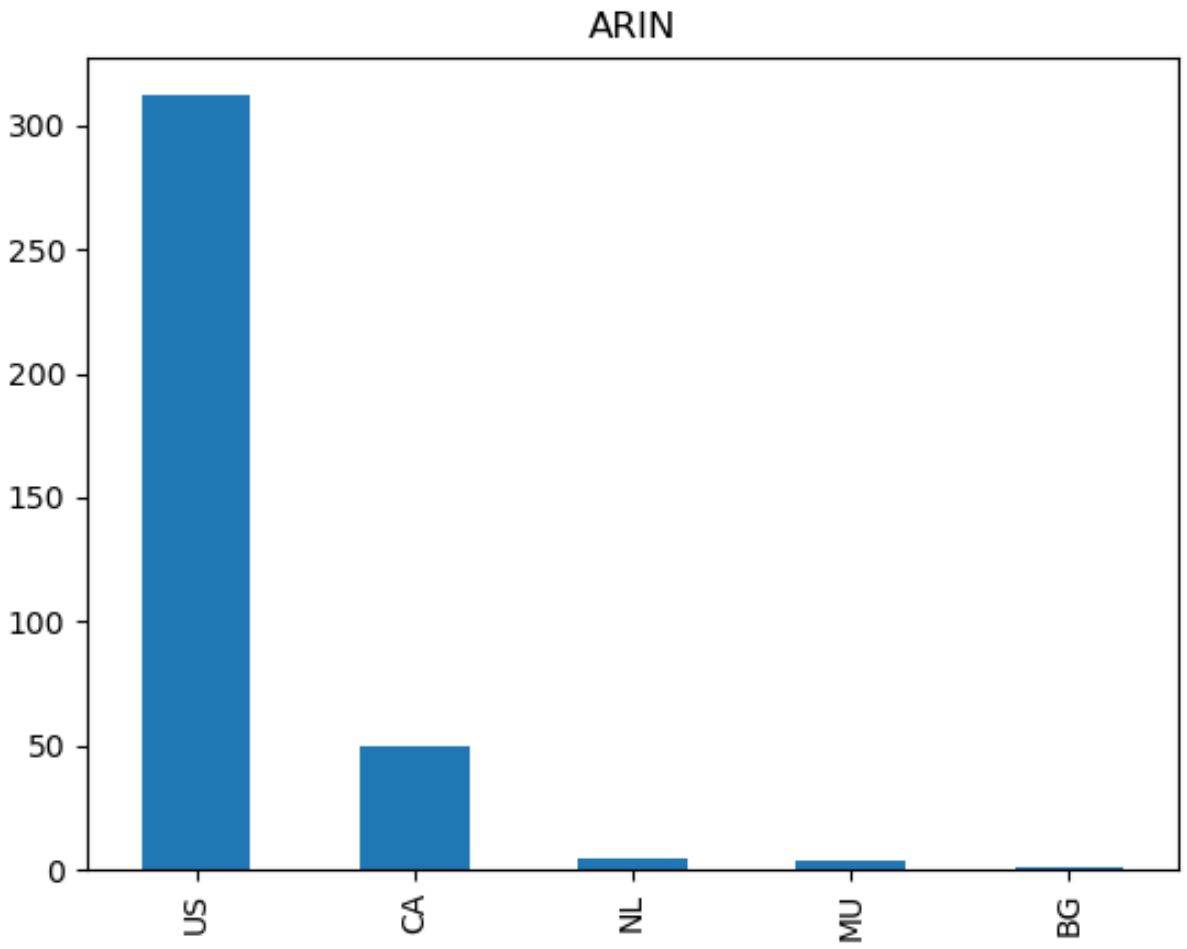
First of all, let us compare the penetration of Regional Internet Registries in the set of our 964 netblocks. The basis of this will be the source field of the WHOIS records which identifies the registry that has provided a certain WHOIS record. If a particular RIR is given here, it at least has to do something with the netblock. We have found the following distribution:

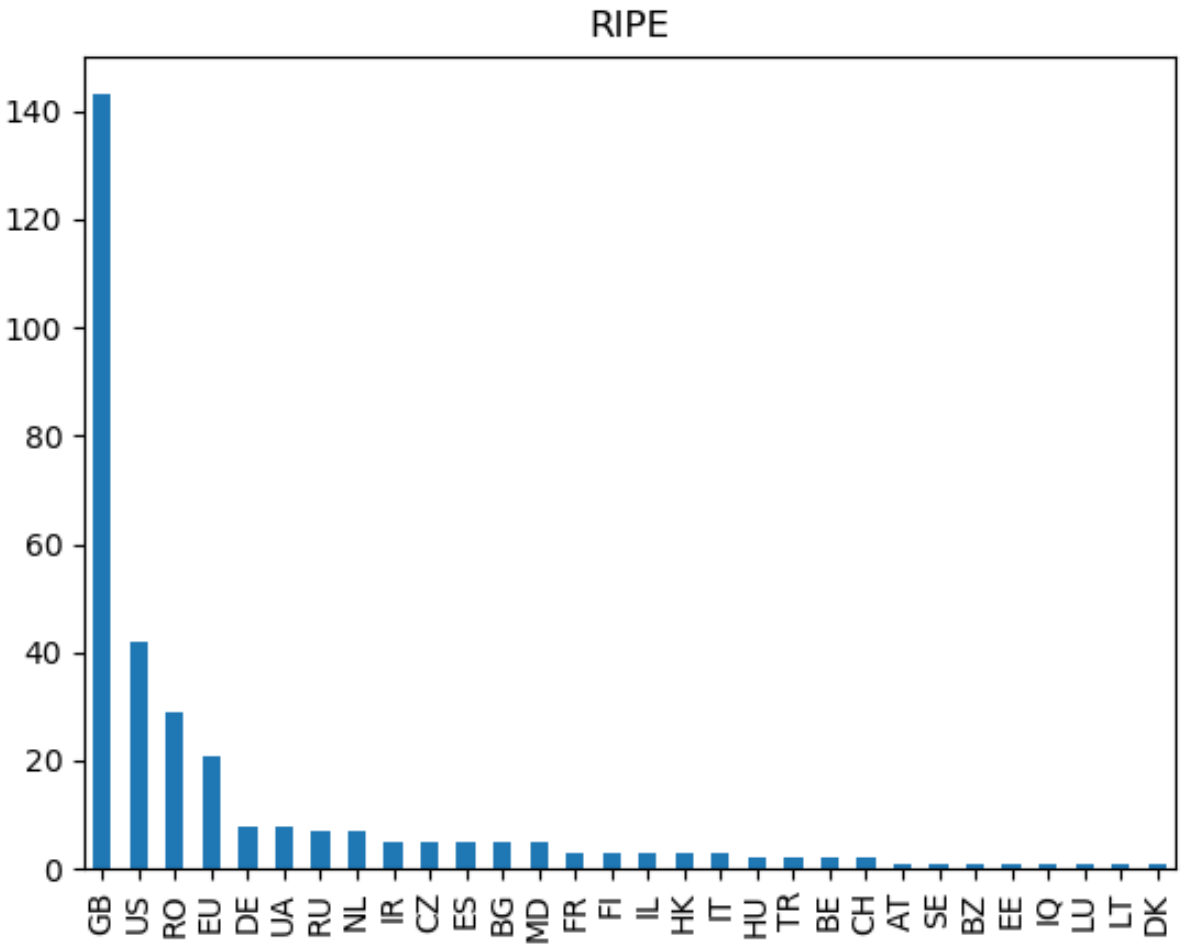


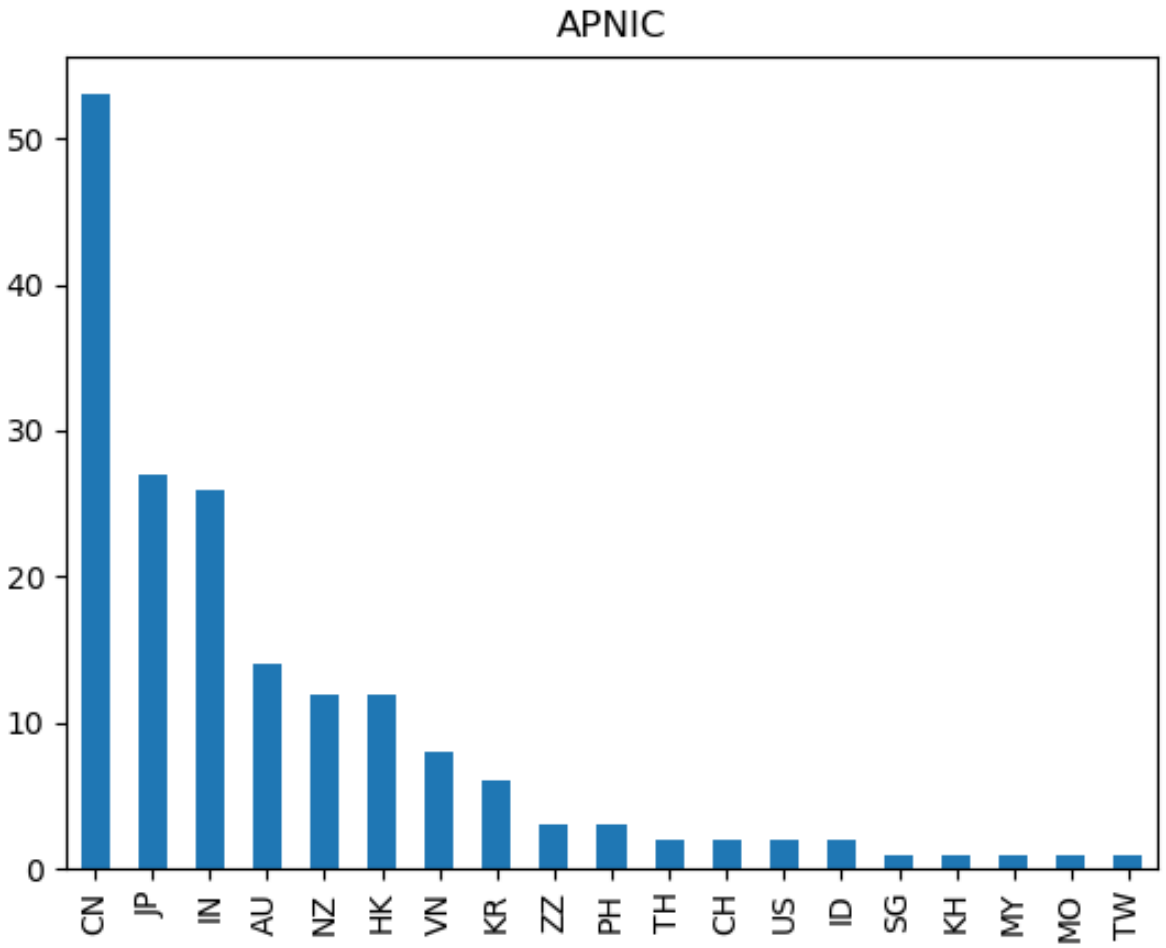
ARIN and RIPE appear to have the biggest share, while APNIC is the third in the list. A reasonable explanation is that a relevant part of malicious infrastructure is located on the cloud, and at hosting providers who have the most of their infrastructure in North America and Europe. Meanwhile there is also a relevant share of this infrastructure in the territory of APNIC. As these providers have their interest in an extensive business, their interest in verifying their clients is limited. The consequences of hosting malicious actors are well-compensated by the benefits of a large-scale business. A detailed look at the WHOIS data confirms this picture since that the Autonomous System names and NETNAMEs (wherever provided) suggest that many of the typical owners are cloud or hosting providers. There is an even bigger share of ISPs and telecom companies; their

large netblocks are ideal places to hide, too.

It is interesting to also check the "Country" field of the WHOIS records; we plot here the country distribution of the top 3 RIRs.




RIPE
140
120
100
80
60
40
20
0
GB
US
RO
EU
DE
UA
RU
NL
IR
CZ
ES
BG
MD
FR
FI
IL
HK
IT
HU
TR
BE
CH
AT
SE
BZ
EE
IQ
LU
LT
DK



The number of appearing countries correlates with the number of countries in a given RIR's territory. However, countries outside the territories also appear. The appearance of "ZZ" in APNIC's case is also rather interesting, as this stands for "Unknown or unspecified country" In case of 2 out of the 3 of these domains, the NETNAME leads to a telecom, network, and cloud service provider having their headquarters in France. In all of the cases, there are a few countries with the largest share, which is not unexpected.

So far, we have obtained an overall picture of the ties to RIRs and countries of the malicious infrastructure identified by Spamhaus's blacklist. This can already be useful in investigating

particular cases and also in preventive action of, e.g., a country's authorities. Moreover, the IP WHOIS data tend to have contact information; they seem to be less affected by the questionable consequences of new data protection regulations like the GDPR. In addition, frequently the AS names and NETNAMEs already reveal the provider where the investigation can be started. In what follows we give a particular example.

3. A particular case

A recurrent AS_NAME in the APNIC portion is "Alpha Infolab Private limited" of India. This Autonomous System has 33 netblocks altogether, all in India. The name clearly leads us to the company with the same name, which is an affiliate of a US company "Alpha Infolab". According to their website, they are a "Managed IT support and services company" with many products and services, including those related to online marketing. Moreover, they are registered or recognized brokers of ARIN, APNIC, RIPE, and LACNIC; they assist in buying and selling netblocks. Hence the appearance of their IPs on Spamhaus' blacklist must be rather unpleasant for them.

In search of the reason for their appearance, simply using Google, one runs into a page where spammers sending Finnish language spam are collected: https://suomispam.net/. On their top list, this India-based company is the 10th. Finnish language spam is something that does not probably have a very broad target audience, so encountering a specialized spam list for this in the context of APNIC is rather unexpected.

It is also instructive to take a look at the excellent and detailed list at https://suomispam.net/ in search of the spamming domain names associated with these IPs they had detected: https://suomispam.net/#!origin/AS133320 (accessed on 10 February 2021). We get the following list of domains, all in the TLD ".com":

bilahtuyip
casinojefe
corontyh
jivaniloi
maritoba

merakyahu
motohuilu
poonamruiy
seatiro
shingoti
sultaniyo
vinyavidedu

Looking at the Domain WHOIS data of these, one finds that those still in existence are typically registered at providers cheap and unverified domain registrations, and all registrant's data are masked. The typical registrant countries (still revealed) are Panama, US, and India, the latter suggesting a confirmation of the ties with the India affiliate of Alpha Infolab. It is also possible to further investigate them in the light of historic domain WHOIS data, which can be obtained using WhoisXML API's respective tools. However, in case of these domains the results are just the same.

It is not likely that a prestigious company such as Alpha Infolab was doing these illicit activities on purpose. It is more plausible that they should keep an eye on their clients to prevent damaging the reputation of their infrastructure.