# CDASH: Community Data Analytics for Social Harm Prevention

Saurabh Pandey, Nahida Chowdhury, Milan Patil, Rajeev R Raje, George Mohler, Jeremy Carter
*Indiana University-Purdue University Indianapolis*
*Indianapolis, Indiana, USA*
{*pandey, nschowdh, mipatil, rraje, gmohler, carterjg*}*@iupui.edu*

***Abstract*—Communities are adversely affected by heterogeneous social harm events (e.g., crime, traffic crashes, medical emergencies, drug use) and police, fire, health and social service departments are tasked with mitigating social harm through various types of interventions. Smart cities of the future will need to leverage IoT, data analytics, and government and community human resources to most effectively reduce social harm. Currently, methods for collection, analysis, and modeling of heterogeneous social harm data to identify government actions to improve quality of life are needed. In this paper we propose a system, CDASH, for synthesizing heterogeneous social harm data from multiples sources, identifying social harm risks in space and time, and communicating the risk to the relevant community resources best equipped to intervene. We discuss the design, architecture, and performance of CDASH. CDASH also allows users to report live social harm events using mobile hand-held devices and web browsers and flags high risk areas for law enforcement and first responders. To validate the methodology, we run simulations on historical social harm event data in Indianapolis illustrating the advantages of CDASH over recently introduced social harm indices and existing point process methods used for predictive policing.**

***Keywords*-social harm; service-oriented systems; CDASH; Hawkes process; Web service.**

## I. Introduction

Crime is highly concentrated in urban communities and hotspot or "predictive" policing efforts aim to apply limited resources to high intensity geographic areas and time intervals to disrupt crime opportunities, leading to aggregate crime rate reductions [1]–[4]. However, police serve other roles in the community beyond crime response and prevention, including traffic enforcement, Emergency Medical Services (EMS) response, and more generally, dealing with events related to social harm [5]. At the same time, the activities police departments employ to address social harm issues in a community (directed patrol, speed traps, community outreach, etc.) have both the potential to decrease the risk of social harm, but may also increase the risk or perception of social harm if the community costs of police activities such as stop-and-frisk reduce trust and increase grievances among disenfranchised groups [5]. Other community stakeholders such as EMS responders, social services, the mayor's office, city prosecutor, and individual citizens also participate to reduce social harm. While collaboration can take place, for example a paramedic riding along on police patrols [6] in high drug overdose hotspots, often data is distributed among several agencies, data analyses are not shared across agencies, and interventions are not coordinated.

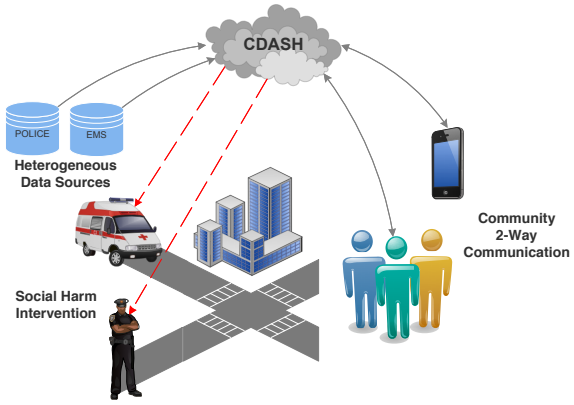


Figure 1. CDASH fuses heterogeneous data sources, estimates risk of social harm, and allocates resources for targeted interventions.

Despite these multiple and disparate daily challenges, existing hotspot and predictive policing algorithms and intervention strategies focus on single or groups of related sub-categories of social harm events and interventions are performed primarily by police in isolation. Given the explosion of data that smart cities are generating, advances in predictive modeling, and the real-time inter-connectedness of citizens through the Internet of Things, smart cities of the future will be able to integrate multiple data streams, detect and predict social harm threats, communicate key information to the general public and allocate resources accordingly. To realize such a capability, new software and analytics methods are needed to facilitate heterogenous data sharing across the various agencies tasked with addressing social harm and to support real-time data driven policing of social harm in collaboration with community stakeholders.

In Figure 1, we illustrate an integrative policing system, called Community Data Analytics for Social Harm (CDASH). CDASH combines historical and real-time data across heterogeneous types of social harm data pulled from police, EMS, and social services databases, along with community feedback (tips and complaints), to prioritize daily activities within each patrol beat in the city. For example, a traffic accident hotspot may be flagged at 7 am for police

This is the author's manuscript of the article published in final edited form as:

Pandey, S., Chowdhury, N., Patil, M., Raje, R. R., Shreyas, C. S., Mohler, G., & Carter, J. (2018). CDASH: Community Data Analytics for Social Harm Prevention. 2018 IEEE International Smart Cities Conference (ISC2), 1–8. https://doi.org/10.1109/ISC2.2018.8656957

intervention and the patrol unit is given a push notification to monitor traffic there when not on a call to service. A community watch group utilizing the application is tasked with providing soft patrols [7] during 9 am - 4:30 pm in their neighborhood that is flagged as a high residential burglary risk. Later at night, a patrol officer is paired with a paramedic [6] in a drug overdose hotspot and positioned to shorten EMS response time. Over longer timescales, beats that receive a higher volume of complaints against officers or are estimated to have higher rates of under-reporting may be flagged for a community meeting to be held in that neighborhood.

In this paper, we provide an overview of CDASH and descriptions of the key components. In Section II, we describe the architecture of the CDASH system. In Section III, we describe a point process-based model for estimating the risk of social harm. In Section IV, we present results from several experiments illustrating the scalability, fault tolerance, and accuracy of CDASH. We run a simulation study using historical social harm event data in Indianapolis to illustrate the potential value of the CDASH system. We conclude the paper by indicating insights learned and possible future directions for research on this topic.

## II. Service oriented architecture of CDASH

### A. System Architecture

As shown in Figure 2, CDASH has a layered architecture and is a distributed Web-based system accessible through Web browsers as well as through mobile hand-held devices. CDASH consists of four layers:

- *Presentation Layer*
- *Middleware Layer*
- *Application Layer*
- *Database Layer*

Below we describe these layers.

- *Presentation Layer*
  The presentation layer, consists of a C#-based Web Server (CWS), handling multiple clients and their views simultaneously. When a client connects, the CWS presents the latest social harm information including the predicted hot-spots and live user feeds (if any) to the client. Also, clients are provided with an option for entering a new incident if they wish to do so.

  For each new feed, the client is required to input certain information including the type of incident and its location. With this, CDASH also provides an option for fetching the location information of the client through the client's device accessing its location service with the client's permission. There are 18 different types of incidences currently supported by the available social harm data in the city of Indianapolis [8] and hence, these 18 options are available in CDASH. Once the incident information is provided as an input, the CWS first updates all the connected clients dynamically. Next, it pushes the data on a Kafka topic [9] as shown in Figure 3. The request is in the JSON (JavaScript Object Notation) format. JSON is desirable as it is fast and light-weight.

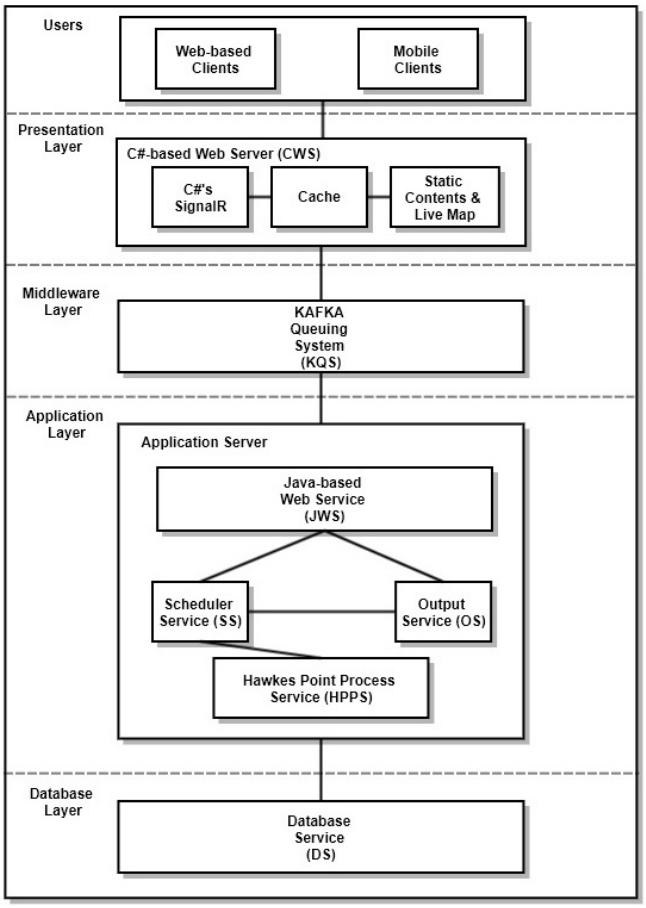


Figure 2. CDASH System Architecture.

- *Middleware Layer*
  The middleware layer of CDASH consists of the Kafka Queuing System (KQS). Apache Kafka® is a distributed streaming platform [9]. Kafka helps in building fast, scalable, and fault tolerant applications. Kafka has its own server that is used in managing the messages passing through it. In CDASH, a live incident fed in by a client is passed on to a Java-based Web Service (JWS). In this, the CWS pushes the data on to a topic which is listened by the JWS. Here, the CWS acts as a data publisher while the JWS acts as a data subscriber.
- *Application Layer*
  The application layer interacts with the presentation layer through the middleware layer. This layer is made of four services and is responsible for handling the business logic of the system. As depicted in Figure 3, on retrieving a live incident from KQS, the JWS checks for

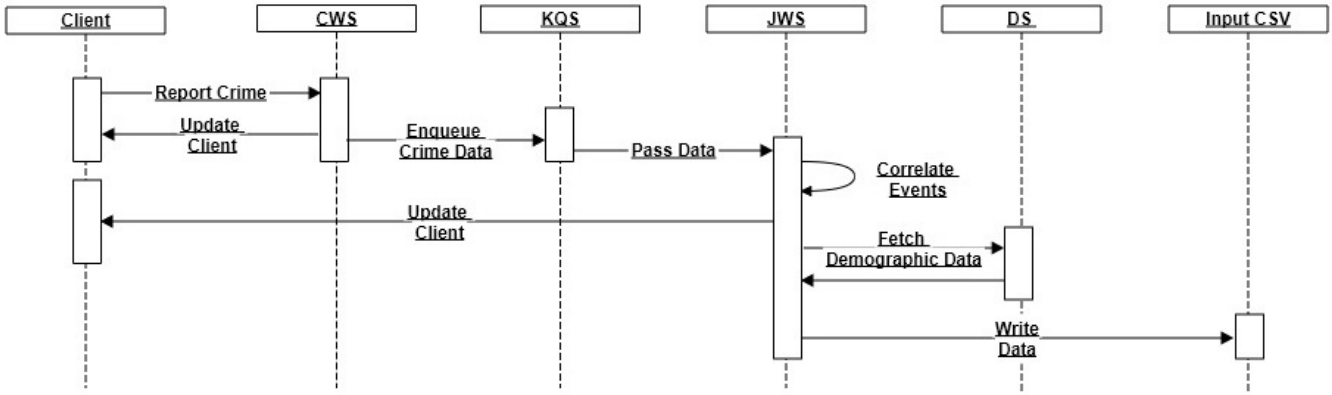


Figure 3. Sequence of Interactions in CDASH System.

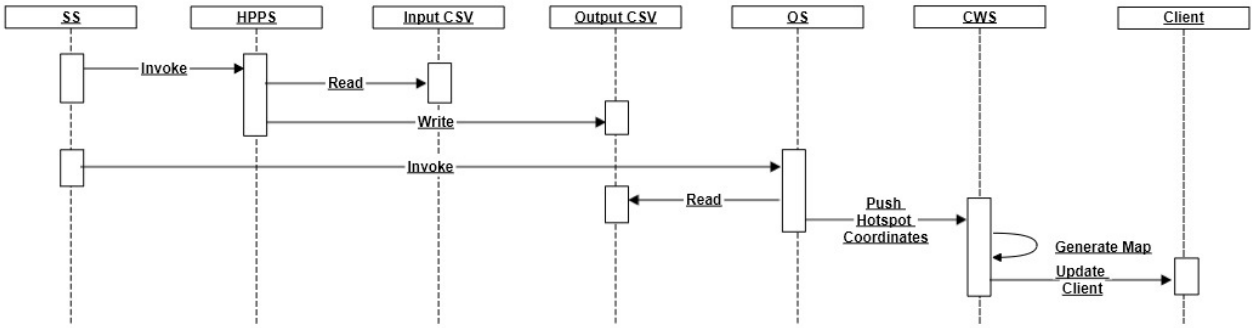


Figure 4. Sequence of events updating the hotspot.

duplication of the reported incident by executing a correlation logic, which attempts to analyze the reported incident on the basis of its location, time and incident-code with the events already reported to the CDASH system. If the event correlates with any pre-existing event in the system, all the clients are accordingly updated by JWS. However, if the event is new and not correlated to any previously reported incidents, the JWS interacts with the MySQL Database Service (DS) in the database layer to fetch the demographic information of the harm location. The database consists of a table including the demographic details of various locations in Indianapolis. Some of the demographic details used are: total population, gender-ratio, income ranges along with literacy, unemployment and poverty rate. The demographic information together with the user input, is staged for the Hawkes Point Process Service (HPPS). The HPPS is a prediction service written in Matlab that takes as input historical incident layer and returns hot-spot predictions. We provide details of the HPPS in Section III.

Periodically, currently every 8 hours, to coincide with a new police shift, the CDASH system runs a Scheduler Service (SS) that invokes the HPPS to read the reported crimes and predict hot-spots. The HPPS requires sufficient amount of new data to generate new and meaningful hot-spots and thus, an interval of 8 hours is chosen to run it. As new hot-spots are generated by HPPS, the SS invokes an Output Service (OS) responsible for pushing the hot-spot information towards the CWS as can be seen in Figure 4. The CWS, on receiving new hot-spots, updates the map accordingly for the clients.

- *Database Layer*
  The database layer of CDASH, as indicated above, consists of a MySQL Database Service (DS). As described in the application layer of the CDASH architecture, the database holds information related to the demography of various locations (on the basis of zipcodes) of Indianapolis metro. Apart from this, the database layer also contains all the live events reported by the users of the system. This helps CDASH in correlating various reported incidents on the basis of the type of incident reported, its time and location, thereby avoiding duplications.

### B. Architectural Patterns

The CDASH system employs an implementation of the Model-View-Controller (MVC) pattern. The CWS in the presentation layer is the view part of the pattern. It helps in interacting with the clients and updating them dynamically as needed. The JWS has an Incident Controller component that handles all the incoming feeds from the CWS through the KQS. The HPPS, SS and DS form the model part of the MVC pattern, holding the application logic for various functionalities within the system. The results returned, after the model executes, are pushed towards the CWS through OS. Having an MVC architecture, makes the design flexible and enhances extensibility of CDASH.

In order to make CDASH interactive, its response needs to be in near real-time and thus, any new updates obtained from the users of CDASH should be pushed on to all connected users dynamically. Hence, the Observer pattern is a perfect fit for the CDASH, where information is being pushed towards the observers instead of a pull model that requires a lot of polling, creating a large network traffic and ultimately slowing down the entire application. In CDASH, we achieve this by using the SignalR technology of C#. The CWS includes a SignalR hub to which all the clients connect automatically when they connect with the application. As soon as any new update is available to the system, SignalR recognizes it, an updated map is generated by the CWS and pushed towards all the connected clients.

## III. Hawkes process model of social harm

A number of algorithmic methods have been proposed for estimating crime hotspot risk including multivariate models [10]–[12], kernel density estimation [13]–[17] and spatio-temporal point processes [18], [19]. While each approach has tradeoffs, marked point processes have the advantage that long-term intrinsic risk [19], short-term dynamic risk [18], and periodic/seasonal trends [20] in the intensity can be handled systematically with only event data as input. In [4], a randomized controlled trial of point process based predictive policing was conducted and this model will form a starting point for our dynamic model of social harm.

### A. Property Crime Hawkes Process

We first review the property crime Hawkes process (also referred to as Epidemic Type Aftershock Sequence or ETAS) defined in [4]. Let a spatial domain be discretized into square cells or "boxes" in which we will estimate the rate of crime incidents. The conditional intensity, or probabilistic rate $\lambda_n(t)$ of events in box $n$ at time $t$ is determined by,

$$\lambda_n(t) = \mu_n + \sum_{t_n^i < t} \theta\omega e^{-\omega(t-t_n^i)}, \tag{1}$$

where $t_n^i$ are the times of events in box $n$ in the history of the process. The ETAS model has two components, one modeling place-based environmental conditions that are constant in time and the other modeling dynamic changes in risk. Rather than modeling fixed environmental characteristics of a hotspot explicitly using census data or locations of crime attractors, long term hotspots are estimated from the events themselves. In particular, the background rate $\mu$ is a nonparametric histogram estimate of a stationary Poisson process [21]. If over the past 365 days a grid cell has a high crime volume, the estimate of $\mu$ will be large in that grid cell. The size of the grid cells on which $\mu$ is defined can be estimated by Maximum Likelihood and in general the optimum size of the grid cell will decrease with increasing data. However, for a fixed area flagged for patrol, a greater number of small hotspots are more difficult to patrol than a small number of large hotspots.

The second component of the ETAS model is the triggering kernel $\theta\omega e^{-\omega t}$ that models "near-repeat" or "contagion" effects in crime data. The exponential decay causes grid cells containing recent crime events to have a higher intensity than grid cells with fewer recent events and the same background rate. The ETAS model estimates both long term and short term hotspots and systematically estimates the relative contribution to risk of each via Expectation-Maximization [18], [19]. Given an initial guess for the parameters $\theta$, $\mu$, and $\omega$, the EM algorithm is applied iteratively until convergence by alternating between the following two steps:

**E-step**

$$p_n^{ij} = \frac{\theta\omega e^{-\omega(t_n^j - t_n^i)}}{\lambda_n(t_n^j)}, \tag{2}$$

$$p_n^j = \frac{\mu_n}{\lambda_n(t_n^j)}, \tag{3}$$

**M-step**

$$\omega = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_{i<j} p_n^{ij}(t_n^j - t_n^i)}, \tag{4}$$

$$\theta = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_j 1}, \tag{5}$$

$$\mu = \frac{\sum_n \sum_j p_n^j}{T}, \tag{6}$$

where $T$ is the length of the time window of observation.

The EM algorithm can be intuitively understood by viewing the ETAS model as a branching process [18]. First generation events occur according to a Poisson process with constant rate $\mu$. Events (from all generations) each give birth to $N$ direct offspring events, where $N$ is a Poisson random variable with parameter $\theta$. As events occur, the rate of crime increases locally in space, leading to a contagious sequence of "aftershock" crimes [18] that eventually dies out on its own, or is interrupted by police intervention; the former occurs naturally so long as $\theta < 1$, while the latter is unaccounted for by the model. In the E-step, the probability that event $j$ is a direct offspring of event $i$ is estimated, along with the probability that the event was generated by

the Poisson process $\mu$. Given the probabilistic estimate of the branching structure, the complete data log-likelihood is then maximized in the M-step, providing an estimate of the model parameters.

### B. A Marked Point Process Model of Social Harm

Now suppose we have $m = 1, ..., M$ social harm event categories. For each event type $m$, we have a secondary mark $c(m)$ representing the average societal cost of an event of type $m$. Given this cost mark, we can then define a dynamic social harm index $SI_n(t)$ in each grid cell $n$ as the expected cost per unit time,

$$SI_n(t) = \sum_{m=1}^{M} c(m)\lambda_n^m(t), \tag{7}$$

where $\lambda^m(t)$ is a point process estimated independently on event data of type $m$. The dynamic social harm index can then be used to rank hotspots over a given time interval, where the top $k$ hotspots are flagged for intervention. Because this type of ranking is common in hotspot analysis and policing, a popular accuracy metric is the Predictive Accuracy Index (PAI). The PAI is the percentage of events captured in the top $k$ hotspots divided by the percentage of city area that the $k$ hotspots comprise. In the case of social harm, we use a modified PAI capturing the proportion of total cost captured in the top hotspots relative to random chance:

$$\text{PAI@k} = \frac{\%\text{ societal cost captured in top k hotspots}}{\%\text{ city area covered by k hotspots}}. \tag{8}$$

The above mentioned model is encapsulated in CDASH as the HPPS. In the next section, we detail how the cost per event can be estimated and present simulation results on applying our point process methodology to social harm data in Indianapolis. We also describe several experiments with the CDASH system. We focus on heterogeneity, scalability, fault tolerance, and predictive accuracy.

## IV. EXPERIMENTS AND ANALYSES

### A. Heterogeneity

Heterogeneity is one of the major challenges faced by any distributed system. We have implemented CDASH in such a way that it can handle heterogeneity in terms of different hardware components and network protocols. To reach a large spectrum of proposed users of CDASH, it is made accessible through all browsers on desktop devices and also from mobile hand-held devices through mobile-based apps.

CDASH ensures that regardless of the device used, the user will always be presented with the most recent view of the global state at any time. It achieves this by updating the views on all the connected devices dynamically as often as needed. This, in turn, ensures that all the users have a consistent view of the global state of current social harm events thereby avoiding any potential confusion and associated chaos.

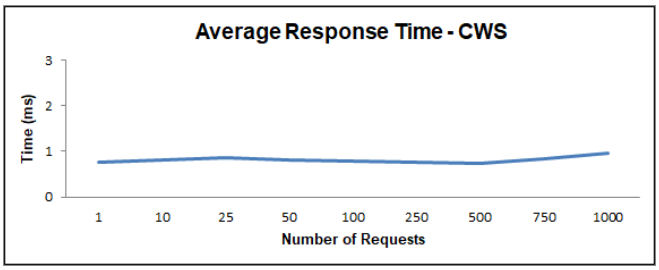


Figure 5. Response Time of CWS.

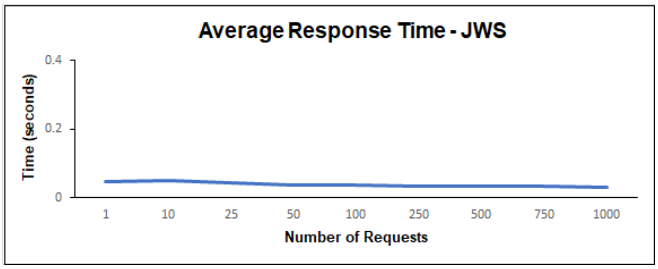


Figure 6. Response Time of JWS.

### B. Scalability

In our experiments, scalability is measured by observing the relation between the number of requests and their average execution time. We have experimented with the scalability-related behavior of the CDASH system by implementing a test module for firing multiple requests. Since the presentation and application layers are decoupled and work as independent units, we analyzed the execution time for the CWS and JWS separately. The performance of the CDASH system is shown in Figures 5 and 6. It is less likely that there would be more than 1000 user requests simultaneously in a real-world scenario. Hence, we have experimented with 1000 as the upper limit on the user requests. The average round-trip time was observed to be in the range of 0.86 milliseconds to 1 millisecond for the CWS while 29 milliseconds to 56 milliseconds for the JWS, which is near real-time and acceptable with respect to the nature of typical social harm events.

We analyzed the above response times for the CWS and JWS separately. Firstly, with the CWS, it was observed that the time taken was shared equally by modules that: i) fetch the geolocation (based on user's location). ii) update the map's markers and legends data to be displayed to the clients and iii) dynamically update all the clients. Next, with the JWS, it was observed that the overall time taken by the JWS was divided almost equally between the JWS, DS and other auxiliary activities (staging data for HPPS). However, as stated above, since the presentation and application layers are decoupled, the overall response time for a user would be equal to that of the time taken by the CWS as the Application layer works asynchronously in the back-end.

### C. Fault Tolerance

Failures can occur in any system. However, distributed systems, having various distributed components working together, are more prone to failures. In CDASH, we dealt with the following failures:

- *CWS Failure*
- *JWS or DS Failure*
- *Client Failure*

Below we describe these failures.

- *CWS Failure*
  If the CWS fails, the point of contact of the users with CDASH is lost. Thus, any user attempting to connect with application will be presented with an error message displaying page not available. The only way of dealing with these failures is restarting the CWS.
- *JWS or DS Failure*
  CDASH is made fault tolerant towards the JWS and/or DS failures by the KQS. Kafka helps in retaining incident details in its server while the JWS or DS is down. The messages are retained in the Kafka server until they are consumed and committed by the consumer. In case of failures, the messages are not committed and hence they are not lost. Once the failed components are up and running, Kafka automatically redelivers the messages thus making these components fault tolerant. Additionally, we have enhanced the fault tolerance of the JWS by running two instances of it at any given time. These instances are configured to operate in active-passive mode running on two different servers. All the requests are directed towards the primary instance (active component). If the primary service instance is down due to any failures, the requests are redirected towards a secondary service instance (passive component). The synchronization between the two instances is configured to be handled automatically in Kafka.
- *Client Failure*
  In the event of a client failure, any of its requests that may have reached the CDASH system will be processed and its effect will be seen in the generated global state of the social harm picture. Later, if the client reconnects, the client can see his input being reflected on the map generated by the CDASH system.

### D. CDASH Accuracy Analysis

In order to assess the accuracy of the CDASH system, we run a historical simulation of the system in Indianapolis. The data we use includes all crime, drug overdose, and vehicle crash data for years 2012-2013 that were provided electronically from the appropriate government agency and included time and data stamp as well as state-plane coordinates for each incident that were converted to WGS84 coordinates. Social harm weights are derived from established crime, drug, and vehicle crash cost estimation studies. Costs for homicide, rape, robbery, aggravated assault, arson, motor vehicle theft, residential burglary, larceny, embezzlement, forgery, fraud, and vandalism were gleaned from estimates of crime costs to society [22]. Vehicle crashes resulting from drugs or alcohol, simple assault, and driving while impaired costs were derived from monetary estimates of crime prevention [23]. Lastly, cost estimates based on per-incident occurrences in the United States were utilized for suicide attempts [24], vehicle crashes not related to drugs or alcohol [25], and drug overdoses [26]. Each of these latter three estimates were calculated by dividing the total annual costs for each incident type by the total number of each incident in a given year. In Table I, we provide summary statistics for Indianapolis social harm including the volume of incidents over 2012 and 2013, the estimated cost per event to society, and the total cost over the two year period attributed to each event category.

We first train the model on a 100x100 grid using Indianapolis social harm data from 2012. We assume that police have fixed resources and can patrol $k$ hotspots each day (see Figure 7). We also assume that if a hotspot is patrolled, then all events are prevented from occurring on that day (an alternative choice would be to allow for a percentage reduction that varies with event category).

Then for each day $t$ in 2013, the simulation proceeds as follows:

- Estimate the expected cost $SI_n(t)$ as in Equation 7 for each grid cell.
- Rank the grid cells in decreasing order according to expected cost $SI_n(t)$.
- Flag the top $k$ grid cells for directed patrol on the next day $t+1$.
- On day $t+1$ record the number of events prevented and the cost associated with those events.

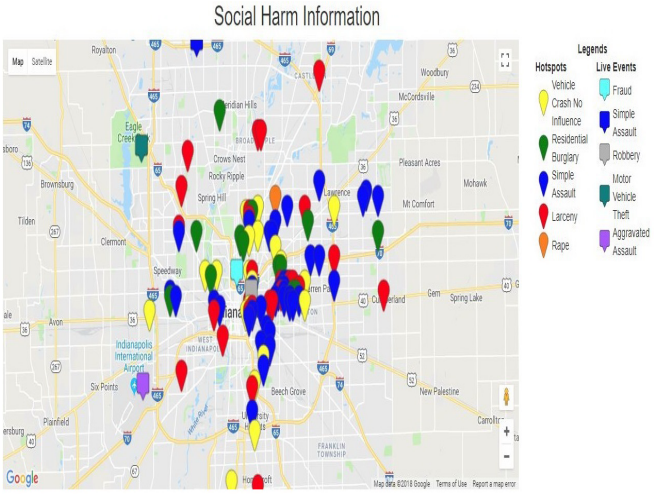


Figure 7. Example CDASH hotspots in Indianapolis.

Table I
SUMMARY STATISTICS FOR INDIANAPOLIS SOCIAL HARM 2012 & 2013

| Type | Count | Cost/Event | Total |
|---|---|---|---|
| Suicide Attempt | 134 | $5,251 | $703,634 |
| DWI Arrest | 3546 | $500 | $1,773,000 |
| Forgery | 481 | $5,265 | $2,532,465 |
| Embezzlement | 876 | $5,480 | $4,800,480 |
| Arson | 723 | $16,428 | $11,877,444 |
| Drug Overdose | 4112 | $3,922 | $16,127,264 |
| Rape | 1160 | $41,247 | $47,846,520 |
| Vehicle Crash Drug/Alcohol | 1610 | $30,000 | $48,300,000 |
| Fraud | 11371 | $5,032 | $57,218,872 |
| Vandalism | 13641 | $4,860 | $66,295,260 |
| Motor Vehicle Theft | 9081 | $10,534 | $95,659,254 |
| Residential Burglary | 21468 | $6,170 | $132,457,560 |
| Robbery | 6386 | $21,398 | $136,647,628 |
| Larceny | 53241 | $3,523 | $187,568,043 |
| Aggravated Assault | 11797 | $19,537 | $230,477,989 |
| Homicide | 220 | $1,278,424 | $281,253,280 |
| Vehicle Crash No Influence | 40718 | $7,864 | $320,206,352 |
| Simple Assault | 30802 | $11,000 | $338,822,000 |
| Total | 211367 | | $1,980,567,045 |

- Repeat for each day in 2013.

We compare our proposed social harm Hawkes process, equation 7, with a property crime Hawkes process [4] and a static harm index [5] using the outlined simulation methodology. In Figure 8, we show the PAI of each method as a function of the fraction of the city flagged for patrol each day in the simulation. Note that a PAI of 1 corresponds to random patrol and all methods perform better than random. Also, PAI values tend to decrease as a larger portion of the city is patrolled, because lower risk cells contain less crime and police interventions have a lesser impact in these areas. The social harm Hawkes process performs the best out of all methods, achieving a PAI of 15 when 50 hotspots are selected each day (comprising 0.5% of the city). In the lower figure we plot the fraction of social harm captured as a function of the fraction of the city patrolled in the simulation. We note that almost $ 200 million (20%) of the social harm cost to Indianapolis in 2013 is captured in 2% of space-time. The top 10% of space-time contains over half of all social harm cost.

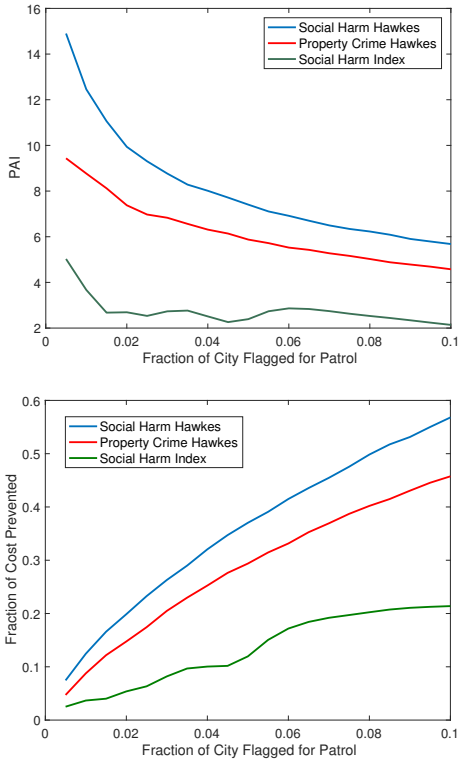


Figure 8. PAI vs. fraction of city selected for patrol (top) and fraction of cost captured in top $k$ hotspots vs. fraction of city selected for patrol (bottom).

## V. DISCUSSION

We introduced CDASH, a system for i) collecting heterogeneous social harm data, ii) modeling space-time social harm risk, and iii) communicating risk to community stakeholders for the allocation of resources. We ran a simulation study using historical data from Indianapolis illustrating the potential impact such a system could have on social harm prevention. Our method captures 20% of social harm cost in 2% of space-time, compared to current social harm indices and predictive policing models of property crime that capture 5-15%.

Future work will focus on several directions. We envision implementing the principles of role-based access control (to provide different privileges and different views to different participants in this effort), and incorporating different trust models associated with different interactions between the users of the system. In addition, while conducting the experiments, we realized that to solve or prevent social harm, civic bodies must create a temporary network and collaborate quickly. This fits in the structure of Virtual Organizations. We will be focusing on building over the concept of Information Technology-based virtual organizations which help decentralized working units in collaborating and coordinating activities.

In terms of predictive modeling of social harm, machine learning and multivariate statistical models may improve upon the predictive accuracy of CDASH and will allow for the incorporation of more data streams (weather data, city sensor data, GIS data, etc). Ultimately these systems need to be tested in field trials to determine what types of tasks are feasible, how can information best be communicated through the application, and what is the impact of interventions on reducing social harm.

## VI. ACKNOWLEDGEMENTS

This project is supported in part by NSF grants CNS-1737585, SES-1343123, and DMS-1737996. G.M. is a co-

founder and serves on the board of PredPol, a predictive policing company.

## REFERENCES


[1] D. Weisburd, L. A. Wyckoff, J. Ready, J. E. Eck, J. C. Hinkle, and F. Gajewski, "Does crime just move around the corner? a controlled study of spatial displacement and diffusion of crime control benefits*," *Criminology*, vol. 44, no. 3, pp. 549–592, 2006.

[2] A. A. Braga and B. J. Bond, "Policing crime and disorder hot spots: A randomized controlled trial," *Criminology*, vol. 46, no. 3, pp. 577–607, 2008.

[3] J. H. Ratcliffe, T. Taniguchi, E. R. Groff, and J. D. Wood, "The philadelphia foot patrol experiment: a randomized controlled trial of police patrol effectiveness in violent crime hotspots*," *Criminology*, vol. 49, no. 3, pp. 795–831, 2011.

[4] G. O. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham, "Randomized controlled field trials of predictive policing," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1399–1411, 2015.

[5] J. H. Ratcliffe, "Towards an index for harm-focused policing," *Policing*, p. pau032, 2014.

[6] R. King, "Perkins and hardwick, a new crime-fighting duo in indianapolis," *Indianapolis Star*, 2016.

[7] B. Ariel, C. Weinborn, and L. W. Sherman, "soft policing at hot spotsdo police community support officers work? a randomized controlled trial," *Journal of Experimental Criminology*, vol. 12, no. 3, pp. 277–317, 2016.

[8] G. Mohler, J. Carter, and R. Raje, "Improving social harm indices with a modulated hawkes process," 2017.

[9] "Apache Kafka," https://kafka.apache.org.

[10] X. Wang, D. E. Brown, and M. S. Gerber, "Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information," in *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*. IEEE, 2012, pp. 36–41.

[11] H. Liu and D. E. Brown, "Criminal incident prediction using a point-pattern-based density model," *International journal of forecasting*, vol. 19, no. 4, pp. 603–622, 2003.

[12] L. W. Kennedy, J. M. Caplan, and E. Piza, "Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies," *Journal of Quantitative Criminology*, vol. 27, no. 3, pp. 339–362, 2011.

[13] K. J. Bowers, S. D. Johnson, and K. Pease, "Prospective hotspotting the future of crime mapping?" *British Journal of Criminology*, vol. 44, no. 5, pp. 641–658, 2004.

[14] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, no. 1, pp. 4–28, 2008.

[15] S. D. Johnson, K. J. Bowers, D. J. Birks, and K. Pease, "Predictive mapping of crime by promap: accuracy, units of analysis, and the environmental backcloth," in *Putting crime in its place*. Springer, 2009, pp. 171–198.

[16] S. D. Johnson, *Prospective crime mapping in operational context: Final report*.

[17] M. Fielding and V. Jones, "'disrupting the optimal forager': predictive risk mapping and domestic burglary reduction in trafford, greater manchester," *International Journal of Police Science & Management*, vol. 14, no. 1, pp. 30–41, 2012.

[18] G. Mohler, M. Short, P. J. Brantingham, F. Schoenberg, and G. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.

[19] G. Mohler, "Marked point process hotspot maps for homicide and gun crime prediction in chicago," *International Journal of Forecasting*, vol. 30, no. 3, pp. 491–497, 2014.

[20] R. D. Peng, F. P. Schoenberg, and J. A. Woods, "A space–time conditional intensity model for evaluating a wildfire hazard index," *Journal of the American Statistical Association*, 2011.

[21] D. Marsan and O. Lengline, "Extending earthquakes' reach through cascading," *Science*, vol. 319, no. 5866, pp. 1076–1079, 2008.

[22] K. E. McCollister, M. T. French, and H. Fang, "The cost of crime to society: New crime-specific estimates for policy and program evaluation," *Drug and alcohol dependence*, vol. 108, no. 1, pp. 98–109, 2010.

[23] M. A. Cohen and A. R. Piquero, "New evidence on the monetary value of saving a high risk youth," *Journal of Quantitative Criminology*, vol. 25, no. 1, pp. 25–49, 2009.

[24] D. S. Shepard, D. Gurewich, A. K. Lwin, G. A. Reed, and M. M. Silverman, "Suicide and suicidal attempts in the united states: costs and policy implications," *Suicide and life-threatening behavior*, vol. 46, no. 3, pp. 352–362, 2016.

[25] N. H. T. S. Administration *et al.*, "The economic and societal impact of motor vehicle crashes, 2010," *Report DOT HS*, vol. 812, p. 013, 2014.

[26] C. S. Florence, C. Zhou, F. Luo, and L. Xu, "The economic burden of prescription opioid overdose, abuse, and dependence in the united states, 2013," *Medical care*, vol. 54, no. 10, pp. 901–906, 2016.