SL-SecureNet: Intelligent Policing Using Data Mining Techniques

M.A.P. Chamikara, Y.P.R. D. Yapa, S.R.Kodituwakku, J. Gunathilake

Abstract*— Many police departments all around the world lack of good and efficient crime recording and analysis systems. The vast geographical diversity and the complexity of crime patterns have made the analyzing and recording of crime data even difficult. According to the Sri Lankan police department, they face these problems for many years. This paper presents an intelligent crime analysis and recording system designed to overcome problems that appear mainly in the Sri Lankan police department. The proposed system is a GIS based system which comprises of data mining techniques such as Hotspot detection, Crime clock, Crime comparison, Crime pattern visualization, Outbreaks detection and the Nearest police station detection. Salient features of the proposed system include a rich environment for crime data analysis and a simplified environment for location based data analysis. It facilitates the identification of various types of crimes in detail and assists the police personals to control and prevent such incident efficiently. The SL-SecureNet was tested for about 1000 crime records. The test results indicated that it functions in an efficient and reliable manner.

***Index Terms*—Crime Analysis, Crime Investigation, Data Mining, Intelligent Policing**

I. INTRODUCTION

Crime analysis has become one of the most essential activities in the world, because the technology development and the high growth of community have resulted a high magnitude of crimes, most of the time with bizarre patterns [12]. Due to the advancement of technology, lot of criminals have become very intelligent and, therefore they conduct crimes in an untraceable manner. Majority of those crimes evolve in a long period of time making them even more difficult to predict. Therefore, manual techniques of analyzing such data with a vast variation have resulted in lower productivity and ineffective utilization of manpower. This is one of the most dominant problems in many law enforcement institutes.

According to the Sri Lankan police department, a manual crime recording and analysis system is used and it gives a little support in decision making. The process is started by recording the complaint of a crime incident as a large description exactly as explained by the complainant. Next, the location of the crime is physically investigated. After that, the complaint is included into the GCR (Grave Crime Record) or MOR (Miner offences Record) depending on the egregiousness or the crime scene. At the same time, the crime scene is spatially mapped on to a paper map by using a pin or a dot with a specific colour assigned to the crime type. The crime scene is also located on a 24 hour paper clock, considering the relevant time of each crime. The data which is gathered during further investigations are then maintained in GCR or MOR in a categorical manner. To record the court data, another book of records is used with GCR or MOR reference numbers on it. The Sri Lankan police department practices this manual method for a long period of time for regional security arrangements and placement of police patrol services.

Due to the inefficiency and the tediousness of this task most of the time the confidence in tracking the exact criminal is low. Moreover, the current system does not facilitate many newer crime analysis techniques such as crime trend analysis, real-time crime outbreak detection, crime pattern identification, crime network analysis, crime association mining etc, because the large dataset of crimes make the human prediction very difficult. One of the main disadvantages in the current manual system is that every time when a police officer from a particular police station other than the main police station wants to locate a crime scene on the regional map, that person has to be physically present in the main police station. The updates made in one police station are also not visible to the other police stations.

An automated web based intelligent crime analysis system with GIS (Geographic Information Systems) composed of an affluent set of data mining tools would annihilate the problems that appear in a manual crime analysis system and improve the efficiency and the accuracy of crime analysis.

Several web based crime mapping systems are available in the internet, but most of them have been custom made for legislative authorities in different countries and those systems are not accessible to outside of those respective authorities [1],[2].

Manuscript Received February 15, 2012

M.A.P. Chamikara, Post Graduate Institute of Science (PGIS), Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka, +94718016715, (e-mail: pathumchamikara@gmail.com).

Y.P.R.D. Yapa, Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka, +94 81 239 (4642), (e-mail: roshany@pdn.ac.lk).

S.R. Kodituwakku, Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka, +94 81 239 (4641), (e-mail: salukak@pdn.ac.lk).

J. Gunathilake, Department of Geology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka, +94777313063, (e-mail: aajkg@yahoo.com).

This paper presents an intelligent crime analysis and recording system named SL-SecureNet which is a web based crime analysis system. Additionally it indicates a set of crime analysis tools such as the postgresql geographic database allowing spatial analysis, Crime locator, Crime clock, Periodic pattern visualizer, Public crime map, Hotspot detection, Nearest police station detection, Crime comparator, Outbreaks detection, etc.

II. MATERIALS AND METHODS

Several existing tools are used to develop the system. This section describes such tools and overall methodology.

A. Materials

The collection of analysis tools of the SL-SecureNet system were successfully designed and implemented with the aid of a collection of some existing APIs (Application Programming Interface), tools, frameworks, and libraries. This section includes a description of each of them.

1. Openlayers

Openlayers is a free and opensource API developed using pure JavaScript for displaying map data in most modern web browsers without any server-side dependencies. Google Maps and MSN (Microsoft Network) Virtual Earth APIs are some of the similar APIs with openlayers [3].

2. GeoServer

GeoServer is an open source software server written in Java that allows users to share and edit geospatial data. It has been designed to achieve interoperability. It publishes data from any major spatial data source using open standards. GeoServer is the reference implementation of the Open Geospatial Consortium (OGC), Web Feature Service (WFS), Web Coverage Service (WCS) standards and a high performance certified compliant Web Map Service (WMS). GeoServer forms a core component of the Geospatial Web [4].

3. GeoExt

GeoExt is a JavaScript library which provides a hybrid API between OpenLayers and ExtJS so that it allows a rich set of tools to powerfully interpret geospatial data with GIS on the web. ExtJS is a JavaScript library for building interactive web applications using techniques such as Ajax, DHTML and DOM scripting [5].

4. Map Layers

Spatial datasets are normally handled as layers in Geographic Information Systems. A map layer represents a specific set of geographical data which is considered under a particular theme of data. For example, roads, buildings, lakes, terrains are considered as separate themes. Therefore, they are represented in separate thematic layers.

5. PostGIS/PostgreSQL database

PostGIS adds support for geographic objects to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS), much like ESRI's SDE or Oracle's Spatial extension. PostGIS follows the OpenGIS "Simple Features Specification for SQL" and has been certified as compliant with the "Types and Functions" profile [6].

6. MySQL database

MySQL is an open source relational database management system. The MySQL database server is very fast, reliable, and easy to use and it works in client/server of embedded systems. It consists of a multi-threaded SQL server that supports different backends, several different client programs and libraries, administrative tools, and a wide range of application programming interfaces (APIs) [10].

7. Spring MVC

MVC is a framework abbreviated for Model-View-Control which was designed to develop the J2EE applications easier. Spring MVC shows how to move beyond Java Server Pages and uses other templating languages such as Velocity and FreeMarker [7]. Spring MVC has been used as the URL routing mechanism for SL-SecureNet and it is depicted in Fig. 1.

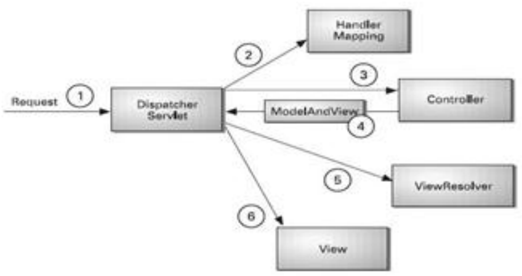

Fig. 01: The life cycle of a request in Spring MVC

B. Methodology

Crime analysis is carried out as a collection of steps: Hotspot detection, Crime clock, Crime comparison, Crime pattern visualization, Outbreaks detection and Nearest police station detection. Each of these steps has been automated as a tool in the SL-SecureNet system. Therefore, the police personals can use different tools in different times according to the situation at hand and decisions can be taken in fast and well organized manner. This section describes each of those analysis tools in detail.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

1. Hotspot Detection

Cluster analysis is the process of identifying groups of a dataset in such a way that the data inside those groups have specific similarities while the relationships among those groups are minimal. Therefore in order to identify hotspots with high crime density, cluster analysis is used for identifying the clusters of crime spots.

The clustering algorithm of the system first accepts the area to be investigated as the input. According to the user's inputs the algorithm measures the Euclidian distances among all the data points with each other within the defined area. Then it clusters the data points into the most suitable number of clusters using the nearest neighbour concept and the calculated Euclidian distances. Finally, the coordinates of the centers of the clusters are identified and the number of crime points inside each of those clusters are returned. Depending on the values returned with a coordinate, each cluster is assigned a colour darkness and a radius according to the magnitude of the cluster. Then using the vector layer concept, those points are graphically displayed on top of the base map as shown in Fig. 6.

2. Crime Clock

A crime clock is a representation of the number of crime scenes that has been taken place within the 24 hours of a day. As depicted in Fig. 5, a crime clock is represented as a bar chart. The 24 hour clock is represented using 24 bars on the graph and the height of each bar represents the number of crime scenes per hour. Three extra bars are used to represent the crime scenes without an exact time of incident. The "day bar" represents the crime scenes which were taken place in the day time, the "night bar" represents the crime scenes which were taken place in the night time and the "unknown bar" represents the crime scenes which cannot be assigned to any time duration.

3. Crime Comparison

Comparing different types of crimes is very important to get an idea about the growth of a particular crime over the other types of crimes. A pie-graph is used to satisfy this requirement by allowing the analyst the maximum freedom to compare the different types of crimes in an optimal way. It is depicted in Fig. 7 and shows the percentage comparison between different crime types.

4. Crime Pattern Visualization

In statistics, signal processing, econometrics and mathematical finance, a time series is a sequence of data points, measured typically at successive times spaced at uniform time intervals. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data [8]. A time series plot is used to represent the changes in frequency of crime occurrence as depicted in Fig. 8. The Y-axis represents the frequency of crimes and the X-axis represents the time.

5. Outbreaks Detection

A crime outbreak is the occurrence of any crime incidents in excess of what would normally be expected in a defined geographical area or a time period. "Crime outbreaks detection tool" is an agent system that observes for number of crimes in different regions. If the number of crimes is increased out of control, an alert will be prompted by the system to all the relevant police stations. In this system, initially the user can define a reference time frame and then the system will calculate the average (μ), and the standard deviation (σ) of the number of crimes per day per each cluster. If, in a particular cluster, number of crimes within a day is greater than $\mu + \sigma$, the system will prompt an alert as depicted in Fig. 9.

6. Nearest Police Station Detection

The J48 decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data [9]. In an emergency like following a suspect on persuit, it is very important to know clearly about the available police support around the current location. To achieve this task, a nearest police station detection tool has been integrated. The "J48" classification algorithm is the methodology used in building this tool. First, the J48 algorithm is trained for about 150 data points per each 400 Km^2 area. Those data points include the coordinates and the nearest police stations . The algorithm was trained several times to adopt the coordinates to the predefined classes (police stations). When the user clicks on a desired point on the map, that coordinate will be analyzed by the algorithm and the most suitable class of that coordinate will be returned. The functionality of this tool is depicted in Fig. 10.

7. *System Architecture*

The system architecture of the SL-SecureNet shown in Fig.2 is implemented on the Spring MVC framework. A particular user can request for any service through the web browser and that standard web request is sent to the Spring MVC core. Then the request is filtered and directed to the relevant service that the user is requesting for. A particular request might be sent to one of the SL-SecureNet analysis tools (Hotspot detection, Crime clock, Crime comparison, Crime pattern visualization, Outbreaks detection, Nearest police station detection, etc) or one of the maps available in the Geoserver. The model of the SL-SecureNet is composed of a MySQL database, a PostGIS/PostgreSQL database and a Map Layers container. SL-SecureNet analysis tools communicate with the two databases, MySQL and PostGIS/PostgreSQL, while the Geoserver communicates with the map layers and the PostGIS/PostgreSQL database. When the user request is for a map, the system communicates with the OpenLayers API. In turn, the API communicates with the Geoserver to resolve the WMS and WFS requests

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

sent by the Geoserver and provides a layered view of maps to the user. The OpenLayers API uses the Google Maps as the base layers while GeoExt API helps the OpenLayers API to view these information in graphically rich environment.

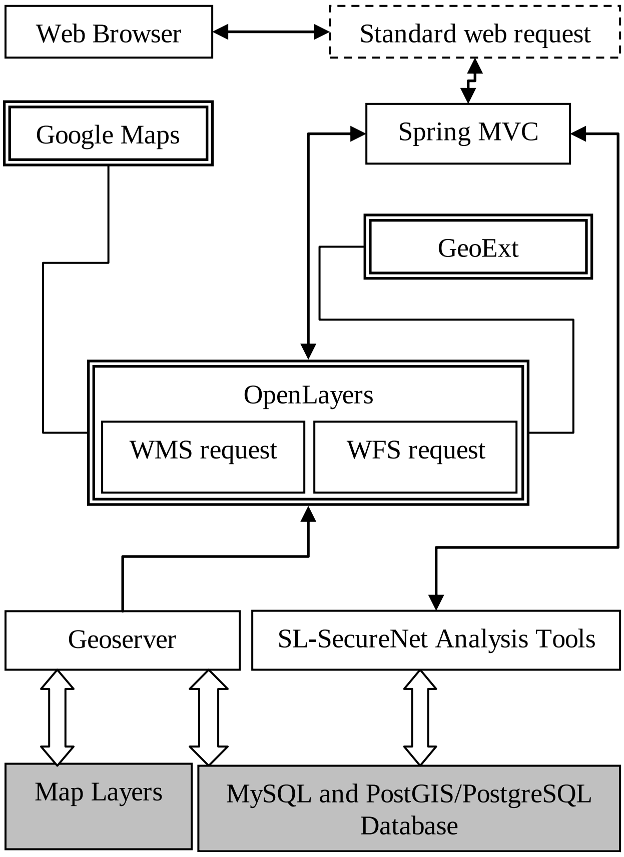

Fig. 2: Architecture of the SL-SecureNet

III. RESULTS AND DISCUSSION

The proposed system was tested with a data set of crime records taken from the Sri Lankan police department. The test results indicated that this system improves the efficiency and the reliability of crime recording and analysis conducted by the police personals. The system also facilitates online crime mapping, recording, analysing and viewing, which ease the process of the existing system by allowing any police station to work with the system online without coming to the regional police station.

Crime Locating Interface which is depicted in Fig. 3 provides a flexible and simple interface to locate the crimes on a geographical interface.

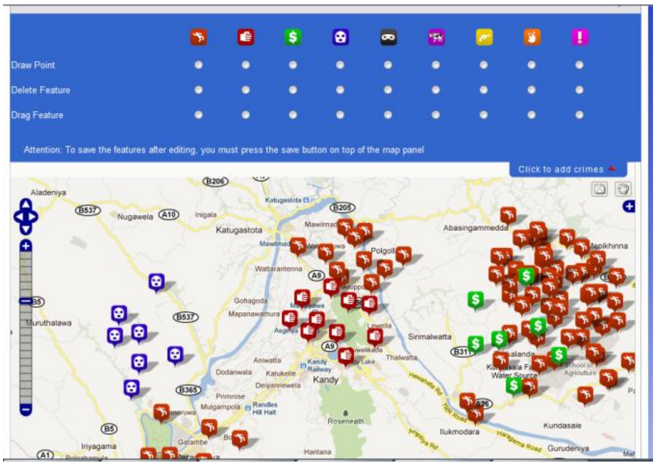

Fig.3: Crime Locating Interface

After locating and inserting details about the crimes, it is necessary to view the crimes in an efficient manner. SL-SecureNet provides a simple interface depicted in Fig. 4 for this purpose. It allows to view the details of each crime scene depending on the user specified query.

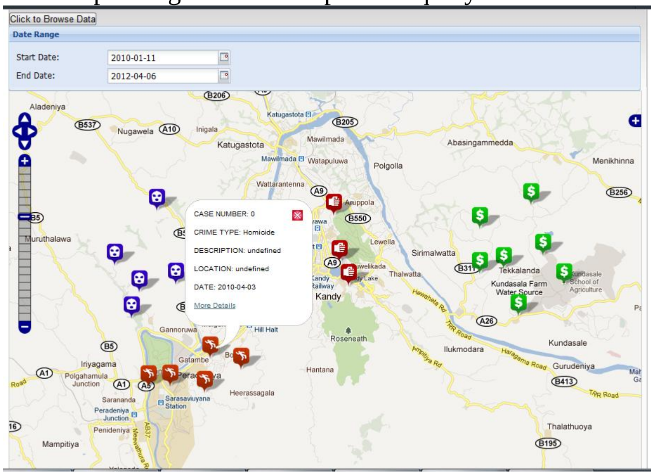

Fig.4: Crime Viewing Interface

SL-SecureNet provides an automated crime clock that over comes the limitations in the manual crime clock. This is shown in Fig. 5. This shows the number of crime scenes in each hour on a 24 hour clock as a bar graph. This tool helps to identify the exact period of time in which the crime frequency is high so that they can apply sufficient amount of security for each hour of a day.

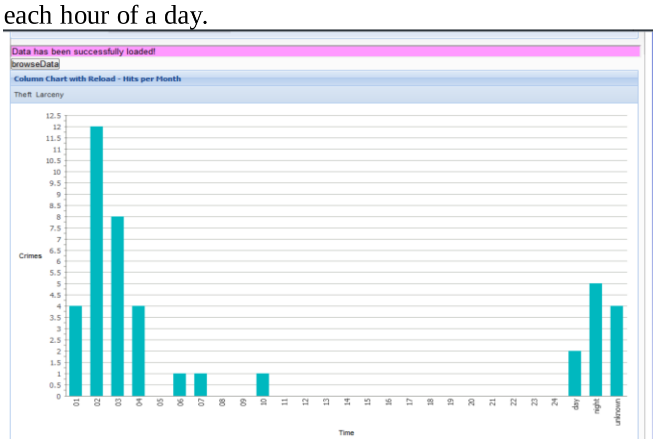

Fig.5: Crime Clock

Hotspot detection which is depicted in Fig. 6 helps to identify the areas with higher crime density than others so that the security can be arranged accordingly. Therefore, hotspot detection helps the police department to utilize the police patrols in an efficient manner in such a way that the patrol services in areas with high crime density can be increased and the patrol services in the other areas can be decreased accordingly.

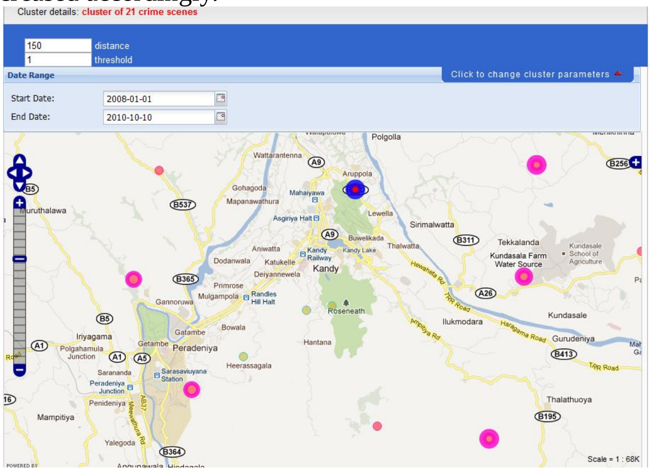

Fig.6: Hotspot detection tool

Depending on the egregiousness of the crime type, its frequency of occurrence may change. For example, number of burglaries within a particular period of time is normally greater than the number of homicides during the same period of time but, there may be some exceptions also [11]. To compare the number of crimes within a particular period of time, a pie chart has been used as shown in Fig. 7. It displays exactly what type of crimes should need more attention, so that necessary actions can be taken by the police officers.

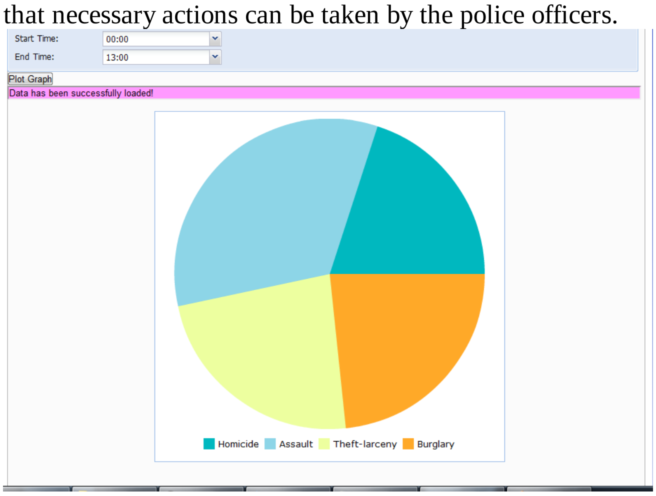

Fig. 7: Crime Comparator

Crime pattern visualizer is depicted in Fig. 8 and it helps the police officers to recognize gradual increments or decrements of crime incidents in different areas and hence they can change security arrangements accordingly. Crime Pattern Visualizer generates time series plots for all the crime types in the system on the same graph making the comparison between those graphs easier. Police personals can use this functionality to visualize the trends involved in different types of crimes so that they can understand whether their security plan is beneficial or not.

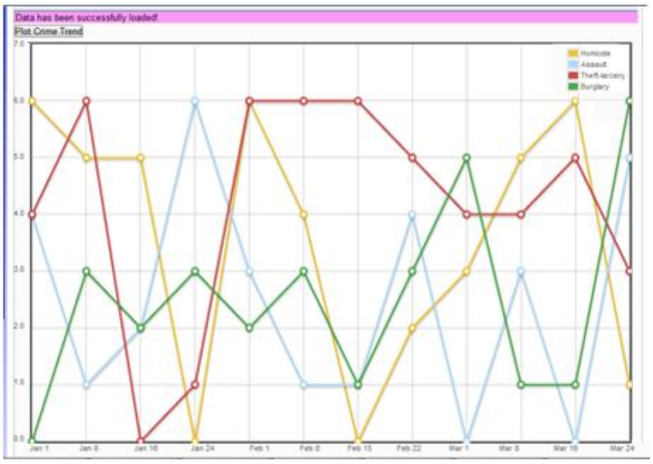

Fig.8: Crime Pattern Visualizer.

Outbreaks detection tool helps the police to be aware of the situation at hand. It is a very important functionality, if there are unusual rates of crime incidents recorded. Outbreaks detection tool which is shown in Fig. 9 is an agent system which can be activated and deactivated according to the user's preference. When it is active it will identify any kind of unusual crime incidents and if so, it will prompt an alert so that the relevant police station can take necessary actions to remove those crime outbreaks from the community.

Fig.9: Outbreaks Detection Tool

Nearest Police Station Detection tool depicted in Fig. 10 is used to identify the nearest police station to a particular location. When a particular police personal wants to know about the nearest backup service due to some emergency situation, that person has to send the correct coordinates of that point of location to the system through GPS or any mobile network service. Then, this tool will instantly respond to that message sending the nearest police station relevant to the given location.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

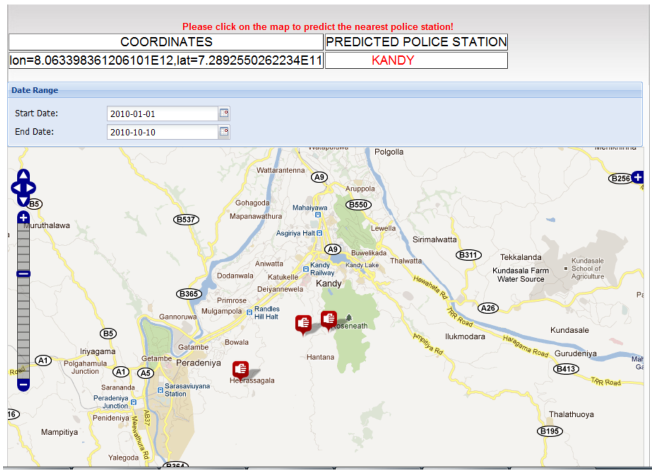

Fig. 10: Nearest Police Station Detection

IV. CONCLUSION

This paper proposed an automated system, SL-SecureNet, which is developed based on a GIS based data mining platform. The proposed system makes Crime locating and recording a very simple and quick task. Therefore, SL-SecureNet would evade most of the problems that appear in the current manual crime recording and analysis system and would improve efficiency making timely decisions on security arrangements. It assists the police department to provide high quality security service. Since the new system is a web based online system, the police personals will no longer have to physically appear in the regional police stations to update their police station wise crime data.

Further improvements and optimizations to the system can be made in order to reduce the response times of the tools and improve the accuracy. System quality could also be increased by introducing more modern frameworks and by introducing advanced data mining techniques.

ACKNOWLEDGMENT

This work was funded by the National Research Council (NRC) of Sri Lanka [Grant number: 11-071]

REFERENCES

1. Crime Mapping and Reporting System. (2011, August 31). [Online]. Available: https://www.crimereports.com/
2. Intelligent Mapping System. (2010, October 16). [Online]. Available: http://maps.met.police.uk/
3. OpenLayers: Free Maps for the Web. (2010, September 15). [Online]. Available: http://openlayers.org/
4. GeoServer. (2010, September 15). [Online]. Available: http://geoserver.org/display/GEOS/Welcome
5. GeoExt. (2010, September 15). [Online]. Available: http://geoext.org/
6. PostGIS. (2010, September 15). [Online]. Available: http://postgis.refractions.net/
7. Craig Walls & Ryan Breidenbach, Spring in Action, 2nd Edition, Manning Publications, USA(2005).
8. Time Series. (2010, September 21). [Online]. Available: http://en.wikipedia.org/wiki/Time_series
9. Classification Methods. (2010, September 21). [Online]. Available:http://www.d.umn.edu/~padhy005/Chapter5.html
10. What is MySQL?. (2010, September 23). [Online]. Available:http://dev.mysql.com/doc/refman/5.0/en/what-is-mysql.html.
11. Grave Crime Abstract for Full Year 2010 for Whole Island From 01.01.2010 To 31.12.2010. (2010, September 26). [Online]. Available: http://www.police.lk/images/others/crime_trends/2010/grave_crime_abstract_for_full_year%202010.pdf.
12. Chen, H.,W.Chung, et al.(2004). Crime data mining: a general framework and some examples. Computer 37 (4):50-56.

AUTHOR PROFILE

M. A. Pathum Chamikara is working as a research assistant at the Post Graduate Institute of Science (PGIS), University of Peradeniya, Sri Lanka. He received his BSc (Special) degree in Computer Science, University of Peradeniya, Sri Lanka (2010). His research interests include Crime analysis, data mining, GIS (Geographic Information Systems), image processing, computer vision and artificial intelligence.

Y.P. R. Darshana Yapa is a senior lecturer at the Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka. His research interests include image processing, computer vision, data mining and GIS (Geographical Information Systems)

S. R. Kodituwakku is an associate professor at the Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka. His research interests include database systems, distributed computing, Role based Access Control systems, and software engineering.

Jagath Gunathilake is a senior lecturer at the Department of Geology, University of Peradeniya, Sri Lanka. His research interests include Natural resource management, GIS (Geographical Information Systems) and RS (Remote Sensing) systems.