Using Whois Based Geolocation and Google Maps API for support cybercrime investigations

Asmir Butkovic*, Fahrudin Orucevic**, Anel Tanovic***
* Sector for Informatics, Police Support Agency of Bosnia and Herzegovina
Aleja Bosne Srebrene bb, Sarajevo 71000, Bosnia and Herzegovina
** Department of Computer Science and Informatics University of Sarajevo, Faculty of Electrical Engineering
Zmaja od Bosne bb, Sarajevo 71000, Bosnia and Herzegovina
*** Department of Computer Science and Informatics University of Sarajevo, Sarajevo School of Science and Technology
Zmaja od Bosne bb, Sarajevo 71000, Bosnia and Herzegovina
asmir.butkovic@psa.gov.ba, forucevic@etf.unsa.ba, atanovic@etf.unsa.ba

Abstract: - A major challenge facing all law-enforcement and intelligence-gathering organizations is accurately and efficiently analyzing the growing volumes of crime data. Cybercrime refers to any crime that involves a computer and network, where computer may or may not play an instrumental part in the commission of the crime. Detection and investigation of cybercrime can likewise be difficult because busy network traffic and frequent online transactions generate large amounts of data, only a small portion of which relates to illegal activities. In this paper, we are focusing on technologies that can help to improve the effective investigation of cybercrime, facilitate police work and enable investigators to allocate their time to other tasks. We have developed an IP mapping tool called MIPA that combines online mapping techniques and IP geolocation technology, and uses application functionality from disparate web sources. The emergence of the Web 2.0 and user-friendly online mapping techniques have created public interest in contributing information through Web-enabled geospatial tools. Moreover, IP geolocation is essential in law enforcement to identify the appropriate jurisdiction to handle enforcement of computer crime statues. Cybercrime investigators must apply a spectrum of techniques to discover associations, identify patterns, and make predictions. We have developed a tool that can be of great help law enforcement agencies during cybercrime investigations and to improve its accuracy and efficiency.

Key-Words: - Cybercrime Investigation, IP geolocation, Maps API, Mapping solutions

1 Introduction

Cybercrime refers to criminal activity that involves the use of a computer and the internet. Never before in history data has been generated at such high volumes as it is today. Sometimes we have data that you need to track to its source. There are several crimes that would require this sort of activity—for example, if the victim receives threatening e-mails or perhaps we wish to trace back the source of a security break-in. It is also true that in most child-pornography cases, it will be important to trace the images back to a specific location[8]. IP geolocation is the process of finding geographic locations of Internet Protocol addresses. Internet geolocation technology (IP geolocation) aims to determine the physical (geographic) location of Internet users and devices. It is currently proposed or in use for a wide variety of purposes, including targeted marketing, restricting digital content sales to authorized jurisdictions, and security applications such as reducing credit card fraud. An IP address is a numerical address that identifies a node on a network. Allocation of IP addresses is not arbitrary, as it is the organization which is responsible for distributing address space.

IANA (Internet Assigned Numbers Authority) is responsible for global coordination of the Internet Protocol addressing systems, as well as the Autonomous System Numbers used for routing Internet traffic. ISPs obtain allocations of IP addresses from a local Internet registry (LIR) or National Internet Registry (NIR), or from their appropriate Regional Internet Registry (RIR): AfriNIC- Africa Region, APNIC - Asia/Pacific Region, ARIN-North America Region, LACNIC-Latin America and some Caribbean Islands, RIPE NCC - Europe, the Middle East, and Central Asia. The IANA's role is to allocate IP addresses from the pools of unallocated addresses to the RIRs according to their needs as described by

global policy and to document protocol assignments made by the IETF [5].

IP address mapping based strategies relies on a controlled mapping process that uses the IP address of a target host to lookup its respective geographic location. Usually, these mechanisms uses support tools, such as Domain Name Server (DNS) traceroute and others [3]. On today's Web, mapping solutions are a natural ingredient. We use them to see the location of things, to search for the position of an address, to get driving directions, and to do numerous other things. Most information has a location, and if something has a location, it can be displayed on a map.

There are several mapping solutions including Yahoo! Maps and Bing Maps, but the most popular one is Google Maps. In fact, according to the website Programmableweb.com, it's the most popular API on the Internet. According to the site's May 2013 statistics, 26 percent of all mashups use the Google Maps API [11].

Application developers utilize Maps API as a platform and combine spatial data from multiple sources to create new customized services – a buzzword commonly called map "mashups". The use of Maps API has revolutionized online mapping applications on the Internet [10].

Section 2. of the paper describes related work of authors from this field. Section 3. of the paper makes the problem formulation. Section 4. of the paper describes the algorithm that was used for the problem solution. In the conclusion of the paper is described the significance of a new developed software tool for cybercrime investigations and its application in industry.

2 Related Work

In paper [2] authors have developed a new information system in Telecom operator by using ITIL recommendations for Service Design phase. This paper attempts to show to the telecom operators that during the realization of their IT processes they do not necessarily have to use the eTOM standard, but may use other standards and concretely they can use ITIL V3 standard.

In paper [1] authors have proposed a new ITIL V3 model that should be used during the implementation of information systems in Telecom operator. In this way Telecom operators now have a new ITSM model with a less number of processes than the standard ITIL V3 model.

In paper [4] authors questioned the reliability of several popular geolocation databases. Given that these databases are frequently used by many services and web sites in the Internet and they do not provide much information about their information sources, the quality of their geolocation information should be checked.

In paper [3] authors concluded that geolocation strategies require the use of hybrid techniques to increase their accuracy, completeness levels and the granularity of location estimates.

In paper [10] has demonstrated an online mapping application that was successfully developed using Google Maps API v3, Google Geocoding, Microsoft SQL Server Express database, and Spry Framework for Ajax. The case study presented in this article provides the advanced functionality to display the locations and state-based summary counts of USDA's thousands of peoples' gardens on the Internet with customized icons and map legend.

3 Problem Formulation

In order to efficiently investigate cybercrime, it is necessary to integrate more techniques, tools and technology for each cybercrime-case. Effective investigation requires the combination of multiple intelligence resources to establish an unify structure that improves data analysis at a low cost. It is also significant that law enforcement resources investigate cybercrime and they are limited compared to the number of cyber intrusions that they would have to address.

Our goal is to develop a tool to support this process, whose task is to combine data from different sources and present them in the form to improve effectiveness and efficiency in the cybercrime analysis and investigative tasks.

To achieve these objectives it is necessary to:

- Select the appropriate technique for geolocation and source IP address data
- Select the appropriate mapping API and source of map data
- Retrieve and extract data from selected Web services
- Combine and present information in a way that can best support the cybercrime investigation process.

During the development of this application we prefer to use databases and web services with free access. The most widely used technique for IP geolocation consists in building a database to keep the mapping between IP blocks and a geographic location.

This approach uses the Whois database to determine the location of the organization to which

ISBN: 978-960-474-308-7

an IP address was assigned. Several databases are available and are frequently used by many services and web sites in the Internet. However, there are several problems with Whois-based approaches. First, the information recorded in the Whois database may be inaccurate or stale. Also, there may be inconsistencies between multiple servers that contain records corresponding to an IP address block. Second, a large (and geographically dispersed) block of IP addresses may be allocated to a single entity and the Whois database may contain just a single entry for the entire block.

Web mapping services such as Microsoft Live Maps, Yahoo Maps, and Google Maps have provided a reliable foundation for building GeoWeb applications. These services provide a transparent and interactive user interface with preloaded maps in combination with open application programming interfaces (APIs).

Google Maps API is one of the most popular API's on the web and popular developer product.
The Google Maps API is free for commercial use, provided that the site on which it is being used is publicly accessible and does not charge for access, and is not generating more than 25 000 map accesses a day [2].

4 Problem Solution

As a development platform we used Microsoft .NET Framework 4. By using the Microsoft Visual Studio 2010 we developed windows based applications called MIPA. To build a service that maps an IP address to the corresponding geographic location we decided to use the Whois database, i.e., database-driven geolocation. Database-driven geolocation usually consists of a database engine (e.g., SQL/MySQL) containing records for a range of IP addresses, which are called blocks or prefixes [4].

There are many commercial companies that provide IP geolocation databases, but in this paper, we only considered public and freely available IP geolocation databases. The disadvantages of Whois approach discussed above is the reason why we decided to use two sources for IP address, merging two databases, primary public RIPE Database and secondarily, freely available, InfoDB database. The Réseaux IP Européens Network Coordination Centre (RIPE NCC) is one of five Regional Internet Registries (RIRs) providing Internet resource allocations, registration services and coordination activities that support the operation of the Internet globally. RIPE NCC is the Regional Internet Registry (RIR) for Europe, the Middle East and parts of Central Asia.The RIPE Database is a public database containing registration details of the IP addresses and AS numbers originally allocated to members by the RIPE NCC [12].

InfoDB is built upon the free Maxmind database version, and incremented by the IANA (Internet Assigned Numbers Authority) locality information. Typically, a geolocation database entry is composed of a pair of values, corresponding to the integer representation of the minimum and maximum address of a block. Each block is then associated with several information helpful for localization: country code, city, latitude and longitude, and Zip code.

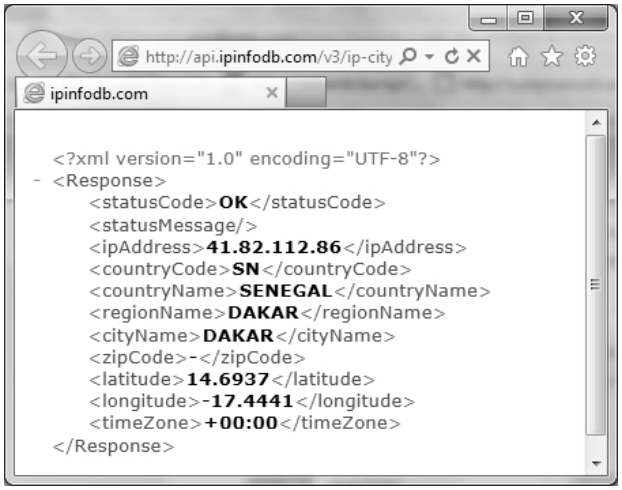

Fig. 1 Results of a Whois search on the InfoDB database

Our choice mapping API is Google Maps API and Google Maps API for .NET. Google Maps (formerly Google Local) is a web mapping service application and technology provided by Google, that powers many map-based services, including the Google Maps website, Google Ride Finder, Google Transit, and maps embedded on third-party websites via the Google Maps API.

Google Maps API is a free service that allows developers to integrate maps and geocoding from the Google Maps service (as well as other content from Google) into their websites or applications, with their own data points.

Maps API for .NET is fast and lightweight client libraries for Google Maps API, that provide all the features available in the Google Maps API. It is being developed in C# for .NET Framework 3.5.
The process searching data from databases RIPE and InfoDB using MIPA application goes through the following steps:

- Enter IP address into the Whois field and clicking on the search button

ISBN: 978-960-474-308-7

- The application query to RIPE Database (URL query string format:" query-string=IPadreess&source=ripe")
- If the result is not found in RIPE Database then the request is forwarded to InfoDB (URL query string format: key= our_api_key &ip= IPadreess &format=xml)
- If none of the above succeeds, MIPA marks the IP address as having an unknown owner.
- Download the query results using WebClient class and its method DownloadString. (WebClient class provides common methods for sending data to or receiving data from any local, intranet, or Internet resource identified by a URI.)
- Extract data from XML string and displaying data in a form.

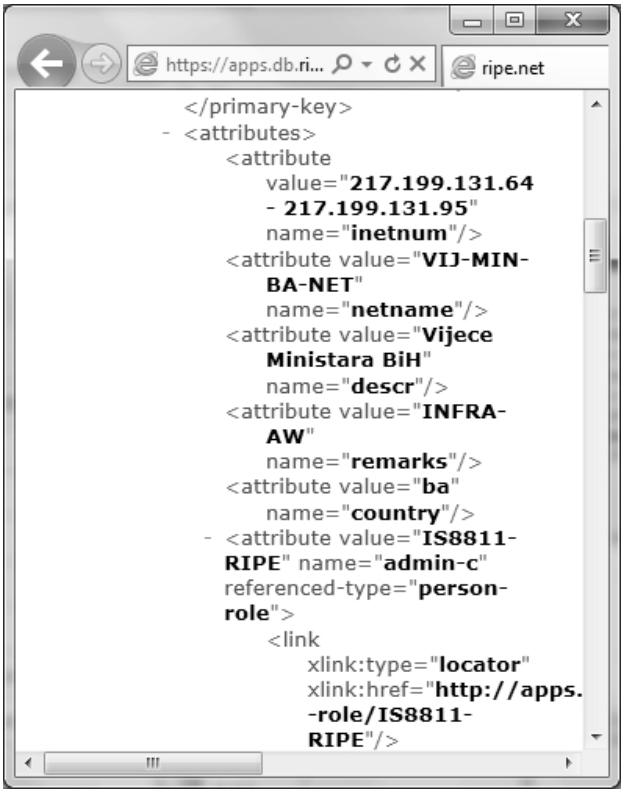

Fig. 2 Results of a Whois search on the RIPE database

Fig. 3 shows how the finished form will look after first phase. When searching for an address location, or route on a Google Map, we have to tap into the Google Geocoding API. The purpose of this API is to convert a written address into a geographic coordinate. A geographic coordinate is expressed in Latitude and Longitude.

If we successfully determine the organization that owns the IP address, in the next phase the application mapping that IP address using information about location provided in previous steps. As mentioned, the input parameter for the mapping is address, and also in application is possible to set some parameters: size of map, map type (Roadmap, Satelite, Terrain, Hybrid), Zoom and other.

MIPA application also provides information about the coordinates of a found location. The coordinates are expressed using latitude and longitude. Latitude measures from south to north, and longitude measures from west to east. At the equator, the latitude is 0. The coordinates are expressed using decimal numbers separated with a comma. The latitude always precedes the longitude value (latitude, longitude). The position for Sarajevo, for example, is 43.85, 18.38.

Fig. 4 shows distance between The Ministry of the Interior of the Republic of Croatia and Police Support Agency of Bosnia and Herzegovina, calculated using their IP addresses.

Table 1. presents a very interesting relation between the blocks of IP addresses and their coordinates in several commonly used geolocation databases.

Database	Blocks	(lat; long)	Countries	Cities
HostIP	8,892,291	33,680	238	23,700
IP2Location	6,709,973	17,183	240	13,690
InfoDB	3,539,029	169,209	237	98,143
Maxmind	3,562,204	203,255	244	175,035
Software77	99,134	227	225	0

Table 1: General characteristics of the some geolocation databases

There is a problem with the coordinates because IP address allocation policies result in blocks that often have common administration, often including physical location. From Table 1, we can notice the strong difference between the number of IP blocks and the number of unique (latitude,longitude) pairs. For example InfoDB contains roughly 3,5 millions of IP blocks, those blocks only refer to 169,000 (latitude, longitude) pairs [4].

5 Conclusion

Cybercrime refers to criminal offenses committed using the Internet or another computer network as a component of the crime. Cyber attacks aren't just an information technology matter. Cyber attacks have extraordinary potential to impact on the economic and social well-being of individuals and businesses.

Even non-computer crimes might still involve computer and Internet resources. For example, a drug dealer might use e-mail to arrange sales and

ISBN: 978-960-474-308-7

purchases, a prostitute and pimp might use Craigslist to facilitate their trade, and a social-networking site can provide clues as to the motive in a violent crime.

Combating cybercrime is especially challenging due to problems of jurisdiction that arise at both the national and international level.

The key to combating cybercrime is to find a way to achieve high effectiveness and efficiency in cybercrime investigations.

In this paper, we presented how mapping techniques and IP geolocation technology can help this process.

In spite of the aforementioned issues, Whois Based Geolocation still may provide useful information.

Our chosen approach geolocation could be improved.

The basic idea is to use the Internet Control Message Protocol (ICMP) to find out the location of next hop(s) routers in the path to a given target IP address. (if I do not know where you live, then at least I know where your neighbors do)

Meaningful visualizations can significantly improve decisionmaking quality and help investigators in taking rapid response.

Some researchers have proposed the use of information visualization techniques to enable efficient discovery of trends and patterns among cybercrime data. These techniques exploit the temporal, textual content similarity, and structural relationships extracted from cybercrime complaint records. Trends presented by data aggregates may provide investigative leads.

Our tool can be of great help law enforcement agencies during cybercrime investigations and to improve its accuracy and efficiency. Using this tool can help the transfer of conventional policing models into an often abstract and technical environment. Governments, law enforcements and corporate security teams worldwide use geolocation as an investigatory tool, tracking the Internet routes of online attackers to find the perpetrators and prevent future attacks from the same location.

We plan to extend this work improvement geolocation strategy using ICMP, using more features available in the Google Maps API and connect our program with a database of suspicious IP addresses.

References:

[1] A. Tanovic, I. Androulidakis, and F. Orucevic, “Development of a new improved model of the ITIL V3 framework for the information system of Telecom operator”, 11th WSEAS International Conference on Data Networks, Communications, Computers (DNCOCO ‘12), pp. 209-215, September 2012.

[2] A. Tanovic and F. Orucevic, “Implementation of the Information System of the Telecom Operators Using the ITIL V3 Methodology for the Service Design Phase”, 2nd International Conferences on Advanced Service Computing, pp. 82-91, November 2010.

[3] Patricia Takako Endo and Djamel Fawzi Hadj Sadok, “Whois Based Geolocation: a strategy to geolocate Internet Hosts”, 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), April 2010.

[4] Ingmar Poese, Mohamed Ali Kaafar, Benoit Donnet and Bamba Gueye, “IP geolocation databases: unreliable?”, ACM SIGCOMM Computer Communication Review, Volume 41 Issue 2, April 2011 Pages 53-56.

[5] IANA (Internet Assigned Numbers Authority) URL http://www.iana.org/.

[6] Jeffrey Carr,"Inside cyber warfare: Mapping the cyber underworld", book, December 2011.

[7] Debra Littlejohn Shinder and Ed Tittel, “Scene of the Cybercrime: Computer Forensics Handbook”, book, August 2002.

[8] Chuck Easttom and Jeff Taylor, "Computer Crime, Investigation, and the Law", book, April 2010.

[9] RADU LIXĂNDROIU, "Customizing Web Advertisements Based on Internet Users’ Location", 11th WSEAS International Conference on Mathematics and Computers in Business and Economics (MCBE '10), June 2010 Pages 273-278.

[10] Shunfu Hu and Ting Dai, “Online Map Application Development Using Google Maps API, SQL Database and ASP.NET”, International Journal of Information Communication Technology Research , February 2013.

[11] Gabriel Svennerberg, "Beginning Google Maps API 3", July 2010.
[12] The Réseaux IP Européens Network Coordination Centre (RIPE NCC) URL https://www.ripe.net/.

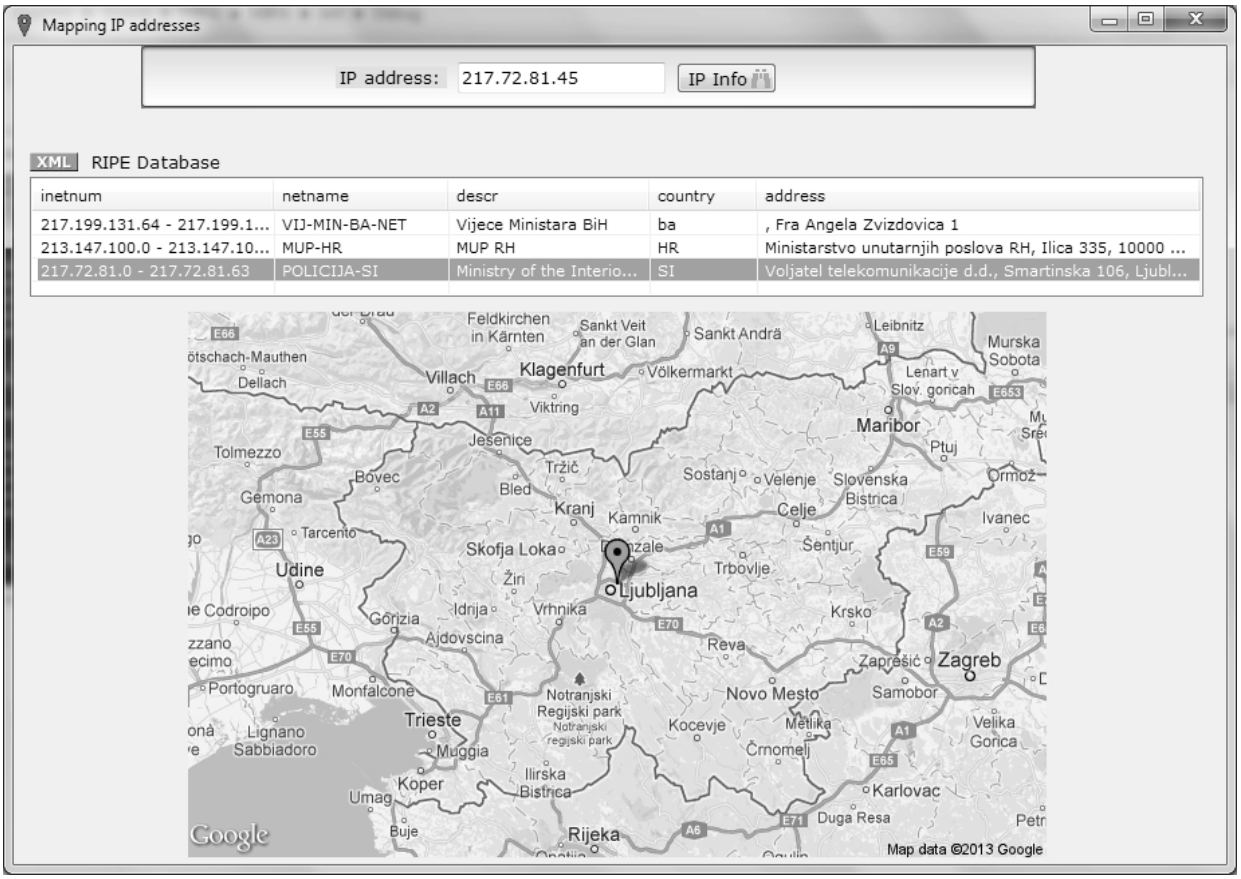

Fig. 3 Shows how the finished form will look

ISBN: 978-960-474-308-7

Fig. 4 Example of calculating distance between two coordinates (two IP addresses of locations)

ISBN: 978-960-474-308-7