STUDIA INFORMATICA 2014
Volume 35 Number 2 (116)

Robert MARCJAN, Marek ŁAKOMY, Michał WYSOKIŃSKI,
Kamil PIĘTAK, Marek KISIEL-DOROHINICKI
AGH University of Science and Technology, Department of Computer Science

LINK – CRIMINAL INTELLIGENCE ENVIRONMENT WITH SPATIAL/GEOGRAPHICAL ANALYSIS TOOLS[1]

Summary. The paper presents the LINK application, which is a decision-support system dedicated for operational and investigational activities of homeland security services. The paper briefly discusses issues of criminal analysis, possibilities of utilizing spatial (geographical) information together with crime mapping and spatial analyses.

Keywords: decision support, data analysis , GIS, criminal analysis

LINK – ŚRODOWISKO ANALIZ KRYMINALNYCH WYKORZYSTUJĄCE NARZĘRZIA ANALIZ GEOPRZESTRZENNYCH

Streszczenie. Artykuł prezentuje system LINK będący zintegrowanym środowiskiem wspomagania analizy kryminalnej przeznaczonym do działań operacyjnych i śledczych służb bezpieczeństwa wewnętrznego. W artykule omówiono problemy analizy kryminalnej, możliwość wykorzystania informacji o charakterze przestrzennym oraz narzędzia i metody analiz geoprzestrzennych.

Słowa kluczowe: wspomaganie decyzji, analiza danych, GIS, analiza kryminalna

1. Introduction

Criminal analysis is a complex process involving information gathered from different sources, such as phone billings, bank account transactions and eyewitnesses testimonies.

[1] The research reported in the paper was partially supported by the project No. O ROB 0008 01 "Advanced IT techniques supporting data processing in criminal analysis", funded by the Polish National Centre for Research and Development.

Most of this data possess spatial (geographical) aspect, what means it can be presented and analysed on maps. Because of massive character of this information, it is hardly possible for it to be processed without the help of sophisticated computer systems. On the other hand, because of crucial role of a human expert in the process, the main role of such a system is to support dealing with so complex data. It means that all the information has to be transformed into a coherent and relatively simple visual form, in which key objects (suspects, events, etc.) and their interrelations can be easily spotted.

The paper describes a system designed to facilitate complex analysis processes, in particular related to operational and investigation activities of homeland security services. It covers the most important tasks and typical analysis steps including: initial data transformation and filtering, data integration and structuring, graphical data representation, visualization on maps and use of geospatial analysis functions.

2. LINK analysis environment and its architecture overview

LINK[2] [2, 3] is a comprehensive data analysis platform aimed to aid criminal analysis which, as described in previous sections, has some specific requirements. The basic version of the system provides a set of tools for integrating, processing and visualizing data which may originate from various sources (e.g. phone billings, bank account statements, address books, etc.). LINK is an extensible platform, based on plug-in architecture [6, 7]. It allows for integrating new tools for automatic or semi-automatic data analysis as well as enhancing existing ones. Such tools may be provided by third party vendors and combined with others in order to extract relevant information from a large volume of data. This approach allows for nearly unlimited ways of extending and adjusting the system to the requirements of particular criminal analysis domain.

2.1. Data acquisition

The first step in the process of data analysis in LINK is importing the data into the system. Based on the experience of professional criminal investigators, it is safe to state that there is no standardized format of analysed data. Not only can it be in any file format (e.g. plain text, Rich Text Format, Comma Separated Values, Microsoft Excel or Word) but also it can be in paper form, which requires additional steps in order to digitalize it.

[2] http://fslab.agh.edu.pl/

LINK provides an advanced data import tool. The import wizard guides the user through the import process in an intuitive manner. During the import the data is validated against various rules, which may be specific for particular data domain (Fig. 1). For example, this may include checking if a phone call record in a billing has both caller and receiver numbers or if bank account numbers in a bank statement are valid. The user can preview each step of the import process and influence it by adjusting parameters like format of time, date, etc. or defining default values. All operations and user modifications are placed in an import log, which is presented to the user once the import is completed.

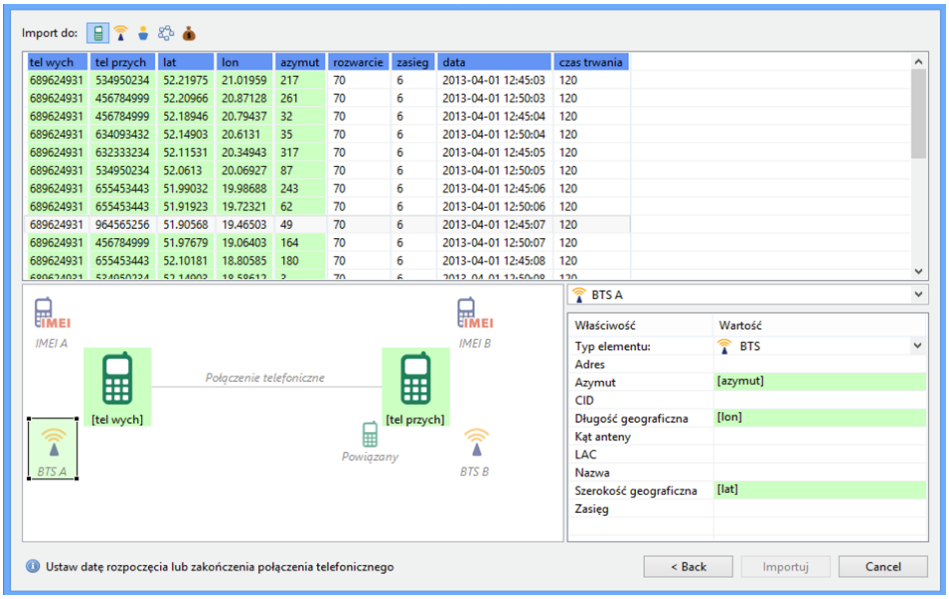

Fig. 1. Data importer wizard
Rys. 1. Graficzny kreator importu danych

The biggest advantage of LINK import tool is its flexibility. It comes in two forms:

- a simplified version for importing data of tabular nature (e.g. phone billings) which is similar to import tools used in spreadsheet applications,
- a generic import tool which is based on parsing trees and may be used to import data in any form (e.g. bank statements).

Both versions of the import tool allow the user to specify the format and the semantics of each portion of the data (e.g. pointing the phone call start time, caller and receiver numbers, etc.) Moreover they support even more sophisticated scenarios. For example, some mobile network providers deliver billings in which the phone call direction is specified by the caller column and receiver column, whereas others provide two columns with numbers and additional column with direction information. LINK import tool is able to handle these cases, as well as others, even more complicated.

Due to the flexibility of the import tool and variety of data formats, the whole import process is tedious and time-consuming. In order to aid that, it is possible to define an import

template which can be later reused for loading data in the same or similar format (templates may be adjusted if needed). Templates are stored as XML files therefore they are easy to share with other users (via email or centralized repository).

2.2. Data processing and analysis

The work of criminal analysts is a form of advanced, context aware, data mining. The goal of analysis process is to find related objects in a network of data, which is achieved by applying mixture of numerous algorithms which are formalized by graph clustering [8], social network analysis and other pattern recognition [10] fields of study. Most of these operations can be automated, at least to some extent, and LINK is designed to facilitate that.

The basic version of LINK provides a set of essential data analysis operators which allow for extraction of vital information from a large volume of data. The most widely used operation is filtering, which may be based on data type, string pattern, value or time range

This allows for quick removal of irrelevant data out of the analysis picture. The user can always see the summary of analysed data in the form of basic statistics, which include the counts of all elements and their relationship occurrence as well as average or aggregated values of their attributes. Such summary is very useful as it allows for quick recognition of data regularities and anomalies, such as frequent events or extraordinary values.

In addition to the basic analysis operators LINK provides more advanced ones. For example, in the phone billing analysis domain users can perform operations such as:

- detecting and removing phone number prefixes,
- merging 2 or more billings into one data set,
- detecting duplicate phone calls (even with time shift).

Since the LINK platform is extensible, new, even more advanced analysis operators and tools can be provided in the future (e.g. for detecting paths in graphs or finding frequent event patterns).

2.3. Visualization

One of the most important features of LINK is its support for exploring and manipulating data in a visual manner, which is in conformance with visual data mining [1] approach.

This is done via editors which allow to analyse different aspects of the data. Editors allow to modify presented data as well as adding new artificial one, therefore they can also be used as general purpose graphical notepads.

The basic version of the system provides two graphical data editors:

- Timeline editor which presents occurrence of events with emphasis on their chronology.
- Schema editor, which presents the data in the form of a graph.

The timeline editor displays horizontal timelines (which may represent people or phone numbers) and events, which are presented as arrows between them. In addition, the timeline diagram displays a horizontal time axis. This form of data analysis is useful for investigating sequences of events between various objects, as well as event patterns.

Data in the schema editor is presented as a graph which consists of nodes and connections with properties. In order to display the graph in a readable form various graph layout algorithms are provided, including:

- Circle layout – nodes are evenly spaced on the border of a circle,
- Group layout – main nodes (with significantly more connections) are placed in a circle, remaining nodes are further outside,
- Peacock [9] layout – main nodes are in the center, remaining nodes are further outside (Fig. 2),
- Grid layout – nodes are evenly spaced on a grid.

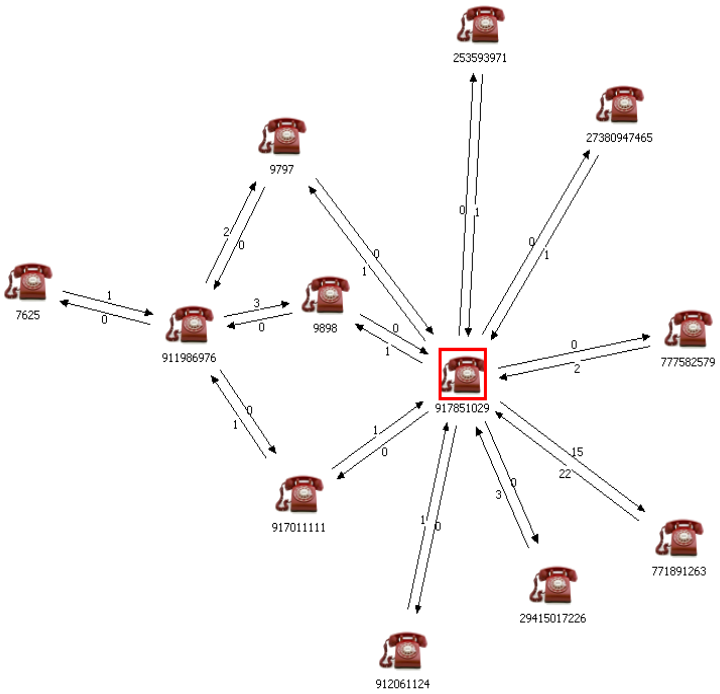

Fig. 2. Peacock layout of schema diagram
Rys. 2. Rozkład „pawi ogon” dla diagramu schematycznego

In addition to above-mentioned editors, others may be provided as extensions for the LINK system.

3. LINK mapping module

Evaluation of standard LINK software data visualization tools, schema editor and timeline editor, has shown the need of special treatment of geographical data visualization. Many cases processed by analysts, involve evaluation of data strictly connected to geographical context. Mobile phone call billings, among other data, provide information about the location and range of BTS[3] stations utilized to establish a call. Sometimes BTS data is incomplete in the file containing billing. In such case data can be data can be supplemented by information from the UKE[4] database. Also bank statements provide information about the addresses of ATM machines and sales points, where the credit or debit card were used. This information visualization is often crucial to understand the whole situation and prove analyzed hypotheses. To achieve this goal analysts often use specialized GIS software [4, 5].

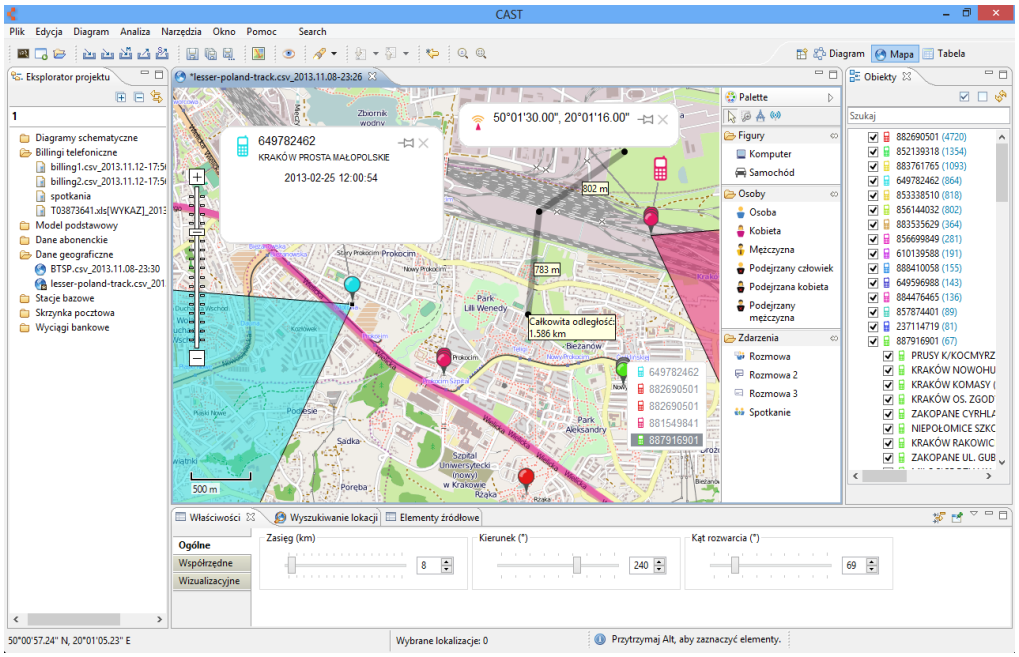

Fig. 3. LINK Map Editor
Rys. 3. Edytor mapowy programu LINK

To help continuing analysis flow, the Map Editor module was developed (Fig. 3). This Map Editor plugin is aimed to increase productivity of analyses based on geographical data. Most operations performed on geographical data during criminal analysis are often repeatable and comes down to few simplified functions. As most day to day analyses would not come beyond that functionality, simple tools for most common data analysis and visualization has been implemented in the LINK Map Editor plugin. If that is not enough, plugin provides tools

[3] Base Transceiver Station
[4] Republic of Poland Office of Electronic Communications *http://en.uke.gov.pl/*

for integration and interoperation with specialized GIS software such as ESRI ArcGIS [5] (briefly described in the next section). The main advantage of this approach is an easy way to export pre-prepared data for the further and more sophisticated GIS analysis.

One of the most important features of every map editor is to support for raster basemaps. LINK software by default uses open access raster map data available through an OpenStreetMap[5] webservice. However the user is not restricted to use the default solution. In order to provide a possibility of access to any type of basemap, LINK defines an extension point, for new providers. Considering the real needs of Polish criminal analysts, and realizing that they often work with internet disabled workstations, application by default comes not only with online OpenStreetMap basemaps, but also with the set of standard basemaps stored on disk files.

Current functionality of Map Editor consist of tools for: visualization of previously gathered data, manual data preparation and object filtering tools.

Base data models available in map editor are points, tracks and ranges (such as BTS stations ranges). Apart of those, it is possible to bind metadata objects with map model. These can be distance measures, events' order indicators, notes or the map legend describing object classes placed on the map.

Apart from the possibility to import basic geographical objects from heterogeneous sources, specialized visualization methods were developed for mobile phone call billings and BTS stations lists, e.g. for objects bound with BTS range data (including mobile phone call events) it is possible to switch display of particular ranges visualization according to the current analysis needs. There is also a tool for data transformation between model of BTS stations with related call events and model in every call event is mapped to a separate pin object (point).

Data filtering is possible via two developed tools – the "timeline dialog" and "objects" view. The timeline allows filtering visible objects by the occurrence time of connected events. It does not only allow the user to filter out events from the specified timeframe, but also to investigate the sequence of events. The "objects" view allows filtering objects by their domain properties' values such as a phone number or BTS station address, including the full text search.

3.1. Integration with ArcGIS[6]

Even though some geo-processing can be done using LINK, it is often necessary to use a more sophisticated analytical software like ArcGIS. It is a set of software tools developed

[5] http://www.openstreetmap.org/about
[6] http://www.esri.com/software/arcgis

by ESRI for creating, processing and analyzing maps and spatial data. The tools are available for desktop and server architectures, as well as for mobile devices. In addition to simple analysis methods like data filtering ArcGIS allows chain processing, which connects individual geo-processing functions and creates very complex models like hot-spot analysis or determination of meeting probability between two suspects (which is presented in the next section).

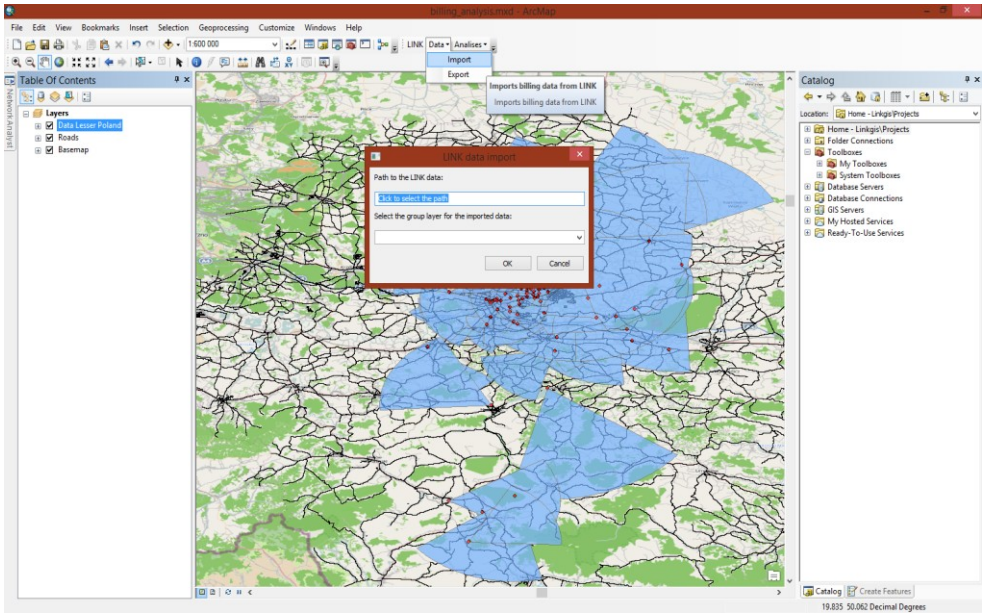

Fig. 4. The LINK data import dialog extension in ArcGis
Rys. 4. Okno importu danych LINK-a jako wtyczka programu ArcGis

The ArcGIS software is designed for end-users, as well as developers. It offers COM based ArcObjects APIs in Java and C#, which can be used for creating applications with geo-processing capabilities and extending functionalities of those provided by ESRI. In addition to that, geo-processing functions can also be accessed via Python modules and put together in a set of scripts or extensions.

To provide interoperability between LINK and ArcGIS, an intermediate layer of data integration was developed. This layer consists of two modules, one on each side (LINK and ArcGis). LINK mapping module has been equipped with a simple export algorithm to shapefile[7] data format and the ArcGIS custom Python extension was developed. It provides tools supporting data import and conducting common analyses. It comes with its own graphical interface (Fig. 4) and allows for data transfer between LINK and ArcGis.

[7] Esri spatial data format: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

3.2. Case study – example of a geographical criminal analysis

To illustrate the use of LINK and ArcGIS software application in criminal analysis considering the geographical data let's walk through a simple case study.

Let's assume, that the series of three gas station robberies had taken place. CCTVs had captured the same registration number of the robbers' car. Car owner has been arrested. Unluckily he has not had the robbed money with him. He has also claimed, that his car had been stolen and he has nothing to do with the robberies. In the meantime another suspect with a large amount of money, matching the stolen amount, has been arrested in the other part of Poland.

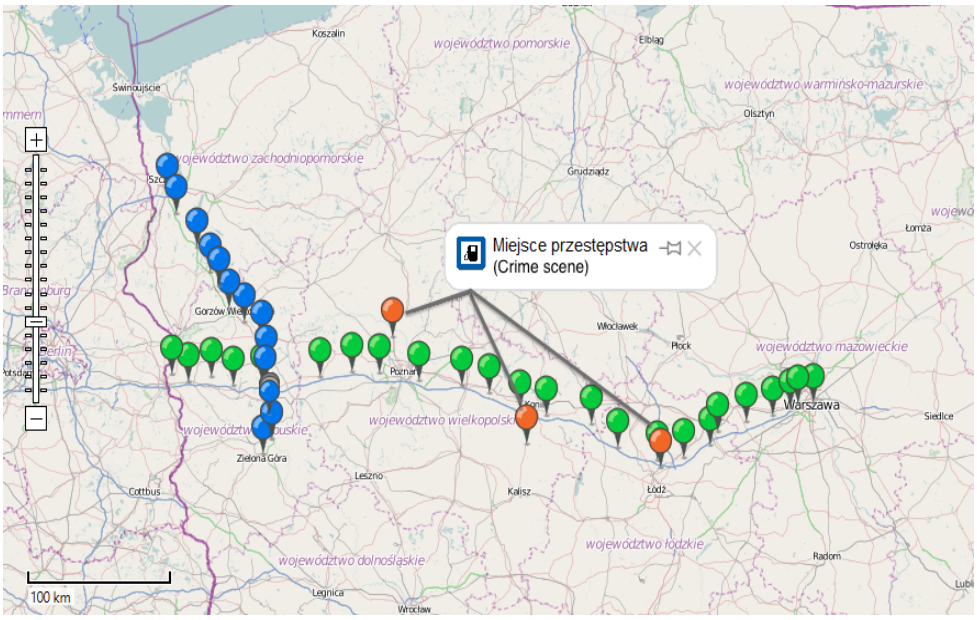

Fig. 5. BTS logins diagram augmented with robbed gas stations' positions
Rys. 5. Diagram logowania do BTS

The analyst has been asked to check, whether one of the suspects could have been at the crime scene. The mobile network operator has provided billings with mobile phone calls of both suspects. The data was imported to LINK, and BTS data was automatically complemented with UKE database contents. Next, map diagram of BTS used for calls was created and augmented with robbed stations positions (Fig. 5).

The diagram showed that only the car owner's mobile phone made calls through BTS stations proximate to the crime scenes (BTS stations indicated by horizontal pins series). Furthermore he is probably lying about the car theft, as he did not claimed his phone has been stolen with it. On the other hand suspect arrested with money could not possibly be with his phone at those places (BTS stations indicated by vertical pins series).

However to exclude all doubt about car owner being able to rob the stations, analyst had to prove that the suspect could manage to get between phone calls to the crime scene and

have enough time to perform the robbery (Fig. 6.). Unfortunately this case is beyond the current LINK software capabilities. However, it is possible to export map data to external GIS application ,such as ArcGIS software, and continue the analysis (Fig. 7).

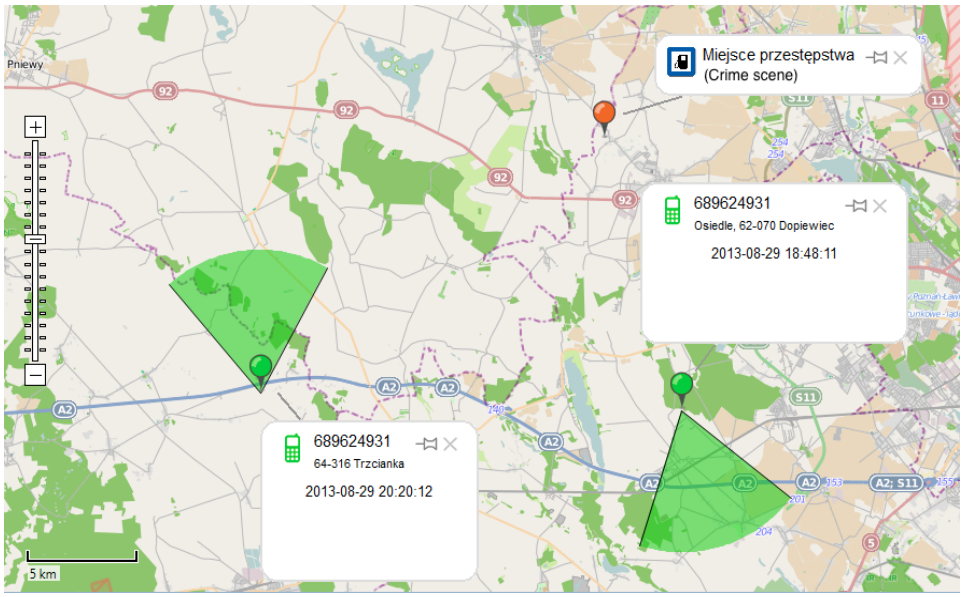

Fig. 6. BTS login times and distance from crime scene
Rys. 6. Czasy logowań do stacji bazowych i odległość stacji od miejsca przestępstwa

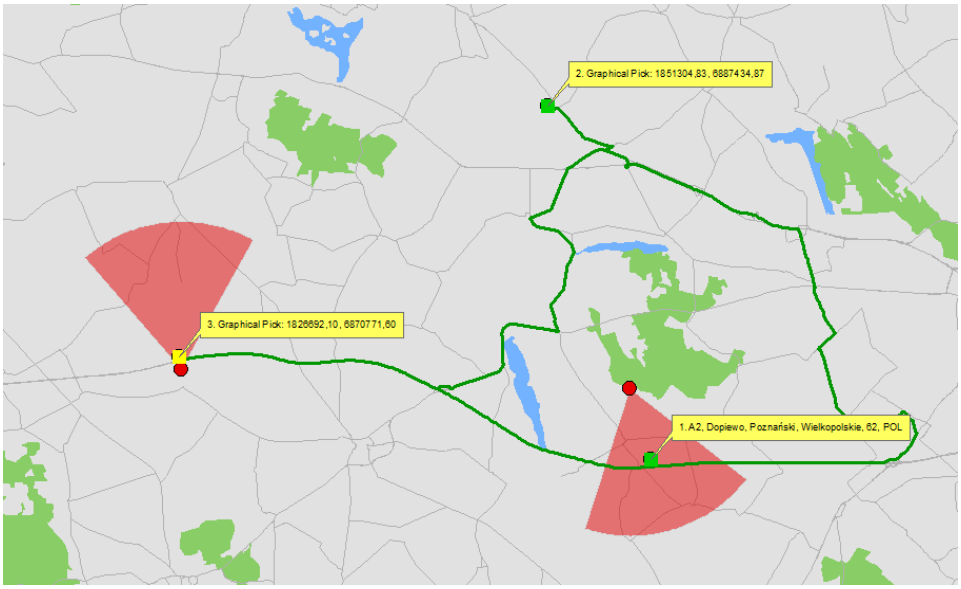

Fig. 7. LINK map diagram data exported to ArcGIS software; network-based analysis visualization
Rys. 7. Dane diagramu mapowego LINK wyeksportowane do programu ArcGIS; wizualizacja analizy sieciowej

The ArcGIS allows for the network-based analysis, and to find the route that could have been taken by the suspect and compare the driving time to the proximate BTS login time difference. The computed driving time is 1 hour 9 minutes (Fig. 8) and the actual login time difference (Fig. 6) is 1 hour 32 minutes. Therefore the suspect could not only get to the crime scene, but also had additional 23 minutes to perform a robbery.

After proving the robberies series performed by the car owner to be possible analyst got another hypothesis to consider: "The person arrested with the money could have met with the car owner and picked up the plunder".

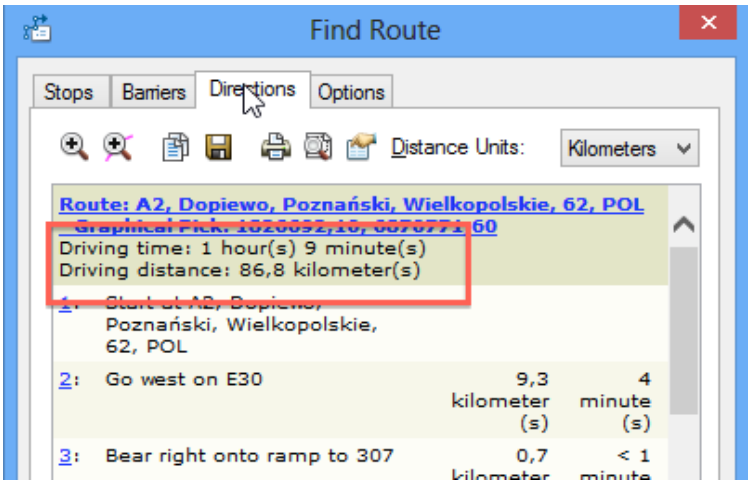

Fig. 8. Computation of the driving time
Rys. 8. Wyliczony czas przejazdu

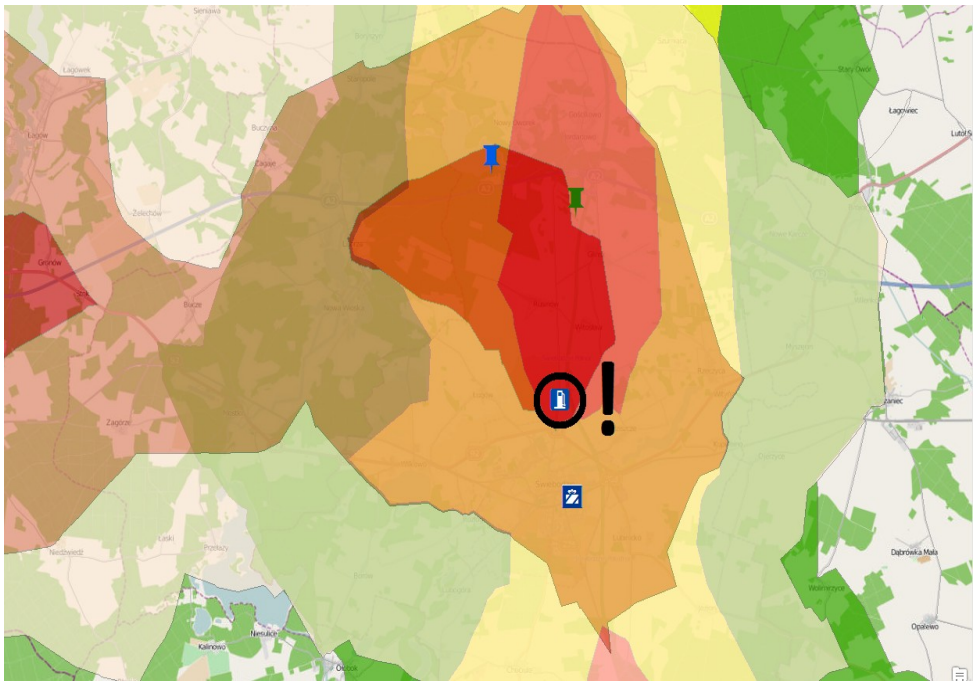

Fig. 9. Diagram of areas possible to reach within the other 5 minutes intervals from BTS login
Rys. 9. Diagram obszarów możliwych do osiągnięcia w ciągu kolejnych 5-minutowych interwałów od logowania do stacji bazowej

The LINK application is equipped with an easy to use tool for potential meeting analysis. For the given case it has found two possible suspects meetings within the 60 minutes time-frame. Positions of BTS stations used for phone calls nearby to the possible meeting area can be easily shown on a map diagram. However, to estimate the actual meeting place, one can again export map data to external ArcGIS application. The network-based analysis can be

performed to form a diagram of areas possible to reach within the defined time intervals from BTS login (Fig. 9). In this case, intervals are set to 5 minutes.

Examinng the given diagram supplemented with additional POI (Points of Interest) information, one can presume that most likely meeting points are the gas station placed in the 5 minute distance rings from both suspects BTS logins, and the gift shop within the 10 minute distance rings. Aided by this information it can be easily concluded that employees at these facilities should be questioned and the CCTVs recordings from these facilities should be captured and examined.

This is of course only the simple case study (fabricated for the purpose of the paper), but it shows the possibilities of spatial data analysis and mapping with the use of LINK and ArcGis and presents the power of these tools.

4. Conclusion and further research

Criminal intelligence supported with spatial analysis tools is a very promising new discipline. It slowly starts beings noticed mainly thanks to the fact that it significantly enhances criminal analysis capabilities. It predicts how suspects might behave in a given environment and helps to coordinate field operations. LINK is one of the very few tools that is made specifically for this purpose. It is a professional tool, which was successfully implemented and is widely used by the Polish security services like Polish Police. The software can operate on general data, however it can be easily adapted to operate effectively on national specific data, like phone billings, maps etc, what is a big advantage on the Polish market. Thanks to these, activities like importing the billing data and furthermore performing various analyses (especially the spatial ones), which is usually a long and arduous process, become effortless. All these tasks could by easily automated with LINK tools. Further research will focus on introducing more types of spatial analyses (which could improve the probability of finding suspects and evidence), together with a convenient user interface

BIBLIOGRAPHY

1. Cox K., Eick S., Wills G., Brachman R.: Visual data mining: Recognizing telephone calling fraud. Data Mining and Knowledge Discovery, Vol. 1, No. 2, 1997, p. 225÷231.
2. Dajda J., Dębski R., Kisiel-Dorohinicki M., Piętak K.: Multi-Domain Data Integration for Criminal Intelligence. [in:] Man-Machine Interactions 3, Advances in Intelligent Systems and Computing, Vol. 242, Springer, 2014, p. 345÷352.

3. Dębski R., Kisiel-Dorohinicki M., Miłoś T., Piętak K.: LINK – a decision-support system for criminal analysis. [in:] MCSS 2010: Multimedia Communications, Services and Security: IEEE International Conference, Kraków 2010, p. 110÷116.
4. Gorr W., Kurland K.: GIS Tutorial for Crime Analysis. ESRI Press, 2010.
5. Hanning R.: Spatial data Analysis, Theory and Practise. Cambridge Press, 2003.
6. Judd C. M., Shittu H.: Eclipse Plug-in Paradigm. Apress, ch. 2, 2005, p. 11÷18.
7. McAffer J., Lemieux J. M. (eds.): Eclipse Rich Client Platform: Designing, Coding, and Packaging Java(TM) Applications. Addison-Wesley Professional, 2005.
8. Schaeffer S. E.: Graph clustering. Computer Science Review, Vol. 1, No. 1, 2007, p. 27÷64.
9. Wills G. J.: Network interactive visualization of very large graphs. Lecture Notes in Computer Science, Vol. 1353, Springer, 1997, p. 403÷414.
10. Theodoridis S., Koutroumbas K.: Pattern Recognition. Academic Press, 2008.

Wpłynęło do Redakcji 30 stycznia 2014 r.

Omówienie

Artykuł prezentuje system LINK będący zintegrowanym środowiskiem wspomagania analizy kryminalnej przeznaczonym do działań operacyjnych i śledczych służb bezpieczeństwa wewnętrznego. Analiza kryminalna jest złożonym procesem wymagającym gromadzenia informacji z wielu różnych źródeł, takich jak billingi telefoniczne, historie transakcji bankowych czy zeznania naocznych świadków. Większość z tych danych ma aspekt przestrzenny, co implikuje możliwość przedstawiania i analizy z użyciem map. Jednocześnie w związku z ogromną ilością przetwarzanych danych ich analiza jest praktycznie niemożliwa bez użycia zaawansowanych systemów informatycznych. W artykule omówiono problemy analizy kryminalnej oraz związane z nimi aspekty przestrzenne (geograficzne), wskazane zostały metody analizy i odpowiednie narzędzia programowe wraz z ich wadami i zaletami oraz omówiono możliwości analizy kryminalnej wspomaganej danymi geograficznymi. Główną część artykułu stanowi opis funkcji systemu LINK, w szczególności procedur zbierania danych oraz ich przetwarzania, analizy i wizualizacji. Artykuł prezentuje również możliwości integracji aplikacji LINK z zestawem narzędzi ArcGIS stworzonych przez firmę ESRI oraz wykonywanie złożonych analiz przestrzennych.

Addresses

Robert MARCJAN: AGH University of Science and Technology, Department of Computer Science, al. Mickiewicza 30, 30 059 Kraków, Poland, marcjan@agh.edu.pl.

Marek ŁAKOMY: AGH University of Science and Technology, Department of Computer Science, al. Mickiewicza 30, 30 059 Kraków, Poland, lakomy@iisg.agh.edu.pl.

Michał WYSOKIŃSKI: AGH University of Science and Technology, Department of Computer Science, al. Mickiewicza 30, 30 059 Kraków, Poland, wysokinski@iisg.agh.edu.pl.

Kamil PIĘTAK: AGH University of Science and Technology, Department of Computer Science, al. Mickiewicza 30, 30 059 Kraków, Poland, kpietak@agh.edu.pl.

Marek KISIEL-DOROHINICKI: AGH University of Science and Technology, Department of Computer Science, al. Mickiewicza 30, 30 059 Kraków, Poland, doroh@agh.edu.pl.