DIGITAL INVESTIGATION 18 (2018) XXXXXXX

Available online at www.sciencedirect.com

Digital Investigation

Journal homepage: www.elsevier.com/locate/diin

Conference Title

Geographic Profiling for serial cybercrime investigation

Asmir Butkovic[a]*, Sasa Mrdovic[b], Suleyman Uludag[c], Anel Tanovic[b]

[a]*Police Support Agency of Bosnia and Herzegovina, Aleja Bosne Srebrene bb, 71000 Sarajevo, Bosnia and Herzegovina*
[b]*University of Sarajevo/Faculty of Electrical Engineering, Zmaja od Bosne bb, 71000 Sarajevo, Bosnia and Herzegovina*
[c]*University of Michigan - Flint, 500 S. State Street, Ann Arbor, MI 48109 USA*

Keywords:
Mapping crime
Geographic profiling
Cybercrime
Cyber profiling
GeoCrime

ABSTRACT

Today's cybercrimes are much more difficult to detect and prosecute than traditional crimes. In the investigation of cybercrimes, law enforcement agencies follow similar techniques to traditional crimes that, however, have to be modified to meet the unique conditions and requirements of virtual space. This paper examines cybercrime profiling techniques prevalent today, and focuses on the feasibility of applying geographic profiling technique to cyber offenders. The primary assumption of the research is that for most types of cybercrime, the steps during the procedure of committing criminal act are not random. For example, the choice of the victim, the choice of crime location, similar characteristics, follow a certain logic, which could provide information about the offender's crime. Testing the utility of a geographical profiling has been carried out on real cybercrime samples obtained by law enforcement agencies. This paper aims to apply the concept of geographic profiling to the issue of cybercrime that involve a physical world, targeting two types of cybercrimes: credit card skimming and spear phishing. Specially developed GeoCrime geographic profiling software designed to assist in the mapping, spatial and statistical analysis of cybercrime patterns was used. The results of the study have shown the possibility of applying geographic profiling to certain types of cybercrimes and under the certain conditions. The importance of geographic profiling is also emphasized, especially in situations where little is known about the offender, such as in cybercrime, where offenders use the Internet to hide their identities and activities.

1. Introduction

Geographic representation and visualization of crime scenes have become very important in the collection of intelligence on crime. It provides a very dynamic and easy way to track crime patterns and analyze them. The analysis of trends within cybercrime have shown a consistent increase in the number and type of crimes, thanks primarily to the rising use of mobile devices and the increasing use of the Internet on such devices [1]. The problem of investigation in the cyberspace is that, it creates unique situations in which conventional investigative practice cannot be fully applied. First of all, the geographical distance between the victim and the attacker is greater (actual attackers can be on the other side of the planet when they commit a criminal offense). Second, the Internet contains a significant amount of personal information about individuals, gives offenders greater access to victims, allows offenders to hide their identity and enable a predator to search for particular types of potential victims. Due to this, the existing traditional investigative techniques have to adapt to new conditions, or use new research approaches. Criminal profiling is an investigative approach based on the assumption that the crime scene provides details about the offense and the offender [2]. The aim of criminal profiling is to prioritize suspects and provide investigators with important case information by identifying the offender's characteristics. Criminal profiling can be broadly divided into two categories: geographic profiling

* *Corresponding author.*
E-mail address: aButkovic@hotmail.com
Peer review under responsibility of xxxxxxxxx-xxxx/$ – see front matter © 2014 xxxxxxxx. Hosting by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aebj.2014.10.012

and profiling personal characteristics of the offender[3]. Geographic profiling is a criminal investigative methodology based on the principles, theories and concepts of environmental criminology for analyzing the locations of a related series of crimes to determine the most probable area of offender residence [4]. Cybercrime profiling is carried out with the same primary goal as for other criminal offenses and it is to make easier to find the offender of a particular criminal offense. Geographic profiling (GP) was originally developed for the investigation of a serial murder, although it was subsequently applied to numerous other serial crime types, such as rape, robbery, arson, burglary, kidnappings and auto theft. The present work explores the utility of proposed geographic profiling model where calibration of the model and testing the accuracy of the forecast of the geographic profile was performed using solved serial cybercrime cases, that included an interaction with a physical world. Before we explain how this is done, we first give a brief description background of the study and research problem.

2. Background

The concept of criminal profiling existed long before the emergence of cybercrime and cyber criminals. The Federal Bureau of Investigation’s (FBI) defines the profiling process as "an investigative technique by which to identify the major personality and behavioral characteristics of the offender based upon analysis of the crime(s) he or she has committed"[5]. However, the basic concepts of such profiling are not significantly different from that of the profiling of cybercrime and cyber criminals. Modern criminal profiling takes two forms: the deductive approach that involves analysis of the evidence found from the case and an inductive approach that uses the statistical analysis of the previous offender’s characteristics in order to construct the offender’s behavioral profile or generate a generalized behavioral pattern of the offender. Geographic profiling is an inductive profiling method that attempts to provide information on the likely “base of operations” (e.g., home address, place of employment, an acquaintance’s residence) of offenders thought to be committing serial crimes. In the mid-nineties of the last century, complex models of prediction of the home address of the offender were developed based on the work of the Brantinghams [6] and other studies of offender travel behavior. Canadian criminologist Dr. Kim Rossmo has particularly improved and popularized this investigation method and developed the Criminal Geographic Targeting (CGT) model, which was later patented and incorporated into the Rigel Geographic Profile Software. Criminal geographic targeting utilizes set of mathematical functions, including a distance decay function and buffer zone to determine the most probable area of offender residence through the production of a jeopardy surface or geographical profile. The distance decay function is an appropriate representation of the strength of spatial relationships between travel behavior of offender and offender travel costs and explains why most crimes occur relatively close to the offender’s home. The buffer zone is an area around the offender's anchor point in which crimes are less likely due to reduced anonymity within the criminal's neighborhood.

Rossmo also defined five minimum requirements for successful geographical profiling [7]:

(1) the offender has committed at least five criminal offenses,
(2) the crimes are related to the same offender and the series is relatively complete,
(3) the offender committing the crimes did not change the area of criminal activity,
(4) the offender has not moved anchor points during his crime series, and
(5) the distribution of suitable targets is relatively uniform around the offender’s residence.

In an effort to simplify the geographic profiling process, geographic profilers have developed various computer programs designed to help the process of calculating data on the crime scene. Next section looks at the related work and similar initiatives in the areas of cyber profiling, geographic profiling and analyzing criminal behavior and victimization in cyberspace.

3. Related work

Interest in geographic profiling has increased with the advancement of mapping software and the increased use of Geographic Information System (GIS) by researchers and practitioners in law enforcement agencies. "A Methodology for Evaluating Geographic Profiling Software" states that geographic profiling is an important step in moving computerized crime mapping beyond the static display of the location of crime sites toward to analytical mapping that help analysts interpret spatial data[8]. However, Derek (2007) points out that most research on geographic profiling in the past 15 years has focused on the development of new software platforms and analysis of case studies, rather than on a critical assessment of the issues in efficiency and accuracy [9]. Unfortunately, a similar approach to the study of the geographic profiling has continued until today, so that this lack of research has created a void in which there are still a lot of unresolved issues relating to geographic profiling, among which is the question of the feasibility of applying the technique of geographic profiling to cyber offenders. When it comes to the application of criminal profiling techniques within a cyber context, most of the efforts so far have been focused on the psychological domain, as opposed to applying the criminological continuum [10]. The majority of the research in this area has focused on the use and application of cyber profiling for the classification and characterization of people behind cyber attacks, as well as in general for analyzing the daily behavior of online users [11]. Particularly popular are the methods of using cyber profiling to predict habits and characteristics of Internet users by interpreting their online behavior as evidence, such as, behavioral evidence analysis of Facebook users [12], or the use of data mining with K-Means technique for analyzing the logs of the activities of Internet users [13], [14]. Another interesting study [15] uses machine learning to profile the cyber attacker based solely on events that are monitored automatically during the actual attack and without any prior information. An automated profiling tool attempts to classify an attacker as a type of human user, claiming that if it does not fit in the any human user's profile, then it is probably a bot.

When talking about a vision for the future research and conceptual development in this research area, Vandeviver and Bernasco [16] in their article state that the particular development of mobile information and communication technologies will transform the geography of crime for three reasons:

(1) offenders use new technologies when committing crime,
(2) law enforcement agencies rely on new technologies to prevent and investigate crime, and
(3) researchers use new technologies to study crime.

What is quite evident is that, there are many papers and research studies that examine the applicability of geographic profiling in the analysis of spatial patterns of behavior of individuals and groups that are not tied to cyber space.

For example, the use of geographical profiling within at the context of terrorist attacks [17] , for understanding the geospatial patterns of terrorist cell behavior [18], for vehicle theft [19] and sex offenders investigations [20], and so on.

Furthermore, geographic profiling, the statistical technique originally developed in criminology has recently been applied to biological data such as the investigation of the point of origin of a biological invasion [21], to predict multiple nest locations of bumble bees [22] and hunting patterns of white shark [23].

Some authors, such as Lucy Mburu & Marco Helbich [24] argue that there has been too large a focus of research on geographic profiling data for developed cities in Western countries, and the lack of analysis and estimates of geographic profiling accuracy for less developed countries with dissimilar socioeconomic and demographic landscapes.

There is a significant academic debate about the most appropriate method for conducting geographic profiling, with some researchers advocating for complex computer algorithms and others for simple statistical and geometric methods. However, despite this disagreement, the basic assumptions of the behavior on which the geographic profiling process is based are simple and indisputable. This study attempts to solve the existing lack of geographic profiling research especially in assessing the efficiency and accuracy of geographic profiling, the comparative assessment of Criminal Geographic Targeting and centrographic technique in this domain, and identifying areas of potential improvement of process profiling. The next section begins by describing some of the mathematical foundations of the geographic profiling problem. We then present our mathematical approach for the geographic profiling problem.

To our knowledge, ours is the first study in the literature that applies the geographical profiling to the cybercrime and presents a preliminary test of the potential for geographical profiling with a sample of the solved cybercrime cases.

4. Mathematical framework

There are several mathematical techniques that underlie existing geographic profiling strategies, classified into two general categories: spatial distribution strategies and probability distance strategies [25]. The most common spatial distribution strategies estimate the anchor point by the center of minimum distance or by the centroid of the crime series. On the other hand, probability distance strategies are currently employed in the major computer programs for geographic profiling and differ almost only in the choice of distance metric and the decay function. Distance metric or distance function is a function that defines how to measure the distance between points. Distance decay function makes mathematical connection between the distance from an offender's residence to a potential target location and the probability that the offender chooses that location to commit crime. For example, Levine within his crime mapping software, CrimeStat provides the option of choosing one of the following distance decay functions[26]:

- Linear $D(r) = \begin{cases} A + Br \; if \; A + Br \geq 0 \\ 0 \quad\; if \; A + Br < 0 \end{cases}$
- Negative exponential $D(r) = Ae^{-Br}$
- Normal $D(r) = \frac{A}{S\sqrt{2\pi}} \exp\left(-\frac{(r-\overline{r})^2}{2S^2}\right)$
- Lognormal $D(r) = \frac{A}{r^2 S\sqrt{2\pi}} \exp\left(-\frac{\left(ln(r^2)-\overline{r}\right)^2}{2S^2}\right)$
- Truncated negative exponential $D(r) = \begin{cases} Br \quad\; if \; r \leq r_p \\ Ae^{-Cr} \; if \; r > r_p \end{cases}$

In geographic profiling concept distance decay and buffer zone are widely used to predict a potential location for a criminal. According to the conclusions of most researchers in the area, criminals prefer to choose places for crimes not very close to the residence (region around the buffer zone), but no more far if there is no necessary (region where the value of the function no longer changes very rapidly). (see Figure 1)

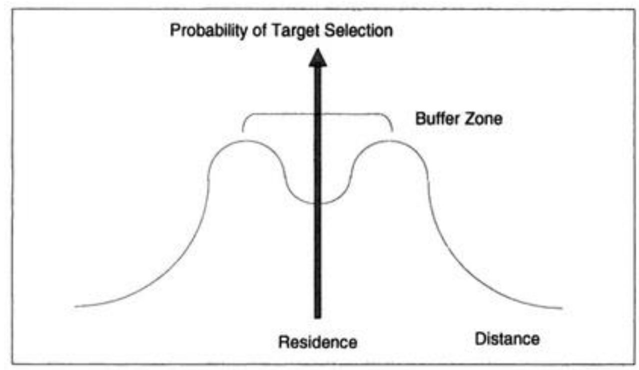

Fig. 1. - Crime distance decay function with buffer zone. [27]

Our model uses the probability distance strategy and the spatial distribution strategy, implemented as algorithms in our computer program for geographic profiling: Criminal Geographic Targeting and centrography.

4.1. Criminal Geographic Targeting (CGT)

Criminal Geographic Targeting (CGT) [28] algorithm describes the mathematical relationship between offender travel and probability of offending and determines the most probable area of offender residence through the production of a jeopardy surface or geographical profile. CGT algorithm analyzes crime site coordinates within the hunting area and produces an offender residence probability surface from the point pattern of the crime locations. The process involves several steps, starting from calculating map boundaries that determines the offender's hunting area, to calculate the function $\boldsymbol{p_{ij}}$ for every point on the map as follows:

$$p_{ij} = k \sum_{n=1}^{C} \left[\frac{\phi}{(|x_i - x_n| + |y_i - y_n|)^f} + \frac{(1-\phi)B^{g-f}}{(2B - |x_i - x_n| - |y_i - y_n|)^g} \right] \qquad (1)$$

where:

- ϕ is equal to 1 if $|x_i - x_n| + |y_i - y_n| > B$, 0 otherwise.
- k is an empirically determined constant;
- B is the buffer zone radius (given as number of grid square units);
- C is the number of crime sites;
- f and g are empirically determined exponents
- (x_i,y_j) are the coordinates of point (i,j) and (x_n,y_n) are the coordinates of the nth site.

The hunting area is defined as the rectangular zone, commonly covered with 40,000-pixel grid (*in our case covered by a grid with 200 by 200 pixels*) and containing all crime locations.

Thus, $\boldsymbol{p_{ij}}$ describes the probability that the location of the offender's base (anchor point) occurs at point (i,j), given the locations of the crime sites.

When the probability for every point on the map is calculated, it produces a probability surface whose three-dimensional representation is referred to as a jeopardy surface, and a two-dimensional is termed a geoprofile.

The jeopardy surface looks like a volcano with the caldera representing the region within the buffer zone and the anchor point in the center of a volcano as shown in Fig. 1.

4.2. Centrography

In our model we also use the technique known as centrography [28] that we applied to a spatial data set of the crime scene locations. Centrographic statistics is a descriptive statistic and the most common method when wanting to analyze spatial distribution of the crime incidents by using standard deviation, spatial mean, median center, or standard distance deviation. The spatial mean (sometimes referred to as the centroid or mean centre) is a method that consists of finding the mean x and y coordinates of the crimes and associating them with the criminal's calculated residence. It is defined as:

$$(\bar{x}, \bar{y})$$

Where:

$$\bar{x} = \left(\sum_{n=1}^{C} x_n\right) \Big/ C$$
$$\bar{y} = \left(\sum_{n=1}^{C} y_n\right) \Big/ C \qquad (2)$$

Similarly, the standard distance deviation defined as:

$$SDD = \sqrt{\left(\sum_{n=1}^{C} ((x_n - \bar{x})^2 + (y_n - \bar{y})^2)\right) \Big/ C - 2} \qquad (3)$$

The standard distance deviation (SDD), or standard distance, is the standard deviation of the distance of each crime location from the mean centre and it is calculated by taking the square root of the summation of the variances of the latitudes and longitudes from the arithmetic mean. This measure is used to determine the radius of the distribution, or to define the degree of the dispersion of the incidents in the series. The standard distance deviation overcomes problems associated with the standard deviation. First of all, it provides a single summary statistic of the dispersion in the locations. Secondly, it is expressed in measurement units. A detailed explanation of our mathematical approach can be found in the next subsection.

4.3. Our approach

In order to experiment with these two geographic profiling techniques to predict an anchor point, we have developed our own application called GeoCrime, designed with a simple user interface and simplified workflow to assist in the process of calculating crime site information. We examine the relevance of geographic profiling to patterns of cyber crimes committed in Croatia and Bosnia and Herzegovina, based on data compiled from a range of different sources. Unfortunately, the test results obtained by examining solved cybercrime cases through the use of this centrographic technique showed a significant deviation of assumed anchor point from the real offender's residence. The use of the CGT algorithm for geographic profiling of cyber crime in our research has shown much better results than centrographic analysis of the spatial distribution, so we decided to build our own approach and mathematical framework for geographic profiling based on this algorithm.

Through the model optimisation procedure, we obtained the values of the constants f and g, and the value for the radius of the buffer zone, B.

The first step in our model testing procedure began with using the error distance and search costs as an accuracy measure to the validity of geographical profile. Error distance is defined as the distance from the offender's actual to predicted home location. Search cost [29], or hit score percentage [28] is the percentage of the total hunting area searched before the offender's residence is found. CGT hit score as accuracy measure is widely used, although there is a problem with the ability of the artificially increased by the selection of an unrealistically large search area. Therefore, we decided to experiment with a new accuracy measure defined as:

$$GC\ Hit\ Score\ = \sum_{i=APx-B}^{APx+B} \sum_{j=APy-B}^{APy+B} p_{ij} \Big/ \sum_{i=1}^{N} \sum_{j=1}^{M} p_{ij} \qquad (4)$$

Where:

- $\boldsymbol{p_{ij}}$ is probability value for (i,j) pixel on the grid
- $\boldsymbol{N}$ and $\boldsymbol{M}$ are the pixel dimensions of the grid of hunting area
- $\boldsymbol{APx}$ and $\boldsymbol{APy}$ are the grid coordinates of the anchor point
- $\boldsymbol{B}$ is the buffer zone radius (given as number of grid square units)

The model performance is better if the CGT hit score percentage value is smaller and GC Hit Score percentage value is bigger. However, in the case of geographic profiling of unsolved crime offences both of these factors do not have a clear criterion when to stop searching and enlarging the proposed search area. Thus, we decided to establish a new method for estimating the efficiency of the search process by defining the index called the Search Process Efficiency Index (PSEI) that represents ratio of CGT hit score and GC Hit Score. This index proved to be a good indicator whether the search area cover the offender's anchor point and helping make decision to stop searching through the cells. The Minimal Search Process Efficiency Index (MPSEI) is calculated in this manner:

$$MPSEI = \text{Min}\left(\frac{CGT\ Hit\ Score}{GC\ Hit\ Score}\right) \qquad (5)$$

In order to verify correctness of our approach we have developed our geographic profiling software GeoCrime that is presented in the next section.

5. Geographic profiling software

The geographic profiling process involves generation of the geographical profile from locations of a connected series of crimes and prioritizing areas around the offender residence. It uses a variety of specialized crime-mapping software designed to assist in the process of calculating crime site information and visualization of the profile. There are several commercial and academic software tools that support different geographic profiling strategies. Four major current systems are: CrimeStat, Dragnet, Predator and Rigel.

In order to study the effectiveness of different profiling strategies we have developed our own application called GeoCrime. The GeoCrime is a Windows-based program for generation of the profile from incident documentation. It combines automatic and manual techniques for the analysis of spatial and probability distribution of crime sites. Use of the program involves inputting data on the crime sites and geocoding and mapping crime locations from address data.

They are then converted to latitude-longitude coordinates using the Google Maps API[1] integrated into the program. Alternatively, user can specify simple text file that contains a list of latitude-longitude coordinates. Based on these data, the program defines the boundaries of the hunting area map and creates a grid over it.

The hunting area is defined as in[28]:

$$\begin{aligned} y_{high} &= y_{max} + \frac{(y_{max}-y_{min})}{2(C-1)} \\ y_{low} &= y_{min} - \frac{(y_{max}-y_{min})}{2(C-1)} \\ x_{high} &= x_{max} + \frac{(x_{max}-x_{min})}{2(C-1)} \\ x_{low} &= x_{min} - \frac{(x_{max}-x_{min})}{2(C-1)} \end{aligned} \qquad (6)$$

where:

- y_{high} is the value of the upper-latitude boundary;
- y_{low} is the value of the lower-latitude boundary;
- y_{max} is the maximum latitude value for any crime site;
- y_{min} is the minimum latitude value for any crime site;
- x_{high} is the value of the right-longitude boundary;
- x_{low} is the value of the left-longitude boundary;
- x_{max} is the maximum longitude value for any crime site;
- x_{min} is the minimum longitude value for any crime site;
- C is the number of crime sites.

Fig. 2 shows a map of the hunting area with defined boundaries and mapped locations of crime sites.

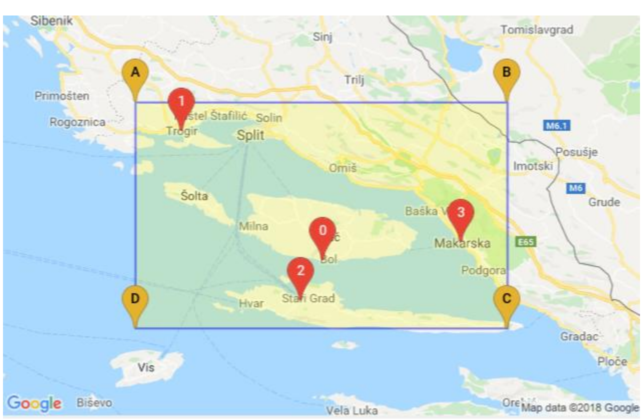

Fig. 2. - Geocoding and mapping of crime sites.

Case OTP - ATM card skimming in the case of OTP bank, 2012 year[2]

The hunting area is automatically defined as a 200x200 rectangular grid of cells, where the Euclidean distance is calculated from every point on the map to each crime location.

Using the mathematical framework previously introduced in the section 4, with a manual adjustment of the model parameters, the program calculates the probability of every point in the map. When the probability is calculated it produces three-dimensional surface maps called jeopardy surface. The GeoCrime displays these probabilities on a two-dimensional map surface, (see Fig. 3).

The value of the probability of each point on the map is normalized in the standard range from 0-255 and encoded as the RGB values for each point on the map.

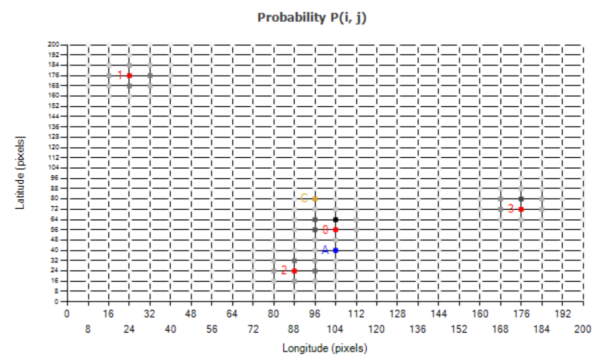

Fig. 3. - Color geoprofile map.

Case OTP: Crime sites: C0(60,101) C1(175,25) C2(25,89) C3(76,175) (in pixels); Hunting area: 4114.84 km2; Cell measured: 0.1036 km2; Centroid at the point:(84,98);

Fig. 3 shows the two-dimensional geoprofile where the crime sites are indicated by red dots and numbers, the offender's anchor point is marked by a blue dot and letter A and the centroid point by a goldenrod color dot and letter C. The color gradient from white (RBG (255,255,55)) to black (RBG (0,0,0)) applied to each point on the map represents probability intensity, where RGB color code values of each point depends on the value of the probability of that point. Points with almost white color indicate the lower the resultant score, while the points of color closer to black represent points with the greater the probability. The program has a unique approach of testing the performance of the model and estimating the efficiency of the search process, as described in subsection 4.3, that helps to minimize search cost and maximize search effectiveness. Also, it is a fast and user-friendly software that provides a variety of data visualization and analysis crime patterns based on geographical crime data.

6. Results

In this section, we present the results obtained in the profiling experiments on dataset that consists of seven solved series of cybercrimes from Bosnia and Herzegovina and Croatia. We evaluate the proposed model by conducting experiments using our cyber profiling software and monitoring the values of the critical features of the model. Initial testing was started with the values of empirical constants and weights like f, g and B based on recommendations and experiences of domain experts from past research[27], [28].

[1] The Google Maps API is not a free service. There is a free allowance of 40,000 calls to the geocoding API per month.

[2] Report of the Minister of the Interior of the Republic of Croatia, of 2013

The assumption of the research is that we are working with a series of n linked crimes, where each crime location C has two coordinate values, latitude and longitude. All cases include series with four or more crimes. Most cyber criminals do not have to appear in physical world while committing the crime, which makes their geographical profiling more difficult, and in some cases even impossible.

There are, however, situations in which offenders must come out of the virtual world and have physical contact with other participants in a crime or victims. Therefore, applying of the concept of geographic profiling focused on two types of cybercrimes: criminal acts of credit card skimming and spear phishing. In our study as crime site we considered location where the offender performed the act, the victim-offender encounter location or the victim's location.

Initially, we adjusted the model by modifying the curve exponents (f and g) to be as similar as possible to the volcano curves shown in Fig. 1. Experimental values for f and g were tried between 0.2 and 3 in increments of 0.2. For most of the cases examined, the best results were obtained for the values of the exponents f=g=1.7 ±0.09 with anchor point, for example, for *Case OTP*, at the point A(39,101) (real offender's residence is at the point R(28,92)) Fig. 4. displays plane section of a jeopardy surface where Fig. 4. (a) shows the level curve in longitude-probability plane at latitude=39, while Fig. 4. (b) shows the level curve in the latitude-probability plane at longitude=101.

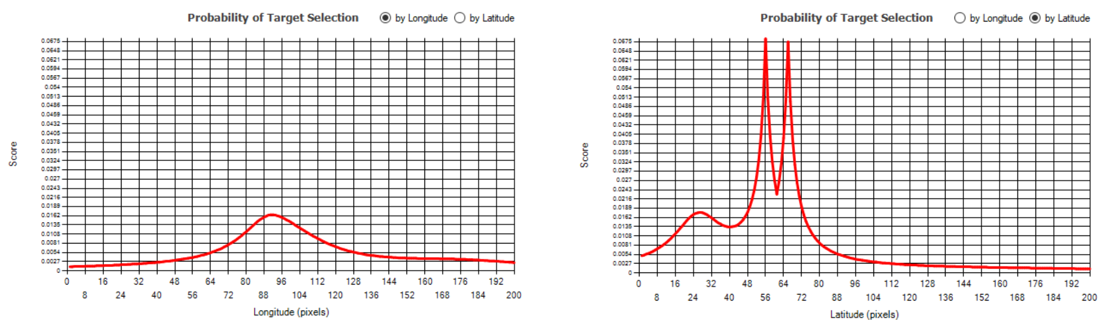

Fig. 4. - (a) at latitude=39; (b) at longitude=101.

Case OTP, plane section of a jeopardy surface (B=5, f=1.7, g=1.7)

On the other side, the buffer zone depends more on the characteristics of the criminal series, and altering the buffer zone radius (B) has significantly more influence on the model than changing f and g parameters. Therefore, we tested values of B starting from the value of 1 to the value for which the search area is outside the hunting area. Testing was followed by measuring the deviation of the predicted anchor point from the actual anchor point as well as calculation of different parameters such as the CGT hit score percentage, GC hit score percentage and PSEI index.

We obtained good results with B=5, while minimum of PSEI proved to be the best criterion for stopping the search process.

Table 1 presents information on crime patterns and CGT test results including number of crime sites, size of hunting and search areas, GC and CGT hit score percentages.

The results of the study have shown the average number of PSEI index of 0.23≈1/4 and a median of 0.14 (standard deviation = 0.24); the average number of crime locations was 5.6 (see Table 1). The PSEI index of 1/4 indicates that by applying this probability distance strategy the percentage of the resulting probability of locating the offender's base is on average 4 times higher than the percentage of the surveyed area, and with the increase in the search area it grows much faster than the percentage of the total search area. The average home to crime distance measured in pixels for all cases was 80 (SD=67.5), while the mean distance between the real offender's residence and the nearest location of the crime in each series was 6, median was 4 (SD=9.4).

Table 1

The Geocrime results *(B=5, f=1.7, g=1.7)*

Case	Crime Sites	Hunting Area (km^2)	Search Area (km^2)	GC hit score (%)	CGT hit score (%)
			Credit card skimming		
CASE T13	5	10227.61	25.59	5.25	0.25
CASE Inspector	7	10000.94	25.27	1.83	0.25
CASE T09	4	58.25	0.15	6.4	0.26
CASE OTP	4	4114.84	10.34	0.87	0.25
			Spear phishing		
CASE Grigory_Moscow_1	6	13917.14	35.27	0.35	0.25
CASE Grigory_Moscow_2	5	12626.85	32.22	0.94	0.26
CASE Mat Shares	8	23947.62	61.12	2.36	0.26

Table 2 gives the results of the application of these two geographical profiling techniques, calculating the deviation of assumed anchor point or centroid point from the real offender's residence. The centrographic technique is used to find the mean $\bar{x}$- and $\bar{y}$- coordinates of the crimes, and then it calculates deviation of obtained centroid point $(\bar{x}, \bar{y})$ from the real offender's residence (shown by the D1 column of Table 2).

The last column in Table 2 gives the deviation of assumed anchor point from the real offender's residence for all cases in sample. The average value for column D1 is 66 (SD=12), while average value for column D2 is 27 (SD=28.9). This is another indication of the fact that the application of the probability distance strategy in relation to the spatial distribution strategy for the geographic profiling of cybercrime gives a much better result for a given sample of research subjects. To help better understand general patterns of crime data, we also calculate the standard distance deviation (SDD) that represents the distance in pixels from the centroid of crime sites and is usually plotted on a map as a circle for a visual indication of dispersion. The gain values for The Nearest Neighbor Index in the range from 1.12 to 1.44 indicate the not random distribution of crime in a given sample.

To evaluate the model's performance, two measures were used: double mean hit score percentage and the Gini coefficient [28]. The double mean hit score percentage was 0.05 and the Gini coefficient value was 0.01.

Table 2
The quantitative data analysis results *(B=5, f=1.7, g=1.7)*

Case	Standard Distance Deviation (pixels)	Nearest Neighbour Index	Centroid (pixels)	Offender's residence (pixels)	D1 (pixels)	Anchor point (pixels)	D2 (pixels)
				Credit card skimming			
CASE T13	87	1.13	(140,70)	(182,19)	66	(180,20)	2
CASE Inspector	78	1.39	(108,95)	(59,44)	71	(102,100)	71
CASE T09	89	1.12	(134,132)	(169,160)	45	(168,173)	13
CASE OTP	77	1.44	(84,98)	(28,92)	84	(39,101)	14
				Spear phishing			
CASE Grigory_Moscow_1	84	1.32	(90,58)	(149,17)	72	(83,21)	66
CASE Grigory_Moscow_2	80	1.34	(82,113)	(20,123)	63	(23,106)	17
CASE Mat Shares	85	1.31	(70,115)	(19,148)	61	(14,144)	6

In the cases we have studied it is interesting to note the way the offenders choose their targets, where, for example, in cases of spear phishing they choose smaller companies (e.g., small construction companies) with a weak IT security infrastructure or in cases of card skimming they choose ATMs at locations with a higher frequency of tourist visits.

A special part of the research was dedicated to the analysis of existing methods and techniques of visualization for easier understanding and analysis of crime patterns. We created more different types of visual graphs, maps and monitoring charts for graphical visualization of crime patterns. In Fig. 5. we have an example of the buffer zone plotted on the street map with all crime locations precisely marked.

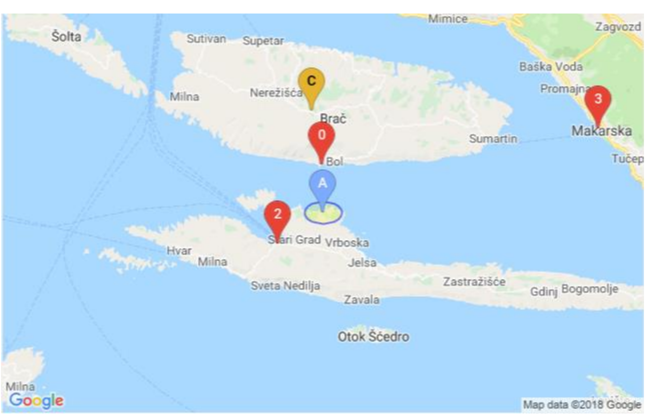

Fig. 5. - Buffer zone.
Case OTP

Given the fact that the buffer zone represents an area of reduced criminal activity, or rather it describes an area surrounding the criminal's anchor point, using this map can help predict the potential locations of the next crime and identify crime hotspots. The buffer zone has the shape of an ellipse with a center at the anchor point because the Geocrime uses the Euclidean measure to measure the distance between the points.

7. Conclusion

In today's online environment, it is necessary to develop new and reliable methods for assessing cybercrime and profiling of cyber criminals. It is time to start looking at cybercrime similar to the other types of traditional crime (fraud, bombings, burglary, rape, auto theft and illegal activities) where a plethora of prevention and detection techniques have been used for a long time. One of them is the geographic profiling, technique originally developed in criminology, where in this study, we have tested the possibilities of geographic profiling of cybercrimes for serial crime location prediction. The profiling process involved generation of the geographic profile from incident reports, verification and optimization of a model, testing the sensitivity of parameters of model, accuracy measures and graphical visualization. All this is done using specialized crime-mapping software that was developed just for this research. Research has shown that certain criminal offenses of cybercrime meet the requirements for successful geographic profiling, that cyber criminals follow a certain logic (nonrandom nature of criminal behavior) and that most crimes have patterns. This study also showed that there are some difficulties in implementing geographic profiling of cybercrime such as diversity of offender data and behavior, offenders change and improve their modus operandi (MO), the non-physical uniqueness of cyberspace, the problem determining the location of the actual crime-scene and the general problem getting real datasets for testing.

However, although there are limitations associated with the use of geographic profiling, the benefits of geographic profiling of cybercrime appear to be promising, especially when the cybercrime has a clear requirement or expectation of a physical dimension.

References

[1] N. Cyber, S. Insights, and G. Results, “2017 Norton Cyber Security Insights Report Global Results,” 2017.

[2] S. Fernando and T. Yukawa, “Internal Control of Information Sharing through user Security Behavioural Profiling,” *Glob. J. Human-Social Sci. Res.*, 2014.

[3] I.-A. Flourish, “Criminal Profiling,” no. Roehampton University, London, 2012.

[4] D. K. Rossmo, “Geographic profiling,” in *Encyclopedia of criminology and criminal justice*, Springer, 2014, pp. 1934–1942.

[5] J. E. Douglas and A. E. Burgess, “Criminal profiling: A viable investigative tool against violent crime,” *FBI L. Enforc. Bull.*, vol. 55, p. 9, 1986.

[6] P. J. Brantingham, P. L. Brantingham, and others, *Environmental criminology*. Sage Publications Beverly Hills, CA, 1981.

[7] K. Emeno, C. Bennell, B. Snook, and P. J. Taylor, “Geographic profiling survey A preliminary examination of geographic profilers? views and experiences,” *Int. J. Police Sci. Manag.*, no. November 2016, p. 1461355715621070, 2016.

[8] T. Rich and M. Shively, “Document Title: A Methodology for Evaluating Geographic Profiling Software,” 2004.

[9] D. Paulsen, “Improving Geographic Profiling through Commuter/Marauder Prediction,” *Police Pract. Res.*, vol. 8, no. 4, pp. 347–357, 2007.

[10] P. R. Stephenson and R. D. Walter, “Toward Cyber Crime Assessment: Cyberstalking,” *6th Annu. Symp. Inf. Assur. (ASIA)*, pp. 1–5, 2011.

[11] N. Al Mutawa, J. Bryce, V. N. L. Franqueira, and A. Marrington, “Forensic investigation of cyberstalking cases using Behavioural Evidence Analysis,” in *Digital Investigation*, 2016, vol. 16, pp. S96–S103.

[12] S. Yu, “Behavioral evidence analysis on Facebook: a test of cyber-profiling,” *Democr. NEED Secur. HAVE TO CONFRONT VIOLENCE*, p. 19, 2013.

[13] M. Zulfadhilah, I. Riadi, and Y. Prayudi, “Log Classification using K-Means Clustering for Identify Internet User Behaviors,” *Int. J. Comput. Appl.*, vol. 154, no. 3, 2016.

[14] M. Zulfadhilah, Y. Prayudi, and I. Riadi, “Cyber Profiling using Log Analysis and K-Means Clustering A Case Study Higher Education in Indonesia,” *Int. J. Adv. Res. Comput. Sci. Appl.*, vol. 7, no. 7, pp. 430–435, 2016.

[15] A. Filippoupolitis, G. Loukas, and S. Kapetanakis, “Towards real-time profiling of human attackers and bot detection,” *Proc. CFET, 2014 - C.*, 2014.

[16] C. Vandeviver and W. Bernasco, “The geography of crime and crime control,” *Applied Geography*, vol. 86, pp. 220–225, 2017.

[17] C. Bennell and S. Corey, “Geographic profiling of terrorist attacks,” in *Criminal Profiling: International Theory, Research, and Practice*, 2007, pp. 189–203.

[18] D. K. Rossmo and K. Harries, “The Geospatial Structure of Terrorist Cells,” *Justice Q.*, vol. 28, no. 2, pp. 221–248, 2011.

[19] M. Tonkin, J. Woodhams, J. W. Bond, and T. Loe, “A theoretical and practical test of geographical profiling with serial vehicle theft in a U.K. context,” *Behav. Sci. Law*, vol. 28, no. 3, pp. 442–460, 2010.

[20] L. Hammond, “Geographical profiling in a novel context: Prioritising the search for New Zealand sex offfenders,” *Psychol. Crime Law*, vol. 20, no. 4, pp. 358–371, 2014.

[21] U. Santosuosso and A. Papini, “Methods for geographic profiling of biological invasions with multiple origin sites,” *Int. J. Environ. Sci. Technol.*, vol. 13, no. 8, pp. 2037–2044, 2016.

[22] Y. Suzuki-Ohno, M. N. Inoue, and K. Ohno, “Applying geographic profiling used in the field of criminology for predicting the nest locations of bumble bees,” *J. Theor. Biol.*, vol. 265, no. 2, pp. 211–217, 2010.

[23] R. A. Martin, D. K. Rossmo, and N. Hammerschlag, “Hunting patterns and geographic profiling of white shark predation,” *J. Zool.*, vol. 279, no. 2, pp. 111–118, 2009.

[24] L. Mburu and M. Helbich, “Evaluating the Accuracy and Effectiveness of Criminal Geographic Profiling Methods: The Case of Dandora, Kenya,” *Prof. Geogr.*, vol. 67, no. 1, pp. 110–120, 2015.

[25] M. O. Leary, “The Mathematics of Geographic Profiling,” vol. 265, pp. 253–265, 2009.

[26] M. O. Leary, “Modeling Criminal Distance Decay,” *Cityscape A J. Policy Dev. Res.*, vol. 13, no. 3, pp. 161–198, 2011.

[27] F. Wang, D. Kim Rossmo, I. Laverty, and B. Moore, “Geographic Profiling for Serial Crime Investigation.” pp. 102–117, 2005.

[28] D. K. Rossmo, *Geographic profiling*. CRC Press, 2000.

[29] D. Canter, T. Coffey, M. Huntley, and C. Missen, “Predicting Serial Killers’ Home Base Using a Decision Support System,” *J. Quant. Criminol.*, vol. 16, no. 4, pp. 457–478, 2000.