Zeszyty Naukowe AON nr 4(93) 2013
ISSN 2299-6753

Robert MARCJAN
Akademia Górniczo-Hutnicza
Kraków

Marek ŁAKOMY
Akademia Górniczo-Hutnicza
Kraków

Michał WYSOKIŃSKI
Akademia Górniczo-Hutnicza
Kraków

ANALIZA KRYMINALNA WSPOMAGANA NARZĘDZIAMI GIS W APLIKACJI LINK[1]

Kamil PIĘTAK
Akademia Górniczo-Hutnicza
Kraków

Marek KISIEL-DOROHNICKI
Akademia Górniczo-Hutnicza
Kraków

Abstract

The aim of the paper is to present the possibilities of utilizing spatial (geographical) information in the criminal analysis and to introduce the LINK software, which is able to perform such analyses. LINK is a decision-support system dedicated for operational and investigational activities of homeland security services. The paper briefly discusses issues of criminal analysis in the Polish reality and focuses on crime mapping and spatial analyses. It also lists available analysis methods and software tools together with their pros and con. The description of LINK system functionality constitutes the main part of the paper. It covers the tasks of data acquisition, processing, analysis and visualization. In addition to that the paper also includes a description of the integration possibilities of LINK and ESRI's ArcGIS application stack and performing complex geospatial analyses.

Key words – decision support, criminal analysis, GIS – geographical information system

[1] Prace zostały sfinansowane z projektu: „Zaawansowane technologie informatyczne wspierające procesy przetwarzania danych w obszarze analizy kryminalnej" Projekt rozwojowy NCBiR nr O ROB 0008 01.

Wstęp

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 charakter danych przestrzennych, co oznacza, że można je przedstawić i analizować z użyciem map. W związku z ogromną ilością przetwarzanych danych ich analiza jest praktycznie niemożliwa bez użycia zaawansowanych systemów informatycznych. Nie oznacza to jednak, że udział człowieka w całym procesie jest marginalizowany. Wręcz przeciwnie, wiedza eksperta jest kluczowym elementem łańcucha analizy, a głównym zadaniem takich systemów jest umożliwienie pracy na olbrzymich zbiorach danych. Oznacza to, iż całość informacji musi zostać przetworzona do spójnej i relatywnie prostej formy wizualnej, w której w łatwy sposób można połączyć ze sobą kluczowe elementy (podejrzani, zdarzenia itd.).

Artykuł opisuje system informatyczny zaprojektowany w celu ułatwienia analizy złożonych procesów, w szczególności tych związanych z działaniami operacyjnymi i dochodzeniowymi służb bezpieczeństwa wewnętrznego. Obejmuje on najważniejsze zadania i typowe etapy analizy kryminalnej takie jak wstępna obróbka i filtrowanie danych, ich graficzna reprezentacja, wizualizacja na mapach oraz wykorzystanie narzędzi analizy przestrzennej.

Analiza kryminalna

Analiza kryminalna [4] jest skutecznie używana już od ponad 25 lat, a obecnie wykorzystują ją zarówno międzynarodowe organy ścigania, jak i instytucje sektora prywatnego. Pośród organizacji zatrudniających w swoich szeregach analityków kryminalnych znajdują się INTERPOL, Europol, a także służby policyjne.

Istnieje wiele definicji tego, czym właściwie jest analiza kryminalna. Definicja, którą przyjęło 12 europejskich państw zrzeszonych w organizacji INTERPOL i która coraz częściej jest używana również przez inne kraje, brzmi:

„Identyfikacja i badanie związków pomiędzy danymi kryminalnymi, a także innymi potencjalnie istotnymi danymi z myślą o wykorzystaniu ich przez policję i w praktyce sądowniczej".

Głównym zadaniem analizy kryminalnej jest pomoc urzędnikom (stróżom prawa, politykom, decydentom) w lepszym radzeniu sobie z niepewnością, co oznacza, że odgrywa jedynie rolę pomocniczą w całym procesie dochodzeniowym.

Analiza jest zazwyczaj robiona na prośbę oficera Policji lub prokuratora, który określa szczegółowy cel analizy i przekazuje analitykowi wszystkie potencjalnie przydatne materiały (dane). Pierwszym krokiem wykonywanym przez analityka jest zwykle import zbioru danych (np. plików z bilingami telefonicznymi, uzyskanymi od operatorów) do bazy danych. Następnie można przejść do procesu wła-

ściwej analizy, która zazwyczaj składa się z następujących etapów, które mogą być powtarzane wielokrotnie (poprzez utworzenie pętli):

- operacje na danych – mogą być wykonywane zarówno przez różne algorytmy analityczne (filtrowanie, usuwanie duplikatów, obliczanie statystyk, łączenie lub dekompozycja źródeł danych), jak i przez osobę je analizującą (opisywanie danych na wykresie),
- wysuwanie hipotez (na podstawie wyników poprzedniego kroku),
- testowanie hipotez.

W powyższym (iteracyjnym) procesie analityk może zażądać dodatkowych danych. Efektem końcowym analizy jest zwykle raport (w formie wydruku), który jest następnie umieszczany w dokumentacji związanej z dochodzeniem.

Zintegrowane środowisko wspomagania analizy kryminalnej LINK[2] – wybrane aspekty funkcjonalne i architektoniczne

Środowisko LINK [3],[4] stanowi kompleksowe rozwiązanie informatyczne przeznaczone do wspierania pracy analityków kryminalnych. Podstawowa wersja systemu udostępnia zestaw narzędzi do integracji, wstępnego przetwarzania oraz wizualizacji danych pochodzących z różnych źródeł. LINK stanowi rozszerzalną platformę opartą o komponentową architekturę (tzw. plug-in architecture [7]) co daje szerokie możliwości integracji różnych metod (pół)automatycznej analizy danych, które mogą zostać dołączone do systemu w postaci niezależnie wytworzonych komponentów. W ten sposób uzyskujemy bardzo duże możliwości rozszerzania funkcjonalności systemu i dostosowywania jego możliwości do potrzeb konkretnych jednostek prowadzących prace z zakresu analizy kryminalnej.

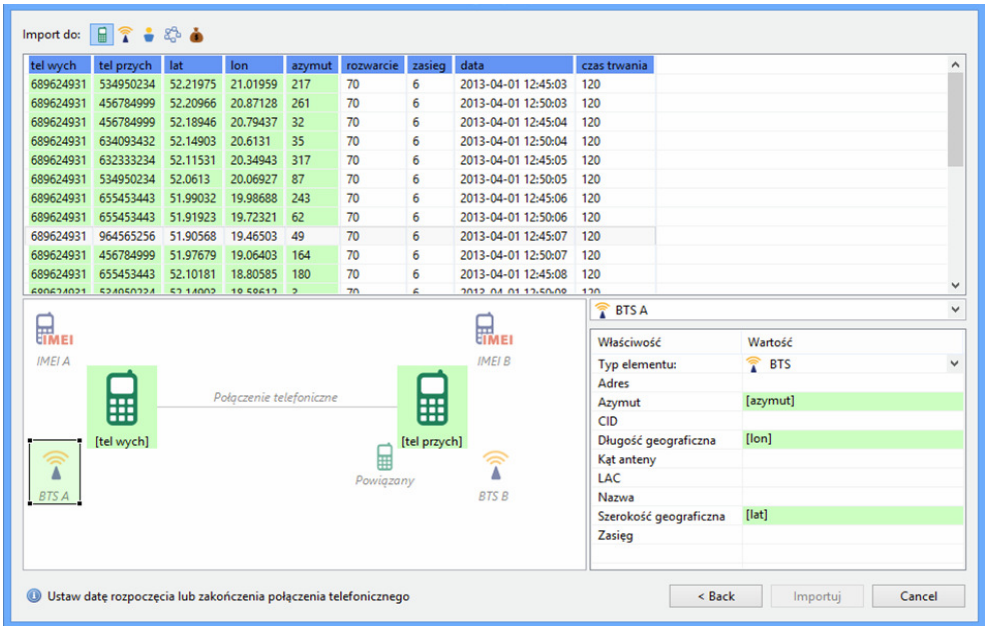
Import danych

Pierwszym krokiem w procesie analizy danych w środowisku LINK jest import danych. Na podstawie doświadczeń oraz opinii analityków kryminalnych z pewnością można stwierdzić, że nie istnieje ustandaryzowana forma danych podlegających analizie. Wynika to m.in. z braku uregulowań prawnych określających poszczególne formaty danych oraz z dużej różnorodności danych podlegających analizie. Dane te dostarczane są do analityków nie tylko w postaci elektronicznej (np. pliki tekstowe, RTF, MS Excel, MS Word), ale również w formie papierowej, co wymaga podjęcia dodatkowych kroków w celu ich digitalizacji.

W odpowiedzi na taką sytuację istotnym i rozbudowanym elementem środowiska LINK jest zestaw zaawansowanych narzędzi wspomagających import danych. Narzędzie te przyjmują formę graficznego kreatora (rys. 1), który prowadzi użytkownika krok po kroku przez cały proces. Podczas importu sprawdzana jest poprawność

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

i kompletność danych w kontekście wybranych dziedzin. Na przykład dla bilingów telefonicznych rekordy reprezentujące połączenie telefoniczne sprawdzane są pod kątem wystąpienia numerów telefonicznych oraz dat wykonania połączenia. W przypadku transakcji finansowych sprawdzana jest na przykład poprawność kont bankowych. W każdym kroku procesu importu użytkownik ma podgląd na dane źródłowe oraz sposób ich interpretacji przez środowisko LINK. Dotyczy to w szczególności formatów czasu i daty, mapowania wartości z pliku źródłowego na wartości predefiniowane przez program, itp. Ze względu na bezpieczeństwo oraz pełną rozliczalność wszystkie operacje i modyfikacje danych dokonane przez użytkownika są zapisywane w dzienniku zdarzeń dostępnym dla każdego procesu importu.



Rys. 1. Graficzny kreator importu danych

Jedną z głównych zalet środowiska LINK jest elastyczność kreatora importu, który dostarczany jest w dwóch odsłonach:

- w uproszczonej wersji przeznaczonej do importu danych tabelarycznych (np. pliki CSV, XLS), który przypomina w swojej istocie narzędzia importu w arkuszach kalkulacyjnych,
- w zaawansowanej, ogólnej wersji przeznaczonej do importu danych w dowolnej formie tekstowej, gdzie niekoniecznie istnieje podział na wiersze i kolumny (przykładem takich plików są często wykazy transakcji bankowych).

Obie wersje kreatora importu pozwalają użytkownikowi na wybranie formatu i zmapowanie każdej porcji danych źródłowych na obiekty i ich atrybuty należące do adekwatnej dziedziny (np. wskazanie daty początku lub końca rozmowy telefonicznej, numeru bankowego właściciela konta). Co więcej, poza prostymi narzędziami do mapowania atrybutów, kreator dostarcza zaawansowane metody ekstrakcji i przypisywania semantyki porcjom danych, które powszechnie występują w danych otrzymywanych przez polskich analityków kryminalnych.

Ze względu na elastyczność oraz bogate możliwości kreatora importu, jego konfiguracja w nietrywialnych przypadkach może być czasochłonna. Aby uniknąć ciągłej konfiguracji kreatora, środowisko LINK dostarcza również tzw. mechanizm szablonów importu, który zapisuje konfigurację raz wykonanego procesu importu, a następnie pozwala ją zastosować do innych plików w tym samym lub podobnym formacie. Szablony importu zapisywane są w przestrzeni roboczej użytkownika w postaci plików XML, dzięki czemu mogą być łatwo współdzielone przez wielu użytkowników (np. za pomocą poczty elektronicznej lub centralnego repozytorium szablonów).

Przetwarzanie i analiza danych

W uproszczeniu analizę kryminalną można przedstawić jako zaawansowaną, kontekstową eksplorację danych. Celem procesu analizy jest stawianie i weryfikacja hipotez, co odbywa się zazwyczaj poprzez odkrywanie istotnych relacji pomiędzy obiektami w dużej ilości powiązanych informacji. W procesie tym pomocne jest wykorzystanie wielu różnorakich algorytmów i heurystyk, które można skategoryzować m.in. jako klasteryzacja grafów [9], analiza sieci społecznościowych oraz odkrywanie i wyszukiwanie wzorców [11]. Większość tych operacji może zostać zaimplementowana w postaci (pół)automatycznych narzędzi wspomagających analityków. Wybrane narzędzia analityczne zostały zrealizowane jako komponenty środowiska LINK, pozostawiając pole rozwoju na inne, przydatne metody w tym obszarze.

Podstawowa wersja środowiska LINK dostarcza niezbędny zestaw operatorów analitycznych przeznaczonych do ekstrakcji istotnych informacji z dużych ilości danych w wybranych dziedzinach. Przykładem takich narzędzi są dziedzinowe mechanizmy filtrowania, która dostowują sposoby filtracji do typów poszczególnych atrybutów (kolumn), dając np. możliwość wyboru rekordów ze względu na datę i czas wystąpienia zdarzenia lub rekordów pasujących do wyrażeń regularnych (rys. 2).

sc-01-era.csv_2009.05.06-10:47:31.0804

Source	Target	Date	Duration [seconds]	Service type
400679	400464	01/08/05 15:0	0	OTHER
400526450	400464200	01/08/05 18:33:00	86	CALL
400464200	400546250	03/18/05 13:04:00	56	CALL
400464200	400123842	03/18/05 13:04:00	56	CALL
400662138	400464200	03/18/05 14:17:00	46	CALL
400464200	400662138	03/18/05 14:37:00	40	CALL
400464200	400662138	03/18/05 17:48:00	1	CALL
400464200	400662138	03/19/05 12:31:00	0	OTHER
400662138	400464200	03/19/05 15:19:00	20	CALL
400679350	400464200	03/19/05 21:16:00	0	SMS
400679350	400464200	03/26/05 11:42:00	0	SMS
400464200	400679350	03/26/05 14:07:00	18	CALL
400679350	400464200	03/26/05 16:55:00	0	SMS
400679350	400464200	03/27/05 08:55:00	0	SMS
400464200	400679350	03/29/05 10:24:00	12	CALL
400679350	400464200	03/29/05 11:14:00	56	CALL
400679350	400464200	03/29/05 16:10:00	0	SMS
400679350	400464200	03/31/05 20:48:00	0	SMS
400464200	400679350	04/06/05 20:14:00	37	CALL
400464200	400679350	04/06/05 20:15:00	14	CALL
400464200	400679350	04/06/05 20:20:00	44	CALL
400679350	400464200	04/07/05 07:04:00	0	SMS
400679350	400464200	04/07/05 09:25:00	11	CALL

Rys. 2. Mechanizmy filtrowania w edytorze tabelarycznym

Narzędzia te pozwalają na szybkie odsianie nieistotnych informacji i skupienie uwagi analityka na istotnych (z pewnego punktu widzenia) danych. Użytkownik może również obejrzeć statystyki i zestawienia, które ukazują zbiorczą informację o analizowanych zbiorach danych. Przykładem takich statystyk jest liczba unikalnych wartości danego atrybutu, suma lub średnia, wartości minimalne i maksymalne wartości liczbowye oraz zaawansowane zestawienia dedykowane wybranym dziedzinom, np. lista najczęściej komunikujących się obiektów, ich aktywność i profile czasowe.

Ponadto środowisko LINK posiada wbudowane operacje umożliwiające przetwarzania danych, które często dostosowane są do poszczególnych dziedzin danych. Przykładem takich operacji dla bilingów telefonicznych są:

- normalizacja numerów telefonów poprzez detekcję i usunięcie prefiksów i infiksów,
- łączenie wielu bilingowych zbiorów danych
- detekcja i usunięcie tzw. dubli, czyli tych samych połączeń telefonicznych, które często rejestrowane są przez różnych operatorów z przesunięciem czasowym.

Dzięki rozszerzalnej architekturze środowiska LINK wzbogacenie środowiska o całkiem nowe operatory i narzędzia analityczne stanowi naturalny kierunek rozwoju środowiska. Przykładem takich operacji jest m.in. wyszukiwanie ścieżek na grafie czy wyszukiwanie wzorców częstych.

Wizualizacja danych

Kolejną cechą środowiska LINK jest wizualizacja danych dostosowana do potrzeb analizy kryminalnej. Metody wizualizacji pozwalają na odkrywanie i edycję danych co wpisuje się w trend wizualnej eksploracji danych [2].

Środowiska LINK dostarcza graficzne edytory ukazujące różne perspektywy danych oraz umożliwiające modyfikacje istniejących i dodawanie nowych obiektów. Dzięki temu mogą być wykorzystane również jako graficzne notatniki wspomagające analizę spraw oraz umożliwiają przygotowanie końcowych raportów przeznaczonych do dalszego postępowania.

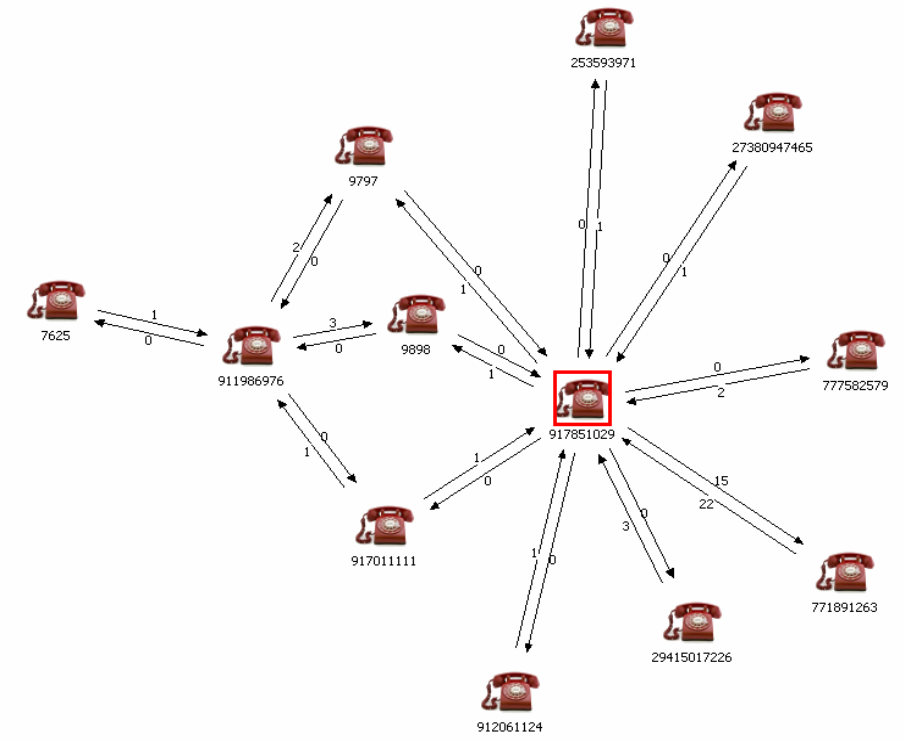
W podstawowej wersji środowiska wbudowane są dwa graficzne edytory:

- edytor schematyczny prezentujący dane na grafie,
- edytor czasowy (w wersji prototypowej), ukazujący dane na osi czasu.

Edytor schematyczny dostarcza szereg narzędzi wspomagających wizualizację obiektów (jako węzłów grafu) oraz powiązań między nimi (jako krawędzi grafu). W szczególności należy zwrócić uwagę na bogaty zestaw zasobów graficznych reprezentujących różne typy obiektów (osoby, telefony, miejsca, itp.), jak również wbudowane metody rozkładu węzłów, pozwalające na przejrzyste ułożenie elementów grafu, w zależności od jego charakterystyki i potrzeb analityka. Do najważniejszych algorytmów rozkładu należą układy znajdujące się w bibliotece GraphViz:

- rozkład „pawi ogon” [10] – główne węzły ułożone są w centrach, pozostałe, powiązane z nimi, rozkładane są naokoło, tworząc tzw. „pawi ogon” (rys. 3),
- rozkład kołowy – węzły ułożone są na okręgach o promieniach odzwierciedlających stopień powiązania z węzłem głównym,
- rozkład hierarchiczny – węzły ułożone są w horyzontalnych liniach, które reprezentują hierarchię występującą w grafie.

Poza tym edytor schematyczny wyposażony jest w zestaw narzędzi wspomagających warunkowe zaznaczanie elementów (np. zaznaczanie węzłów-liści, selekcja na podstawie zakresu wartości wybranych atrybutów węzłów lub krawędzi), rozpoznawanie i scalanie podobnych węzłów oraz tworzenie raportów (m.in. notatki, legenda, edycja wizualnych właściwości elementów grafu).



Rys. 3. Rozkład „pawi ogon" dla przykładowego diagramu schematycznego

Z kolei prototyp edytora czasowego ukazuje chronologię zdarzeń na osi czasu oraz pokazuje relacje między obiektami (np. spotkanie, połączenie telefoniczne, transakcja finansowa). Diagramy czasowe składają się z osi czasu (każda dla jednego obiektu), zdarzeń reprezentowanych przez notatki osadzone na osi czasu danego obiektu oraz relacji przedstawianych jako linie łączące osie czasu obiektów, których dotyczy dana relacja. Diagram czasowy wzbogacony jest o linię czasu prezentującą kontekst czasowy (zakresy daty lub czasu). Taki sposób wizualizacji danych jest szczególnie użyteczny w analizie sekwencji zdarzeń oraz analizie korelacji pomiędzy zdarzeniami dotyczącymi różnych obiektów (co jest typowe np. dla analiz w obszarze badania przestępstw gospodarczych).

Śledzenie operacji

Zazwyczaj proces analizy kryminalnej dotyczący jednej sprawy jest skomplikowany i złożony z wielu kroków. Liczne transformacje danych (np. łączenie, filtrowanie, podział zbiorów) mogą prowadzić do trudności z późniejszym przedstawieniem wykonanych kroków analitycznych oraz wskazaniem pochodzenia obiektów kluczowych ze względu na rezultaty analizy. Jest to jednak niezmiernie

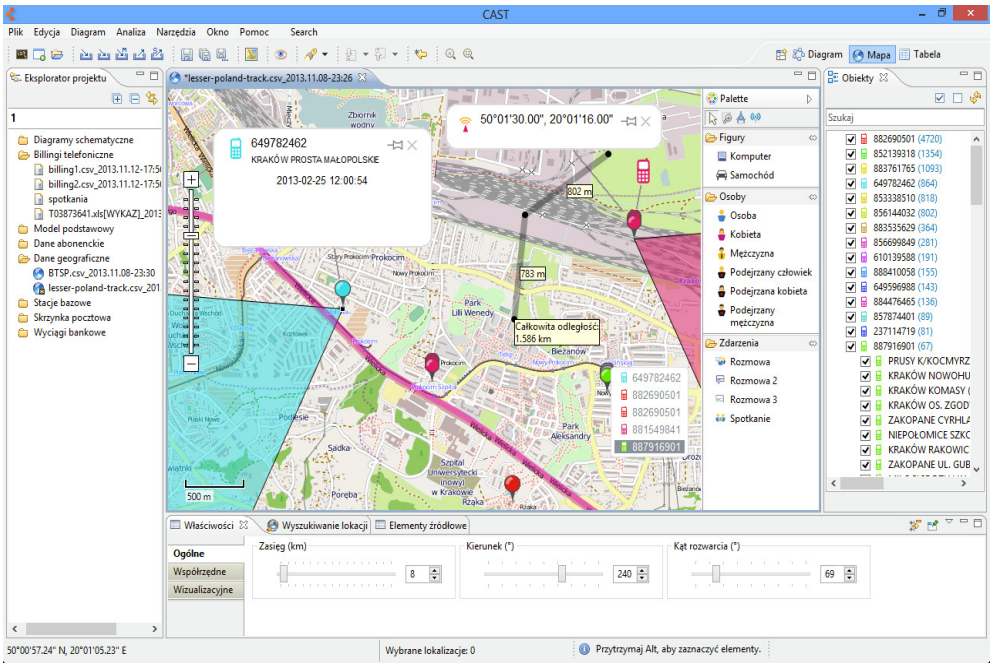
istotne w procesie dowodowym ze względu na wiarygodność przedstawionych wniosków, dlatego raporty analityków zawierają często opis wszystkich kroków, które doprowadziły ich do postawienia wniosków. Aby ułatwić im zadania, środowisko LINK dostarcza mechanizmy śledzenia wykonanych operacji, począwszy od importu danych, poprzez operacje analityczne, aż do wizualizacji danych i przygotowania diagramów. Historia operacji wyświetlana jest w formie grafu przetwarzania, który ukazuje każdy zbiór danych jako węzeł, a każda operacja przedstawiona jest jako skierowana krawędź grafu, transformująca jeden lub wiele zbiorów w zbiór wynikowy.

Edytor mapowy programu LINK

Ocena standardowych narzędzi wizualizacji dostarczanych przez oprogramowanie LINK, w postaci edytora schematycznego i chronologicznego, wygenerowała potrzebę indywidualnego podejścia do wizualizacji danych geograficznych. Wiele spraw prowadzonych przez analityków kryminalnych wymaga przetwarzania danych ściśle związanych z kontekstem geograficznym.

Billingi telefoniczne dostarczają (pośród innych danych) informacji o lokalizacji i zasięgu stacji bazowych telefonii komórkowej, użytych do zestawienia połączenia. Również wyciągi bankowe stanowią źródło informacji na temat adresów bankomatów i punktów handlowo-usługowych, w których dokonywane były transakcje kartami płatniczymi. Wizualizacja tych informacji jest często kluczowa dla zrozumienia całości analizowanej sprawy i poparcia stawianej hipotezy. Aby osiągnąć te cele, analitycy często stosują specjalistyczne oprogramowanie GIS [5], [6].

W celu umożliwienia ciągłości analiz aplikacja LINK dostarcza dedykowanego edytora mapowego (rys. 4). LINK posiada modularną architekturę bazującą na systemie Eclipse (ang. Eclipse Plugin Architecture)[1], co pozwoliło na realizację edytora mapowego w postaci rozszerzenia środowiska. Edytor stworzony jest z myślą o zwiększeniu wydajności analiz prowadzonych w oparciu o dane geograficzne. Większość podstawowych operacji, wykonywanych na tych danych podczas analiz kryminalnych, ma charakter powtarzalny i daje się sprowadzić do kilku uproszczonych funkcji. Biorąc pod uwagę fakt, że większość analiz wykonywanych na co dzień nie wykracza poza te funkcjonalności, proste narzędzia dla najbardziej typowych analiz i sposobów wizualizacji została zaimplementowana bezpośrednio w rozszerzeniu edytora mapowego. W przypadku złożonych analiz LINK dostarcza narzędzi pozwalających na integrację i współpracę ze specjalistycznym oprogramowaniem GIS, np. ESRI ArcGIS [5] (zagadnienie szerzej opisane w kolejnej sekcji). Główną zaletą takiego podejścia jest możliwość wykorzystania potencjału, jaki daje użycie dedykowanego do importu i manipulacji danymi przetwarzanymi w analizach kryminalnych programu LINK, do wstępnego przygotowania danych geograficznych na potrzeby prowadzenia dalszych, bardziej zaawansowanych analiz GIS.



Rys. 4. Edytor mapowy LINK

Dane geograficzne

Dane przestrzenne (geograficzne) przechowywane są niezależnie od pozostałych danych programu. System, w zależności od potrzeb, umożliwia przechowywanie tych danych przy wykorzystaniu standardowego dla systemu LINK mechanizmu wykorzystującego plikową bazę SQLite, zoptymalizowanego na operacje na dużych zbiorach danych przestrzennych mechanizmu, wykorzystującego serializację danych przestrzennych do formatu JSON i przechowującego dane w formie skompresowanych plików oraz mechanizmu wykorzystującego bazę danych PostgreSQL z rozszerzeniem przestrzennym PostGIS. Zastosowanie bazy PosGIS znacznie ułatwia integrację środowiska LINK ze specjalistycznym oprogramowaniem GIS. Ponadto, dzięki zestawowi wbudowanych funkcji dostępnych z poziomu języka zapytań SQL, umożliwia ono także łatwiejszą i efektywniejszą implementację operacji manipulujących danymi geograficznymi.

Jedną z najważniejszych cech każdego edytora mapowego jest wsparcie dla obsługi rastrowych podkładów mapowych. Oprogramowanie LINK domyślnie wykorzystuje w tym celu publicznie dostępne dane mapowe, udostępniane przez serwis OpenStreetMap. Możliwa jest także konfiguracja, jako źródła podkładu mapowego, dowolnego serwera TMS (ang. Tiled Map Server). Biorąc pod uwagę prawdziwe potrzeby polskich analityków kryminalnych, należy również zdać sobie sprawę z faktu, że często zmuszeni są oni (ze względów bezpieczeństwa) pracować na stacjach roboczych celowo pozbawionych dostępu do sieci Internet. Z tego powodu

aplikacja pozwala nie tylko na wykorzystanie danych dostępnych zdalnie, ale również przygotowanych odpowiednio map podkładowych przechowywanych na lokalnych nośnikach.

Narzędzia dostępne w edytorze mapowym

Aktualnie edytor mapowy dostarcza narzędzi przeznaczonych do:

- wizualizacji wcześniej zebranych danych,
- manualnego przygotowania danych,
- przeszukiwania i filtrowania obiektów.

Podstawowe modele danych dostępne w edytorze to punkty, trasy i zasięgi (takie jak np. zasięgi BTS – stacji bazowych sieci telefonii komórkowej). Możliwe jest również powiązanie z modelem mapowym obiektów reprezentujących metadane. Należą do nich pomiary odległości, elementy wskazujące na kolejność zdarzeń, notatki czy legenda mapy, pozwalająca na dodanie opisów do klas obiektów użytych na mapie.

Oprócz możliwości pozwalających na import podstawowych danych geograficznych z heterogenicznych źródeł, wypracowane zostały wyspecjalizowane metody wizualizacji dostosowane do natury danych pochodzących z billingów telefonicznych i wykazów stacji bazowych telefonii komórkowej. Wprowadzony został specjalny typ obiektu reprezentujący stacje bazowe powiązane ze zdarzeniami odpowiadającymi połączeniom. Dla obiektów zawierających dane na temat zasięgu stacji bazowej możliwa jest automatyczna generacja powiązanych obiektów wizualizujących zasięg oraz ich selektywne pokazywanie i ukrywanie. Opracowane zostało również narzędzie do transformacji modelu agregującego zdarzenia logowania i dane stacji bazowych w model składający się z obiektów reprezentujących stacje bazowe i zależnych obiektów odpowiadających logowaniom.

Przeszukiwanie i filtrowanie danych możliwe jest dzięki dwóm stworzonym na te potrzeby narzędziom – oknom dialogowym „Filtruj elementy po datach" i widokowi „Obiekty". Okna dialogowe pozwalają użytkownikowi na filtrowanie widocznych obiektów względem czasu wystąpienia połączonych z nimi zdarzeń. Umożliwia to nie tylko wybranie konkretnego okna czasowego, z którego obiekty mają zostać odfiltrowane, ale także na śledzenie następstw czasowych pomiędzy zdarzeniami. Uzupełnieniem tej funkcjonalności jest widok „Obiekty", który pozwala na filtrowanie obiektów według ich właściwości dziedzinowych, takich jak numer telefonu czy adres stacji bazowej.

Warto również zauważyć, że w początkowej fazie rozwoju edytor mapowy był dużo bardziej rozbudowanym komponentem i dostarczał wielu funkcjonalności pozwalających na wykonywanie innych podstawowych operacji na danych mapowych. Aktualnie zaprzestano utrzymywania dużej części z nich na rzecz lepszego wsparcia dla narzędzi najczęściej wykorzystywanych w prowadzeniu analiz.

Pozwoliło to w znaczny sposób uprościć interfejs edytora i skrócić czas potrzebny na naukę obsługi oprogramowania. Zaawansowane analizy możliwe są poprzez integrację ze specjalistycznym oprogramowaniem analitycznym GIS.

Integracja z ArcGIS[3]

LINK oferuje pewne możliwości geoprzetwarzania, jednak często są one niewystarczające i zachodzi potrzeba skorzystania z bardziej zaawansowanego oprogramowania analitycznego takiego, jak ArcGIS.

ArcGIS to zestaw narzędzi programowych, opracowany przez firmę ESRI do tworzenia, przetwarzania oraz analizy danych przestrzennych i map. Narzędzia te dostępne są w wersjach dla komputerów osobistych, serwerów oraz urządzeń mobilnych. Oprócz prostych metod analitycznych, takich jak przeszukiwanie, filtrowanie danych i wizualizacja danych, ArcGIS pozwala na przetwarzanie łańcuchowe, za pomocą którego można łączyć ze sobą podstawowe operacje geoprzetwarzania i tworzyć złożone modele, jak np. analizy hot-spot, analizy sieciowe (wykorzystujące dane o sieci dróg), analizy wykorzystujące cyfrowy model terenu, szacowanie prawdopodobieństw zdarzeń, analizy prognostyczne itp. ArcGIS nie jest jedynym produktem na rynku, który oferuje zaawansowane możliwości geoprzetwarzania. Konkurencyjne rozwiązania istnieją zarówno w sektorze aplikacji płatnych, jak i bezpłatnych, jednak ArcGIS wyróżnia się kilkoma cechami:

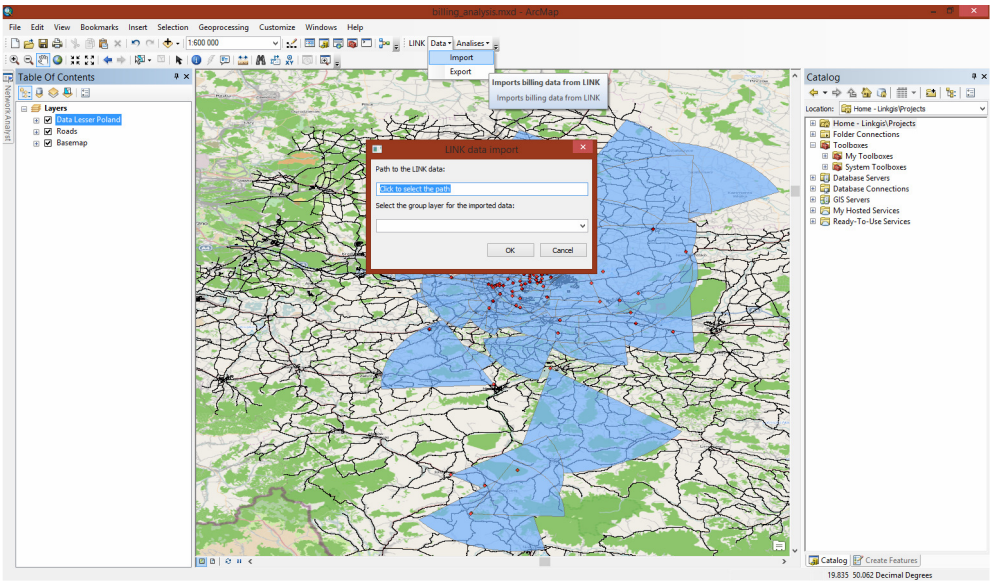
- wspieranie kompletnego łańcucha przetwarzania, od tworzenia i przechowywania danych w bazach danych do wyświetlania rezultatów analiz,
- duża liczba różnorodnych funkcji do przetwarzania danych przestrzennych,
- obszerna i łatwa w użyciu dokumentacja,
- duża społeczność użytkowników i profesjonalne wsparcie techniczne.

Oprogramowanie ArcGIS jest tworzone zarówno dla deweloperów, jak i użytkowników końcowych. Oferuje ono API w językach Java i C#, bazujące na obiektach COM, za pomocą którego można tworzyć nowe aplikacje, jak i rozszerzać funkcjonalność tych stworzonych przez ESRI. Funkcje służące do geoprzetwarzania są również udostępniane przez wysokopoziomowe moduły języka Python i mogą być ze sobą łączone w skryptach i wtyczkach.

LINK i ArcGIS integrują się za pomocą pośredniczącej warstwy danych obsługiwanej przez wtyczki w obu systemach. Moduł mapowy w LINK-u został wyposażony w prosty algorytm eksportu do formatu shapefile (format danych przestrzennych rozwijany przez ESRI). Po stronie ArcGISa integracja odbywa się za pomocą modułu przygotowanego w języku Python. Umożliwia on import danych, a także przeprowadzanie często wykonywanych analiz, które są udostępniane za pomocą dedykowanego graficznego interfejsu użytkownika (rys. 5). Dane geograficzne eksportowane z LINK-u do plików musiały zostać uproszczone z uwagi na

[3] http://www.esri.com/software/arcgis.

pewne ograniczenia formatu shapefile, takie jak możliwość użycia tylko jednego typu danych przestrzennych dla danych przechowywanych w pliku, czy brak możliwości powiązania przechowywanych danych relacjami. Wymusiło to przekształcenie eksportowanych danych i ich uproszczenie do tego stopnia, że po przetworzeniu ich w ArcGISie utrudniony został ich import z powrotem do LINK-u.



Rys. 5. Okno wtyczki programu ArcMap służące do importu danych z LINK-u

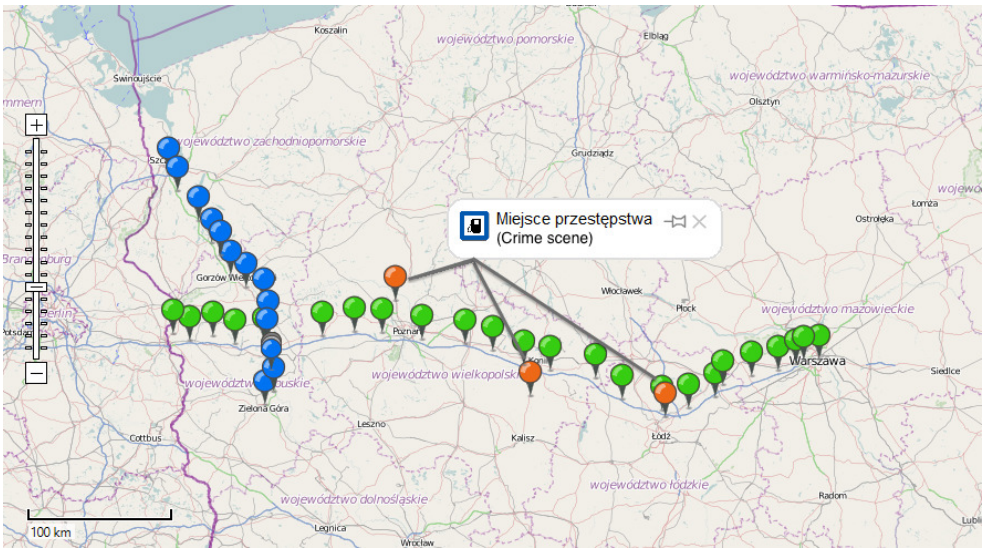
Aby pokonać te trudności i zapewnić ściślejszą integrację z ArcGISem, prowadzone są obecnie prace w kierunku wymiany danych poprzez bazę danych przestrzennych PostGIS. Pozwala to uniknąć utraty informacji w trakcie przenoszenia danych, a także wyklucza potrzebę ręcznego importu i eksportu danych oraz zapewnia możliwość równoczesnego dostępu do danych z poziomu systemów LINK i ArcGIS.

Wykorzystanie ArcObjects SDK for Java przy tworzenia wtyczki do LINK-u pozwala na pełną integrację pomiędzy systemami w jednym oknie aplikacji, co ułatwi pracę analityka, likwidując potrzebę przełączania się pomiędzy kontekstami dwóch aplikacji.

Studium przypadku – przykładowa analiza z użyciem danych geograficznych

Aby zilustrować wykorzystanie oprogramowania LINK i ArcGIS w prowadzeniu analiz kryminalnych z wykorzystaniem danych geograficznych, przeanalizowany zostanie prosty przykład.

Załóżmy następującą hipotetyczną sytuację: wystąpiła seria trzech napadów na stacje benzynowe, kamery przemysłowe zarejestrowały te same numery rejestracyjne pojazdu użytego przez sprawców, zatrzymany został właściciel samochodu, niestety nie znaleziono przy nim zrabowanych pieniędzy. Zatrzymany twierdził, że jego samochód został skradziony i nie ma on nic wspólnego z napadami. W międzyczasie zatrzymano również innego mężczyznę z dużą ilością pieniędzy niewiadomego pochodzenia, wartość znalezionej gotówki odpowiada skradzionej sumie, mężczyzna został zatrzymany w innej części Polski.



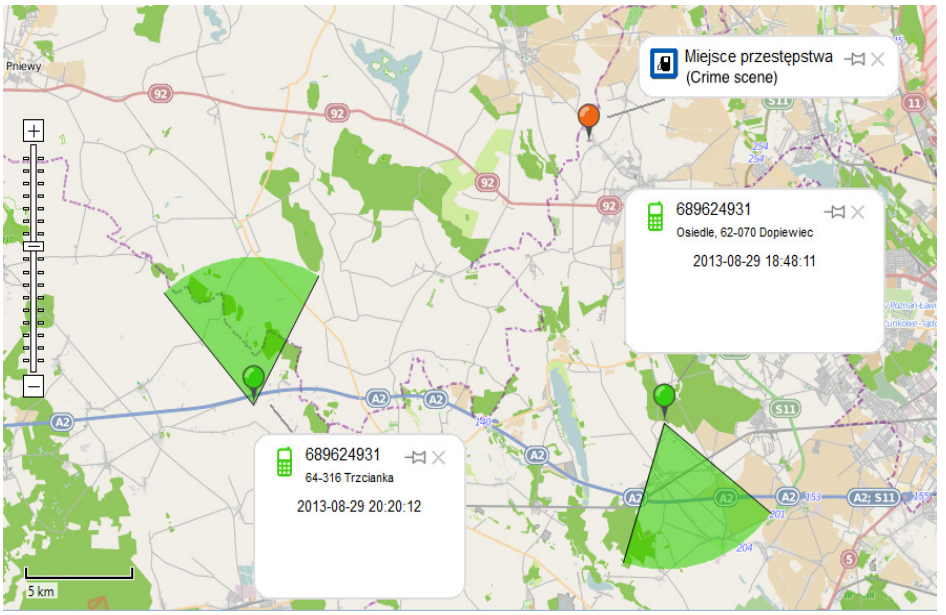
Rys. 6. Diagram logowań do stacji bazowych

Analityk kryminalny został poproszony o sprawdzenie, czy jeden z podejrzanych mógł przebywać w okolicy miejsc przestępstw. Operator sieci komórkowej dostarczył billingi dla telefonów komórkowych obu podejrzanych. Dane zostały zaimportowane do programu LINK a dane na temat stacji bazowych zostały automatycznie uzupełnione na podstawie danych dostępnych bazie UKE[4]. Następnie stworzono diagram mapowy zawierający położenia stacji bazowych, używanych do zestawienia połączeń przy użyciu telefonów podejrzanych (rys. 6).

[4] Urząd Komunikacji Elektronicznej http://uke.gov.pl.

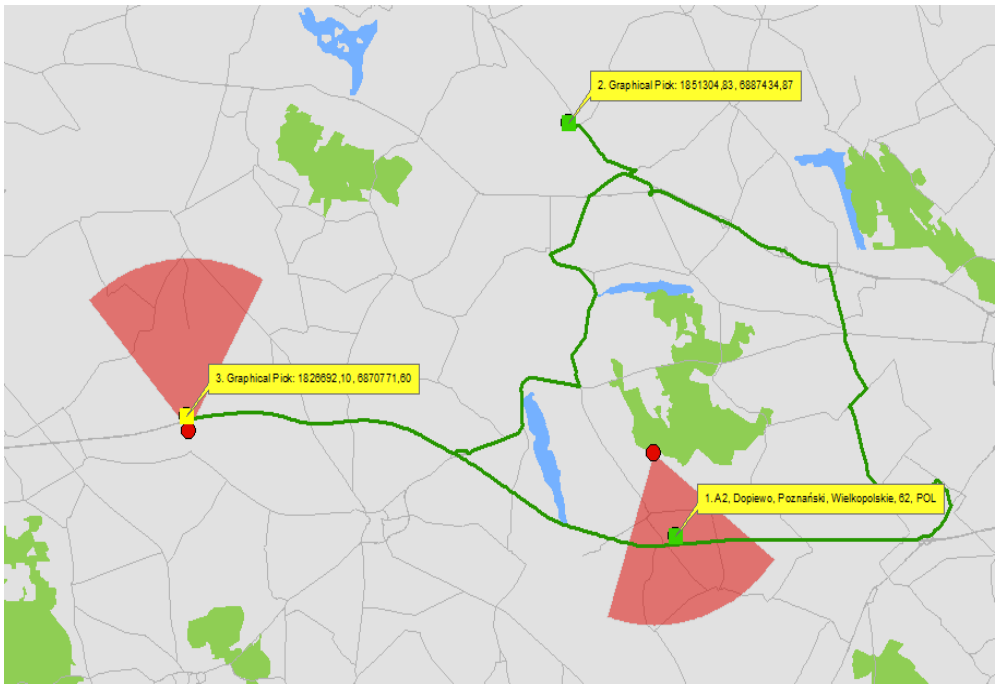
Diagram pokazał, że jedynie właściciel pojazdu wykonywał połączenia telefoniczne w okolicy miejsc przestępstw (stacje bazowe zaznaczane jako pozioma seria danych). Ponadto prawdopodobnie kłamał w sprawie kradzieży samochodu, gdyż nie zgłosił, że wraz z samochodem skradziony został jego telefon komórkowy. Z drugiej strony podejrzany zatrzymany z dużą ilością pieniędzy nie mógł znaleźć się w okolicy miejsc przestępstw (logowania do stacji bazowych oznaczonych jako pionowa seria danych.).

Jakkolwiek, aby wykluczyć wszystkie wątpliwości na temat możliwego obrabowania stacji benzynowych przez właściciela samochodu, analityk musi udowodnić, że podejrzany był w stanie pomiędzy wykonywanymi połączeniami dotrzeć do miejsc przestępstw, oraz że miał tam wystarczająco dużo czasu, aby dokonać napadów (rys. 7). Niestety ta analiza wykracza poza możliwości obecnie dostępne w programie LINK. Możliwy jest jednak eksport danych mapowych do specjalistycznego oprogramowania GIS, jak np. ArcGIS i kontynuowanie analizy (rys. 8).

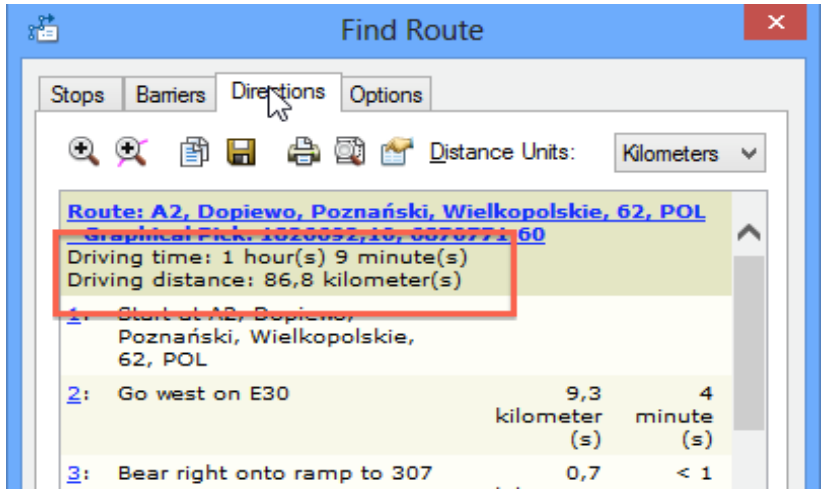


Rys. 7. Czasy logowań do stacji bazowych i ich odległość od miejsca przestępstwa

ArcGIS pozwala na przeprowadzenie analizy sieciowej i obliczenie prawdopodobnej trasy przejazdu, którą mógł poruszać się podejrzany. Tak obliczony czas można następnie porównać z różnicą w czasie pomiędzy logowaniami do najbliższych stacji bazowych. Obliczony czas potrzebny na przemieszczenie się pomiędzy zadanymi punktami wyniósł 1 godzinę i 9 minut (rys. 9), natomiast rzeczywista różnica w czasie pomiędzy logowaniami (rys. 7), to 1 godzina 32 minuty. Z powyższego obliczenia wynika jasno, że podejrzany nie tylko mógł dotrzeć do miejsca przestępstwa, ale także prawdopodobnie zostały mu 23 minuty na dokonanie rabunku.

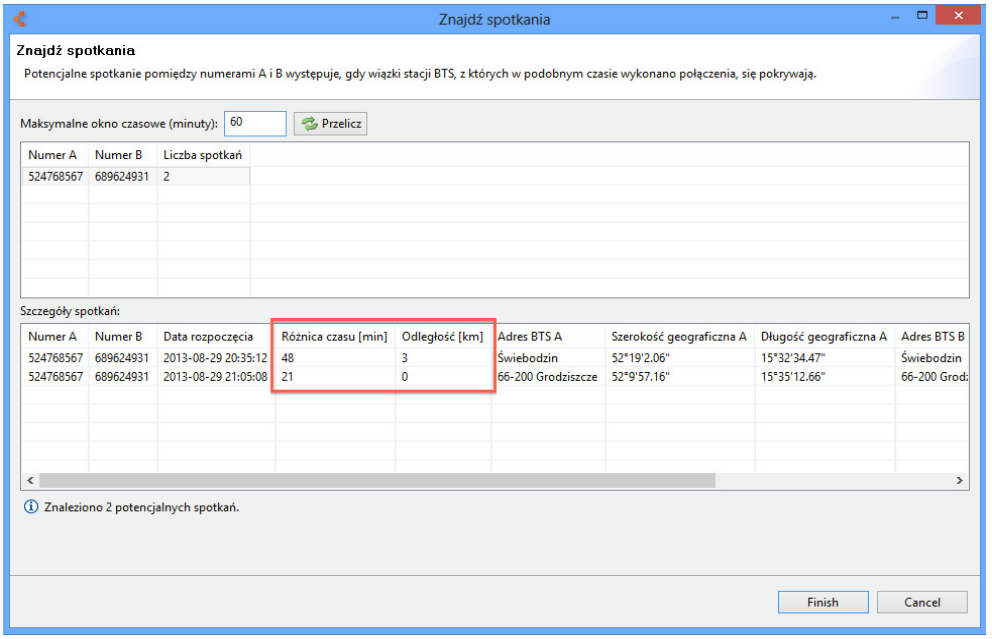


Rys. 8. Dane z diagramu mapowego LINK wyeksportowane do programu ArcGIS; wizualizacja analizy sieciowej



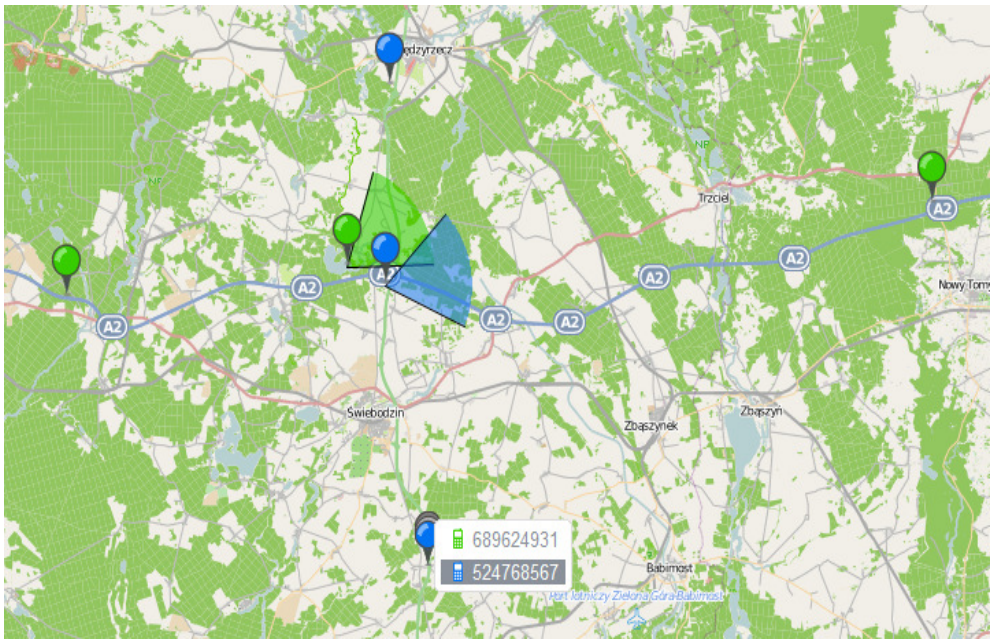
Rys. 9. Wyliczony czas przejazdu

Po wykazaniu, że seria napadów mogła prawdopodobnie być przeprowadzona przez właściciela samochodu, analitykowi zadana została kolejna hipoteza do weryfikacji: „Osoba zatrzymana z pieniędzmi mogła spotkać się z właścicielem pojazdu i odebrać łup".

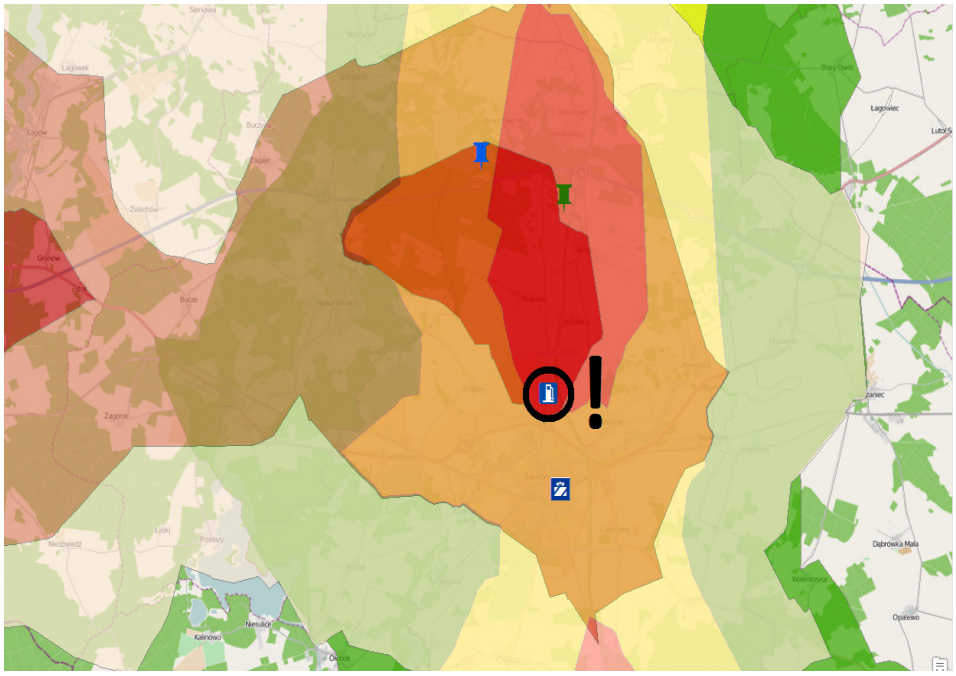


Rys. 10. Narzędzie wyszukiwania spotkań

Program LINK wyposażony został w łatwe w użyciu narzędzie analityczne do wyszukiwania potencjalnych spotkań. Dla rozpatrywanego przypadku przy jego pomocy, z użyciem 60-minutowego okna czasowego, znaleziono dwa prawdopodobne miejsca spotkań podejrzanych (rys. 10). Pozycje stacji bazowych użytych do wykonania połączeń w okolicy miejsca prawdopodobnego miejsca spotkania mogą zostać zwizualizowane na diagramie mapowym (rys. 11). Dla określenia dokładniejszego przybliżenia dane te mogą po raz kolejny zostać wyeksportowane do zewnętrznej aplikacji ArcGIS. Następnie dzięki analizie sieciowej stworzony może zostać diagram obszarów możliwych do osiągnięcia w ciągu kolejnych następujących po sobie interwałów czasowych od logowania do stacji bazowej. W tym przypadku interwał został ustawiony na pięć minut (rys. 12).



Rys. 11. Diagram mapowy ilustrujący logowania do stacji bazowych w okolicy prawdopodobnego miejsca spotkania



Rys. 12. Diagram obszarów możliwych do osiągnięcia w ciągu kolejnych pięciominutowych interwałów od logowania do stacji bazowej

Analizując wynikowy diagram, uzupełniony o dodatkowe informacje na temat punktów POI (ang. Point of Interest), takich jak sklepy czy stacje benzynowe, można zauważyć, że najbardziej prawdopodobnymi lokalizacjami, w których mogło dojść do ewentualnego spotkania i przekazania łupu są stacja benzynowa znajdująca się w obszarze możliwym do osiągnięcia w ciągu pięciu minut przez obydwóch podejrzanych oraz sklep zawierający się w obszarach możliwych do osiągnięcia w ciągu 10 minut. Bazując na tej informacji, może zostać podjęta decyzja o konieczności zabezpieczenia i sprawdzenia nagrań monitoringu tych obiektów oraz przesłuchania ich pracowników.

Analizowany przypadek jest oczywiście znacznie uproszczony i spreparowany na potrzeby przykładu. Pokazuje on jednak możliwości, jakie daje analiza danych geograficznych przy użyciu oprogramowania LINK i ArcGIS.

Podsumowanie oraz dalsze kierunki badań

Analiza kryminalna wspomagana narzędziami analizy przestrzennej to nowy, obiecujący kierunek. Powoli zaczyna on być dostrzegany, ponieważ otwiera przed analitykami zupełnie nowe możliwości. Z jego pomocą można przewidywać zachowania podejrzanych w konkretnym środowisku, a także wspomagać koordynację działań operacyjnych. LINK jest jednym z niewielu narzędzi koncentrujących się na właśnie takim przeznaczeniu. Jego wielką zaletą jest to, iż tworzony jest do pracy na krajowych billingach i mapach. Dzięki temu czynności takie jak import danych billingowych czy przeprowadzanie na nich analiz (szczególnie tych przestrzennych), które są zazwyczaj długotrwałe i żmudne, stają się łatwe i wygodne. Te i inne zadania mogą być w prosty sposób zautomatyzowane przy użyciu narzędzi LINK. Dalsze badania skoncentrują się na wprowadzaniu nowych analiz przestrzennych (które mogłyby poprawić prawdopodobieństwo znalezienia podejrzanych lub dowodów) oraz na tworzeniu wygodnego interfejsu użytkownika, który umożliwiałby przeprowadzanie skomplikowanych operacji w prosty sposób.

Bibliografia

[1] A. Bolour. (2003) Notes on the eclipse plug-in architecture. [Online]. Available: http://www.eclipse.org/articles/Article-Plug-in-architecture/plugin_architecture.html.

[2] K. C. Cox, S. G. Eick, G. J. Wills, and R. J. Brachman, “Visual data mining: Recognizing telephone calling fraud,” Data Mining and Knowledge Discovery, vol. 1, no. 2, pp. 225–231, 1997.

[3] J. Dajda, R. Dębski, M. Kisiel-Dorohinicki, K. Piętak, Multi-Domain Data Integration for Criminal Intelligence, in: Man-Machine Interactions 3, Advances in Intelligent Systems and Computing, vol 242, pp-345–352, Springer, 2014.

[4] R. Dębski, M. Kisiel-Dorohinicki, T. Miłoś, K. Piętak, „LINK – a decision-support system for criminal analysis" in MCSS 2010 : Multimedia Communications, Services and Security : IEEE International Conference, pp 110–116 : Kraków, 2010.
[5] W.L. Gorr, K. S. Kurland. GIS Tutorial for Crime Analysis. ISBN 978-1-58948-214-2, ESRI Press, 2010.
[6] R. Hanning, Spatial data Analysis, Theory and Practise, ISBN 0-521-77319-9, Cambridge Press, 2003.
[7] C. M. Judd and H. Shittu, Eclipse Plug-in Paradigm. Apress, 2005, ch. 2, pp. 11–18.
[8] J. McAffer and J.M. Lemieux, Eds., Eclipse Rich Client Platform: Designing, Coding, and Packaging Java(TM) Applications. Addison-Wesley Professional, 2005.
[9] S. E. Schaeffer, "Graph clustering," Computer Science Review, vol. 1, no. 1, pp. 27 – 64, 2007.
[10] G. J. Wills, Network interactive visualization of very large graphs, Lecture Notes in Computer Science, vol 1353, pp 403–414, Springer, 1997.
[11] S. Theodoridis and K. Koutroumbas, Pattern Recognition. ISBN-10: 1597492728 Academic Press, 2008.