III. RESULT AND DISCUSSION

To effectively contend with crime from happening, a Bi-layer Crime Prevention Framework (BCPF) is proposed. BCPF aims to reduce the crime rate by involving law enforcement agency (LEA) through the utilization of various significant data sources and techniques effectively. The main concept is to maximize the LEA's operational efficiency and in the meantime, gradually deter criminal behavior through environmental design.

A. Framework Overview

BCPF, as the name implies, is a framework that prevents crime in two layers.

- Layer 1: It discerns crime hotspot areas by analyzing big data through text mining. Short-term and long-term crime prediction techniques are used jointly to predict crime, which supports the planning of LEA's daily operation.
- Layer 2: It selectively activates for further crime analytics when the first layer fails. It performs crime analytics on additional data sources from the telecommunication company to allocate the most suspicious individual, which enables LEA's operation planning in a dynamic environment.

Simultaneously, both layers will disclose valuable information with the goal of deterring crime through environmental design [6]. Fig. 1 illustrates the overview of BCPF framework.

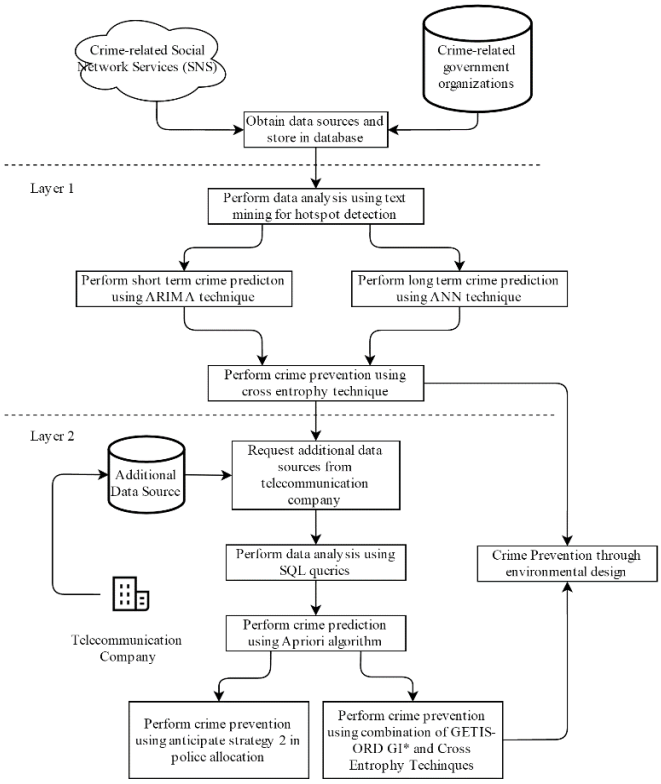

Fig. 1 BCPF Framework Overview

B. Data Collection & Preprocessing

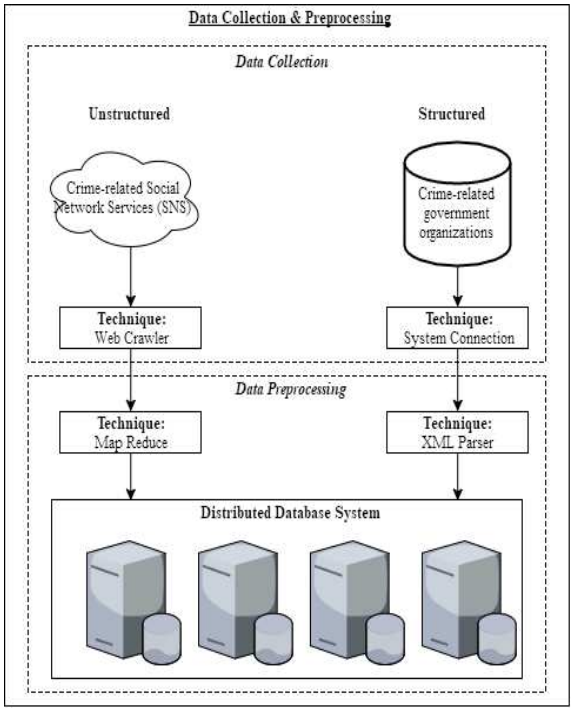

Fig. 2 Data Collection & Data Preprocessing of BCPF

Fig. 2 illustrates the data collection and preprocessing processes and techniques used in BCPF. Crime data is collected from Social Network Services (SNS) and government organizations. The formal crime data is unstructured, and the recent crime data is structured. The techniques used are depending on its data types. SNS data is collected by a web crawler while the government organization data is collected by system connections [5].

The collected SNS data is preprocessed using map reduce technique, which aims to reduce and eliminate the duplication of data. This can be reasoned by the design of crawler, which iteratively crawls data from the World Wide Web where different web pages with the same content are crawled and duplicated [27]. As for collected government organization data, the data is parsed with XML parser to the desired format before storing in the database.

C. BCPF Layer 1

From the crime data collected and preprocessed, text mining analysis technique is used for hotspot detection. Text mining is used to omit words and to determine crime and location of related words. With the crime and location classified, the hotspot detection will be identified according to the frequency of the words appearing throughout the data in text mining. The highest frequency of crime appearing will serve as a benchmark for assigning color on the map. The input will be the preprocessed data above for the crime prediction. Predictions fall into two categories: long-term and short-term. For the long-term, Artificial Neural Network (ANN) method will be used. The input is the crime cluster dataset consisting of crime type, the number of occurrences, and the cluster hotspots map with centroid [15]. This will

output accurate annual prediction of crime, which can help the LEA to focus on dividing the task among them.

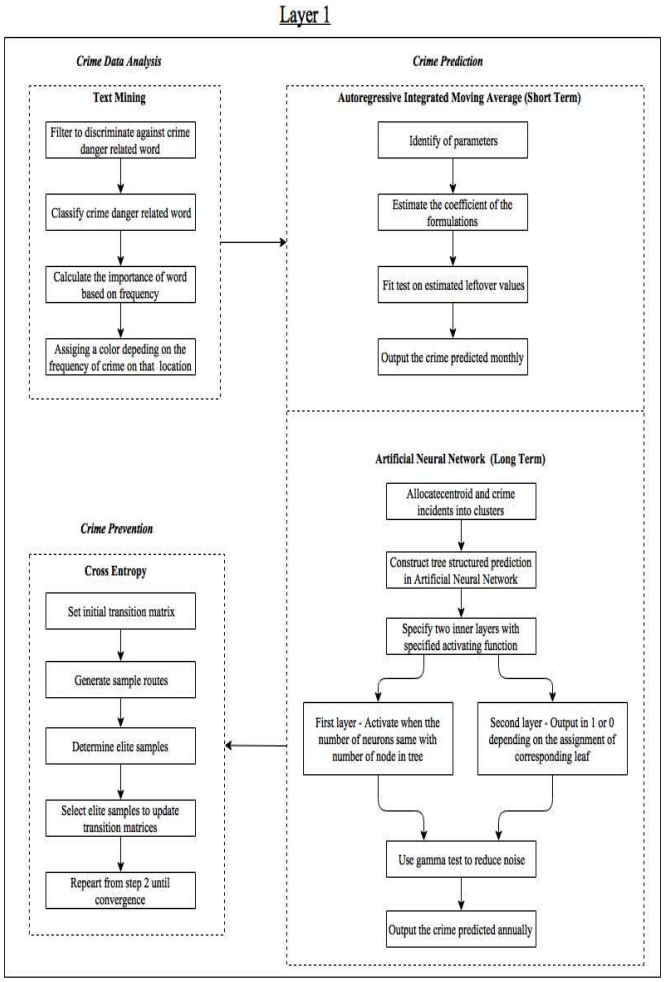

Fig. 3 BCPF Layer 1

LEA cannot depend on the annual prediction and keep the plan unchanged throughout the year. Changes might happen to affect the prediction, which leads to short-term prediction. For the short-term prediction, Autoregressive integrated moving average (ARIMA) method is used. ARIMA model with the input of the previous month of crime cluster data will predict the sudden changes in the dataset which will provide a more accurate prediction for the following month [16], [17].

Predicted data will be utilized by integrating them as the guidance for crime prevention. Crime prevention strategy will encompass two components, which focus on daily operational planning and long-term environmental design. The formal engages cross-entropy method and randomize strategy with the data predicted in crime prediction for patrol routes planning, while the latter will utilize the data collected in daily operation that observe the attributes of environment design of the crime scene and use it for city planning.

D. BCPF Layer 2

Fig. 4 illustrates the detail techniques and the steps proposed in big data analytics as well as some prevention strategies in Layer 2. For further data analysis, the LEA requests the higher authority for the Call Data Records (CDR) of the mobile numbers of the area where the crime is committed [8].

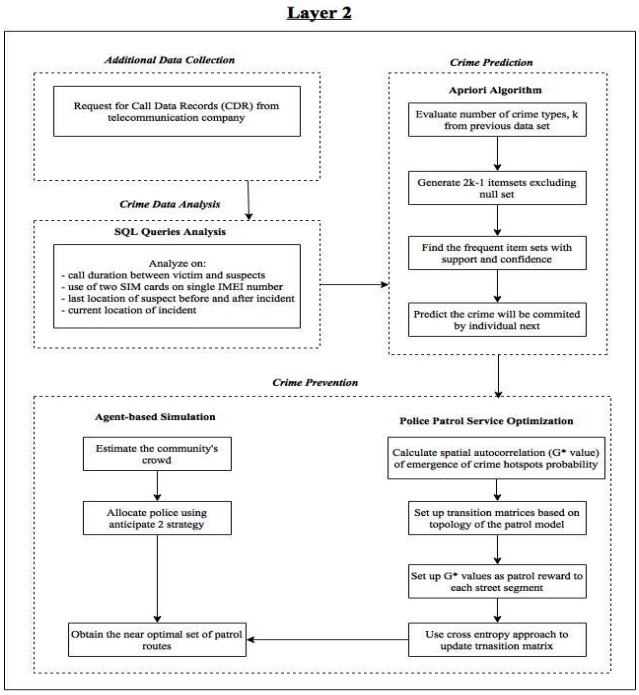

Fig. 4 BCPF Layer 2

SQL queries analysis is applied to the CDR collected. The details of CDR are shown in Table 1.

TABLE I
INFORMATION IN CDR

Call Data Record (CDR)
Phone Number of Subscriber (on who's name the Number is Registered)
Phone Numbers of the parties on the other end of the Calls.
Time and Date of Calls
Call Durations
Sequence Number
Route by which Call Entered the Exchange
Route by Which the call left the Exchange

With the required criterions, the result of the suspicious target is then given as output. The output will then be used to predict the possibility of crime an individual will be committing in future through an Apriori algorithm. Apriori algorithm is an algorithm used to reduce the number of the item set in the sense that if the item set is frequent for the individual, then the individual must also be frequent to all the subsets [21].

In order to enhance the crime prevention strategy, layer 2 will have a broader focus area, which involves the allocation of police. Some police will be allocated according to the crime hotspots predicted. Meanwhile, for the patrol routes planning, a combination of Getis-ORD Gi* score and cross entropy methods will be used and applied in a dynamic environment. Getis-ORD Gi* score will be given according to the predicted probability of crime emerging in each street [24]. Turnout data from the Apriori algorithm will not only be used to predict crime but also be used to provide insight for the community's estimation crowd. The safety of location with an estimated increment in the crowd will be enforced by increasing the number of police [26]. Data collection will continue to feed and exploit for city planning.

E. Discussion

This BCPF is proposed after courteous consideration from each of the techniques used by other researchers in big data analytics. It envelopes from the crime data collection and preprocessing stage, to crime data analysis, followed by crime prediction and prevention.

The rationale of having the second layer in the BCPF is to aid the LEA operational planning in an immediate and dynamic environment. There is no absolute prediction when the prediction is based on the historical dataset. As long as there is a new criminal or crime-taking place, the prediction is no longer feasible. At this moment, the second layer serves as a countermeasure where it performs additional data analytics to counter this problem.

IV. Conclusion

Crime is expected to increase by the improvement of technology that provides the convenience for the criminal to commit crime. There is always a challenge to prevent crime using big data analytics by handling the enormous volume of data generated from various sources.

In this paper, a bi-layer crime prevention framework using big data analytics is presented. The uniqueness of this framework is the proper strategy designed to address the challenges, whereby the first and second layers not only ease the LEA's operational planning, but also continuously deliver valuable data for city's environmental design. Hence, a safe and secure city is molded in the future.

Further investigations would need to drive artificial intelligence techniques into big data analytics. To contend with criminal minds, big data analytics with artificial intelligence might shape in the future. Besides, future research will be considered to focus on preventing cybercrime as recent WannaCry ransomware attack that affects more than 230,000 individuals across 150 countries is believed to be a nightmare and might resurface in the future [28].

Acknowledgment

We would like to express our gratitude to Universiti Sains Malaysia (USM) for supporting this research.

References

[1] (2017) UrbiStat AdminStat in Demography [Online]. Available: https://ugeo.urbistat.com/AdminStat/en/fr/demografia/dati-sintesi/paris/75/3

[2] (2017) "IoT: number of connecting devices worldwide 2012-2025 | Statista", Statista [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[3] G. Marc, *Future crimes: Everything is connected, everyone is vulnerable and what we can do about it*, 2015.

[4] (2017) TechTarget Search Business Analytics [Online]. Available: http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics

[5] J. Jeon, and S. Jeong, "Designing a Crime-Prevention System by Converging Big Data and IoT," in *Journal of Internet Computing and Services (JICS)*, 2016, pp. 115-128.

[6] D. W. Sohn, "D. (2016). Residential crimes and neighborhood built environment: Assessing the effectiveness of crime prevention through environmental design (CPTED)," 2016). Cities, 52, pp.86-93.

[7] V. Luca, B. Elena, and T. Politecnico, "A spectral analysis of crimes in San Francisco," 2016.

[8] K. S. Er., A. Firoz, A. D. Hamza, and Sabigua, "Criminal Investigation Using Call Data Records (CDR) through Big Data Technology," in *International Conference on Nascent Technologies in the Engineering Field*, 2017.

[9] F. B. B. Nasution, N. E. N. Bazin, Daliyusmanto, and A. Zulfikar, "Big Data's Tools for Internet Data Analytics: Modelling of System Dynamics," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 3, pp. 745-753, 2017. [Online]. Available: http://dx.doi.org/10.18517/ijaseit.07.3.1088

[10] J. Isuru, S. Chamath, L. Sampath, W. Tharindu, P. Indika and W. Adeesha, "Crime Analytics: Analysis of Crimes Through Newspaper Articles, 2015.

[11] Y. Zhang, "Analysis of Crime Factors Correlation Based on Data Mining Technology," in *International Conference on Robots & Intelligent System*, 2016.

[12] R. Umar, I. Riadi, and G. M. Zamroni, "Mobile Forensic Tools Evaluation for Digital Crime Investigation," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no.3, pp. 949-955, 2018. [Online]. Available: http://dx.doi.org/10.18157/ijaseit.8.3.3591

[13] L. Teng, "Criminal Behavior Analysis Method based on Data Mining Technology," in *International Conference on Smart City and System Engineering*, 2016.

[14] A. Turki, S. Daming, W. David, and W. William, "Criminal Pattern Identification Based on Modified K-means Clustering," in *International Conference on Machine Learning and Cybernetics, Jeju, South Korea*, 2016.

[15] Soumya and A. S. Baghel, "A Predictive Model for Mapping Crime using Big Data Analytics," vol. 04, no. 04, pp. 344–348, Apr. 2015.

[16] P. Chen, H. Yuan, and X. Shu, "Forecasting Crime Using the ARIMA Model," in *Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, Nov. 2008.

[17] N. M. M. Noor, A. Retnowardh, M. L. Abd, and M. Y. M. Saman, "Crime Forecasting using ARIMA Model and Fuzzy Alpha-cut," I in *Journal of Applied Sciences*, vol. 13, no. 1, pp. 167–172, Feb. 2013.

[18] A. Ghazvini, M. Z. B. A. Nazri, S. N. H. S. Abdullah, M. N. Junoh and Z. Abidin bin Kasim, "Biography commercial serial crime analysis using enhanced dynamic neural network," *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, Fukuoka, 2015, pp. 334-339.

[19] J. Xu, "Predict Future Events with Point-process Modeling," IBM Big Data & Analytics Hub, 18-Sep-2015. [Online]. Available: http://www.ibmbigdatahub.com/blog/predict-future-events-point-process-modeling

[20] M. A. Awal, J. Rabbi, S. I. Hossain, and M. M. A. Hashem, "Using Linear Regression to Forecast Future Trends in Crime of Bangladesh," in *5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 2016.

[21] A. Agarwal, D. Chougule, A. Agarwal, and D. Chimote, "Application for Analysis and Prediction of Crime Data using Data Mining," in *International Journal of Advanced Computational Engineering and Networking*, ISSN: 2320-2106., vol. 4, no. 5, pp. 9–12, May 2016.

[22] (2017) "United Nations Office on Drugs and Crime [Online]. Available: https://www.unodc.org/unodc/en/justice-and-prison-reform/CrimePrevention.html

[23] C. Xu and T. S. P. Yum, "Cross Entropy approach for patrol route planning in dynamic environments," *in 2010 IEEE International Conference on Intelligence and Security Informatics, Vancouver*, BC, 2010, pp. 114-119.

[24] L. Li, Z. Jiang, N. Duan, W. Dong, K. Hu, and W. Sun "Police patrol service optimization based on the spatial pattern of hotspots," in *Service Operations Logistics and Informatics (SOLI) 2011 IEEE International Conference*, on, pp. 45-50 July 2011.

[25] R. Danilo, M. Adriano, L. V. C. André and F. Vasco, "Towards Optimal Police Patrol Routes with Genetic Algorithms," in *S. Mehrotra et al. (Eds.): ISI 2006*, 2006.

[26] T. Bosse and C. Gerritsen, "Comparing Crime Prevention Strategies by Agent-Based Simulation," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Milan, Italy, 2009, pp. 491-496.

[27] (2012) "Web Crawling: Data Scraping vs. Data Crawling [Online] Available: https://www.promptcloud.com/data-scraping-vs-data-crawling/

[28] CERT-MU, *The WannaCry Ransomeware: White Paper*, May 2017.