Evaluation of Unsupervised Classification on Police Patrol Zone Design Problem

1st Semih Dinc*
Department of Math and Computer Science
Auburn University in Montgomery
Montgomery, AL
sdinc@aum.edu

2nd Imren Dinc*
Computer Science Department
Troy University
Montgomery, AL
idinc@troy.edu

***Abstract*—Police patrols are one of the effective ways to respond a recent incident or to prevent a likely crime. However, due to limited number of available officers, not all streets can be patrolled continuously. To minimize the potential crimes, available forces should be systematically assigned to their zones, where an incident is most likely to happen. In this study, we propose a new technique that automatically determines the patrol zones and shifts, then assigns available officers to the zones. Our technique utilizes the previous crime data of the region and employs unsupervised classification to automatically identify zones. Then, it estimates a crime weight for each zone based on several factors, such as the probability and type of an incident. Finally, it assigns officers to the zones by prioritizing crime weight estimations.**

As a case study, we tested our system on the city of Montgomery, AL. We used the crime dataset including all reported 19687 incidents in 2017. We have built our scheduling system using 11 months of data and then tested the system using the incidents happened in the last month. Test results show that, with the proposed technique, scheduled officers are most likely to present nearby the incident before it happens. We measured 7 minutes of response time on the average, which is lower than the national average.

***Index Terms*—Patrol Scheduling, Patrol Zone Design, Crime Prevention, Maximal Coverage Locations, Montgomery**

I. Introduction

Patrol officers are assigned to designated districts (or zones) for monitoring possible incidents or crimes. This duty is one of the essential parts of public safety as the officers can provide a rapid response to calls for service, protection, crime prevention, and a good public relation. However, responding to the incidents, or to prevent a likely crime in a US metropolitan area is a challenging and demanding task for the local police force [1]. Limited resources and the manpower must be utilized efficiently to minimize the average response time to the incidents.

The response time is a vital factor to reduce the harmful consequences of an incident. And a rapid response depends on several factors, such as the size of the patrol zone, distance to the incident, time of the day, and traffic and weather conditions [2]. Although some of these factors cannot be controlled (e.g. traffic and weather), smart decisions on others can significantly reduce the response time. Considering all these factors, this problem can be transformed into an optimization problem where the objective function is the minimum average response time to incidents.

In the history of police patrol, a good number of techniques have been tried to optimize the utilization of resources. Different studies focused on different aspects of the problem, such as length of patrol shifts [3], patrol type (car patrol vs. foot patrol) [4], or the patrol zone design [5]. This study however, focuses on optimizing the patrol zones and shift schedules for motorized patrols.

Patrol zone design is a complicated task. Each zone is expected to be equivalent in terms of size, population, crime rate and type, etc. However, in reality, every district in a city may have different characteristic to consider individually. A traditional zoning approach is manual (hand drawn) zoning that is a time consuming process, and cannot easily adjust to changes on crime patterns [6]. Computer based methods are good alternatives to include many complex factors and goals into the system. These methods consider the problem as a general zoning problem called the Maximal Covering Location Problem (MCLP). In MCLP, the goal is to maximize the coverage of available data points (incidents) by determining the optimal locations of zones. Number of zones can be variable or predetermined. And zone design can involve multiple constraints, such as max size, population, and density [7].

One of the early implementations of computer based methods is [8] that used p-Median clustering to minimize the total weighted travel distance. Amico et al. [9] consider the problem as a constrained graph-partitioning problem and propose a simulated annealing algorithm to search for a good partitioning of command districts. Curtin et al. [6] applied a variation of maximal covering formulations to determine optimal patrol zones. Zhang et al. [10] proposed an algorithm, where a small atomic units are selected as seeds from data points. The patrol zones are determined by incrementally growing the seed areas.

Majority of previous studies consider the zoning problem as whole. However, several other studies suggest that crime rates of patrol zones change during the day [11], [12]. In other words, an optimal patrol zone design may be different on different hours of the day. We started from this motivation and made a similar analysis for our case as reported in Figure 1. As can be seen from the figure that number of incidents occurred

*Both authors contributed equally to this manuscript

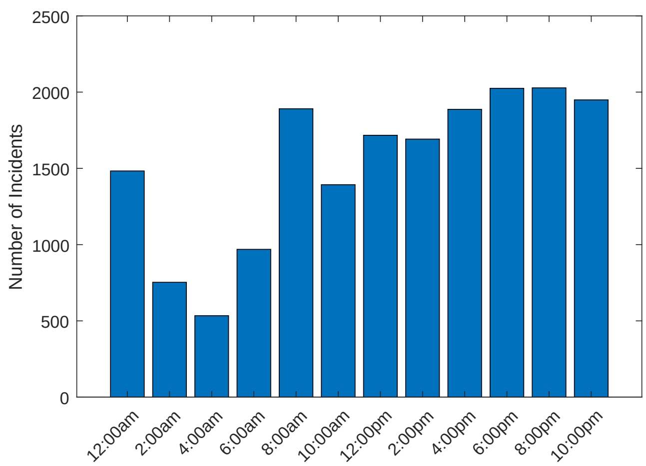

Fig. 1. Number of incidents for each time period. (City of Montgomery, AL, 2017)

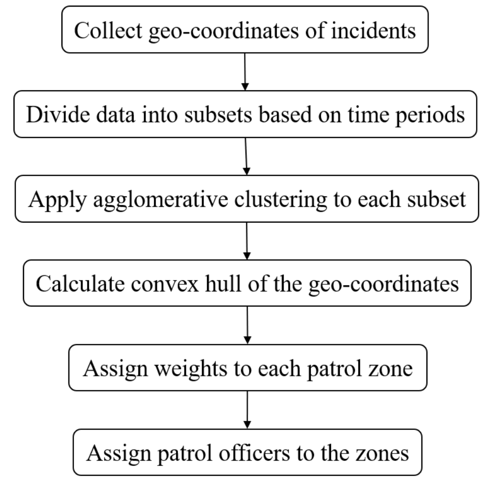

Fig. 2. The overview of proposed method.

in different hours is significantly different.

This paper investigates the benefits of hourly zone determination and scheduling for police patrols. Our solution allows more dynamic zoning system, where different patrolling zones can be formed for different hours of the day. In this way, patrol officers can be utilized more efficiently as they can be re-assigned to important hourly zones during the day. To determine the patrol zones, we use an unsupervised classification technique called "Hierarchical Clustering" that efficiently group data points based on their distances to each other. Hierarchical clustering is beneficial for this particular problem because, by its nature, it can easily adopt the changes on the number of patrol zones.

To evaluate the efficiency of our approach, we performed a simulation using a real dataset (all reported incidents in 2017) published by Montgomery Police Department (MPD). We used 93% of the data points for determining the hourly zones and reserved 7% of the points for testing in our simulations. Test results yield low average response times as reported in Section IV.

II. Proposed Approach

In this paper, we propose a new approach to design patrol zones using hierarchical clustering technique. Unlike the manual zone design and existing computerized techniques, our methodology determines the zone based on given time periods, which allows more dynamic scheduling thus a better resource utilization. A crime weight is assigned to each patrol zone so that available patrol officer can be assigned more "hot zones" first. An overview of our system is provided in Figure 2.

In the following, first a brief information is given about the hierarchical clustering method and the distance measure we adopted. Later, a detailed description is provided about how each module is integrated, and data is processed in our method.

A. Hierarchical Clustering

Hierarchical clustering (HC) is one of the widely used data analysis methods in unsupervised classification [13]. In HC, a dendrogram (hierarchical tree) of clusters is built using two alternative approaches: 1) Agglomerative (bottom-up) clustering 2) Divisive (top-down) clustering. The bottom-up approach starts building the tree by assuming all individual data points are separate clusters, and combines them into larger clusters through the upper levels of the tree. However, the top-down approach starts from the top of the tree by assuming all data points belong to single big cluster, and divides the them into smaller clusters through the bottom of the tree.

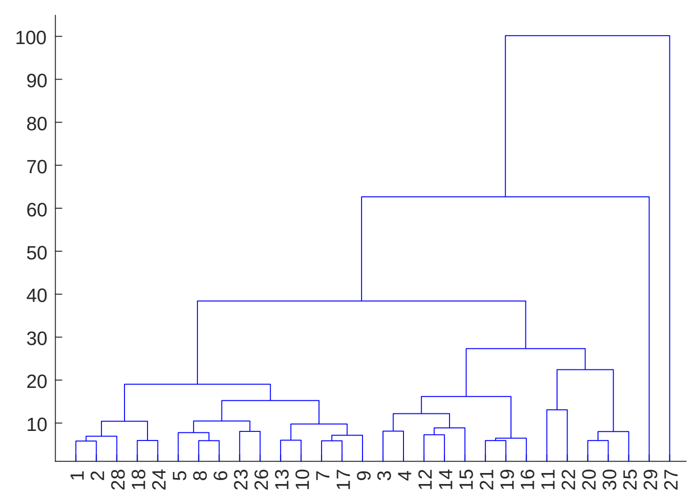

Fig. 3. Dendrogram of the hierarchical agglomerative clustering.

Divisive clustering is usually more complex and requires exhaustive search without using heuristics. Therefore, in this paper, we utilize agglomerative hierarchical clustering (AHC) for zone design. At the beginning, each incident is considered as a separate patrol zone, and two closest zones are iteratively merged until we reach the pre-selected number of zones.

Figure 3 shows a sample tree built by agglomerative clustering technique.

B. Distance Measure

An important decision on agglomerative clustering is to determine the proximity between the clusters. There are several approaches to calculate proximities such as single (MIN) link, complete (MAX) link, and the group average. Figure 4 shows min and max link approaches. In single link, the distance between two clusters is accepted as the minimum distance between any two data points between the clusters. On the other hand, the complete link calculates the distance only considering the farthest points between the clusters. In this paper, we use complete link approach on our problem not to have wide clusters.

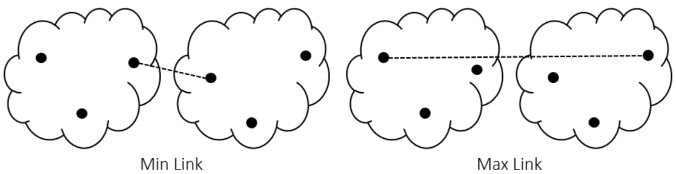

Fig. 4. Distance Measures. Single (Min) link is on the left, and Complete (Max) link is on the right.

Clustering methods rely on proximity measures and finding the right measure for the problem is crucial to have reliable results. In this work, the dataset contains the physical addresses of the crime entities, thus we first converted these addresses into the geographical coordinates (i.e. longitude, latitude) using Google Maps Application Program Interface (API). Then, we used the Haversine formula [14] to calculate distance d between two crime entity c_1 and c_2 as follows:

$$d(c_1, c_2) = 2r\,arcsin\left(\sqrt{sin^2(\tfrac{\phi_2-\phi_1}{2}) + cos(\phi_1)cos(\phi_2)sin^2(\tfrac{\lambda_2-\lambda_1}{2})}\right) \quad (1)$$

where r is the radius of the earth, ϕ and λ are the is latitude and longitude of the crime entity, respectively.

C. Patrol Zone Design using HAC

We adopted the following notation in our calculations. First, we divide a 24-hour day into a set of m equal sized time periods,

$$T = \{t_i | \quad 1 \leq i \leq m\}, \quad (2)$$

where t_i is the i^{th} time period, and each period is p hours long. Let C_i be a crime dataset including all the incidents in t_i such as

$$C_i = \left\{c_j^{t_i} | \quad 1 \leq j \leq n_i\right\}, \quad (3)$$

where $c_j^{t_i}$ is the j^{th} incident happened in time t_i, and n_i is the number of incidents in the dataset C_i. Every $c_j^{t_i}$ consists of following attributes:

$$c_j^{t_i} = \{a_j, b_j, c_j\}. \quad (4)$$

Attributes a_j, b_j, and c_j represents the address, type, victim features (e.g. age, gender, etc.) of the incident j, respectively. Let Z_i be a set of dynamically determined patrol zones for time period t_i such that

$$Z_i = \left\{z_k^{t_i} | \quad 1 \leq k \leq s_i\right\} \quad (5)$$

where $z_k^{t_i}$ is the k^{th} patrol zone, and s_i is the number of patrolling zones selected for t_i. Each patrol zone is comprised of a set of coordinates that form a region on the map.

In the first stage of the proposed method, the overall dataset is divided into $m = 24/p$ number of subsets. In other words, for each distinct time period t_i, a small dataset C_i is generated. Later, we applied agglomerative hierarchical clustering to each subset C_i using the complete link approach.

We create s_i number of clusters using the clustering tree, which is the output of the hierarchical clustering. Here, the clusters having few number of incidents are ignored for further analysis. Then, using the geographical locations of the incidents in every zone $z_k^{t_i}$, a convex hull is calculated. In this step, only the longitude and latitude of the incident locations are employed to determine the zone.

Since the available number of patrol officers is limited, a prioritization of patrol zones is necessary. Therefore, a total crime weight is calculated for each $z_k^{t_i}$ using the attributes and the frequency of the incidents.

TABLE I
CRIME WEIGHTS USED IN THE CASE STUDY.

Crime Type	**Weight**
Violent Crime	1
Rape and Sexual Assault	0.8
Robbery	0.7
Burglary	0.6
Theft	0.5
Traffic Violations	0.3
Disturbance	0.4
Anything related to an investigation	0.1
Other	0.2

The total crime weight of a zone ω_{z_k} is calculated as follows:

$$\omega_{z_k} = \sum_j \alpha_j H(z_k^{t_i}, c_j^{t_i}) \quad (6)$$

where α_j is the crime weight of incident from the Table I. $H(z_k^{t_i}, c_j^{t_i})$ is the function that indicates whether incident $c_j^{t_i}$ is in zone $z_k^{t_i}$ or not. It is defined as follows:

$$H(z_k^{t_i}, c_j^{t_i}) = \begin{cases} 1, & if \quad c_j^{t_i} \in z_k^{t_i} \\ 0, & otherwise \end{cases} \quad (7)$$

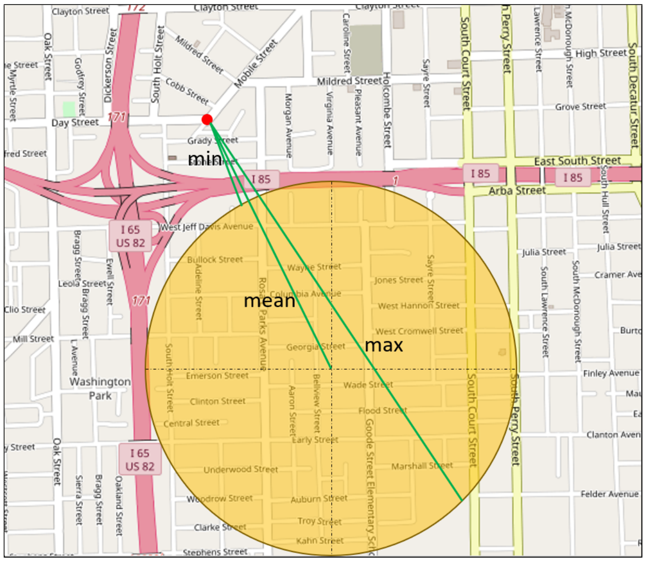

Fig. 5. Estimating the average response time. Red point represents the incident location and yellow circle is the closest patrol zone. Three different distances are calculated to estimate average response time to the incident.

Please note that an expert help would be beneficial to assign appropriate α_j values. Therefore, we adopted similar crime weight values reported in [2].

D. Evaluation

A numeric and objective evaluation of patrol zones is not a straightforward task. The ideal solution is to test new patrolling zone with real officers and incidents. However, a simulation based evaluation must be made primarily to observe general behavior of the system. The simulation should reflect similar results to a real case. To achieve this goal, we reserved a set of real incidents to test our patrolling zones. We generated a simulated environment on MATLAB with Open Street Maps API to estimate an average response times to the incidents. We simulated the all test incidents by running on their real recorded time.

When an incident occurred, we measured distance between the incident address and the officer in the closest patrol zone. The officer can be in any location in the patrol zone. Therefore, as shown in the Figure 5, we calculated the closest, the farthest, and the mean distance between the zone and the incident. In this way, a more realistic observation can be made. Later, we estimated a response time based on the travel time between the addresses. We performed this simulation for all test incidents and recorded response times. Finally, we reported an average response time for each zone, and one average response time for all incidents in all zones.

III. Case Study: Montgomery Patrol Scheduling

City of Montgomery is the historic capital of US State of Alabama. City has a total area of 156.2 square miles with the population of around 202,000 residents as well as 150,000 others who travel to or through the city each day. Majority of the population lives in single family houses located suburban areas and commute to the downtown where the government and large commercial buildings are located. According to the 2016 FBI report [15], violent crime rates are above the US average by 4.72 per 1000 residents.

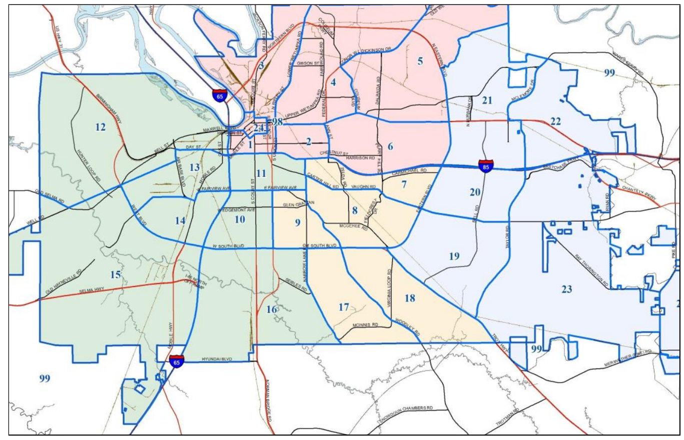

Fig. 6. Current patrol sectors and districts in Montgomery [16]. Red and green districts belong to sector A, yellow and blue districts belong to sector B.

The Montgomery Police Department was established in 1820 and currently has 524 officers. As shown in Figure 6, city is currently divided into 2 main patrol sectors and total of 4 (north, south, east, west) commander districts. Every district is manually divided into patrol zones. A special attention is paid to the hot zones, where the crime rates are usually high. Patrol officers are assigned as two man patrols and work 12 hours shifts.

A. Dataset

To design and train our model, we used the crime data [17] published by the Montgomery Police Department. This dataset is publicly available and daily updated. Each reported incident has a detailed information such as occurred day and time, crime type, address, responded officer, etc. For this study, we acquired a year of data including total of **19687** incidents from Dec 15th 2016 to Dec 15th 2017. Figure 7 shows the Montgomery City map with all occurred incidents in 2017. Among all, 11 months of data **(18396 incidents)** is used for building the our model. The incidents occurred in the last month **(1291 incidents)** are used to test our patrol zones.

IV. Simulation Results

We performed simulations to evaluate the effectives of the proposed approach. Adopting the reported number in Montgomery Police Department website, we assigned a total 50 officers to patrol duty for a 12 hour shift. Using a two-man patrolling schedule, 25 patrol vehicles assigned to patrol zones. In other words, 25 patrol zones ($s = 25$) are determined for each time period. We used a two-hour time period ($p = 2$) for patrol zone design. Thus, there are 12 time periods to consider ($m = 12$). Patrol zones in each time period are designed individually, which means patrol zone sizes and locations change for different time periods. A crime weight is assigned

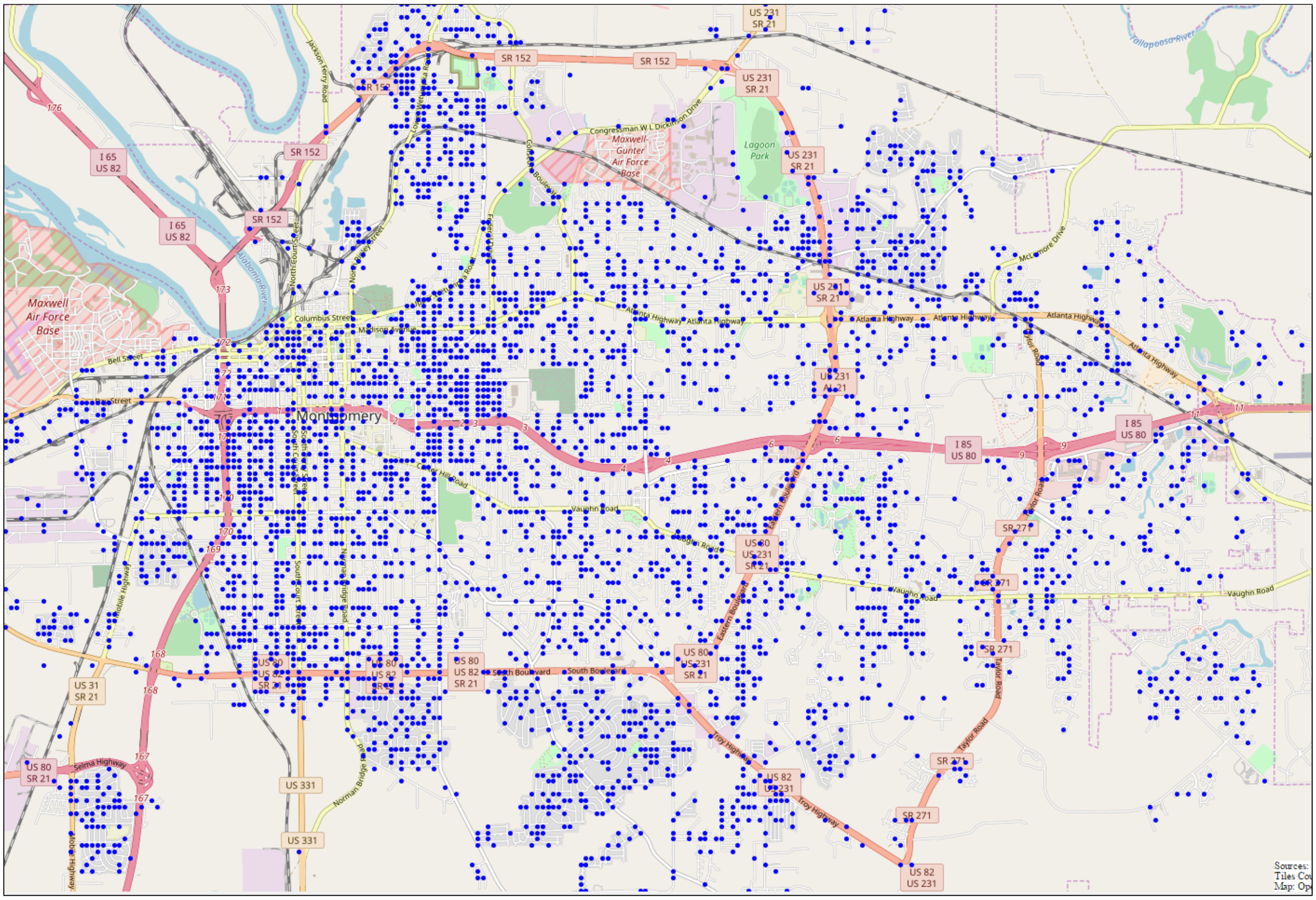

Fig. 7. All call for service incidents happened in City of Montgomery, AL in 2017.

to each zone by employing the crime types reported in Table I. Available patrol officers are assigned starting from the zone with the highest crime weight.

In the first part of the simulations, we focus on the efficient patrol zone design and estimating the response times of test incidents. In the second part, we focus on resource allocation.

A. Average Response Times

Average response time for a test incident is estimated by calculating a travel time between the test incident location and the closest zone. Since the patrol vehicle can be in any location in the zone, three travel times (minimum, mean, and maximum) are calculated for more reliable results. All test incidents examined individually, and response times were estimated. The results reported in Table II shows only the travel times. We can assume an additional one minute to answer the call and inform closest officer about the incident. Therefore, it is an acceptable assumption that the officer can reach to the incident location in $(6.09 + 1)$ minutes. We are unable to access the current average response times in Montgomery. However, according to American Police Beat Magazine, national average is around 10 minutes.

TABLE II
AVERAGE RESPONSE TIMES OF ALL TEST INCIDENTS

Min (minutes)	**Mean** (minutes)	**Max** (minutes)
4.78	6.09	8.81

Since we make a separate analysis, a zoning map is designed for each time period. Figure 8 shows an example zoning map, where $t_i = 6$ (10:00am to 12:00pm). Each zone is color coded in the blue-red spectrum. Blue color tones represent low crime rates, while red tones represent high crime rates. Our zoning approach targets to cover "hot zones" in the city, so it does not cover complete city limits. If an incident occurs outside of a zone, the closest officers in the near zones are assigned to the incident.

We evaluated each individual time periods, and estimated average response times for each time period. This way, time periods can be examined and different number of patrol zones can be determined for different time periods. For instance, as Figure 1 shows, in time period 4:00am to 6:00am $(t_i = 3)$, it is less likely to happen an incident. So, less number of patrol officers (or less number of zones) can be assigned for

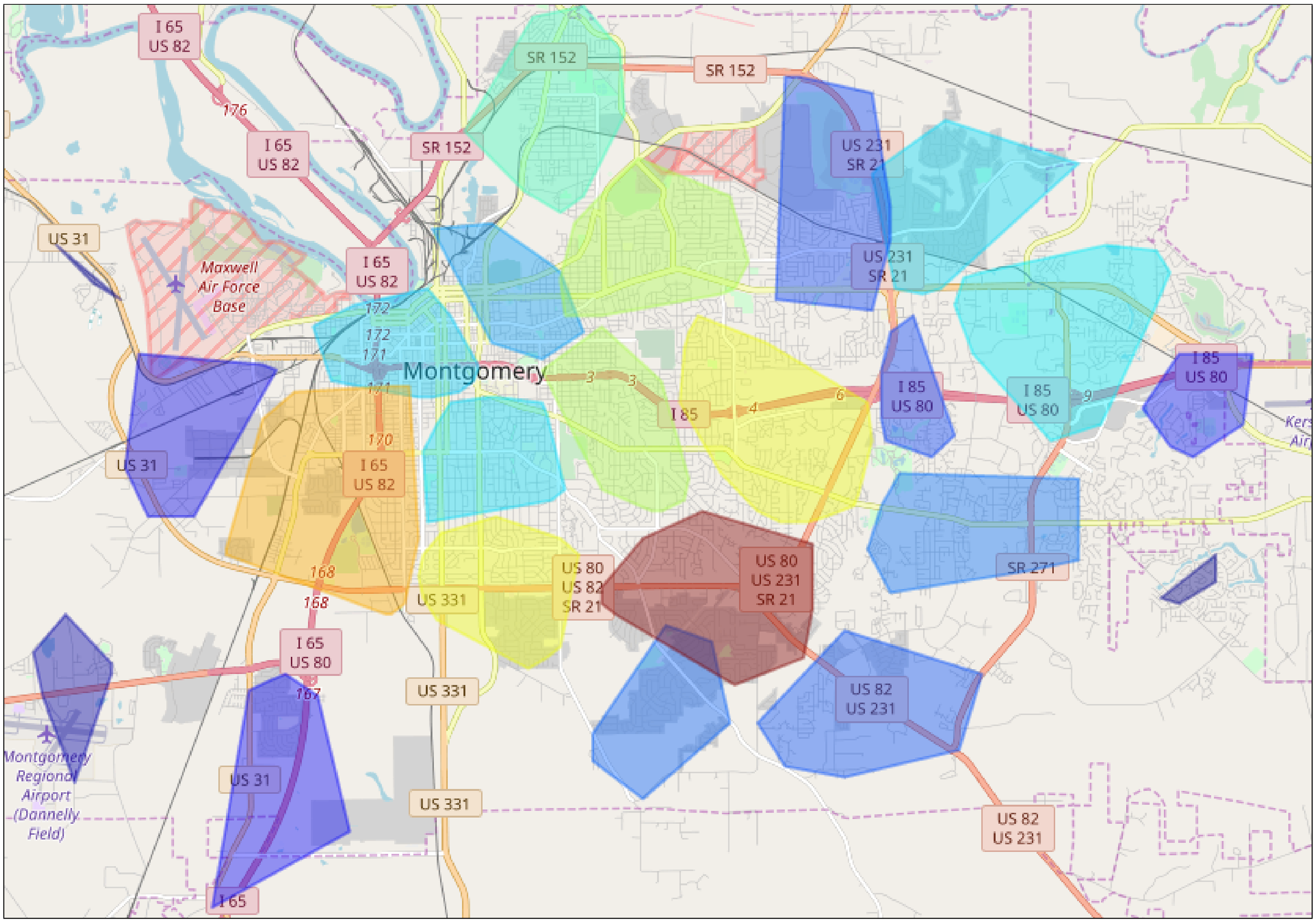

Fig. 8. All Patrol zones for the time period 10:00am to 12:00pm. Blue-Red color spectrum shows the crime weights of zones, where blue represents low crime rate, and red represents high crime rate.

this time period. According to the results in Table III, response time varies around one minute between time periods. Including the additional minute, assigned officer is expected to arrive incident location between 6.8 minutes to 7.2 minutes on the average.

B. Resource Allocation

Proposed methodology can also be employed on deciding the minimum number of patrol officers to achieve a targeted average response time. It is an acceptable assumption that increasing number of patrol officers (or patrol zones) decreases the average response time, since the officers will most likely to be closer to an incident. Nevertheless, finding the optimum number of patrol officers requires further analysis. For this purpose, we calculated the average response time for varying number of patrol officers.

The results in Figure 9 shows that above assumption is correct. According to the results, officers can arrive the incident location in around 40 minutes if there are only a few patrol vehicles in the city. At least 20 patrol vehicles must be available to achieve less than 10 minutes response time. Increasing number of patrol vehicles to 50 helps reducing the response time down to 6 minutes; however, does not yield significant change. 7 minute response time can be achieved with only 25 patrol vehicles.

V. Future Work

This study presents our preliminary results on patrol zone design problem. Our goal was the introduce an alternative machine learning based approach to the problem, and we received promising results. Nevertheless, this paper does not intend to offer a final solution to such a complex problem. Various different conditions must be considered before testing on a real patrol duty. In the future, we plan to extend this work by collaborating with the Montgomery Police Department. Our ultimate goal is to implement and test the proposed system on real patrolling duty. In this way, real response times can be measured, and the change on crime prevention rates can be observed.

A real experimentation requires more reliable clustering model. Also, several other factors must be integrated into the system to prevent unexpected outcomes. First of all, we plan to start with utilizing more historic crime data. We will extend our dataset by including last 10 years crime records in the

TABLE III
AVERAGE RESPONSE TIMES OF EACH TIME PERIOD

	Min (minutes)	**Mean** (minutes)	**Max** (minutes)
t_1	4.78	6.15	9.15
t_2	4.79	5.88	8.15
t_3	5.00	5.96	8.11
t_4	4.98	6.19	8.56
t_5	4.69	6.02	8.90
t_6	4.77	6.00	8.67
t_7	4.74	6.13	9.17
t_8	4.75	6.12	8.89
t_9	4.71	6.15	9.13
t_{10}	4.72	6.00	8.93
t_{11}	4.71	5.92	8.62
t_{12}	4.78	6.01	8.80

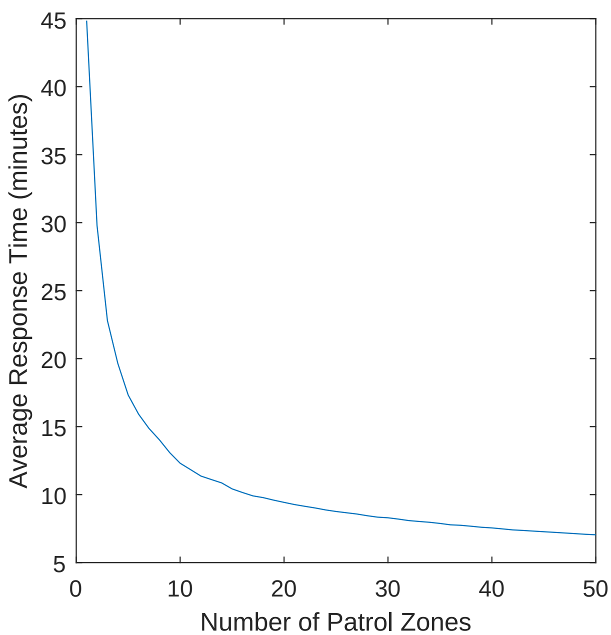

Fig. 9. The average response time for varying number of patrol zones.

Montgomery. Second, we plan to improve our patrol zones by considering natural boundaries or obstacle in the city, such as highways, rivers, etc. Third, we plan to consider the response time of the backup officers in the case of necessary conditions. And last, we plan to develop a simulation tool to implement more realistic simulations.

VI. CONCLUSION

In this paper, we investigate the police patrol design and scheduling problem for metropolitan cities. We propose a machine learning based methodology to automatically design the patrol zones using historical crime data. But unlike other computerized techniques that employs the data as whole, we made our analysis based on different time periods. In this way, we aimed to achieve more dynamic and efficient scheduling of patrol officers. To determine the patrol zones, we preferred agglomerative hierarchical clustering technique that allows a flexible design for changing number of patrol zones.

We performed a simulation using the crime dataset provided by Montgomery Police Department and evaluated our approach using real incidents. Simulation results show that, using the proposed approach, available patrol officers can be efficiently assigned to their zones. Moreover, after employing the dynamic zoning system, the average response time in our simulations is measured around 7 minutes, where the national average is around 10 minutes.

REFERENCES

[1] Peter Kolesar, Kenneth Lloyd Rider, and Thomas B Crabill, “Guidelines for scheduling police patrol cars,” 1976.

[2] Brooke Allen Saladin, *A methodology for the allocation of police patrol vehicles*, Ph.D. thesis, The Ohio State University, 1980.

[3] Gay Schaffer et al., “Patrol scheduling alternatives,” 2005.

[4] James Q Wilson and George L Kelling, “Broken windows,” *Atlantic monthly*, vol. 249, no. 3, pp. 29–38, 1982.

[5] John M Carroll and Pierre G Laurin, “Using simulation to assign police patrol zones,” *Simulation*, vol. 36, no. 1, pp. 1–12, 1981.

[6] Kevin M Curtin, Karen Hayslett-McCall, and Fang Qiu, “Determining optimal police patrol areas with maximal covering and backup covering location models,” *Networks and Spatial Economics*, vol. 10, no. 1, pp. 125–145, 2010.

[7] Richard Church and Charles R Velle, “The maximal covering location problem,” *Papers in regional science*, vol. 32, no. 1, pp. 101–118, 1974.

[8] Phillip S Mitchell, “Optimal selection of police patrol beats,” *J. Crim. L. Criminology & Police Sci.*, vol. 63, pp. 577, 1972.

[9] Steven J D’Amico, Shoou-Jiun Wang, Rajan Batta, and Christopher M Rump, “A simulated annealing approach to police district design,” *Computers & Operations Research*, vol. 29, no. 6, pp. 667–684, 2002.

[10] Yue Zhang and Donald E Brown, “Police patrol districting method and simulation evaluation using agent-based model & gis,” *Security Informatics*, vol. 2, no. 1, pp. 7, 2013.

[11] Philip E Taylor and Stephen J Huxley, “A break from tradition for the san francisco police: Patrol officer scheduling using an optimization-based decision support system,” *Interfaces*, vol. 19, no. 1, pp. 4–24, 1989.

[12] Max Woolf, “Analyzing san francisco crime data to determine when arrests frequently occur,” http://minimaxir.com/2015/12/sf-arrests/, Dec 2015, (Accessed on 01/17/2018).

[13] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, *Introduction to Data Mining, (First Edition)*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[14] Nitin R Chopde and Mangesh K Nichat, “Landmark based shortest path detection by using a* and haversine formula,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 2, pp. 298–302, 2013.

[15] Federal Bureau of Investigation, “Crime in the united states, 2016 table 4,” https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/tables/table-4, 2016, (Accessed on 01/17/2018).

[16] Montgomery Police Department, “Find your zone — city of montgomery, alabama,” http://www.montgomeryal.gov/city-government/departments/police/find-your-zone, (Accessed on 01/17/2018).

[17] Montgomery Police Department, “Crime data — city of montgomery, alabama,” http://www.montgomeryal.gov/city-government/departments/police/crime-data, (Accessed on 01/17/2018).