IP WHOIS lookups vs. an IP Netblocks WHOIS database

Posted on February 20, 2019

In many of the aforementioned applications, it is equally important to find out who an actual IP address is assigned to and which part of the network it belongs to. Technically, it necessary and sufficient for a device to have an IP address to be able to communicate on the network. As it is sufficient, there are nodes which are not assigned a domain name. However, in every communication it is necessary for the IP address to be able to be tracked back at least. This makes IP WHOIS data useful in many of the aforementioned applications, and indeed essential for IT security. In a typical server log, for instance, we have IP addresses whose ownership can be identified via its IP WHOIS record obtainable by the WHOIS protocol.

But the ownership of a single IP address is not the only relevant question. The addresses form networks, which are sets of contiguous netblocks with the same physical entity behind. Thus, when analyzing the structure of network traffic, this structure of ownership tells more than the information on individual IPs. With IP WHOIS lookups alone, however, this structure is very hard to reveal.

In what follows, after a brief review of IP WHOIS and related concepts, we shall introduce the netblocks database by WhoisXML API.

Table of Contents

1. IP WHOIS

As an example, let's see the IP WHOIS record of one of our web servers, www.bestwhois.org, with IP 104.28.11.139 at the time of writing this blog. The data can be obtained primarily from the WHOIS servers of the respective regional Network Information Centers. These are high-level ones and there are currently 5 of them. We shall list them along with the corresponding region and WHOIS server:

- ARIN, North America: whois.arin.net
- APNIC, Asia-Pacific: whois.apnic.net
- AfriNIC, Africa: whois.afrinic.net
- RIPE NCC, Europe: whois.ripe.net
- LACNIC, Latin America/Caribbean: whois.lacnic.net

There are other services, too, which we shall mention later. Now, as we are a US-based company, let's ask whois.arin.net. In a BASH shell (Linux, Mac OS X, or Windows 10 with bash on Ubuntu on Windows), open the command prompt and type:

```
whois -h whois.arin.net 104.28.11.139
```

resulting in the following information:

#
ARIN WHOIS data and services are subject to the Terms of Use
available at: https://www.arin.net/whois_tou.html
#
If you see inaccuracies in the results, please report at
https://www.arin.net/resources/whois_reporting/index.html
#

#

NetRange: 104.16.0.0 - 104.31.255.255
CIDR: 104.16.0.0/12
NetName: CLOUDFLARENET
NetHandle: NET-104-16-0-0-1
Parent: NET104 (NET-104-0-0-0-0)
NetType: Direct Assignment
OriginAS: AS13335
Organization: Cloudflare, Inc. (CLOUD14)
RegDate: 2014-03-28
Updated: 2017-02-17
Comment: All Cloudflare abuse reporting can be done via https://www.cloudflare.com/abuse
Ref: https://rdap.arin.net/registry/ip/104.16.0.0

OrgName: Cloudflare, Inc.
OrgId: CLOUD14
Address: 101 Townsend Street
City: San Francisco
StateProv: CA
PostalCode: 94107
Country: US

RegDate: 2010-07-09
Updated: 2018-10-10
Comment: All Cloudflare abuse reporting can be done via https://www.cloudflare.com/abuse
Ref: https://rdap.arin.net/registry/entity/CLOUD14

OrgAbuseHandle: ABUSE2916-ARIN
OrgAbuseName: Abuse
OrgAbusePhone: +1-650-319-8930
OrgAbuseEmail: abuse@cloudflare.com
OrgAbuseRef: https://rdap.arin.net/registry/entity/ABUSE2916-ARIN

OrgTechHandle: ADMIN2521-ARIN
OrgTechName: Admin
OrgTechPhone: +1-650-319-8930
OrgTechEmail: rir@cloudflare.com
OrgTechRef: https://rdap.arin.net/registry/entity/ADMIN2521-ARIN

OrgNOCHandle: NOC11962-ARIN
OrgNOCName: NOC
OrgNOCPhone: +1-650-319-8930
OrgNOCEmail: noc@cloudflare.com
OrgNOCRef: https://rdap.arin.net/registry/entity/NOC11962-ARIN

RNOCHandle: NOC11962-ARIN
RNOCName: NOC
RNOCPhone: +1-650-319-8930
RNOCEmail: noc@cloudflare.com
RNOCRef: https://rdap.arin.net/registry/entity/NOC11962-ARIN

RTechHandle: ADMIN2521-ARIN
RTechName: Admin
RTechPhone: +1-650-319-8930

RTechEmail: rir@cloudflare.com
RTechRef: https://rdap.arin.net/registry/entity/ADMIN2521-ARIN

RAbuseHandle: ABUSE2916-ARIN
RAbuseName: Abuse
RAbusePhone: +1-650-319-8930
RAbuseEmail: abuse@cloudflare.com
RAbuseRef: https://rdap.arin.net/registry/entity/ABUSE2916-ARIN

#
ARIN WHOIS data and services are subject to the Terms of Use
available at: https://www.arin.net/whois_tou.html
#
If you see inaccuracies in the results, please report at
https://www.arin.net/resources/whois_reporting/index.html
#
Copyright 1997-2019, American Registry for Internet Numbers, Ltd.
#

(If you do not want to bother with the command-prompt, you can obtain the same information by entering the IP address in the box "SEARCH WhoisRWS" on the webpage of ARIN: https://whois.arin.net. We recommend this approach in a native Windows environment.)

What can you learn from this operation? First of all, we have just revealed one of our web service providers, along with some relevant contact information. But what do these two lines:

NetRange: 104.16.0.0 - 104.31.255.255
CIDR: 104.16.0.0/12

account for? This is the netblock our address belongs to as we shall explain right now.

2. IP addresses and Netblocks

The IP address is essentially a 4x8 digit number. Usually it is written in the form that each 8 bits are converted to decimals and separated with dots. Hence, for instance, our web server, 104.28.11.139 is 01101000.00011100.00001011.10001011. Omitting the dots we can convert it to decimal: 1746668427. This identifies our server amongst the 536,870,911 possible IP addresses. How do we align this with a hierarchical structure of blocks which can then be assigned to various entities?

The solution is called Classless Inter-Domain Routing (CIDR). The main idea is variable-length subnet masking (VLSM): a netblock is a set of IP addresses in which the first (i.e. most significant) n bits, the CIDR prefix bits are kept fixed. These identify the netblock. The rest then distinguishes between the IPs in the given block. Note that in this way we define hierarchy contiguous intervals.

IP netblocks can be denoted in two ways:

- The beginning and the end of the intervals, in our example of the last Section: "NetRange: 104.16.0.0 - 104.31.255.255"

- In the CIDR notation, where a representative address is given, and the number of the fixed significant digits is specified. In our example, "CIDR: 104.16.0.0/12" says from the binary representation of 104.28.11.139, the first 12 bits, 01101000.0001 are fixed, so the last 12 specify the block. Setting them all to zero, we obtain the beginning of the netblock: 01101000.0001|0000.00000000.00000000 is 104.16.0.0, whereas setting them all to 1, 01101000.0001|1111.11111111.11111111 will be 104.31.255.255, just as given in NetRange.

An even smaller block within this one can be defined by fixing additional bits. These blocks are then assigned to a hierarchy of entities, ranging from the aforementioned regional Internet registries to end-users.

So which is the parent of our example block? The record

Parent: NET104 (NET-104-0-0-0-0)

helps us find it out, we may perform

whois -h whois.arin.net 104.0.0.0

to find further details.

3. Problems with IP WHOIS lookups

So far it seems that IP WHOIS can reveal everything we may need, doesn't it? Well, not really. As we shall see, even obtaining the information can much more cumbersome by just using WHOIS, and there are tasks for which more information on netblocks would be required.

3.1 Different record structure

The first bad news is: the record structure of these servers is not standardized. So not only do you need to know the region in advance to decide which region to query but also the record will look different. As an example let us take look at the following record from whois.apnic.net:

% [whois.apnic.net]
% Whois data copyright terms http://www.apnic.net/db/dbcopyright.html

% Information related to '202.12.29.0 - 202.12.29.255'

% Abuse contact for '202.12.29.0 - 202.12.29.255' is 'noc@apnic.net'

inetnum: 202.12.29.0 - 202.12.29.255
netname: APNIC-SERVICES-AU
descr: Asia Pacific Network Information Centre
descr: Regional Internet Registry for the Asia-Pacific Region
descr: 6 Cordelia Street
descr: South Brisbane
country: AU
org: ORG-APNI1-AP
admin-c: AIC1-AP
tech-c: AIC1-AP
mnt-by: APNIC-HM
mnt-irt: IRT-APNIC-IS-AP
status: ASSIGNED PORTABLE
last-modified: 2018-06-02T00:26:15Z
source: APNIC

irt: IRT-APNIC-IS-AP
remarks: APNIC Infrastructure Services
address: South Brisbane, Australia
e-mail: noc@apnic.net
abuse-mailbox: noc@apnic.net
admin-c: AIC1-AP
tech-c: AIC1-AP
auth: # Filtered
mnt-by: MAINT-APNIC-IS-AP
last-modified: 2018-11-04T23:43:29Z
source: APNIC

organisation: ORG-APNI1-AP
org-name: Asia Pacific Network Information Centre
remarks: APNIC Infrastructure Services
country: AU
address: 6 Cordelia Street

phone: +61-7-3858-3100
fax-no: +61-7-3858-3199
e-mail: noc@apnic.net
mnt-ref: APNIC-HM
mnt-by: APNIC-HM
last-modified: 2018-06-06T05:06:58Z
source: APNIC

role: APNIC Infrastructure Contact
address: 6 Cordelia Street
address: South Brisbane
address: QLD 4101
country: AU
phone: +61 7 3858 3100
fax-no: +61 7 3858 3199
e-mail: noc@apnic.net
admin-c: HM20-AP
tech-c: NO4-AP
nic-hdl: AIC1-AP
mnt-by: MAINT-APNIC-IS-AP
last-modified: 2018-10-08T02:52:19Z
source: APNIC

% Information related to '202.12.29.0/24AS4608'

route: 202.12.29.0/24
descr: APNIC Network
country: AU
origin: AS4608
mnt-by: MAINT-APNIC-IS-AP
last-modified: 2018-11-20T03:20:12Z
source: APNIC

```
% This query was served by the APNIC Whois Service version 1.88.15-46 (WHOIS-UK3)
```

So in an application where you require information for automated processing, it needs to be parsed first.

3.2 Non-contiguous blocks

Turning to the second issue: a network of some entity can consist of multiple non-contiguous blocks. These are linked together with the attribute "ASN", the globally unique "Autonomous System Number". In our first example, RC sh OriginAS: AS1031

```
OriginAS:       AS1031
```

will tell us this identifier. (Note that e.g. in APNIC's format it is "origin" instead…) We may do a WHOIS lookup for this:

```
whois -h whois.arin.net  AS13335
```

resulting in

```
#
# ARIN WHOIS data and services are subject to the Terms of Use
# available at: https://www.arin.net/whois_tou.html
#
# If you see inaccuracies in the results, please report at
# https://www.arin.net/resources/whois_reporting/index.html
#
# Copyright 1997-2019, American Registry for Internet Numbers, Ltd.
#
```

ASNumber: 13335
ASName: CLOUDFLARENET
ASHandle: AS13335
RegDate: 2010-07-14
Updated: 2017-02-17
Comment: All Cloudflare abuse reporting can be done via https://www.cloudflare.com/abuse
Ref: https://rdap.arin.net/registry/autnum/13335

OrgName: Cloudflare, Inc.
OrgId: CLOUD14
Address: 101 Townsend Street
City: San Francisco
StateProv: CA
PostalCode: 94107
Country: US
RegDate: 2010-07-09
Updated: 2018-10-10
Comment: All Cloudflare abuse reporting can be done via https://www.cloudflare.com/abuse
Ref: https://rdap.arin.net/registry/entity/CLOUD14

OrgAbuseHandle: ABUSE2916-ARIN
OrgAbuseName: Abuse
OrgAbusePhone: +1-650-319-8930
OrgAbuseEmail: abuse@cloudflare.com
OrgAbuseRef: https://rdap.arin.net/registry/entity/ABUSE2916-ARIN

OrgNOCHandle: NOC11962-ARIN
OrgNOCName: NOC

OrgNOCPhone: +1-650-319-8930
OrgNOCEmail: noc@cloudflare.com
OrgNOCRef: https://rdap.arin.net/registry/entity/NOC11962-ARIN

OrgTechHandle: ADMIN2521-ARIN
OrgTechName: Admin
OrgTechPhone: +1-650-319-8930
OrgTechEmail: rir@cloudflare.com
OrgTechRef: https://rdap.arin.net/registry/entity/ADMIN2521-ARIN

RAbuseHandle: ABUSE2916-ARIN
RAbuseName: Abuse
RAbusePhone: +1-650-319-8930
RAbuseEmail: abuse@cloudflare.com
RAbuseRef: https://rdap.arin.net/registry/entity/ABUSE2916-ARIN

RTechHandle: ADMIN2521-ARIN
RTechName: Admin
RTechPhone: +1-650-319-8930
RTechEmail: rir@cloudflare.com
RTechRef: https://rdap.arin.net/registry/entity/ADMIN2521-ARIN

RNOCHandle: NOC11962-ARIN
RNOCName: NOC
RNOCPhone: +1-650-319-8930
RNOCEmail: noc@cloudflare.com
RNOCRef: https://rdap.arin.net/registry/entity/NOC11962-ARIN
#
ARIN WHOIS data and services are subject to the Terms of Use
available at: https://www.arin.net/whois_tou.html
#
If you see inaccuracies in the results, please report at
https://www.arin.net/resources/whois_reporting/index.html

```
#
# Copyright 1997-2019, American Registry for Internet Numbers, Ltd.
#
```

We have further contact information, but how do we find the other blocks with the same ASN?

3.3 Other obstacles

There are tasks for which you really need IP WHOIS information in bulk. WHOIS servers, however, are not designed for frequent or bulk queries. You may run into a limitation quickly, which could disable further lookups for a while.

It is not only the format of the data but also the possible structure of the queries which is different for different WHOIS servers. ARIN, for instance, has a very handy set of options (c.f. https://www.arin.net/resources/services/whois_guide.html#using), but the others have different options with different syntax. This system is not designed for efficient complex queries or data mining.

4. The solution: an IP netblocks database

More complex queries of IP WHOIS data may be needed in many cases. Notably, for IT security experts who have to track IP addresses to reveal correlated attacks against their system and identify the opponents it is essential to get information on the whole network of IP addresses. An up-to-date and complete local database, either relational (e.g. MySQL) or NoSQL facilitates the querying of IP netblocks data. In principle, any kind of information that is hidden in the structure of IP netblocks can be revealed this way. The fast availability of the data opens the possibility of implementing real-time filtering rules based on IP netblocks queries, which can be a part of firewall systems, e-mail filters and any other network security solution. When built on a reliable local database, it will be independent of the various WHOIS services and insensitive to their specialties.

4.1 WhoisXML data

WhoisXML API, Inc. offers the opportunity to download the entire IP range ownership information in house for all IP ranges. These are available both in JSON and CSV formats and are therefore readily suitable for storing and processing with a large variety of tools ranging from traditional relational databases through noSQL solutions through advanced big data analysis tools.

In addition to the data of the aforementioned five NIC IP WHOIS services, the database includes all data from the following country internet registries:

- APJII (Indonesia)
- CNNIC (China)
- IRINN (India)
- JPNIC (Japan)
- KISA (Republic of Korea)
- TWNIC (Taiwan)
- VNNIC (Viet Nam)

4.2 Downloading data and further documentation

The specifications of downloadable data can be found on the web page of WHOIS XML IP Netblocks WHOIS database product:
https://ip-netblocks-whois-database.whoisxmlapi.com

There you can find further documents on downloading and using these data with various technologies. We conclude the present blog by a short demonstration on what is doable with these

data.

4.3 A demonstration

We shall use a MySQL database built from CSV files using a script which creates the respective database (to be found on github: https://github.com/whois-api-llc/whois_database_download_support/tree/master/netblocks_csv_to_mysqldb). In particular, we shall implement the MySQL database described in the manual of this script. (NoSQL examples are available in another blog: https://ip-netblocks-whois-database.whoisxmlapi.com/blog/who-owns-the-internet-ip-netblocks-whois-data-will-tell-you).

Let us remain with the IP 104.28.11.139 we started our considerations with. First, we shall reproduce the WHOIS query:

```
mysql> SELECT inetnum,netname,as_number,as_name,contacts.name,contacts.country,contacts.city
FROM netblocks
LEFT JOIN contacts ON netblocks.org_id=contacts.id
WHERE MBRCONTAINS(ip_poly, POINT(INET_ATON('104.28.11.139'), 0));
+---------------------------+----------------------------------+----------+--------------+--------------------------------
| inetnum                   | netname                          | as_number | as_name      | name
+---------------------------+----------------------------------+----------+--------------+--------------------------------
| 104.16.0.0 - 104.31.255.255 | CLOUDFLARENET                  |    13335 | CLOUDFLARENET | Clo
| 104.0.0.0 - 104.255.255.255 | NET104                         |        0 | NULL         | American Registry fo
| 104.0.0.0 - 104.153.83.255  | NON-RIPE-NCC-MANAGED-ADDRESS-BLOCK |        0 | NULL
| 0.0.0.0 - 255.255.255.255   | IANA-BLK                       |        0 | NULL         | Internet Assigned Nu
+---------------------------+----------------------------------+----------+--------------+--------------------------------
4 rows in set (0.00 sec)
```

The query is rather self-explanatory, for the details of the database structure we refer to the script documentation. We have chosen a few fields to be more or less able to present the results, but all relevant data are available in various fields. The script creates an r-tree index, which is a trick to query more efficiently: the WHERE clause is equivalent to

```
WHERE INET_ATON('104.28.11.139') BETWEEN inetnumFirst AND inetnumLast;
```

but it is much more efficient on some subsystems.

Needless to say, all other contacts related to the records are also available in the database. Let's look for the admin contact:

```
mysql> SELECT * FROM admin_contacts LEFT JOIN contacts ON admin_contacts.id=contacts.id WH
+---------------------------+--------------+------+--------------+------+------------------+----------------+---------+--
| inetnum                   | id           | type | id           | name | email            | phone          | country | city
+---------------------------+--------------+------+--------------+------+------------------+----------------+---------+--
| 104.16.0.0 - 104.31.255.255 | NOC11962-ARIN | role | NOC11962-ARIN | NOC  | noc@cloudflare.co
+---------------------------+--------------+------+--------------+------+------------------+----------------+---------+--
1 row in set (0.10 sec)
```

So far you might well say this was all known from the WHOIS query. So apart from having our own database and being able to do all these queries also from our favorite programming environment, what else do we win? To illustrate, let us look at the other netblocks with the netname "CLOUDFLARENET", to see how trivially the question of finding non-contiguous blocks gets answered:

```
mysql> SELECT inetnum FROM netblocks WHERE netname='CLOUDFLARENET';
+-----------------------------+
| inetnum                     |
+-----------------------------+
```

```
| 104.16.0.0 - 104.31.255.255     |
| 108.162.192.0 - 108.162.255.255 |
| 162.158.0.0 - 162.159.255.255   |
| 172.64.0.0 - 172.71.255.255     |
| 173.245.48.0 - 173.245.63.255   |
| 198.41.128.0 - 198.41.255.255   |
| 199.27.128.0 - 199.27.135.255   |
+---------------------------------+
7 rows in set (0.00 sec)
```

This was something which is hardly as simple to do with direct WHOIS. It is either impossible or it relies on heavily server-specific options.