Designing Offload Zones to reduce Offload Delay

Young European Queueing Theorists workshop
Stochastic Service Systems

Peter T. Vanberkel, PEng, PhD
Dalhousie University
Halifax, NS, Canada
Outline

Designing Offload Zones to reduce Offload Delay?

➤ What are Offload Zones & what is Offload Delay?
➤ Failure modes at CDHA’s Offload Zone
➤ Initial Offload Zone model
➤ Future plans
Offload delay

- Offload delay is a prolongation of the interval between an ambulances arrival in the emergency department (ED) and transfer of patient care.
- Typically occurs when EDs are crowded and there are no beds available.
Offload delay

- Impacts on patients include:
  - delays in patient care
  - poor pain control
  - delayed time to antibiotics
  - increased morbidity and potentially increased mortality

- Impacts on EHS systems:
  - decreases the availability of ambulances for the community
  - could result in a need for additional EHS staffing to compensate
Offload delay

- It is estimated that EHS spends
  - 29,000 hours / year on offload delay
  - ~$754,000 / year

- The 90th percentile of delivery intervals in NS
  - 2002: 24 minutes 37 seconds
  - 2007: 109 minutes 2 seconds
  - 2010: 114 and 142 minutes (Halifax EDs only)

- The Nova Scotia Gov’t benchmark is 20 minutes 90% of the time
Without the offload zone:
Ambulances queue waiting to offload their patients

Benefits

Reduced offload delay
Better ambulance coverage

If nothing changes then benefits are easy to quantify

With the offload zone:
Multiple patients wait with a paramedic & nurse, relieving the ambulances
Offload Zone Anecdotes

- Offload Zones are often at full capacity
- Offload delays still occur
Studies

- **HFMEA - Healthcare Failure Mode and Effect Analysis**
  - Alix JE Carter MD MPH FRCPC, Medical Director of Research EHS, Assistant Professor, Dalhousie Dept. of Emergency Medicine
  - James B Gould BSc., Student Researcher, Dalhousie Dept. of Emergency Medicine
  - Peter Vanberkel, PEng, PhD., Assistant Professor, Department of Industrial Engineering
  - Jan L Jensen ACP MAHMR, Research Leader EHS
  - Jolene Cook MD, EMS Fellow, Dalhousie Dept. of Emergency Medicine
  - Steven Carrigan MPA, Research and Statistical Officer, EHS
  - Mark R Wheatley ACP, Operations Manager, EHS
  - Andrew H Travers MSc MD FRCPC, Provincial Medical Director, EHS, Associate Professor, Dalhousie Dept. of Emergency Medicine and Community Health & Epidemiology

- **Modelling**
  - Peter Vanberkel, PEng, PhD., Assistant Professor, Department of Industrial Engineering
  - Corine Laan, MSc Student, Stochastic Operations Research, University of Twente
  - Richard J. Boucherie, PhD., Professor, Stochastic Operations Research, University of Twente
HFMEA

1. Develop detailed process flowchart
2. Identify potential failures (patient care and throughput)
3. Rate severity
4. Rate likelihood
5. Prioritize potential failures
Main steps

1. Patient transported by ambulance
2. Arrival in ED
3. Transfer of patient Care from EHS to OZ
4. Patient assessment in OZ
5. Patient care in OZ
6. Patient transfer out of OZ to ED
4. Patient assessment in Offload Zone
   - EKG required?
   - Blood work required?
   - Imaging required?
   - Consult MD
   - Transport patient to receive diagnostic

5. Patient care in Offload Zone
   - Intervention required?
   - MD sees patient
   - MD writes orders
HFMEA - Results

- “High Risk Failures” related to patient care
  - Delay in assessment/patient care because of limited equipment
  - Patient not properly assessed/cared for because lack of privacy
  - ...

- “High Risk Failures” related to throughput
  - Patient not placed in ED from OZ because patient already receiving care in OZ
  - Patient not placed in ED from OZ because ED waiting for patient to receiving some level of care from the OZ
  - ...

DALHOUSIE UNIVERSITY
Inspiring Minds
Original intent of OZ was to reduce ambulance offload delay, however...

- OZ provides extensive care, similar to an ED
- Incentive to admit an OZ may be lost
- The OZ has become an extension of the ED

Summed up: OZ may cause offload delay
How does patient selection affect the performance of the Offload Zone?
<table>
<thead>
<tr>
<th>Without OZ</th>
<th>With OZ</th>
<th>Metric: Ambulances waiting (X)</th>
<th>Patient selection</th>
<th>Tiebreaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queue 2</td>
<td>Max(0, Queue 2 – OZ capacity)</td>
<td>Lowest CTAS</td>
<td>Control: Queue 2</td>
<td></td>
</tr>
<tr>
<td>Lowest CTAS</td>
<td>Lowest CTAS</td>
<td>Policy 1: Queue 2</td>
<td>Policy 2: Queue 1</td>
<td>Policy 3: Queue 2 with probability $p_{oz}$</td>
</tr>
</tbody>
</table>

• Ignore shorter service times due to starting treatment in OZ
Polling Model - single server, cyclic order

- M/G/c with multiple priority classes
- ~M/M/c with multiple priority classes
  - Patient class specific E[W] and E[L] in closed form
  - Queue Length distributions from Markov Chain
Assumptions

- Steady state
- Stationary arrival process
- System is stable
- Poisson arrival process
Continuous Time Markov chain

- Define state space: $S$
- Define rates from state $s$ to $s'$
- Solve limiting distribution $\pi$
  - Balance flow equations
  - Normalization equations
  - System of linear equations solver
Continuous Time Markov chain

\[ S = [N_b, N_b^H, N_1, N_2, N_3] \]

Where:

- \( N_b \) is ED beds in use
- \( N_b^H \) is ED beds in use by low CTAS patients
- \( N_1 \) waiting ambulance arrivals (low CTAS)
- \( N_2 \) waiting walk-in arrivals (low CTAS)
- \( N_3 \) waiting walk-in arrivals (high CTAS)

Assumes

- 3 patient types
- 2 service times: low CTAS and high CTAS
Continuous Time Markov chain

Generator matrix, Q:

\[
q_{(N_b, N_b^H, N_1, N_2, N_3), (N_b+1, N_b^H, N_1, N_2, N_3)} = \lambda_3 \quad (N_b^H \leq N_b < c; N_1, N_2, N_3 = 0)
\]

\[
q_{(N_b, N_b^H, N_1, N_2, N_3), (N_b+1, N_b^H+1, N_1, N_2, N_3)} = \lambda_1 + \lambda_2 \quad (N_b^H \leq N_b < c; N_1, N_2, N_3 = 0)
\]

\[
q_{(N_b, N_b^H, N_1, N_2, N_3), (N_b-1, N_b^H, N_1, N_2, N_3)} = (N_b - N_b^H)\mu_L \quad (N_b^H \leq N_b < c; N_1, N_2, N_3 = 0)
\]

\[
q_{(N_b, N_b^H, N_1, N_2, N_3), (N_b+1, N_b^H, N_1-1, N_2, N_3)} = p_{OZ} N_b^H \mu_H \quad (N_b^H \leq N_b < c; N_b^H > 0; N_1, N_2 > 0)
\]

\[
q_{(N_b, N_b^H, N_1, N_2, N_3), (N_b, N_b^H, N_1, N_2-1, N_3)} = (1 - p_{OZ}) N_b^H \mu_H \quad (N_b^H \leq N_b < c; N_b^H > 0; N_1, N_2 > 0)
\]

... where:

- \(c\) is beds available
- \(\lambda\) is mean arrival rate
- \(\mu\) is mean service rate
Continuous Time Markov chain

- Limiting distribution, $\pi$:
  - linear balance flow equations:
    - $\pi_s \sum_{s' = s} q_{s,s'} = \sum_{s' = s} \pi_{s'} q_{s',s} \forall s \in S$
    - $0 = \pi Q$
  - and normalizing equation:
    - $\sum_{s \in S} \pi_s = 1$

- Successive over-relaxation approximation computed with Matlab BiCGSTAB
Continuous Time Markov chain

- **Metrics:**

  1. \( E[X] = \sum_{s \in S} [N_1^s - b_{OZ}]^+ \pi_s \)
  
     Where \( b_{OZ} \) is the capacity of OZ

  2. \( E[W] = \frac{E[X]}{\lambda_1} \)

- Queue length distribution

  X is ambulances waiting
Data

<table>
<thead>
<tr>
<th></th>
<th>CTAS 1</th>
<th>CTAS 2</th>
<th>CTAS 3</th>
<th>CTAS 4</th>
<th>CTAS 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient type distribution(^1)</td>
<td>1%</td>
<td>15%</td>
<td>45%</td>
<td>32%</td>
<td>7%</td>
</tr>
<tr>
<td>Fraction arriving by ambulance(^2)*</td>
<td>78%</td>
<td>20.6%</td>
<td>20.6%</td>
<td>3.8%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Average service time(^3) (minutes)</td>
<td>155</td>
<td>190</td>
<td>110</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

*12% in total

\(^1\) Capital District Emergency Services Council. Quarterly report, Oct 2012
\(^3\) Canadian Institute for Health Information. Understanding emergency department wait times, 2005.
# Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Control</th>
<th>$\lambda_1, \lambda_2, \lambda_3$</th>
<th>$\mu_H, \mu_L$</th>
<th>$b_{OZ}$</th>
<th>ED beds</th>
<th>$m_1, m_2, m_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No OZ</td>
<td>Control</td>
<td>0.40, 1.53, 1.25</td>
<td>0.45, 2</td>
<td>2</td>
<td>5</td>
<td>10, 10, 25</td>
</tr>
<tr>
<td>OZ has priority</td>
<td>$p_{oz}=1$</td>
<td>0.40, 1.53, 1.25</td>
<td>0.45, 2</td>
<td>2</td>
<td>5</td>
<td>10, 10, 25</td>
</tr>
<tr>
<td></td>
<td>$p_{oz}=0.95$</td>
<td>0.40, 1.53, 1.25</td>
<td>0.45, 2</td>
<td>2</td>
<td>5</td>
<td>10, 10, 25</td>
</tr>
<tr>
<td></td>
<td>$p_{oz}=0.05$</td>
<td>0.40, 1.53, 1.25</td>
<td>0.45, 2</td>
<td>2</td>
<td>5</td>
<td>10, 10, 25</td>
</tr>
<tr>
<td>OZ does not have priority</td>
<td>$p_{oz}=0$</td>
<td>0.40, 1.53, 1.25</td>
<td>0.45, 2</td>
<td>2</td>
<td>5</td>
<td>10, 10, 25</td>
</tr>
</tbody>
</table>

* $\rho = 0.98$
### Scenarios

<table>
<thead>
<tr>
<th></th>
<th>$E[X]$</th>
<th>$P(X \geq 1)$</th>
<th>$P(X \geq 2)$</th>
<th>$P(X \geq 3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No OZ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.153</td>
<td>0.13</td>
<td>0.02</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>OZ has priority</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{oz}=1$</td>
<td>0.004</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{oz}=0.95$</td>
<td>0.005</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$p_{oz}=0.7$</td>
<td>0.013</td>
<td>0.01</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>$p_{oz}=0.5$</td>
<td>0.036</td>
<td>0.02</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>$p_{oz}=0.3$</td>
<td>0.164</td>
<td>0.08</td>
<td>0.041</td>
<td>0.022</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$p_{oz}=0.05$</td>
<td>0.884</td>
<td>0.252</td>
<td>0.189</td>
<td>0.142</td>
</tr>
<tr>
<td><strong>OZ does not have priority</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{oz}=0$</td>
<td>0.992</td>
<td>0.27</td>
<td>0.21</td>
<td>0.16</td>
</tr>
</tbody>
</table>

$X$ is ambulances waiting
OZ benefits are only achieved if patients are given priority at least 35% of the time.

Benefits increase only slightly after patients are given priority 60% of the time.
\[ p'_{oz} = 0.35 \text{ (Breakeven point)} \]

\[ p''_{oz} = 0.6 \text{ (Diminishing returns threshold)} \]

\[ \Delta E[X] < 0.005 \text{ per 0.05 increase in } p_{oz} \]
Larger problems

<table>
<thead>
<tr>
<th>Run Time (seconds)</th>
<th>ED beds</th>
<th>OZ beds</th>
<th>$m_1, m_2, m_3$</th>
<th>$p'_{oz}$</th>
<th>$p''_{oz}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>687</td>
<td>5</td>
<td>2</td>
<td>10, 10, 25</td>
<td>0.35</td>
<td>0.6</td>
</tr>
<tr>
<td>25,606</td>
<td>5</td>
<td>2</td>
<td>20, 20, 50</td>
<td>0.35</td>
<td>0.6</td>
</tr>
<tr>
<td>14,161</td>
<td>30</td>
<td>2</td>
<td>10, 10, 25</td>
<td>0.3</td>
<td>0.55</td>
</tr>
<tr>
<td>500,540</td>
<td>30</td>
<td>2</td>
<td>20, 20, 50</td>
<td>0.3</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Larger problems

\[ E[X] \]

- 10,10,25
- 20,20,50
- 10,10,25*
- 20,20,50*

* No OZ
OZ bed capacity

\[ E[X] \]

\[ p_{OZ} \]

Graph showing the expected value \( E[X] \) for different numbers of OZ beds from 1 to 5, along with the no OZ case. The x-axis represents \( p_{OZ} \) and the y-axis represents \( E[X] \). The lines indicate the expected value for each case, with the no OZ case being the horizontal line at the bottom.
Less busy EDs

\[ \rho \in \{0.98, 0.95, 0.90, 0.85\} \]

\[ E[X] \text{ vs } \rho \text{ and } \rho_{oz} \]
Conclusions

Model

- Insensitive to reasonable queue length limits
- Sensitive to OZ capacity
- Sensitive to service load

Offload Zone

- When priority is disproportionately given to walk-in patients implementing an offload zone will actually increase offload delay
- Depending on several factors, such as service load, number of beds and the patient mix:
  - Offload delay can be decreased if priority is given to offload zone patients at least a certain fraction of time \( (p'_{OZ}) \)
  - Decreases in offload delay becomes negligible when priority is given to offload zone patients at least a certain fraction of time \( (p''_{OZ}) \)
Next steps

- **Model**
  - Approximate $p'_{OZ}$ and $p''_{OZ}$
  - Validate with simulation*
  - Service rate distribution

- **Case study**
  - Detailed historical data
  - Validation with QEII or DGH ED (Actual baseline)
  - $p_{OZ}$ to achieve NS Gov’t benchmark of 20 minutes 90% of the time

- **General**
  - OZ bed sharing policy

- **Holy Grail**
  - Reserving ED beds in place of OZ
HFMEA: The Offload Zone to Mitigate EMS Offload Delay in the Emergency Department: A Process Map and Hazard Analysis

Poster Accepted at:
- Canadian Association of Emergency Physicians Conference
- Maritime Trauma and Emergency Medicine conference
- National Association of EMS Physicians
- Dalhousie Division of Emergency Medical Services Research Day

Journal Article:
- Canadian Journal of Emergency Medicine Special Issue for EMS (Under revision)

Modelling: The offload zone as a solution to the offload delay of the emergency medical services

Poster Accepted at:
- First International Workshop on Planning of Emergency Services
Announcement

The 17th Annual Conference for Canadian Queueing Theorists and Practitioners will be held at Dalhousie University, (Late) August 2015

General goal: promote research and applications of queueing theory
Specific 2015 goal: Attract European speakers!

Details / Registration: email peter.vanberkel@dal.ca
Questions?