Fraud Mapping and GIS Technologies

11/9/2010

CSCS5980

Amy Casuscelli and Steve Hawkins

Contents

[Introduction 2](#_Toc277007513)

[Current Applications 2](#_Toc277007514)

[Further Development 4](#_Toc277007515)

[Scenarios for Use 5](#_Toc277007520)

[Challenges Ahead 7](#_Toc277007521)

[Success to date 8](#_Toc277007522)

[Conclusion 8](#_Toc277007523)

[Glossary 8](#_Toc277007524)

[Bibliography 9](#_Toc277007525)

# Introduction

We hope to address the potential application of GIS systems in fraud management systems for authentication.  Most fraud analysis has a strong temporal component and we would like to understand how this could be displayed with maps.  We will be bringing together Oracle products used in authentication of a user to a website and Quova and geo intelligence database that contains among other attributes the coordinates of visitors IP address. We have spoken to many expert integrators and fraud analysts in the field and there is an interest in seeing this type as a commercial offering.

# Current Applications

 How is GIS applied in fraud investigation today? Most applications are centered on physical acts that can be mapped and then correlations can be understood using GIS and analysis. This project is different in that the acts of fraud are virtual but we provide these as examples of how crime is mapped with GIS systems and is used to help predict or solve crime.

Crime mapping is an often-misused term typically associated with the visual display of crime statistics on a map. In reality, crime mapping includes not only the visualization of statistics, but the geographic principles, criminological theories and statistical techniques. The idea of visually displaying crime statistics on a map is not new. It first began in the 1930’s, but its usefulness was limited due to the limited means of data analysis. With the development of computers and ever increasing computational capacity, crime mapping is becoming increasingly useful. The merging of economic, demographic and social data with crime statistics allows a GIS system to plot the statistical data and enables analysis of potential causal factors.

The geographic principles involved in crime mapping include the identification of places, Tobler’s First Law of Geography, and spatial process. The criminological theories involved in crime mapping are the theory of rational choice, routine activities, the analysis of crime patterns and social disorganization theory. There are several different techniques for mapping criminological data. They range in complexity and effectiveness, depending upon the end user requirements. Thematic mapping can be used to show distribution of numeric values, for instance, a geographic distribution of robberies. The usage of non-graphical indicators can show clustering of events and can be useful in identifying relationships between crime and place. Hot spot (or cold spot) mapping is commonly used by law enforcement to identify more efficient resource allocation.

Many different spatial statistical techniques are utilized in mapping of hot spots. Spatial Regression mapping techniques can be useful as a more explanatory model, taking into account dependence in data. The technique of geographic profiling can be used to analyze the statistical data and, using other profiling methods, make presumptions about the location of offenders. As research continues and technology develops further, crime mapping will prove to be increasingly utilized. In the near future, crime mapping can be used to better identify resource allocation, like law enforcement or community services. Crime mapping can also be analyzed to evaluate the effect of various policy decisions. (Wilson, 2008)

The visual representation of crime statistics could potentially lead to an exploration of the predictive capabilities of GIS technologies and crime analysis. Through the analysis of historical data, the identification of trends and patterns could potentially uncover predilections through predictive analysis. Predictive analysis can be useful in many areas, including fraud analysis and law enforcement. Forecasting identifies trends and makes assumptions of future activity by assuming those trends will continue. Propensity analysis utilizes statistical techniques like Decision Trees, Bayesian Networks and Regression Analysis to calculate a predisposition. Effective predictive analysis requires a lot of secure, multi-source data. This data would be best stored in a dimensional data warehouse. Not only would this dimensional data warehouse provide a unified location for large amounts of historical data, it would enable the detection of criminal patterns, correlation between suspects and crimes and profiling a target victim or premises via GIS technologies. (Oakland County IT, 2006)

There have been several states that have created programs merging GIS technologies and business intelligence with the goal of identifying fraud in various public delivery programs, namely the Food Stamp program and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). In these examples, mapping functioned as a simple visualization tool for the analysis of the business intelligence. These programs attempt to model consumer store choice by construction of vendor propensity measures in order to identify and monitor fraud. Through the use of various modeling techniques, the programs aim to provide a method to identify high risk vendors and the spatial patterns that may indicate interactions between vendors and participants. However, these programs are mostly in developmental stages. Due to the complexity of human behavior and the dangerous implications of making assumptions about behavior and propensity, further analysis is needed before overarching assumptions can be made. These models do not allow for uncertainty, randomness, learned behavior or other aspects of human decision making. (Kim, 2007)

A recent example of the potential use of GIS technologies and fraud analysis is the software developed by ESRI for the Recovery Accountability and Transparency Board (RATB). This tool was specifically developed to monitor the money distributed as part of the 2009 American Recovery and Reinvestment Act. The RATB tool uses maps, tools and templates to analyze data from the 50 states to show where and to whom the federal funds have gone. The user can drill down for more detailed information about spending locally. The first government agency to test this tool will be the Centers for Medicare and Medicaid Services (CMS). CMS will use the tool to identify potential fraud and abuse as well as inaccurate payments to Medicare and Medicaid providers, which together accounted for $65 billion in improper payments last year, according to Office of Management and Budget director Peter Orzsag. In a June 18 blog post announcing the CMS project, Orzsag said the tool, “gathers enormous quantities of information – in real time – and then analyzes the data and helps connect the dots to identify indicators of possible fraud or error.” In addition, there will be a pilot project in which this data analytics tool is used to identify suspect providers among a list of providers in a particular geographic region of the country where fraud has been identified. The results will help to validate those providers already identified as high-risk and to identify additional high-risk providers who had previously been deemed legitimate by the current agency process. Once these efforts are underway and evaluated, the results will be assessed as well as plans to roll this tool out to other areas across the federal government.

The largest user base of the OAAM product, as of this writing, is the financial services industry. They currently do not have a widespread use of GIS fraud analysis in place. Their fraud detection solutions beyond real-time fraud prevention products like OAAM are to use Anti-Money Laundering suites that evaluate transactional data once captured to determine if there are patterns of abuse. Integrated a GIS solution with AML suites is likely another integration worth investigating but is beyond the scope of this effort. Ultimately a full fraud investigation system would incorporate a multi-channel information input into the GIS layers for the fraud analyst to investigate and correlate the potential fraud alerts.

# Further Development

In order to further develop the idea of mapping fraud data, the goal is to determine the feasibility of marrying the Geo Informatics from Quova and the fraud data from OAAM datasets and then represent in a meaningful way in a GIS context. We have gathered some information from Oracle sources on possible configurations.

**OAAM – Oracle Adaptive Access Manager**

**Strong Authentication and Proactive, Real-Time Fraud Prevention**
Rapid growth in online commerce has brought increasing sophistication of Internet fraud. Threats from Phishing, Pharming, Trojans, Key Logging, and Proxy Attacks, combined with regulations governing online data privacy such as FFIEC, HIPAA, and PCI, have placed online security at a premium. Customers must feel protected for online business channels to grow. Oracle Adaptive Access Manager 11g provides superior protection for businesses and their customers through strong yet easy-to-deploy multi-factor authentication and proactive, real-time fraud prevention. (Oracle.com, 2010)

**Oracle Fusion Middleware MapViewer**
MapViewer is a J2EE service for rendering maps using spatial data managed by [Oracle Spatial](http://www.oracle.com/technetwork/database/options/spatial/index-086235.html%22%20%5Ct%20%22_blank) . MapViewer provides services and tools that hide the complexity of spatial data queries and cartographic rendering, while providing customizable options for more advanced users. MapViewer is designed to integrate with Location-Based services and applications. (Oracle.com, 2010)

# Oracle BI Publisher

Oracle Business Intelligence Publisher (BI Publisher, formerly XML Publisher) is an enterprise reporting solution to author, manage, and deliver all types of highly formatted documents eliminating the need for costly point solutions. Built on open standards, IT staff and developers can create data models against practically any data source and use BI Publisher APIs to build custom applications leveraging existing data sources and infrastructure. End users can easily design report layouts using familiar desktop tools, dramatically reducing the time and cost needed to develop and maintain reports. Extremely efficient and highly scalable, BI Publisher can generate tens of thousands of documents per hour with minimal impact to transactional systems. Reports can be viewed online or scheduled for delivery to a wide range of destinations.

# Quova

IP geolocation is the science of determining the physical location and Internet connection characteristics of a Web visitor, and then leveraging this data to deliver the optimal user experience and determine business strategy. IP geolocation is the technology upon which our Internet Location Intelligence platform is based. (www.quova.com, 2010)

The Quova dataset is a flat file that has a row for each IP address that is being tracked. It contains the following information: Continent, Country, Country CF, Region, State, State CF, City, City CF, DMA, MSA, Carriet, ASN, Connection Type, Connection Speed,Country Special Routing Type, Time Zone, Postal/Zip Code, Telephone Prefix/Area Code, Latitude, Longitude, IP Routing Type, Network Connection Speed and Proxy detection. The CF fields are Confidence Factor, this can be used as a gauge for uncertain datasets.

The OAAM product is typically integrated with the authentication process for a web application to evaluate the risk of the user access the site. Further some site will also integrate OAAM for evaluation of risk at transaction runtime. The system will evaluate information gathered from the users sessions, the past history of site interaction, and their geo location information to create a risk score. These risk scores can alert a fraud analyst to investigate a possible fraud.

Lukáš Svoboda (Svoboda, 2009) cites 4 ways in which geo information could be visualized:

* *table*
* *chart*
* *raster map*
* *dynamic map*

This project will endeavor to use the dynamic map capabilities of the Oracle Mapviewer technology. This will allow the fraud analyst to interact with the captured data from the risk engine (OAAM) and get an overview of the locations that a potential fraudster was accessing the site from.

# Scenarios for Use

Sample Use Cases we are going to investigate an implementation for:

1. A fraud analyst investigating a particular user could view a map of all locations from which the user has logged in over the past week.
2. Login location points can be colored or sized based on the risk score for the session.
3. Details about the session can pop-up when you click on a point.
4. Lines could connect a previous login location to the next, to show the travel path.
5. Thicker lines would represent greater velocities. (distance/time)
6. A fraud manager’s dashboard could include a world map of suspected fraudulent logins over the past 24 hours.
7. Countries could be colored to represent the total number of fraudulent login incidents for the time period.
8. If in-session rules are being run, the country color could represent the total value of suspected fraudulent transactions.

We interviewed fraud analysts from select users of OAAM to determine if they would have use for this type of data representation. They suggested other simple mappings that would be of use to the Fraud Analyst:

1. What devices are nearby?
2. What IPs are nearby?
3. What ISPs are nearby?
4. What alerts are nearby?
5. Timeline of alerts on a map

We anticipate a few more responses and may update this section with additional information if received with in the next few weeks.

LiuJian (L J) Qian is the Development Manager for Oracle MapViewer. A colleague contacted him and he advised us in some of the steps to consider in doing this type of integration.

Our interpretation of L J’s response, we would do the following:
1.      Turn on the Locator or Spatial (not sure which) feature of the Oracle Database.
2.      Load initial mapping metadata into the database.
3.      Install and deploy MapViewer onto WebLogic server (maybe into its own WebLogic domain.)
4.      Create a column of type MDSYS.SDO\_GEOMETRY in any OAAM tables that have latitude & longitude columns, such as VCRYPT\_CITY.
5.      Create an index on the SDO\_GEOMETRY column.
6.      Write a SQL query to update each city with its SDO\_GEOMETRY value constructed using the existing latitude & longitude values.
7.      Update the Quova import utilities to populate the SDO\_GEOMETRY column in a similar fashion.
8.      Use Map Builder (thick client) to create map style and theme metadata.
9.      Configure the data source for MapViewer.
10.  Integrate MapViewer into BI Publisher.

 Additional sources for investigation:
Demo of MapViewer integrated with BI Publisher
<http://www.oracle.com/technology/products/spatial/pdf/orabi_ee_mapviewer_int_demo.pdf>

Oracle MapViewer 11g FAQ
<http://www.oracle.com/technology/products/mapviewer/pdf/mapviewer_faq_11gr1.pdf>

Oracle MapViewer 11g
Product and User’s Guide
<http://www.oracle.com/technology/software/products/mapviewer/index.html>

Oracle Spatial documentation is included in the Database 11g documentation library.

Discussion of using geo spatial analysis for insurance fraud investigation: <http://linkanalysisnow.com/2010/01/using-geospatial-analysis-to.html>

# Challenges Ahead

There are several difficulties ahead that may present challenges to our project goal. They include:

1. Have not worked with Mapviewer
2. Oracle spatial is a new area for our team
3. Lack of examples of website fraud being investigated with geo spatial analysis
4. Limited academic articles identified on using ip geo intelligence for fraud analysis visualization
5. Data integration concerns

Data integration is a problem that has faced most, if not all, business enterprises in recent history. Effective data integration would give the user the benefit of access and reuse of data through a unified access point. It would also afford the user a more comprehensive answer to questions through the usage of multi-source data. However, data integration is problematic on many different levels. Most systems are not designed for integration due to the functionality of the systems, the type of data stored, the autonomy requirements, the architectural view of the system and limited resources. Several types of heterogeneity also need to be addressed, such as data management software, operating requirements, existing data models, user interfaces and business rules. There are several approaches to effectively integrate data. They include: Manual integration, Common user interface, Integration by applications, integration by middleware, uniform data access and common data storage.

In the past, data integration only had more well-structured data models to consider. With the increase of data from open sources, like the internet and other web based technologies, data integration must now also consider semi and unstructured data. Additionally, data integration must also consider the question of semantics. The same word can mean different things in different contexts. Semantic integration attempts to address this issue by ensuring that only data that is closely related to the same concept is integrated. With the relative nature of semantics, it proves to be near impossible to completely define all data. In an attempt to overcome semantic heterogeneity, the concept of single domain models was introduced. This single domain model would force the user to adopt a single view of the world, which is not realistic.

Some advantages our team has that may help us to overcome these potential difficulties include:

1. OAAM is core competency of Integral team
2. BI Publisher has been installed before
3. GIS class resources
4. Some access to Oracle expertise for assistance

# Success to date (11/8/2010)

Have established a VM image with Oracle Enterprise Linux, Oracle 11G db, OAAM 11G, Arm Automator and 3rd party product that will permit us to generate data into the db, and a partial Quova load for representative data. Our VM currently needs to be resized to accommodate the BI publisher application and the mapviewer app as well. We have made positive progress so far.

# Conclusion

 It appears that it may be feasible to create this type of integration with the discussed products. It is also reasonable based on interviews that there exists a potential client base for this functionality. As part of the project we will attempt to perform the integration and have a working sample to demonstrate the capability.

# Glossary

Action – Message passed back to calling application when a rule evaluates to true.

Alert – Message displayed in the OAAM console when a rule evaluates to true.

ARM Automator – 3rd Party tool that permits the user to create fraud scenarios in the OAAM rules engine.

Checkpoint – OAAM evaluates the users digital fingerprint as well as any historic data to Alert, determine an Action, and provide a Risk score.

Device – OAAM defines a device as the computer a user is using to access the site.

Digital Fingerprint – OAAM captures the User Agent String from the browser and combines that with a Flash LSO to create a digital fingerprint

ISP – Internet Service Provider

OAAM – Oracle Adaptive Access Manager

Oracle BIPublisher – Business Intelligence reporting tool that integrates with mapview and OAAM.

Oracle MapViewer – Fusion Middleware product that will render maps from geo spatial data.

Quova – company that provides geo intelligence data that has been integrated into the OAAM solution.

Risk Score – A score assigned by the fraud analyst in the rules that run and evaluate the session at a checkpoint.

User Agent String – Each browser is considered a user agent, this string contains information that helps the website that it is communicating with recognize capabilities in the browser (user agent)

# Bibliography

(2010). Retrieved 10 10, 2010, from Oracle.com: www.oracle.com/us/products/middleware/identity-management/oracle-ada-access-mgr/index.html

(2010). Retrieved 10 10, 2010, from Oracle.com: www.oracle.com/technetwork/middleware/mapviewer/overview/index.html

(2010). Retrieved 10 10, 2010, from www.quova.com: www.quova.com/what/faq/

Kim, Y. (2007). *Using Spatial Analysis for Monitoring Fraud in a Public Delivery Program.* Social Science Computer Review.

Mosquera, M. (2010 йил 21-June). *Legislative News Item*. Retrieved 2010 йил 5-November from Government Health IT: http://govhealthit.com/newsitem.aspx?tid=79&nid=74039#

Oakland County IT. (2006 йил 15-June). *Using Predictive Analysis with Crime Data.* Retrieved 2010 йил 9-October from www.oakgov.com: http://www.oakgov.com/dataware/assets/doc/Predictive%20Analysis-Crime.pdf

Svoboda, L. (2009, 1 1). *Possibilities of Visualisation Geolocation of Web Pages Visitors.* Retrieved 10 10, 2010, from www.gis.vsb.cz: http://gis.vsb.cz/GIS\_Ostrava/GIS\_Ova\_2009/sbornik/Lists/Papers/071.pdf

Wilson, R. (2008). Crime Mapping and Analysis. In S. Shekhar, & H. Xiong, *Encyclopedia of GIS* (pp. 180-186). New York: Springer.

Ziegler, P., & Dittrich, K. *Three Decades of Data Integration - All Problems Solved?* Zurich: Database Technology Research Group, Department of Informatics.