Load Balancing at Emergency Departments using ‘Crowdinforming’

M.R. Friesen and R.D. McLeod
Electrical & Computer Engineering
University of Manitoba
Winnipeg, Canada
Marcia_Friesen@UManitoba.ca

T. Strome
Winnipeg Regional Health Authority
Winnipeg, Canada

S.N. Mukhi
Canadian Network for Public Health Intelligence
Winnipeg, Canada

Abstract—Emergency Department overcrowding is an important healthcare issue garnering significant public and regulatory scrutiny in Canada and around the world. Many approaches to alleviate excessive waiting times and lengths of stay have been considered and implemented to various degrees. In theory, a more optimal emergency department patient flow may be assisted via balancing patient loads between emergency departments within the region – in essence spreading patients more evenly throughout this system. This investigation uses available data for one regional health authority, and utilizes simulation to explore a process control strategy built on ‘crowdinforming’, aimed to balance patient loads between six Emergency Departments within a mid-sized Canadian city. The preliminary model uses patient arrival rates as an indicator of ‘busyness’ and as the basis for crowdinforming. Intuitively expected metrics such as the number of patients waiting, the average length of stay, and time waiting to be seen were not found to be meaningful metrics of emergency department ‘busyness’ with this particular data set for the purpose of a process control strategy.

Keywords- Patient oriented re-direction, Load balancing, Patient flow, Emergency department wait time.

I. INTRODUCTION

Patient wait times, excessive lengths of stay, and overcrowding at hospital Emergency Departments are chronic issues that receive considerable attention within healthcare administration and from the general public. Debates on how best to address these problems have been ongoing for several decades [1][2][3] and will likely continue. Factors related to emergency department overcrowding and excessive wait times are known to be complex. They include internal factors such as emergency departments that are too small, procedural bottlenecks in transferring admitted patients out of the emergency department, staff shortages, and delays in consultative and diagnostic services. External factors include increasing patient volumes and increasing complexity and acuity of cases [4][5][6][7].

Many approaches have been attempted to reduce emergency department wait-times and lengths of stay. Familiar approaches include minor treatment streams (‘fast-tracks’) for low-acuity patients [8][9][10], and re-conceptualization of the treatment model [11], such as team triage [12]. It is also known that patient perceptions of wait times and service delivery are important factors to consider as the basis for patients’ satisfaction with healthcare delivery [13].

Within the health authority that served as the context for this work, a degree of patient diversion between emergency departments already takes place for those who arrive by ambulance. In these cases, the ambulance dispatcher diverts patients based on locale, severity, and facility specialization. More recently, a pilot program has been initiated to transport patients to an urgent care facility as opposed to emergency departments, for the stable patients who previously waited in an emergency department at the low-complexity end of the triage scale. A significant benefit is also anticipated for the delivering paramedics [14]. The present policy at emergency departments is for paramedics to wait until their patient is seen by a doctor before they leave. The policy change mentioned is intended to both reduce the pressure at emergency departments and free up paramedics sooner. This is not the specific type of patient redirection considered in this work, but it illustrates the multifaceted nature of problem and solution.

Modelling and simulation have been available to the study of the dynamics of emergency care for several decades, yet the resultant efforts are sparse and relate to many aspects of emergency department operations, including staff scheduling, patient flow, and departmental performance metrics [15][16][17]. Several studies also address patient loads in particular, although the underlying modelling methodologies and approaches are dissimilar from this work [18][19][20].

The patient redirection simulated here is associated with ambulatory patients and those with sufficient acumen to consult a service (from home or elsewhere) that provides a measure of the ‘busyness’ of an emergency department, prior to deciding upon which emergency department to attend. The work develops a simulation model based on a process control strategy that relies on this concept of ‘crowdinforming’ (proactive information dissemination) as an input. In practical manifestation, such a strategy would give patients some means of control over their decision to visit a hospital emergency department. In theory, actual patient loads at and between emergency departments may be balanced (resulting in streamlined patient flow), and patients’ perceptions of their emergency department experience are enhanced as a result of crowdinforming.
The concept of crowdinforming refers to this delivery of relevant information (emergency department-specific ‘busyness’ metrics) to the general public, which can be accessed by the patient from home or another location, including a mobile device, prior to a pending emergency department visit. The patient then uses this information to decide which emergency department to attend.

Crowdinforming is derived as a variant of the better-known concepts of crowdsourcing and crowdcasting, which recently have been used in biosurveillance applications [21][22]. One of the fundamental tenets of crowdsourcing is that the feedback loop needs to be closed, as information mined through crowdsourcing flows back to the crowd that generated it (crowdcasting), presumably to accrue benefit.

There are healthcare examples of the crowdinforming concept in practice. These include pre-hospital ambulance redirection based on internet-accessible real-time emergency department loads in Western Australia [23], and a service by which cellular phone users in Texas, USA can text a common number and receive a return text indicating the expected wait time at the nearest emergency department [24]. The crowdinforming simulation presented here would augment the types of mobile device apps that are being developed to assist people in selecting the closest emergency department [25].

The current study is one of the first illustrating the use of emergency department data in a feedback control manner in an effort to assess the impact of making real-time emergency department data available to the general public. Simulations are suggested as an appropriate means of gaining this insight. Another important outcome, albeit somewhat obvious is that each region will likely have differing characteristics that need to be evaluated [26][27]. In [20], the authors point out that ‘the presence of autocorrelation in the models for wait time and Length Of Stay is consistent with our intuition that process times are affected by prior patient arrivals’ and that ‘future work will focus on discrete event simulation modeling’. This is similar to the case presented here; we have data that appear consistent with intuition and are using the data as input to simulation. Nonetheless, some of the findings proved to be counterintuitive.

The balancing of emergency department loads is a complex multifaceted problem. Redirection of patients is just one piece of the solution. The aspect considered here explores the uses and value of leveraging available data, such that an informed decision can be made in the case of lower acuity patients self redirecting their ER visit. The rationale for considering lower acuity patients is that they are arguably the patients with greater decision capabilities or flexibility than high acuity patients.

II. Methods

The objective of this study was to first analyze existing data collected from various emergency department information systems associated with a mid size urban centre. The objective of the analysis was assess how emergency department efficiencies may be potentially improved if ‘busyness’ metrics could be provided to the general population. As such, a person would then be in a position to consult a web service via a web application or a more traditional information service, in determining which emergency department they may choose to visit.

The interest then lies in a model and simulation that would demonstrate how emergency department loads could conceivably be balanced once patients were provided with emergency department ‘busyness’ information and subsequently allowed to self-redirect to alternative facilities. Data for the study initially came from six months of patient visit data taken from the Emergency Department Information System of the health authority of a mid-sized Canadian city of approximately 650,000 people. The source included approximately 120,000 patient visits from six emergency departments. Data included gender and age of the patient, and timestamps of various milestones related to arrival, triage and registration, and discharge time. The data essentially comprise a trace file of events that have taken place. Initially, the data that one intuitively expected to be most meaningful as an input into a strategy for self-initiated patient redirection were patient waiting times and length of stay. The six months of data were augmented with one month of more detailed data that, in addition to the above, included timestamps of the beginning and the duration of individual treatment events, specialty consultation, and/or diagnostic services. Initially, only the more rudimentary data set was available, which nonetheless was the appropriate starting point for modeling efforts aimed at improving the efficiencies between emergency departments (inter-institutional). The expanded data which included specific trajectories through a particular emergency department will potentially facilitate intra-institutional modeling efforts. Data were fully anonymized and blinded to ensure patient anonymity and data handling and storage security measures were taken to ensure patient confidentiality. Any patient identifiers were removed prior to receipt of the data, and emergency department sites were masked by labels A through F.

For purposes of this initial investigation, all sites were treated equal in terms of bed capacity, although it is known that the health region consists of two larger tertiary centres and four community centres. This was done for modeling purposes, the conjecture being that if load balancing was not achievable in this simplified modeling scenario, it is unlikely to be effective once varying capacities were included.

This work also applied concepts from the telecommunications domain and from queueing theory. Telecommunications has a long tradition of optimizing loads on various resources, including priority buffers, probabilistic dropping of packets to shape network traffic, and load balancing at web servers. The application of concepts from the telecommunications domain adds an additional degree of novelty to healthcare performance modeling.

The primary means of analysis was simple statistical observations and calculations of mean wait times and correlations between different measures of waits times. The main means of illustrating ‘crowdinforming’ was through simulation using trace files or logs of arrivals as one indication of emergency department ‘busyness’.
III. RESULTS AND DISCUSSION

A. Initial Data Analysis and Model Development.

The first data intuitively expected to be associated with ‘busyness’ were the average Length Of Stay and Waiting To Be Seen durations. Determining a ‘busyness’ measure is undoubtedly (and surprisingly) very difficult. In fact, evidence within one regional health authority may not be strong in terms of its universal applicability or generalizability [26][27].

Certainly the root cause of crowding will be multifaceted, with an aspect being both urgent as well as non urgent cases. One of the reasons for not using the urgent cases here (although one could) would be that the demographic likely to consult a smart phone application or a web service would not likely be the high-priority patients (triage class 1 or 2). The majority of high-priority cases arrive by ambulance within this particular health authority and are a-priori subject to an ambulance diversion policy. In addition, if we used only urgent cases as an indicator of ‘busyness’, the self redirection of non-urgent cases would probabilistically increase the attendance at less busy emergency departments (as indicated by a low number of urgent cases) but would not add to that emergency department’s ‘busyness’ in the model. This clearly would not be the case in reality. We have tried to emphasize that in a given case, that context-specific data be used when possible.

The ‘busyness’ metric is buried in the data, but may not take the same form across data files, as similarly-labeled metrics may nonetheless capture different parameters. If various regions were to consider a similar service for patient self redirection, it would have to be guided by data from that emergency department regional area.

Following similar lines of reasoning, the average Length Of Stay and Waiting To Be Seen was considered as potentially providing insight into ‘busyness’. Figure 1 illustrates a running average Waiting To Be Seen and Length Of Stay at facility B with a 10-patient running average (for display).

All six emergency departments were analyzed and after plotting, their correlation was investigated. In all cases the correlation between Waiting To Be Seen and Length Of Stay was between 0.42 and 0.62, implying that for this data, they are moderately correlated. The data shown in Figure 1 is for the weakest correlation.

Other interesting and somewhat surprising results were associated with the Length Of Stay and number of patients in the emergency department. This was conjectured to be a reasonable control parameter, as we had initially expected there to be a statistically significant relationship between these two parameters. Figure 2 illustrates the average Length Of Stay vs. number of patients waiting in the emergency department. A similar analysis was done for all six facilities. Surprisingly, the average Length Of Stay was not influenced by the number of patients present in the emergency department. The figure here is for patients triaged at level 3-5. A similar analysis was done for Time Of Day considerations, and again, Length Of Stay was not demonstrated to be a function of Time Of Day. This result was consistent for all six facilities.

Based on these initial investigations we discounted the average Length Of Stay as a control parameter, and the average Waiting To Be Seen time was similarly discounted on the basis of similar analysis. Neither average Length Of Stay or Waiting To Be Seen would be useful as a real time control parameter (in this data set), as these averages tend to lag behind data that would be available at a given instant of time. Based on this observation, we also investigated the average Length Of Stay and average Waiting To Be Seen up to a point in time (running average) and have included this representation in Figure 3.

![Figure 1: Average Waiting To Be Seen and Length Of Stay (running average) over a one month period.](image)

![Figure 2: Average Length of Stay vs. number of patients in the emergency department, triage levels 3-5, one facility](image)
Here, the running average Waiting To Be Seen and Length Of Stay were calculated to the beginning of the hour for reference. This type of data is already available and is displayed on a ‘dashboard’ at one of the emergency departments within the regional health authority. The correlation coefficient here was 0.04. This type of information is likely useful when combined with a more meaningful measure or estimates of ‘busyness’.  

There may still be open questions regarding the impact of various triage levels and their impact on emergency department ‘busyness’. Figure 4 presents data collected over a one month period at Facility B, indicating average waiting time (Waiting To Be Seen) up to the hour for triage group (1+2) as well as triage group (3+4+5). These data are also correlated with a correlation coefficient of 0.45 (again, moderately correlated). As such, if these measures were to be included in a ‘busyness’ metric, their correlation would have to be accounted for, and either measure would suffice.

In a further exploration of a meaningful metric of ‘busyness’, patient arrival rates and departure rates from an emergency department were considered and found to have some degree of exploitable structure. Figure 5 illustrates the near instantaneous patient arrival and departure rates as seen by facility A for a representative 180-hour slice of data. This plot becomes slightly less erratic if the arrivals are shifted by the average Length Of Stay, which in the case of Facility A is approximately 4 hours. As expected, the arrival rates and departure rates are correlated to one another across all six facilities, albeit shifted in time.

B. Process Control Strategy Development Built upon Crowdinforming Application.

In conceptualizing ‘busyness’ in terms of patient arrival rates, some of the anticipated effects of patient-initiated redirection were observed. The redirection policy model was based on five minute intervals of data availability and updates. The redirection policy in this simulation was straight-forward: a percentage of patients (e.g. 75%) were set to consult the posted arrival rates and then self-redirect to the facility with the lowest present arrival rate. Figure 6 illustrates the raw arrival rates at the six regional emergency departments.

The main observation is that without any form of patient redirection modeled, there is a fair degree of diversity in the arrival rates at any particular facility. Figure 7 then illustrates the stochastic decision process modeled.
The primary objective was to simulate the effect of a stochastic weighted fair decision. In this case each patient makes a weighted fair decision as to which emergency department to attend. At the time of writing, only one of the hospitals included in this study has an emergency department wait-times display ‘dashboard’ deployed in the emergency department waiting room. Currently, this dashboard displays the number of patients in the emergency department waiting room, average waiting time of patients in the waiting room, and expected wait time to see a physician, proxied by a lagging indicator of how long those patients currently in treatment waited. Information on this dashboard is updated in near-real-time from the institutional Emergency Department Information System database, and the dashboards are expected to roll-out to all emergency departments in the region within the next 12 months.

Our model extends this notion of a wait-times dashboard and conjectures that in a reasonably short period of time, these data could be made available through a web service, such that an individual would be able to query the hospital ‘dashboards’ from an ‘emergency department wait times’ website over the internet with any browser, be it mobile cellular or wired. In addition, a text service could be established where a user would text a notable number (for example, 311EDWait) and receive a list informing of anticipated waits or load at various facilities.

For our modelling purposes, in this crowdinforming scenario investigated here, a proactive decision is made by a patient who intends to visit a hospital. In addition to the stochastic process of deciding upon going to an emergency department, each patient is also provided with estimates of load, proxied by patient arrival rates. These data are available and easily instrumented within the controls being modelled.

In (1), \( N \) represents the number of emergency department facilities. A weighting has been assigned to reflect a probabilistic decision that incorporates both travel time and arrival rates. Here \( \gamma \) reflects a person’s decision to consider redirection. The term \( \alpha \) weights the probabilistic component associated with travel time associated with distance \( d_i \), where \( d_i \) represents the distance to facility \( i \). The parameter that would be crowdinformed from a regional hospital authority would be the term associated with ‘busyness’. In this data set, the ‘busyness’ metric is patient arrival rates denoted \( \lambda_i \), where \( \lambda_i \) represents the arrival rate at facility \( i \).

For example, if one were to discount travel time and only use the measure associated with busyness (i.e. patient arrival rates), \( \alpha = 1 \) and \( \beta = 0 \). If a person would choose emergency department \( A \) with a probability of \( 0.45 \), with \( \gamma = 0.75 \) (i.e., there was a 75% chance that the person would consider self redirection) the probability of the person choosing emergency department \( A \) would be \( 0.75 \times 0.45 = 0.33 \). This would model approximately a 33% chance that the person would redirect to emergency department \( A \).

This also presents an opportunity to introduce a capacity parameter as in (2). An appealing feature about this type of measure is that it is relative between facilities. The capacity parameter would be arrived at experientially for each institution, although back-casting from available data may be a starting point.

\[
P(\text{Redirect}_j) = \gamma \left( \frac{1}{\sum_{i=1}^{N} \lambda_i} \right) + \beta \left( \frac{1}{\sum_{i=1}^{N} \lambda_i} \right) \]

Using such a method, the patients make a probabilistic decision weighted by the least anticipated wait, inferred from the busyness of the emergency departments as derived from the near real-time patient arrival rates. As a consequence, the overall surge seen between hospitals is dampened by the behaviour of informed individuals.

\[
\frac{C \sum_{i=1}^{N} \frac{1}{\lambda_i}}{C \sum_{i=1}^{N} \frac{1}{\lambda_i}}
\]

A statistical measure associated with mean and variance indicates that the load balancing is statistically significant given our behavioural assumptions, where an inference of estimated ‘busyness’ probabilistically influences a person’s decision in attending a particular emergency department. The overall effect is that of a low pass filter, smoothing out the peaks and valleys in space and, to some extent, in time. Filtering in time is expected to result from incorporation of an estimate of travel time as indicated in (1).

As the data are time varying, samples from weighted fair redirection, sorting redirect, and the raw trace file were transformed to the frequency domain and filtered, essentially removing the circadian fundamental. This allows one to estimate the variance from the periodic moving average. The standard deviation arising from the inverse transform was approximately 0.41 for the raw trace file and 0.50 for the weighted fair redirection for Hospital B (selected arbitrarily). It was somewhat unexpected that the weighed fair redirection
performed more poorly in terms of variance than the original trace file, even though an element of balancing appears evident from Figure 8. Figure 8 re-adjusts the trace file over several passes through the trace file to more fully represent the steady state associated with this control strategy.

Although preliminary, this work represents one of the first modelled instances of the potential of crowdinforming in providing a policymaker with a simulation to aid decision support. The decision would be to create public access – ‘crowdinforming’ – to ‘dashboard’ type information that is currently available in emergency departments or electronic records. The public available ‘dashboard’ would be accessible via the web or a smart phone app. Crowdinforming serves to filter the simulated patient surges at a given facility by balancing patient load between facilities (to each facility’s respective capacity). As the ‘dashboards’ are a relatively new concept within the regional health authority considered in this work, it is unlikely that exactly the same data would be presented to a patient considering self redirection.

The model is preliminary and currently uses one metric – patient arrival rates – as an indicator of ‘busyness’ and as the basis for crowdinforming. Intuitively expected metrics such as the number of patients waiting, the average Length Of Stay, and Waiting To Be Seen were not found to be meaningful in this particular data set, and thus were also ineffective as an input into a process control strategy for patient redirection.

C. Limitations

This work did not set out to develop a model to fit observed data, nor to develop a ‘busyness’ model of an emergency department. The objective was to develop a process control strategy built on crowdinforming, and in doing so, a meaningful metric of ‘busyness’ is required. Thus, this work explored the data in order to define a meaningful metric for this data set, as a pre-requisite to the development of a meaningful process control strategy for crowdinforming. The work implied that the input parameters (in this case, a ‘busyness’ metric) may be ‘data-unique’ and must be defined and verified for each respective data set. The validity of patient arrival rates as an input parameter has not been established definitively; however, the work did indicate that other parameters, surprisingly, are not a good metric for this particular data set. Additional statistical analysis needs to be carried out on the trace file data, to facilitate a refinement of the model to account for the interplay of multiple variables.

A limitation associated with research of this nature is that stochastic models of behaviour have to be estimated. Modelling social dynamics is a difficult and messy problem; however, insights garnered from simulation and modelling are nonetheless useful and would be used in support of an otherwise best-guess decision.

Also, in reality, not all patients can easily redistribute themselves to all sites equally. For example, the most acute patients (i.e., triage levels 1 or 2) need to get to the nearest or specialized emergency department, and most likely by ambulance. Additionally, these cases wait very little once they arrive at a hospital, regardless of which site. To further complicate this, certain types of conditions get directed to a specific site. For example, in the regional health authority represented by the data, trauma and stroke cases are most often directed to site B whereas cardiac cases are most likely sent to site F. Patients assessed at triage level 3 are most likely sick enough prefer the nearest emergency department, whereas the patients assessed at CTAS level 4 and 5 (i.e., least acute) are likely the most sensitive to wait times and length of stay issues. In addition, several of the larger and busier sites have resources/services not available at smaller sites, so a length of stay at one site may not be same if a patient chooses a different site with a different scope of services available. Future models need to take these considerations into account, extending the objective of these initial considerations of determining an appropriate metric for ‘busyness’.

The present framework appropriately considers systemwide impacts with simplified assumptions on respective emergency department capacity, prior to intra-institutional simulation and other institutional refinements. The framework developed here allows policies to be simulated and evolved, adding a qualitative assessment to decisions that may otherwise be experiential or best intent. This work is one of the first demonstrations of the crowdinforming intervention as a variant of the broader concept of crowdsourcing, and it demonstrates the role that similar technologies which provide crowdinformed decisions will continue to play in the future.

This study was limited to data from one regional health authority. Specific results of similar analyses at other regions would likely be different. A limitation of modeling is that the assumption of ‘busyness’ is consistent with our intuition that process times are affected by prior patient arrivals.

These limitations imply that data collected from a multiple-emergency department system have to be carefully considered as regional specifics may differ. However, as data become more widely available, there is increased opportunity for analysis and improved interpretation of causality concerning ‘busyness’. Similar considerations also apply to the feedback control strategy one may be interested in. Fortunately, there is an increasing number of modeling and simulation tools and methods readily amenable to this type of policy modeling. One means of providing a meaningful ‘busyness’ metric is streaming video of each facility. The technology is in place, although implementation has political challenges. On the other hand, one can exploit Stein’s paradox where an estimate of ‘busyness’ is enhanced when multiple parameters, apparently and superficially unrelated, are estimated simultaneously.
IV. CONCLUSION

This work illustrated the role that data-driven modelling can play in developing policy or supporting public health decisions. The work derives a preliminary process control strategy upon which to model the effects of self-initiated patient redirection on hospital patient loads. The work illustrates how one specific intervention— that of proactive information dissemination, or crowdingform— can potentially provide a degree of process control in the form of patients' self-initiated re-direction, resulting in load balancing at emergency departments. The work illustrates that meaningful strategies rely on appropriate data- or context-specific choices in the variable or variables chosen as metrics of 'busyness' and consequently to their broadcast through crowdingform. The results are not intended as specific recommendations for a regional health authority, but rather illustrate a flexible framework around which applications of this type can be evaluated and potential outcomes assessed.

The preliminary model was derived from simple statistical analyses of the available Emergency Department Information System data in the regional health authority, in order to test support for preliminary conjectures. As the initial conjectures were discounted by the data, the model was refined. The application of a stochastic patient decision process representative of crowdingform was demonstrated to contribute to load balancing at individual hospital emergency departments. The evidence for the findings was the trace file data associated with real emergency department visits within a regional health authority, and the balancing of arrival rates as a consequence of self redirection when a metric of emergency department 'busyness' is presented to the public. It was noted however, that although loads were being balanced across the facilities, the simulation did not indicate any alleviation in the variance associated with the arrival rates. In this regard, an increase in variability may detrimentally contribute to the patient flow within the various emergency departments.

Future work will include the utilization of running averages of the Waiting To Be Seen time and Length Of Stay once more data become available. Preliminary analysis indicates that these are uncorrelated and both could be included in a more detailed estimate of 'busyness'.

ACKNOWLEDGMENTS

The authors would like to thank the Winnipeg Regional Health Authority for access to some of the preliminary data from their emergency department information system.

REFERENCES


