Dispatch Guided Allocation Optimization for Effective Emergency Response

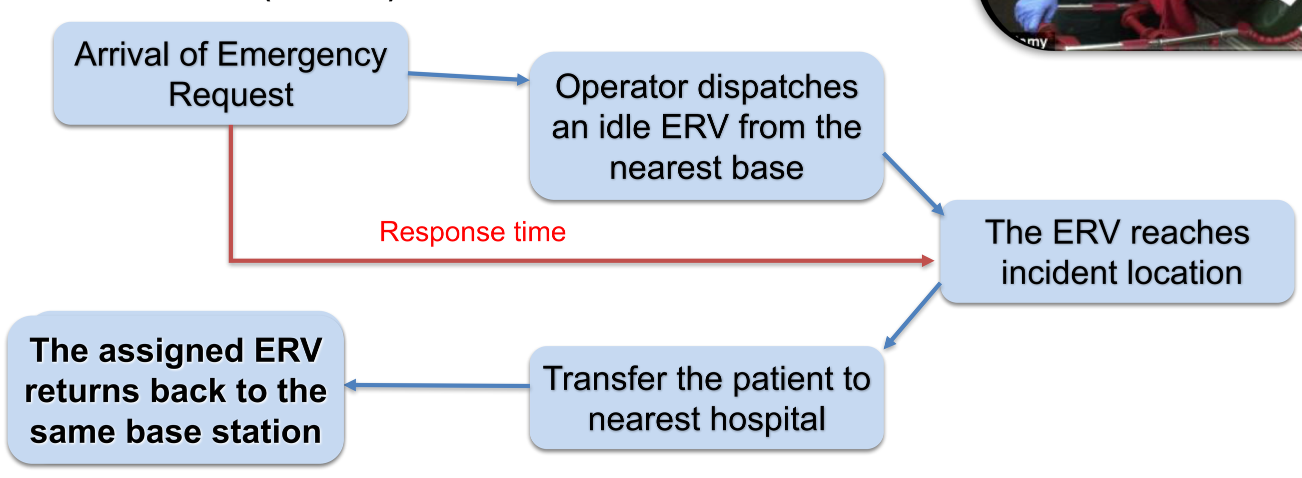
Supriyo Ghosh and Pradeep Varakantham

School of Information Systems, Singapore Management University

Email: supriyog.2013@phdis.smu.edu.sg, pradeepv@smu.edu.sg

Motivation

- Emergency Response Systems:
  - E.g., Medical, Fire or Criminal management.
  - Important for public safety and security.
  - Response times for Emergency Response Vehicles (ERVs) are crucial.

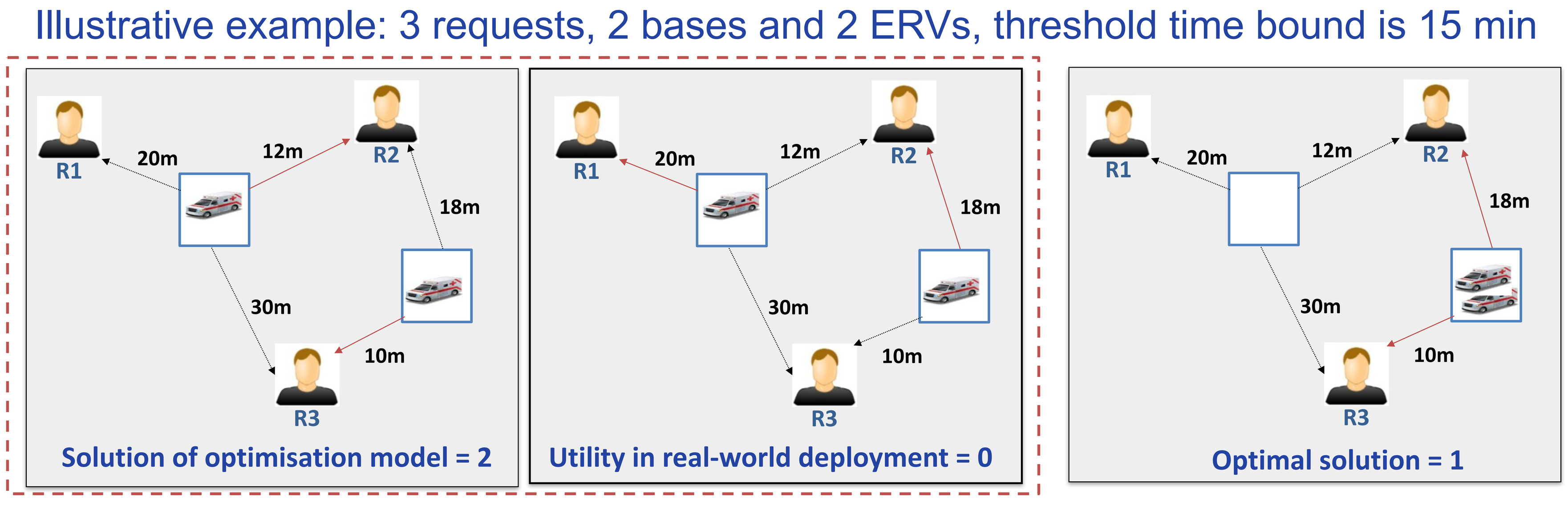


Challenges & Contributions

Goal: Find efficient allocation for an entire fleet of ERVs to a given set of bases so as to optimise **Bounded time response** (e.g., maximise the number of requests served within a threshold time bound) metric.

- **Existing Approaches:**
  - Greedy approach (Yue *et. al.*, 2012):
    - Iteratively allocate one ERV at a time with highest marginal gain value.
    - Greedy solutions might be far away from optimal as ERV allocation problem is non-submodular.
  - Optimisation approach (Saisubramanian *et. al.*, 2015):
    - Minimise response times for fixed percentage of requests.
    - Optimisation model does not incorporate nearest available ERV dispatch rules.

Illustrative example: 3 requests, 2 bases and 2 ERVs, threshold time bound is 15 min



Optimized solution without ERV dispatch strategy and the ground truth

Optimized solution of original problem

- **Our contributions:**
  - A novel Integer Linear Programming (ILP) model for dynamic allocation of ERVs.
    - Incorporates real-world ERV dispatch strategies as linear constraints.
    - Exactly imitates the real dynamics of EMS when optimizing the allocation.
  - Two novel heuristic approaches for efficiently solving large-scale problems
    - A relaxation based heuristic approach
    - Two-stage hierarchical optimisation approach.
  - Evaluate solutions on an event-driven simulator build on two real-world datasets.

Optimization Model for ERV Allocation

- Inputs: ERV allocation problem are defined using tuple: $\langle \mathcal{B}, \mathcal{A}, \mathcal{R}, \boldsymbol{T}, \boldsymbol{C}, L \rangle$

$\mathcal{B}$: set of bases, $\mathcal{A}$: Fleet of ERVs, $\mathcal{R}$: Set of requests, $\boldsymbol{C}$: Capacity of bases.

$\mathbf{T}$: A 2-D matrix that provides travel time between any two base locations

$$L_{rl} = \begin{cases} 1 & \text{if } T_{l,r.source} \le \Delta \text{ minutes} \\ 0 & \text{Otherwise} \end{cases}$$

Bounded time response objective

- Outputs: Number of ambulances, a_l allocated to each bases $l \in \mathcal{B}$
- Objective: Maximize number of requests served within threshold time bound.
- Assignment decision variables: $x_{rl} = \begin{cases} 1 & \text{if request } r \text{ is served from base } l \\ 0 & \text{Otherwise} \end{cases}$
- $\mathcal{B}_r$ denotes set of feasible bases for request r, sorted based on their response time.

i.e., if $\mathcal{B}_r = \{l_1, l_2, ...l_n\}$, then $T_{l_i,r.source} \le T_{l_{i+1},r.source}$

- Integer Linear Programming (ILP) formulation (without dispatch constraints):

$$\max_{a,x} \sum_{r \in \mathcal{R}} \sum_{l_i \in \mathcal{B}_r} x_{rl_i} L_{rl_i}$$

Maximize bounded time response

$$\sum_{l_i \in \{\mathcal{B}_r \cup \perp\}} x_{rl_i} = 1, \qquad \forall r \in \mathcal{R}$$

Serve request from one base only; $\perp$ represents null assignment

$$x_{rl_i} + \sum_{j \in P_r^{l_i}} x_{jl_i} \le a_{l_i}, \qquad \forall r \in \mathcal{R}, l_i \in \mathcal{B}_r$$

Request can be served only if an ERV is available

$P_r^{l_i}$ is a set of requests that arrives before r, finishes after r arrives and can be served from base l_i

$$\sum_{l \in \mathcal{B}} a_l = |\mathcal{A}|$$

$$a_l \in \{0, ..., C_l\}, x_{rl_i} \in \{0, 1\}$$

Ensures all the ERVs are allocated and respect the base capacity constraints

Representation of Dispatch Constraints

- Non-linear dispatch constraints: assign **nearest available** ERV to a request

$$\sum_{k \le i} x_{rl_k} \ge 1 \ \textbf{if} \ \underbrace{a_{l_i} - \sum_{j \in P_r^{l_i}} x_{jl_i}}_{\#\text{ERV available at base } l_i} \ge 1, \qquad \forall r \in \mathcal{R}, l_i \in \mathcal{B}_r$$

Dispatch from base l_i or other nearer bases.

Base l_i is not empty when request r arrives in the system.

- Linearization of complex dispatch constraints:

$$\sum_{k \le i} x_{rl_k} \ge \frac{1}{C_{l_i}} [a_{l_i} - \sum_{j \in P_r^{l_i}} x_{jl_i}], \qquad \forall r \in \mathcal{R}, l_i \in \mathcal{B}_r$$

Capacity of base l_i

Heuristic 1: Relaxation

- Solve the global IP as MILP with continuous assignment variables.
  - Integer allocation, but continuous assignment, i.e., $a_l \in \{0, ..., C_l\}, x_{rl_i} \in [0, 1]$
  - Provides relaxed solution, but a valid allocation of ERVs.
  - Finally, we execute the resulted allocation on a real-world event-driven simulator (from Yue *et. al.,* 2012) to obtain the actual utility.

Observation 1: *If all the base stations have single capacity, then **Relaxation** approach provides an optimal and integral solution.*

Heuristic 2: Two-stage Optimisation

- **First stage:** Solve the global MILP as LP relaxation.
  - Round down the fractional allocation, $\hat{\boldsymbol{a}}$ to integral one, $\bar{\boldsymbol{a}}$.

$$\bar{a}_l = \begin{cases} \lceil \hat{a}_l \rceil & \text{if } \hat{a}_l - \lfloor \hat{a}_l \rfloor \ge 0.95 \\ \lfloor \hat{a}_l \rfloor & \text{Otherwise} \end{cases}$$

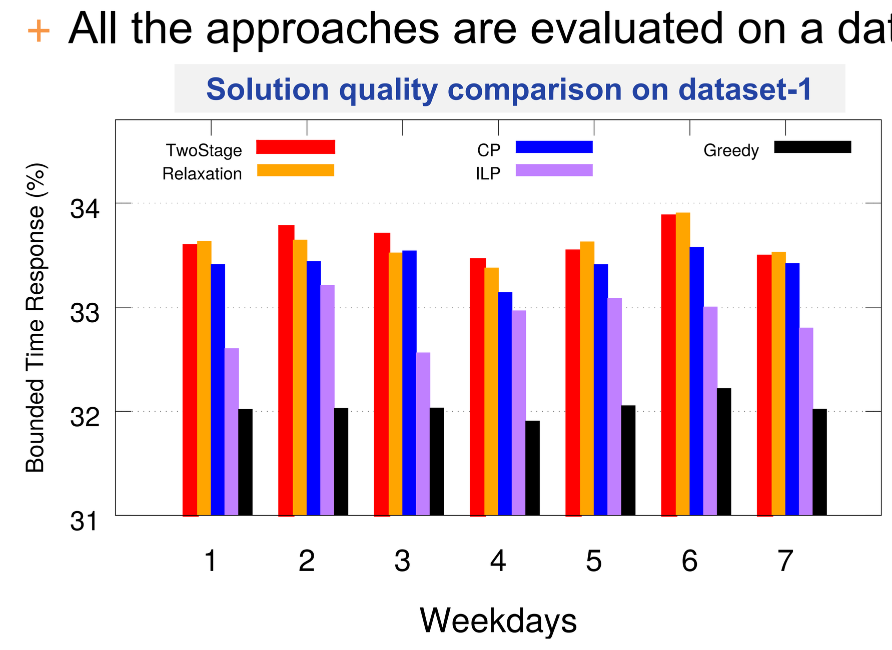
- **Second stage:** Solve the global MILP with following additional set of constraints
  - A fixed allocation for a *subset of ERVs*, $\bar{\boldsymbol{a}}$ (1st stage solution).

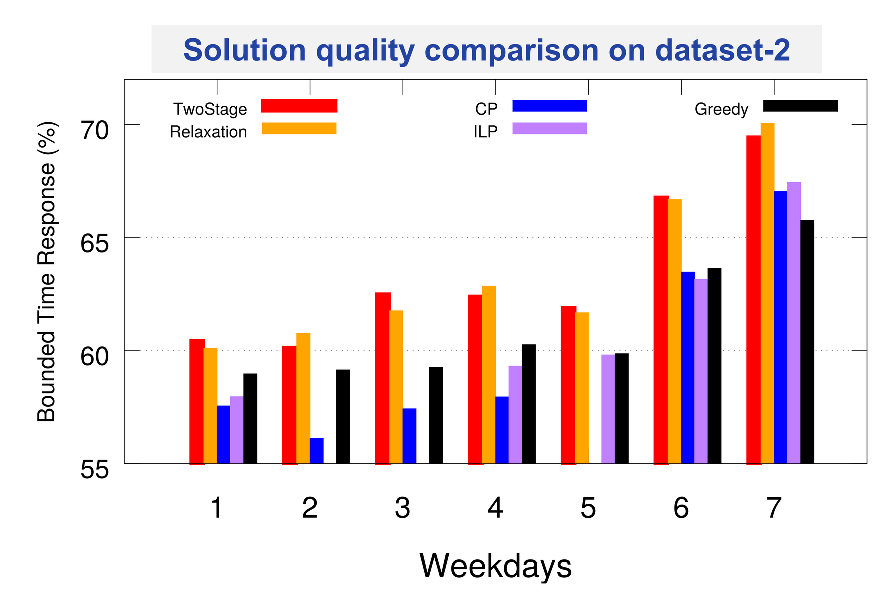
$$a_l \ge \bar{a}_l, \qquad \forall l \in \mathcal{B}$$

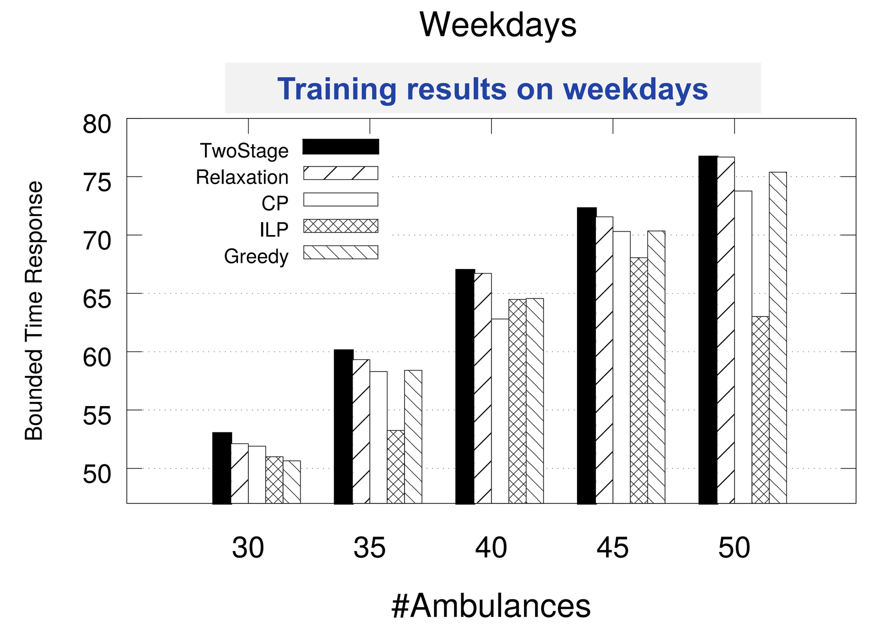
Ensure a lower bound on the number of ERVs at base *l*

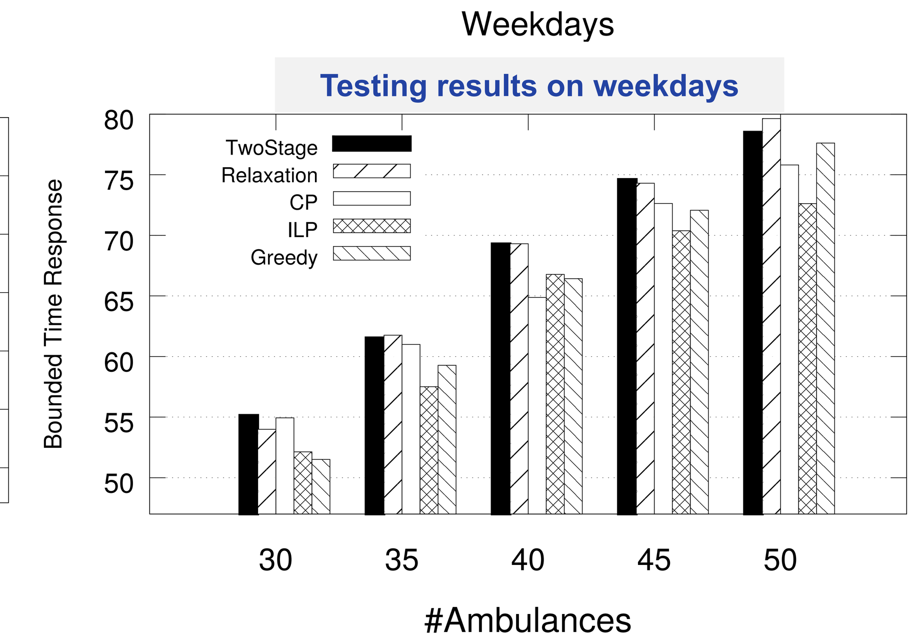
Experimental Results

- **Data sets:** Two real EMS data sets from large Asian cities.
  - **Dataset-1:** 58 bases, 58 ambulances (threshold time bound is 15 minutes).
    - 1500 weeks of request samples – divided into training, validation and test set.
  - **Dataset-2:** 35 bases and 35 ambulances (threshold time bound is 8 minutes).
    - 6 months of request samples – 3 months of training and 3 months of testing set.
- **Approaches**
  - Baseline – **Greedy** approach for optimising Bounded Time Response (Yue *el. al.,* 2012)
  - Global Integer Linear Program (**ILP**) – terminated after 2 hours.
  - Constraint Programming (**CP**) model of the ILP – terminated after 2 hours.
  - **Relaxation** – Our first heuristic approach.
  - **TwoStage** – Our second heuristic approach using two stage optimisation.
- All the approaches are evaluated on a data-driven simulator (Yue *et. al.,* 2012)









- Heuristic approaches serves 2.4% additional requests within 15 minutes over the existing greedy approach (*Yue et al., 2012*) on *dataset-1*.
- TwoStage optimization serves 2.5% extra requests within 8 minutes on *dataset-2*.
- Performance of our heuristics improves for tight resource constraint and therefore, suitable for EMSs with limited number of ambulances.

Summary

- We propose a complete optimisation model for ERV allocation in EMS.
- Two efficient heuristic approaches are provided for solving large-scale problems.
- Observation 1 can be exploited to further improve solution using classical SAT solver.

Ministry of Education SINGAPORE

Acknowledgement: This research was supported by the Singapore Ministry of Education Academic Research Fund (AcRF) Tier 2 grant under research grant MOE2016-T2-1-174.