Developing Personalized Security Information Service Using Open Data

Takuhiro Kagawa, Sachio Saiki, Masahide Nakamura
Graduate School of System Informatics, Kobe University
1-1 Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan
Email: kagawa@ws.cs.kobe-u.ac.jp, sachio@carp.kobe-u.ac.jp, masa-n@cs.kobe-u.ac.jp

***Abstract*—Local governments in Japan recently provide security information services for residents, which deliver regional incident information using Email or Web. However, since the conventional services usually provide "one-for-all" information. users tend to miss important incidents within the flood of information. In this paper, we propose a new security information service, called *PRISM (Personalized Real-time Information with Security Map)*. For given incident information and user's living area, PRISM first computes severity of the incident, based on distance, time, and type of the incident. It then visualizes the incidents with the severity on a heat map. Thus, PRISM provides real-time personalized information adapted to individual situation of users. To illustlate the feasibility, we implement PRISM as a Web application using Hyogo Bouhan Net, and Kobe city facility open data.**

***Index Terms*—open data, security information service, smart city, visualization, street crime, Web application**

I. Introduction

Street crimes are a major factor that threatens safety and security of people living in the region. There are many incidents reported every day, including murder, robbery, assault, snatching, groping, exposure, suspicious act of speaking, and observation of dangerous animals. In order to achieve the safety and security of the community, individual residents are recommended to protect themselves by understanding the street crimes and incidents, spontaneously.

Based on such circumstances, local governments have implemented various policies[1]. In Japan, many local governments recently start providing *security information service* for residents. The security information service distributes information of crimes and incidents to residents using the Internet. Residents can make use of the information for avoiding crimes. The typical security information service shows the list of recent incidents and a security map in a web site, or delivers the incident information by e-mail. For example, Hyogo Prefectural Police in Japan provides "Hyogo Bouhan Net"[2]. The service publishes incident information that Hyogo prefectural Police recognize on the Web. By registering a personal email address, a user can receive the information by e-mail. Similarly, Tokyo Metropolitan Police Department provides the e-mail delivery service, called "Mail Keishicho". The Department also publishes "Tokyo Crime Map"[3], which is a security map showing where and when every suspicious person appeared.

In these existing security information services, every incident information is uniformly delivered to all users. Various types of incidents occur every day at various locations in the region. However, user's living area varies from one person to another. Therefore, even if an incident is critical for a user, it may not be so serious for another user who is living at distant place. Thus, how the incident is severe depends on individual users. However, this fact is not taken into account in the existing security information services. All information of incident is distributed uniformly to all users. Hence, when much information is delivered in a day, a user may miss important information. Also, it is a time-consuming task for an ordinary user to identify only relevant information from the list of a lot of incidents within the Web portal.

In this paper, we propose a new security information service, called *PRISM (Personalized Real-time Information with Security Map)*, which personalizes the incident information based on living area of individual users. For every incident information provided by the existing security information services, PRISM computes *severity* of the incident according to the living area of a user. More specifically, based on the distance between the living area and the incident, the time elapsed from the occurrence, and the type of the incident, PRISM adds a weight to the incident, so that closer and newer incidents become more serious for the user. It then visualizes the weighted incidents on a *heat map*. Since the weight of severity varies depending on user's living area, the resultant heat map becomes a personalized and real-time security map.

One of technical challenges in implementing PRISM is how to manage incident information in unstructured text. PRISM applies *text mining* to the incident information originally described in natural language. It then extracts attributes such as date, time, location, and the type of the incident, and stores the attributes in a relational database (RDB). On top of the database, we develop Web-API with which external applications can easily query and retrieve incident data. Furthermore, in PRISM, we exploit *facility open data* published by local government so that users can easily register own living areas. The facilities include schools, stations, and parks in the region. Even if a user cannot read a map, the user can specify the living area by using the facilities as landmarks.

To show the practical feasibility, we implement PRISM as a Web application and a mobile application. In the implementation, we use "Hyogo Bouhan Net" to obtain incident

978-1-5090-5504-3/17/$31.00 ©2017 IEEE
SNPD 2017, June 26-28, 2017, Kanazawa, Japan

Fig. 1. Example of security map(Kobe city. From January to May, 2016)

information, and "Kobe City Facility Open Data" as the facility open data. Finally, we deploy PRISM on the Internet. Using the application, users can browse personalized and real-time security information within Hyogo prefecture.

II. Preliminaries

A. Security Information Service

Security information service is an information service that distributes incident information recognized by a police to residents. The service aims to promote residents to understand the situation for self-protection against incidents. Typical security information services use e-mail and Web-based security maps to deliver the incident information.

The e-mail-based services send incident information in text message to the subscribers who registered e-mail addresses. The delivered information is usually archived on the web site, where users can browse the past information. In some services, users can register the ward or division of a city to specify the area range to be notified. For example, in Hyogo Bouhan Net, if a user wants information within Nada ward and registers the range, then incidents under the jurisdiction of Nada Police Station are notified.

Security map is an annotated map showing where incidents occurred. It visualizes geographical occurrences of incidents that are hard to grasp by textural information only. As an example, we show a security map describing street incidents within Kobe City from January to May, 2016. in Figure1. In the map, a pin represents a place where an incident occurred.

B. Open Data

Open data is machine-readable data that anyone can freely use and share without limitation of copyright. Recently, various government and municipal bodies have published public data as open data, in order to improve quality of life of people as well as performance of business activities of enterprises[4]. For example, Kobe City in Japan discloses city administration data such as population, information of facilities, time table of subway[5]. These data are represented in CSV or RDF format. Japanese government also has published a data catalog site[6]. We can use various data in this site.

C. Problems

In the existing security information services described in **II-A**, any incident information is uniformly distributed to all users. Thus, the *degree of how severe a given incident is* for a user (we call it *severity*), which varies from one person to another, is not taken into account. As a result, users tend to miss relevant incidents within the flood of information. For this, we try to cope with the following three challenges.

Challenge P1 (Considering living area): If an incident occurs nearby living area of a user, then the incident is very important for the user to be alerted. The existing services, however, do not take how *close* the incident is. For example, in Hyogo Bouhan Net, incident information of Nada Police Station is distributed to all subscribers of the Nada ward in the same way. However, comparing a user living 200 meters from an incident with the one living 5 kilometers away, the incident should have higher severity for the former user.

Challenge P2 (Considering real time) Although each incident information has date and time of occurrence, the current services do not consider how *fresh* the incident is. For example, in Tokyo Crime Map, an incident that occurred yesterday and one that occurred two weeks ago are shown in the same marker on the map.

Challenge P3 (Considering type of incident) There are various types of incidents reported from serious ones to just informative ones. In the existing services, however, all incident information are delivered in the same way. For example, witness information of a man with a knife and arrest information of past incident are delivered in the same e-mail format. Since the arrest information is about a resolved incident, it should have lower severity than the information of a man with a knife.

III. Security Information Service PRISM

In this paper, we propose a new security information service, called **PRISM(Personalized Real-time Information with Security Map)**, which personalizes the incident information based on living area of individual users.

A. Overview

In order to cope with the three challenges described in **II-C**, we exploit two ideas in PRISM. The first idea is to put a *weight of severity* on every incident, based on the living area, the current time, and the type of the incident. The second idea is to visualize the weighted incidents on a *heat map*.

A user of PRISM first registers his/her living areas. The living areas represent places where the user often visits in the daily life, such as a house, a station, a working place, a school of children, and a shopping center. Every user can register multiple places as living areas.

Then, for each delivered incident information, PRISM calculates the *distance* between the point of incident occurrence and the living area of the user. PRISM also calculates the *elapsed time* from date and time of occurrence to the current time. Based on the distance and the elapsed time, PRISM adjusts a weight of the *default severity* pre-determined for each type of incident. The resultant weight is the severity of the incident personalized for the user. Finally, PRISM generates a heat map based on the personalized severity, which achieves a personalized and real-time security map.

B. Architecture

Figure2 shows the system architecture of PRISM. In the figure, the dotted rectangle represents the system boundary of PRISM. A *crawler*, at the bottom left of the figure, periodically obtains incident information from an existing security information service. Then, the crawler analyzes the retrieved text, extracts attributes, and inserts the attributes into an *incident DB*. A user at the top right of the figure registers his/her *living areas data*. In the registration, the user can refer to a *facility DB*, where facility open data of the local government is imported. The user can use various facilities in the region as landmarks of the living area. Based on the incident DB and the living areas data, PRISM *computes the severity* of every incident. Finally, PRISM generates a heat map based on the weighted incident information, and presents the map to the user. We will explain the details in the following subsections.

C. Obtaining Incident Information

PRISM periodically obtains incident information from an existing security information service using the crawler. In general, the incident information is written in natural language. So it is difficult to make queries or take statistics for the original text. Therefore, PRISM analyzes the original text in the crawler, structures the text into attributes, and inserts the attributes into the RDB (i.e., the incident DB).

We explain the flow of the text analysis. First, the crawler applies pre-processing to the original text, where orthographic variants in characters and sentences are corrected. Then, applying a text mining to the pre-processed text, the crawler extracts the following eight attributes.

- **id:** an identifier of an incident
- **datetime:** date and time of occurrence of the incident
- **title:** a title of the incident information
- **content:** a content of the incident information
- **severity:** a default severity of the incident
- **address:** an address of the place where the incident occurred
- **lat:** latitude of the place where the incident occurred
- **lng:** longitude of the place where the incident occurred

Although the original incident information is given in the natural language text, it basically follows a strict *writing convention*. Therefore, by applying pattern matching using regular expressions, we can extract datetime, title, content, and address.

The values of lat and lng can be obtained from address using geographic information web services such as Geolocation API. The value of id can be extracted from the article number or URL of the security information service.

The value of severity represents a *default severity* that the incident has. It is defined by the following four categories, determined by *keywords* contained in the title and content.

Alert (severity 3) the most serious incident that can threaten life of citizens. The keywords include murder, robbery, shooting, assault, gun, knife, etc.

Warning (severity 2) incidents that may cause physical damage to citizens. The keywords include snatching, pickpocket, theft, stalking, groping, etc.

Caution (severity 1) incidents to be paid attention, not directly linked to life or physical damage. The keywords include scam, animal, wild boar, etc.

Notice (severity -1) other information from the police. The keywords include arrest, resolution, notice, etc.

Note that the above default severity represents a default value of each incident, where the situation of individual users is not yet counted. It will be adjusted in the subsequent process based on the living areas of individual users.

D. Incident DB

The incident DB is a relational database that stores and manages incident information with the eight attributes described in **III-C**. The data schema is as follows:

[id, datetime, title, content, severity, address, lat, lng]

Figure3 shows an example that inserts an incident information text into the incident DB. In this example, PRISM analyzes an incident "Grouping occurred", extracts the eight attributes, and inserts them as a record.

On top of the incident DB, we develop Web-API, so that external applications can easily retrieve incident data. The Web-API has the following three commands:

- **latest:** For a given incident ID, the command returns the latest incidents that occurred after the incident specified by the ID.
- **search:** For a given query of ID, keyword, date, time range, and severity, the command returns incidents that matched the query.
- **range:** For given coordinates and distance, the command returns incidents that occurred within the range defined by the coordinates as a center and the distance a radius.

We have implemented Web-API as CGI handling JSON. The command and parameters are given in a form of URL, and the result is returned in a JSON format.

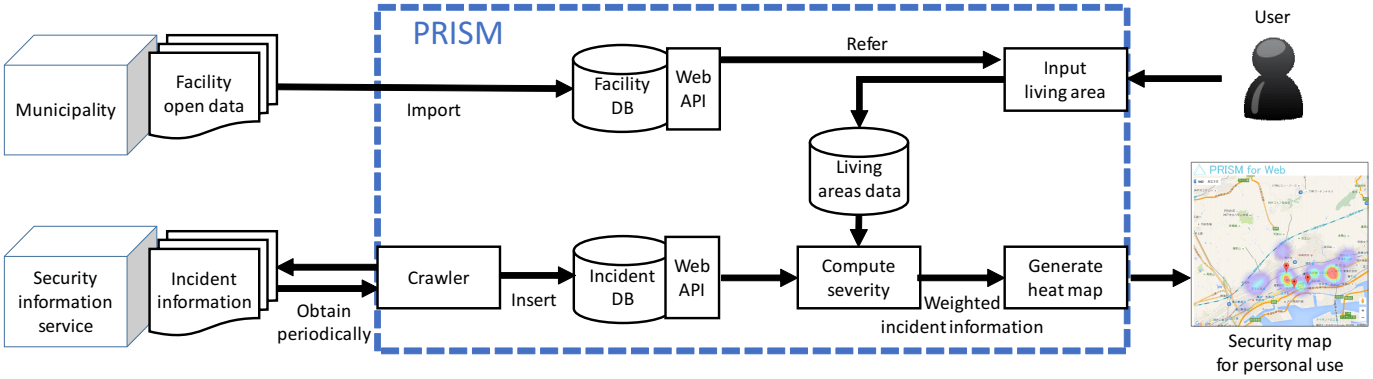


Fig. 2. Architecture of PRISM

http://hyogo-bouhan.net/criminalCase/detail.do?...&topicId=21939

Groping occurred

On May 12 (Fri) around 10:05, groping occurred on the street in the vicinity of 6-1, Shinkaichi, Hyogo-ku Kobe-shi, Hyogo. The suspect is a man who was wearing white cap and dark clothes.

Insert

Id	datetime	title	content	severity	address	lat	lng
21939	2017-05-12 10:05:00	Groping occurred	On May 12 (Friday) around 10:05, groping occurred ...	2	6-1, Shinkaichi, Hyogo-ku Kobe-shi, Hyogo	34.676265	135.175308

Fig. 3. Example of inserting incident information to incident DB

E. Registering Living Area with Open Data

PRISM manages user's living area as a set of coordinates. Accordingly, the user needs to register each living area with coordinates (latitude, longitude). With the help of API of map information service (e.g., Google maps), PRISM allows the user to input the coordinates just by double-clicking a point of a map in a Web browser.

On the other hand, a user, who are not good at reading the map, may not be able to point the exact coordinates on the map. To cope with such cases, we utilize *facility open data* published by local governments. Using nearby facilities as landmarks, the user can easily input places of living areas. The facility open data contains location information of various facilities in the region, including government offices, fire stations, police stations, stations, nursery schools, kindergartens, schools, parks, and cultural facilities. The facility open data is generally structured in the form of CSV or RDF.

In PRISM, we import the facility open data into *facility DB*, which is referred when registering the living area. The data schema of the facility DB is as follows:

- **id:** an identifier of a facility
- **name:** a name of the facility
- **postalCode:** postal code of the facility
- **address:** an address of the facility
- **phone:** a phone number of the facility
- **lat:** latitude of a place where the facility is located
- **lng:** longitude of a place where the facility is located
- **ward:** a ward where the facility is located
- **type:** a type of the facility (e.g. elementary school)

Similar to the incident DB, we develop Web-API for the facility DB as CGI handling JSON. External applications can search the facilities with the attributes and distance.

F. Computing Personalized Severity

Next, PRISM computes the severity of incidents according to individual circumstances, based on the living areas data. More specifically, PRISM adjusts the default severity of each incident according to the following two viewpoints.

Distance: The closer the distance between the place of incident and user's living area is, the more serious the incident is for the user, since a new incident may happen again nearby. Thus, higher weight is given to the incident. On the other hand, the longer the distance is, the smaller the severity is. When the distance exceeds a threshold Th_d, the severity is set to zero.

Time: The shorter the elapsed time from the incident occurrence, the more serious is for the user, since a new incident may happen again soon. Thus, higher weight is given to the incident. On the other hand, the longer the time is, the smaller the severity is. When the time exceeds the threshold Th_t, the severity is set to zero.

Now, for an incident x and a user u, let d be the distance from living area of u to the place where x occurred. Also, let t be the elapsed time from the time when x occurred. Then, we define the severity of x for u, denoted by $severity(x, u)$, as follows:

$$severity(x, u) = 1/3 * (WD(d) + WT(t) + severity(x) * 1/3)$$

where $WD(d)$ and $WT(t)$ are weight functions with respect to distance d and time t, respectively. $severity(x)$ represents the

default severity of x (ranging over -1, 1, 2 or 3) as defined in **III-C**. As for the definition of functions $WD(d)$ and $WT(t)$, there are various methods. In the current version of PRISM, we use the following functions.

$$WD(d) = \begin{cases} 1.0 & (0 \leq d < 0.5Th_d) \\ -2(d - Th_d)/Th_d & (0.5Th_d \leq d < Th_d) \\ 0.0 & (Th_d \leq d) \end{cases}$$

$$WT(t) = \begin{cases} 1.0 & (0 \leq t < 0.5Th_t) \\ -2(t - Th_t)/Th_t & (0.5Th_t \leq t < Th_t) \\ 0.0 & (Th_t \leq t) \end{cases}$$

The above functions maintain the weight 1.0 until the given distance (or time) reaches the half of the threshold, and decrease the weight linearly from 1.0 to 0.0 up to the threshold. Currently, Th_d is set to 4.0km, Th_t is set to 14 days. However, they are freely customized according to characteristics of the area as well as the crime situation of the region.

G. Heat Map

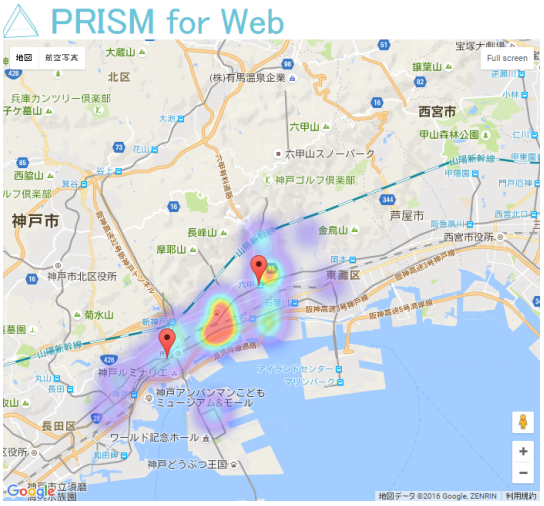
Finally, PRISM visualizes the incident data with personalized severity on heat map. The value of $severity(x, u)$, which is the severity of incident x for user u, is between 0.0 and 1.0. PRISM creates a heat map such that the severity value is scaled into the seven colors, [purple, indigo, blue, green, yellow, orange, red]. For each incident x, PRISM puts a data point on coordinates(lat, lng) of x using a color associated with $severity(x, u)$. This generates a heat map adapted to individual living area and the current time.

Figure4 shows two heat maps generated for two users A and B, where the incident information within Kobe City at a certain date is visualized. The pins in the map indicate locations of living areas registered by the users. The colored points indicate the places where incidents occurred. In this example, user A registered Kobe Sannomiya Station and Hankyu Rokko Station as living areas. On the other hand, user B registered Hankyu Rokko Station, Shukugawa Station, and Hanshin Fukae Station. We can see that completely different heat maps are generated depending on the living area, even though the map area and the time are the same.

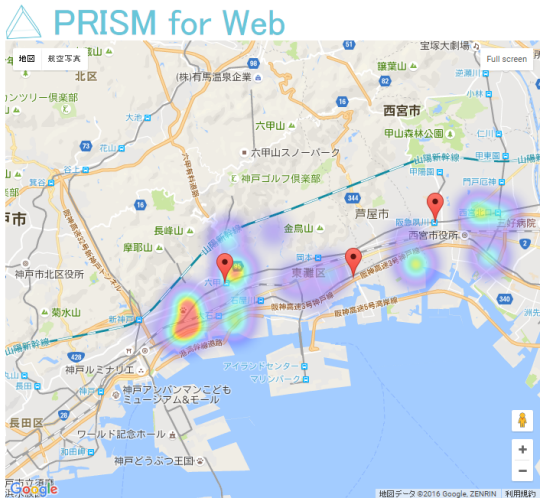
IV. Implementation

We have implemented the proposed PRISM as a web application and a mobile application. Figure5(a) and (b) show screenshots of PRISM for Web and PRISM Android, respectively. In the proposed architecture (see Figure2), the crawler, the incident DB, the facility DB and their Web-API were implemented in the server side. Whereas, the registration of living area and the generation of the heat map were implemented in the client side. Technologies used for the implementation are as follows.

- **Security information service:** Hyogo Bouhan Net
- **Facility open data :** facility open data of Kobe City
- **Crawler:** perl, cron, Yahoo!GeoCoder API [7]
- **Incident DB, facility DB:** MySQL



(a) User A's heat map



(b) User B's heat map

Fig. 4. Example of outputting heat map

- **Web-API:** perl, CGI.pm, JSON.pm
- **Client (Web application) :** Chrome browser, JavaScript, Google Maps, Google Heatmap Layer
- **Client (mobile) :** Android, Google Maps, Android Heatmap Utility

V. Discussion

A. Sufficiency of Requirement

We here discuss how PRISM cope with Challenges P1, P2, and P3 in **II-C**. As mentioned in **III-F**, PRISM calculates the severity of incidents based on the living area and the elapsed time of the incident, and visualizes incidents on a heat map. Thus, a security map adapted to every user is generated. A closer and newer incident is shown to be more serious in the map. Thus, P1 and P2 are resolved. Moreover, as mentioned in **III-C**, four types of default severity based on keywords in

(a) PRISM for Web

(b) PRISM Android

Fig. 5. Screenshots of implemented PRISM

title of incident information in PRISM, which corresponds to P3. However, it is necessary to verify the validity of the value, which is left for our future work.

B. Limitations in Using Data

Regarding the incident information provided by existing security information services, the crawler converts unstructured text data into structured data by text analysis (see **III-C**). For this, if we switch to another security information service, we have to re-create the document parser. Such tight coupling of the crawler and the security information service limits the extensibility of the system. However, this problem will be alleviated if the incident data is published as structured open data.

We used facility open data of Kobe City in this study. However, the data did not contain the sufficient number of familiar facilities (e.g., railway stations). More facilities are necessary to cover more users living in various places.

C. Related Work

"Hyogo Anzen Anshin Map" [8] is a Web application quite similar to PRISM. The application analyzes incident information of "Hyogo Bouhan Net", and display the incident on Google map. However, it does not consider living area or personalization.

VI. Conclusion

In this paper, we have proposed a personalized and real-time security information service, called PRISM. Using the incident information provided by the existing security information services, PRISM adapts the incident to individuals based on living area of the user. We introduced a metric, called *severity*, which quantifies how the incident is serious for the user. The severity is computed based on the distance from the living area and the elapsed time. Incidents weighted with the severity are visualized on a heat map. We also implemented proposed PRISM as a web application and a mobile application.

In our future work, we will verify the validity of the value of the severity, as well as the weight functions described in **III-F**. We also plan to have many users actually use PRISM, and evaluate the usability.

Acknowledgments

This research was partially supported by the Japan Ministry of Education, Science, Sports, and Culture [Grant-in-Aid for Scientific Research (B) (No.16H02908, No.15H02701), Challenging Exploratory Research (15K12020)], and Tateishi Science and Technology Foundation (C) (No.2177004).

References

[1] S. Tulumello, "The multiscalar nature of urban security and public safety," *Urban Affairs Review*, vol. 0, no. 0, p. 1078087417699532, 0. [Online]. Available: http://dx.doi.org/10.1177/1078087417699532
[2] Hyogo Prefectural Police, "Hyogo Bouhan Net," https://hyogo-bouhan.net/.
[3] Tokyo MetroPolitan Police Department, "Tokyo Crime Map," http://www2.wagmap.jp/jouhomap/Portal?langmode=1.
[4] J. C. Molloy, "The open knowledge foundation: Open data means better science," *PLOS Biology*, vol. 9, no. 12, pp. 1–4, 12 2011. [Online]. Available: https://doi.org/10.1371/journal.pbio.1001195
[5] Kobe city, "Open Data Kobe," https://data.city.kobe.lg.jp/.
[6] Japanese government, "DATA.GO.JP," http://www.data.go.jp/?lang=english.
[7] "Yahoo!Geocoder API," https://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/geocoder.html.
[8] "Hyogo Anzen Anshin Map," http://anzn.net/hyogo/safety/index.html.