Location and Capacity Optimization of Points of Dispensing for a Decision Support System

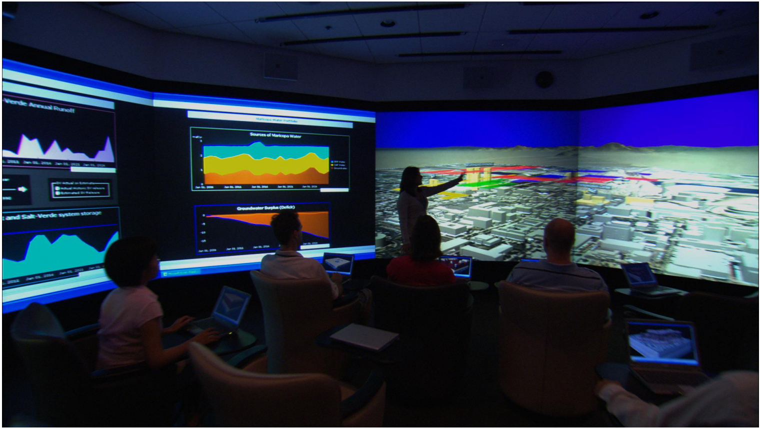
Adrián Ramírez Nafarrate
Instituto Tecnológico Autónomo de México (ITAM)
Department of Industrial & Operations Engineering

January 7, 2016
U.S.-Mexico Workshop on
Optimization and its Applications

Collaborative project:

Ramirez-Nafarrate, A., J.D. Lyon, J.W. Fowler and O. Araz. 2015. "Point-of-Dispensing Location and Capacity Optimization for a Decision Support System". *Production and Operations Management* 24:8. 1311-1328.

ASU Decision Theater:



dt.asu.edu

Outline

Outline

PODs

Points-of-Dispensing (PODs)

- Mass dispensing operations (prophylaxis).
- Typically located in public facilities (e.g., schools, parking lots).
- POD classes: Open vs Closed, Walk-in vs Drive-Thru.
- Distribuition of antibiotics must be completed within a frame time.
- For instance: anthrax emergency requires distribution of antibiotics to 100% of the target population within 36-48 hours.

Problems

- How many PODs should be open?
- Where should PODs be located?
- How big should the PODs be?
- Scarce Resources! (volunteers, time, etc.)

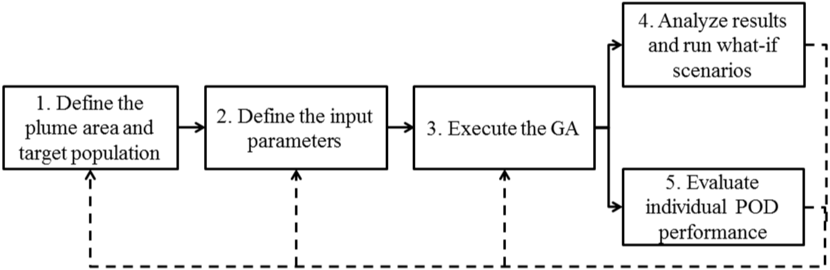
Literature Review

- POD Management Literature:
  - Evaluation of dispensing plans using simulation (Lee 2008, Whitworth 2006, Aaby et al. 2006).
  - Comparison of dispensing strategies: walk-in PODs, drive-thru PODs, mail and pharmacy dispensing (Richter and Khan 2009).
  - Decision support system: Real Opt©(Lee 2006 and 2009).
- Our Contributions:
  - Solve the location and capacity problem simultaneously.
  - Consider POD capacity as a decision variable, not an input.
  - Consider not only geospatial information, but also demographics.
  - Provide valuable information by relaxing limiting assumptions/constraints.

Outline

Decision Support System

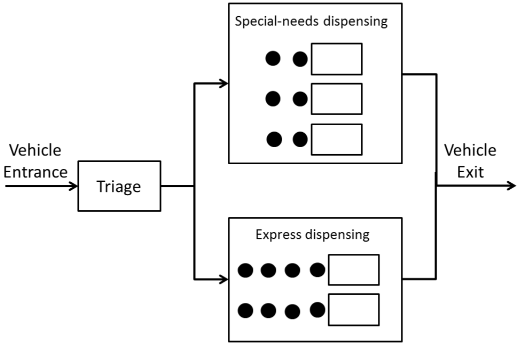
- Decision Support System (DSS) for emergency response:
  - Module 1: GIS Model.
  - Module 2: POD Location and Staffing Optimization.
  - Module 3: $What - if$ Evaluation.
  - Module 4: Individual POD Simulation.
- Planning process for anthrax emergency using the DSS:



Outline

Problem Objective and POD Layout

- Optimize POD location and staffing.
- Objective: minimize the average travel time to PODs plus the average waiting time within PODs (HtoP: House-to-Prophylaxis).
- POD layout:



Model Assumptions

- Modified version of the p-median problem.
- Demand points: census tracts.
- Each household represents a vehicle to serve in a POD.
- Rectilinear distance is used to compute travel time.
- Average waiting time in a POD is computed as a weighted average of the dispensing areas using $M/G/1$ queueing formulations.
- Queueing formulations are approximations.
- Triage station and other support staff are not included in the model.
- Number of PODs to open and candidate POD list is obtained *a priori*.

Model Formulation

Let $I = \{1, 2, ..., M\}$ be the set of census tracts in the area under study, $J = \{1, 2, ..., N\}$ be the set of candidate PODs and $K = \{1, 2, ..., L\}$ be the set of dispensing station types required for each POD (in our description, L=2, to cover express and special-needs). The problem consists of (i) determining the set $S = \{1, 2, ..., \delta\}$, which is the list of PODs that should be open ($S \subseteq J$); (ii) for each census tract $i \in I$ determining the assignment of i to POD $j \in S$; and (iii) for each POD $j \in S$ determining the number of servers assigned to each dispensing station $k \in K$. The objective function consists of minimizing the sum of the average vehicle travel time from all households to the assigned POD ($\overline{T}$) and the average vehicle waiting time across all PODs ($\overline{W}$).

Model Formulation

Decision Variables:

$$x_{ij} = \begin{cases} 1 & \text{if demand point } i \text{ is assigned to POD } j, \\ 0, & \text{otherwise,} \end{cases}$$

$$y_j = \begin{cases} 1 & \text{if candidate POD } j \text{ is open,} \\ 0, & \text{otherwise,} \end{cases}$$

c_{kj}: Number of servers in dispensing area k at POD j.

λ_{kj}: Mean arrival rate per server of dispensing area k at POD j.

Model Formulation

Parameters:

δ: Number of PODS to open determined *a priori* by the decision makers.

T: Target dispensing time.

m_i: Number of households in demand point i.

α_i: Compliance factor in demand point i, $\alpha_i \in (0, 1)$.

P: Total number of households (vehicles) to serve by open PODs, $P = \sum_{i \in I} \alpha_i m_i$.

p_{ki}: Fraction of households (vehicles) from demand point i that require service from dispensing area k, (p_{2i}=1-p_{1i}).

Model Formulation

Parameters (continued...):

μ_k: Mean service rate of each server in dispensing area k.

σ_k^2: Variance of service time of each server in dispensing area k.

cv_k: Coefficient of variation of the service time in dispensing area k, $cv_k = \sigma_k \mu_k$.

ρ_k^{max}: Maximum desired utilization of dispensing area k (0 $\leq \rho_k^{max} \leq 1$).

Model Formulation

Parameters (continued...):

${c_k}_j^{min}$: Minimum number of servers for dispensing area k to assign to POD j.

${c_k}_j^{max}$: Maximum number of servers for dispensing area k to assign to POD j.

C_k: Total number of available servers for dispensing area k.

ν: Factor determining average travel speed as a function of distance.

d_{ij}: Distance from demand point i to POD j.

Model Formulation

$$Min \frac{1}{P}\left[\sum_{i\in I}\sum_{j\in J}\alpha_i m_i x_{ij}(t_{ij}+\overline{W}_j)\right] \quad (1)$$

Subject to:

$$\sum_{j\in J} x_{ij} = 1 \qquad \forall i \in I \quad (2)$$

$$\sum_{j\in J} y_j \leq \delta \quad (3)$$

$$x_{ij} \leq y_j \qquad \forall i \in I \forall j \in J \quad (4)$$

$$\sum_{j\in J} c_{kj} \leq C_k \qquad \forall k \quad (5)$$

$$y_j c_{kj}^{min} \leq c_{kj} \leq y_j c_{kj}^{max} \qquad \forall j \in J \forall k \quad (6)$$

Model Formulation (continuación...)

$$\lambda_{kj} = \frac{\sum_{j \in J} \alpha_i m_i p_{ki} x_{ij}}{T c_{kj}} \qquad \forall j \in J \forall k \qquad (7)$$

$$\lambda_{kj} \leq \mu_k \rho_k^{max} \qquad \forall j \in J \forall k \qquad (8)$$

$$\overline{W}_j = f(\lambda_{kj}, \mu_k, cv_k, c_{kj}) \qquad \forall j \in J \qquad (9)$$

$$t_{ij} = \frac{d_{ij}}{\nu \sqrt[3]{d_{ij}}} \qquad \forall i \in I \forall j \in J \qquad (10)$$

$$x_{ij}, y_j \in \{0, 1\} \qquad \forall i \in I \forall j \in J \qquad (11)$$

$$c_{kj} \in Z^+ \qquad \forall j \in J \forall k \qquad (12)$$

$$\lambda_{kj} \in R^+ \qquad \forall j \in J \forall k \qquad (13)$$

Average Waiting Time Calculation

We used two equations to approximate the average waiting time:

- Assuming that arrivals are evenly distributed (ED) across the servers:

$$\overline{W}_j = \sum_{k\in K} \left(\frac{\lambda_{kj}}{\sum_{k\in K} \lambda_{kj}}\right) \left(\frac{1+cv_k^2}{2}\right) \frac{\left(\frac{\lambda_{kj}}{\mu_k}\right)}{\mu_k-\lambda_{kj}}$$

- Assuming that arrivals always choose the shortest queue (SQ)

$$\overline{W}_j = \sum_{k\in K} \left(\frac{\lambda_{kj}}{\sum_{k\in K} \lambda_{kj}}\right) \left(\frac{1+cv_k^2}{2}\right) \frac{\left(\frac{\lambda_{kj}}{\mu_k}\right)}{c_{kj}(\mu_k-\lambda_{kj})}$$

Validation of Queueing Formulations

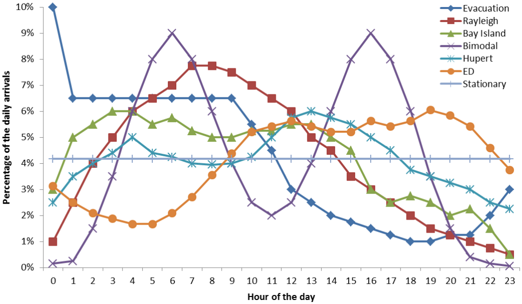
Two problems arise when using queueing formulations (QF) for this problem:

- Arrivals to PODs could be non-stationary.
- The system is transient.

We used a simulation model (Sim) of one POD under different arrival patterns and compared the results with QF.
We assume that the station has 4 servers. The mean arrival rate in QF (λ) is set to 78.4 vehicles/hr (station utilization = 98%).

Non-stationary Arrival Patterns

We implemented the following patterns in the simulation model:



In order to determine the mean arrival rate per hour in the Sim model, the mean arrival rate used in QF (78.4 v/h) could represent the mean, median, maximum or other quantile of the arrival pattern (θ).

QF vs Sim

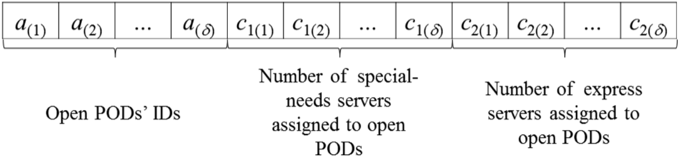
Average waiting time (min) using QF and Sim.

			Simulation Results (θ statistic with rate of 78.4 vehicles per hour)									
Pattern	Rule	QF	Mean	Median	62.5%	66.7%	70.8%	75%	79.2%	83.3%	87.5%	Max
Evacuation	SQ	19.7	218.2	351.8	18.8	18.8	18.8	18.8	18.8	18.8	18.8	0.7
Rayleigh	SQ	19.7	195.5	181.3	97.3	33.4	33.4	17.0	17.0	7.7	7.7	2.4
Bay Island	SQ	19.7	127.5	31.8	16.2	16.2	16.2	8.0	8.0	8.0	8.0	2.6
Bimodal	SQ	19.7	172.0	275.4	33.2	33.2	33.2	33.2	5.0	5.0	5.0	1.9
Hupert	SQ	19.7	52.8	59.6	32.4	32.4	27.1	11.2	11.2	11.2	4.1	1.8
ED	SQ	19.7	92.8	31.9	15.7	9.1	9.1	9.1	9.1	5.5	5.5	2.5
Stationary	SQ	19.7	14.6	14.6	14.6	14.6	14.6	14.6	14.6	14.6	14.6	14.6
Evacuation	EvD	78.9	228.4	350.3	26.1	26.1	26.1	26.1	26.1	26.1	26.1	2.7
Rayleigh	EvD	78.9	206.0	190.8	105.4	41.6	41.6	25.4	25.4	15.7	15.7	7.9
Bay Island	EvD	78.9	140.0	41.6	27.5	27.5	27.5	18.2	18.2	18.2	18.2	9.4
Bimodal	EvD	78.9	179.3	277.6	38.7	38.7	38.7	38.7	10.5	10.5	10.5	6.1
Hupert	EvD	78.9	72.3	77.1	44.8	44.8	38.1	18.1	18.1	18.1	10.2	6.2
ED	EvD	78.9	102.5	40.6	26.2	18.7	18.7	18.7	18.7	14.5	14.5	8.6
Stationary	EvD	78.9	35.4	35.4	35.4	35.4	35.4	35.4	35.4	35.4	35.4	35.4
Min difference			5.14	1.81	0.95	0.95	0.95	0.95	0.95	0.95	0.95	5.14
Avg difference			89.81	108.67	29.91	27.13	27.23	30.38	32.48	34.39	35.47	42.00
Max difference			198.44	332.07	77.53	60.28	60.28	60.88	68.46	68.46	68.76	76.25

Outline

Solution Approach

- Traditional p-median problem is NP-hard.
- Model includes nonlinear equations.
- We propose a Genetic Algorithm to solve the problem.
- The objective of the GA is to find Pareto-optimal solutions of the tuple $(\overline{W}, \bar{t})$.
- Chromosome length: 3δ genes.



GA Process

- The GA is based on a greedy assumption: assign each demand point to the nearest POD.
- The initial population is randomly generated.
- Each chromosome is checked for feasibility: Available servers, target dispensing time and stability conditions.

Outline

Constraint Relaxation

- The GA may not find feasible solutions depending on population size and available resources.
- It is necessary to have an approach to deal with infeasible solutions.
- We implemented an approach that allows decision makers relaxing three constraints:
  1. Relax the assignment of demand points to nearest POD.
  2. Relax capacity constraints.
  3. Relax the target dispensing time constraint.

Relax Assignment to Nearest POD

Algorithm 1 Relax nearest POD assumption

Rank the census tracts from the highest to the lowest population.
Make $\lambda_{k_j}^{current} = 0, \; \forall k \in K, \forall j \in J$.
for $i = 1 \rightarrow M$ **do**
    Assign census tract i to the nearest open POD j that satisfies: $\lambda_{k_j}^{current} + \frac{\alpha_i m_i p_{ki}}{c_{kj} T} < \mu_k \rho_k^{max}, \forall k \in K$
    Update $\lambda_{k_j}^{current} = \lambda_{k_j}^{current} + \frac{\alpha_i m_i p_{ki}}{c_{kj} T}, \forall k \in K$
end for

Relax Capacity Constraints

Algorithm 2 Relax capacity constraints

Assign each census tract to the nearest open POD.
for $j = 1 \to \delta$ **do**
  for $k = 1 \to L$ **do**
    if $\frac{\sum_{i \in I} \alpha_i m_i p_{ki} x_{ij}}{c_{kj} \mu_k T} > \rho_k^{max}$ **then**
      $c_{kj} = \left\lceil \frac{\sum_{i \in I} \alpha_i m_i p_{ki} x_{ij}}{\mu_k T \rho_k^{max}} \right\rceil$
    end if
  end for
end for

Relax Target Dispensing Time

Algorithm 3 Relax target dispensing time constraint

1: Assign each census tract to the nearest open POD.
2: Assign servers to each POD using the GA.
3: **for** $j = 1 \rightarrow \delta$ **do**
4: $\quad T_j^{adj} = T \times max_k \left\{ \frac{\sum_{i \in I} \alpha_i m_i p_{k\,i} x_{i\,j}}{c_{k\,j} \mu_k \rho_k^{max}} \right\}$
5: **end for**
6: Return $max_j \left\{ T_j^{adj} \right\}$ and $\overline{T_j^{adj}}$.

Outline

Anthrax Emergency in the Phoenix Metro Area

- All census tracts within 25 miles from the Capitol building must receive prophylaxis within 48 hours.
- We compare the proposed co-optimization method with the sequential method (first optimizes location, then capacity).
- Input parameters:

Parameter	Value	Parameter	Value
M	695 census tracts	μ_1	20 vehicles/hr
α_i	1, $\forall i$	σ_1^2	2400 sec^2
p_{1i}	Uniform(0.35, 0.45), $\forall i$	cv_1	0.2721
p_{2i}	1-p_{1i}, $\forall i$	μ_2	80 vehicles/hr
P	1,194,251 vehicles	σ_2^2	37.5 sec^2
T	48 hours	cv_2	0.1360
N	200 candidate PODs	c_{kj}^{min}	1 servers, $k=1,2$, $\forall j$
ν	15 mph	c_{kj}^{max}	60 servers, $k=1,2$, $\forall j$

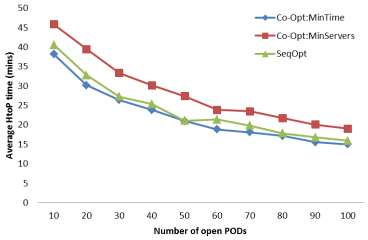
Relaxing the Capacity Constraints

Using the proposed co-optimization, we obtain two solutions:

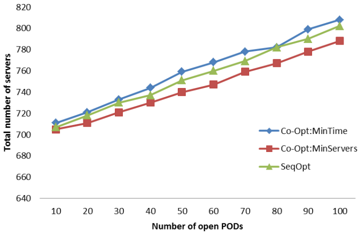
- Co-opt:MinTime: Solution that minimizes the average HtoP time with co-optimization.
- Co-opt:MinServers: Solution with the minimum total number of servers amongst the last population of chromosomes.

Relaxing the Capacity Constraints

Average HtoP time and total number of servers relaxing the capacity constraints.



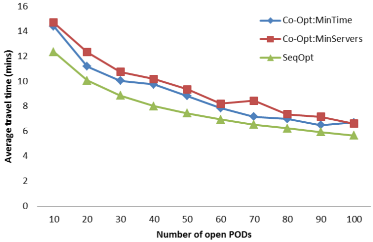
(a) Average HtoP time



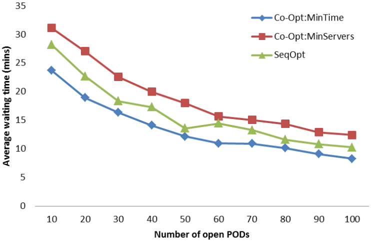
(b) Total number of servers

Relaxing the Capacity Constraints

Average travel time and average waiting time relaxing the capacity constraints.



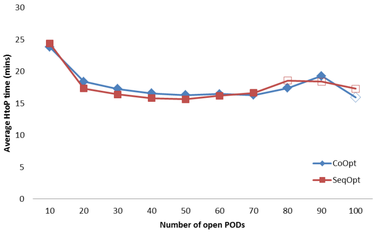
(c) Average travel time



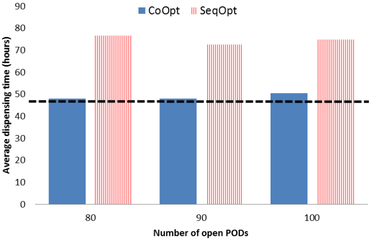
(d) Average waiting time

No Relaxation

POD performance with fixed capacity (780 servers). The empty markers are infeasible scenarios with relaxation of target dispensing time.



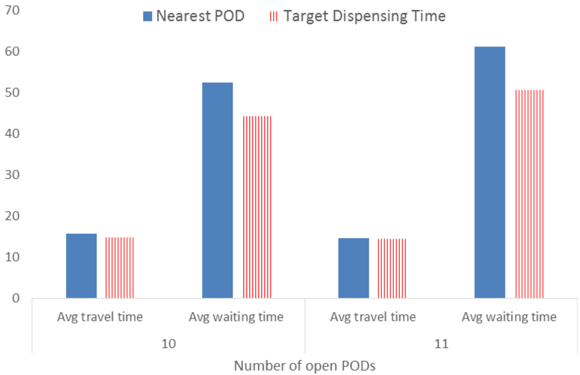
(e) Average HtoP time



(f) Average dispensing time

Relaxing the Assignment to the Nearest POD

Average travel and waiting time (min) relaxing the assignment to the nearest POD and the target dispensing time when the total number of servers is 700. The sequential method did not find feasible solutions for these scenarios.



Validation with RealOpt©(Lee 2006 and 2009)

RealOpt finds the optimal number of PODs to open, their location and the amount of people required to serve. The input parameters of RealOpt include the maximum distance to PODs and the capacity of PODs (throughput).

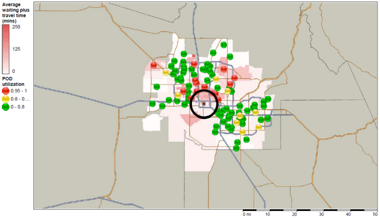
- Case study in the Atlanta metro area.
- RealOpt: P=1868444, DSS: P=1890208.
- T = 36 hours.
- DSS: N=1000, δ=100, $p_{1i} = 0.5$.
- Service time according to Lee et al. (2009).

Validation with RealOpt©(Lee 2006 and 2009)

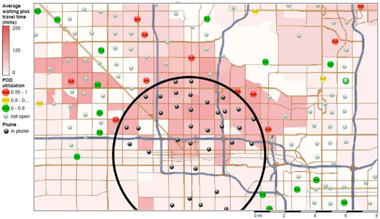
Results:

- RealOpt: PODs=53-128, DSS:PODs=68-78 (Co-Opt:MinServers)-(Co-Opt:MinTime), respectively.
- RealOpt: $\overline{T}$=13-16 min, DSS: $\overline{T}$=11.6-13.3 min.
- RealOpt: Staff=2400-2600/12 hr-shift, DSS: Staff=2600-2626/shift, (servers*1.5).
- RealOpt: $\overline{W}$=?, DSS:29.4-35.4 min adjusting λ_{kj} to the 62.5%-quantile.

Results using GIS



(g) Target area



(h) PODs around the plume

Color of the POD is related with server utilization and color of the census tract is related with average HtoP time.

Outline

Conclusions

- Emergency response plans require to solve complex problems in a very short period.
- We developed a mathematical model and a solution approach based on a GA to assist decision makers.
- The solution approach allows decision makers relaxing limiting assumptions/constraints.
- The proposed co-optimization method is superior to the sequential method in large scenarios.
- Decision makers can make adjustments and obtain effective solutions quickly.

Extensions

- Design a solution approach that allows decision makers relaxing several constraints simultaneously.

Thank you!

ICORD 2016

ITAM
Mexico City, June 9 & 10, 2016
ifors.org/icord2016
Paper submission deadline: January 31st