Feature-variant Clustering Methods for Tolling Zone Definition and Their Impact on Distance-based Toll Optimization

Antonis F. Lentzakis, SMART   Ravi Seshadri, SMART   Arun Akkinepally, Intelligent Transportation Systems Lab, MIT   Vinh-An Vu, Relativ, Tokyo   Moshe Ben-Akiva, Intelligent Transportation Systems Lab, MIT

SMART Singapore-MIT Alliance for Research and Technology

Massachusetts Institute of Technology

ABSTRACT

- We propose the use of feature-variant clustering methods, OPTICS and HDBSCAN*, as a systematic approach for tolling zone definition to operationalize distance-based tolling schemes.
- We develop a framework for predictive distance-based toll optimization to evaluate network performance for various tolling zone definitions derived from feature-variant density-based clustering methods.
- Tolling function parameters are optimized using a simulation-based DTA model that integrates guidance generation with predictive optimization
- Experiments on the real-world Expressway and Major Arterials network of Singapore demonstrate improved effectiveness of distance-based toll optimization from density-based clustering-derived tolling zone definitions, when compared to fixed/adaptive cordon-based pricing.

PREDICTIVE TOLL OPTIMIZATION PROBLEM FORMULATION

$$\max_{\boldsymbol{\theta}} \left[\sum_{v=1}^{V} \frac{\text{U}^v}{|\beta_c^v|} + (1-\alpha) \times \sum_{v=1}^{V} tc^v \right]$$

s.t. Demand/Supply Simulator Output Constraint

$$G(\mathbf{x}^{\mathbf{P}}, \gamma^{\mathbf{P}}, \mathbf{tt}^{\mathbf{g}}, \boldsymbol{\theta}) = \mathbf{tt} \quad \text{Toll Bound Constraint}$$

$$\tau_{LB} \leq \phi_l(\boldsymbol{\theta}_l^{\boldsymbol{h}}, D_l^v) \leq \tau_{UB}, \forall v = 1, 2, ..., V; l = 1, 2..., L; h = 1, 2, ..., H$$

- GA approach is employed for distance-based tolling function parameter optimization, due to non-linearity of objective function and constraints.

HIERARCHICAL DENSITY-BASED CLUSTERING METHODS

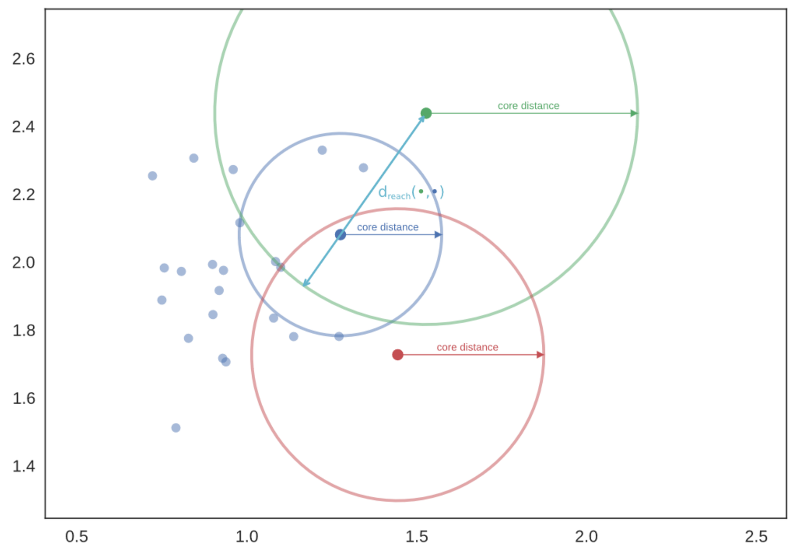
OPTICS (Ordering Points To Identify Clustering Structure)

Core distance

$$d_{\text{core}}^{\epsilon,\kappa}(x) = \begin{cases} \text{Undefined} & \text{if } |N_\epsilon(x)| < \kappa \\ \kappa\text{-th smallest distance to } N_\epsilon(x) & \text{otherwise} \end{cases}$$

Reachability distance

$$d_{\text{reach}}^{\epsilon,\kappa}(x_i, x_j) = \begin{cases} \text{Undefined} & \text{if } |N_\epsilon(x_i)| < \kappa \\ \max(d_{\text{core}}^{\epsilon,\kappa}(x_i), d(x_i, x_j)) & \text{otherwise} \end{cases}$$



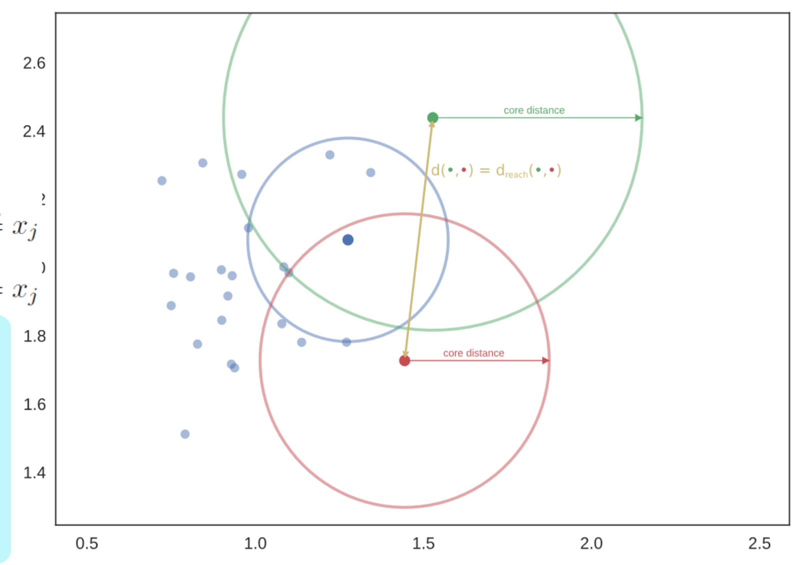
HDBSCAN (Hierarchical DBSCAN*)*

Core distance

$$d_{\text{core}}^{\kappa}(x_i) = \text{Distance of the } \kappa\text{-th nearest neighbor of } x_i$$

Mutual Reachability distance

$$d_{\text{mreach}}^{\kappa}(x_i, x_j) = \begin{cases} \max(d_{\text{core}}^{\kappa}(x_i), d_{\text{core}}^{\kappa}(x_j), d(x_i, x_j)) & \text{if } x_i \neq x_j \\ 0 & \text{if } x_i = x_j \end{cases}$$

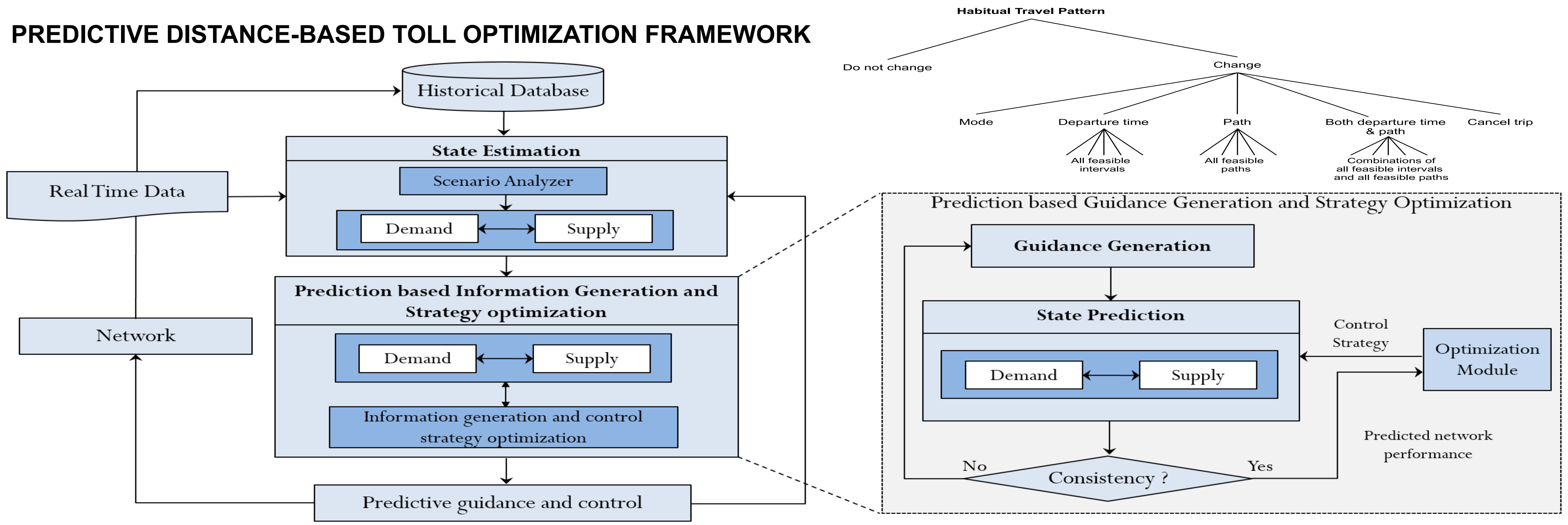


Clustering Performance Indices

$$s(d) = \frac{b(d) - a(d)}{\max\{a(d), b(d)\}}, \boldsymbol{SC} = \sum_{i=1}^{N} s(d_i)/N$$

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}}, D_i = \max_{j \neq i} R_{i,j}, \boldsymbol{DB} = \sum_{i=1}^{N} D_i/N$$

PREDICTIVE DISTANCE-BASED TOLL OPTIMIZATION FRAMEWORK



TOLL OPTIMIZATION OBJECTIVE DESCRIPTION

- Network $G = (N, A)$ is divided into $l = 1, 2, \dots, L$ tolling zones
- We define tolling function $\phi(\boldsymbol{\theta}_l^t, D_l)$ where $\boldsymbol{\theta}_l^t$ is a vector of parameters, D_l is travel distance
- The problem is to determine tolling function parameters over prediction horizon H, $\boldsymbol{\theta} = (\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots, \boldsymbol{\theta}^H)$, that optimize Social Welfare SW:

Consumer Surplus, Total Producer surplus, Fixed Cost, Variable Cost, Total Revenue

$$\begin{aligned} SW &= CS + TP \\ &= CS + (TR - FC - VC) \\ &= \sum_{v=1}^{V} \frac{U^v}{|\beta_c^v|} + \left[(1-a) \times \sum_{v=1}^{V} tc^v \right] \end{aligned}$$

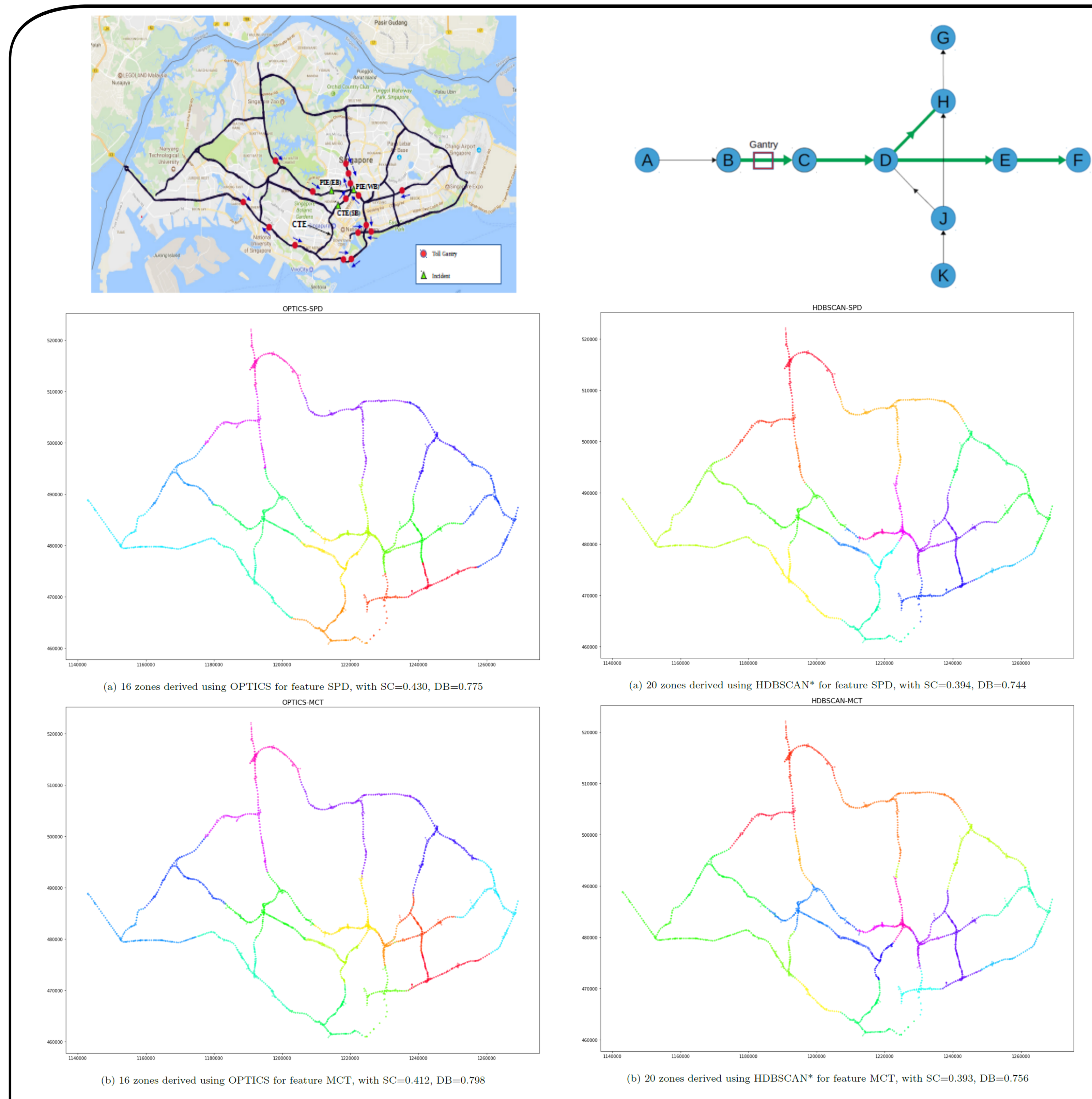


Fig. 1 TOLLING ZONE DEFINITIONS FOR DISTANCE-BASED TOLLING SCHEMES (AD-HOC, OPTICS, HDBSCAN*)

Scenario	Tolling Scheme	Description
S0	Cordon-based	*Non-adaptive*, based on real toll rates
S1	Cordon-based	*Adaptive*, optimized predictive tolls for fixed gantries
S2	Distance-based	*Optimized predictive: ad-hoc*
S3	Distance-based	*Optimized predictive: OPTICS-SPD*
S4	Distance-based	*Optimized predictive: OPTICS-MCT*
S5	Distance-based	*Optimized predictive: HDBSCAN*-SPD*
S6	Distance-based	*Optimized predictive: HDBSCAN*-MCT*

	Scenarios					
Metrics	$S1$	$S2$	$S3$	$S4$	$S5$	$S6$
SW ($)	106780.8	241119.1	139581.9	505341.3	235575.3	510405.3
CS ($)	76793.8	146989.7	12814.3	250802.7	30317.2	256896.2
TT (s)	586.3	564.6	577.8	511.2	559.4	512.5

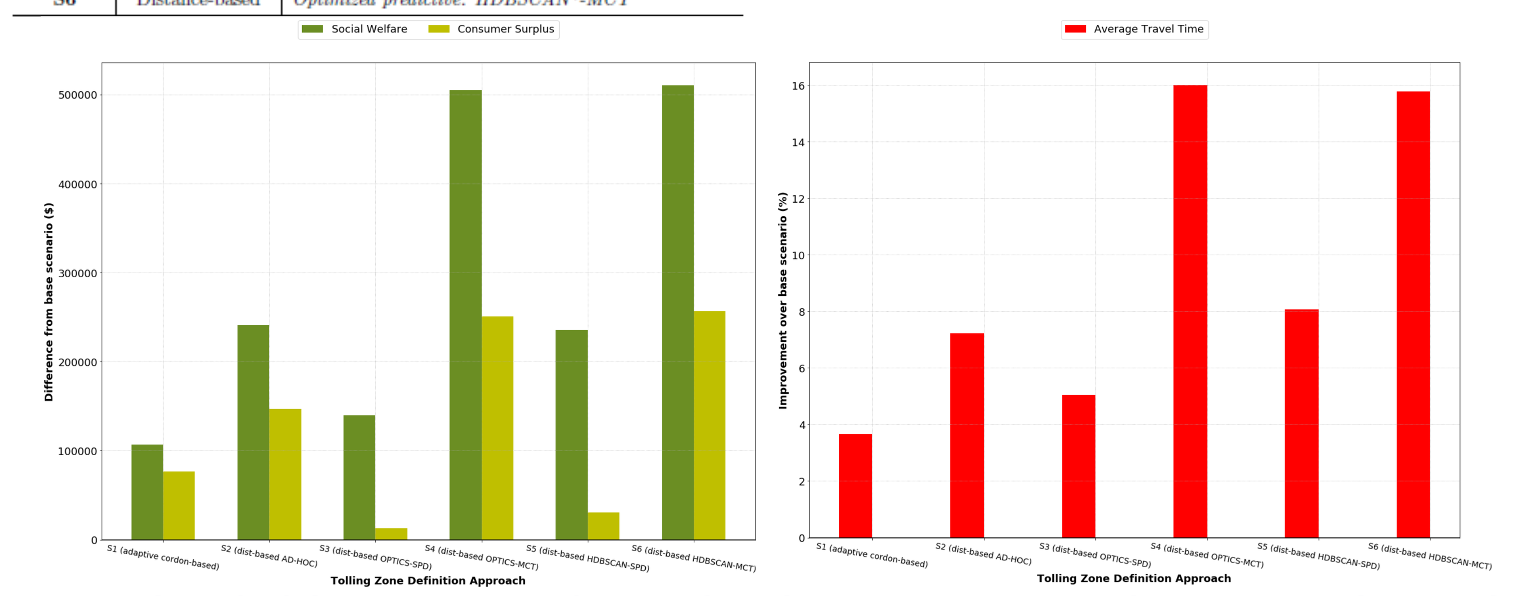


Fig. 2 SW & CS DIFFERENCE FROM S0 AND TT IMPROVEMENT OVER S0

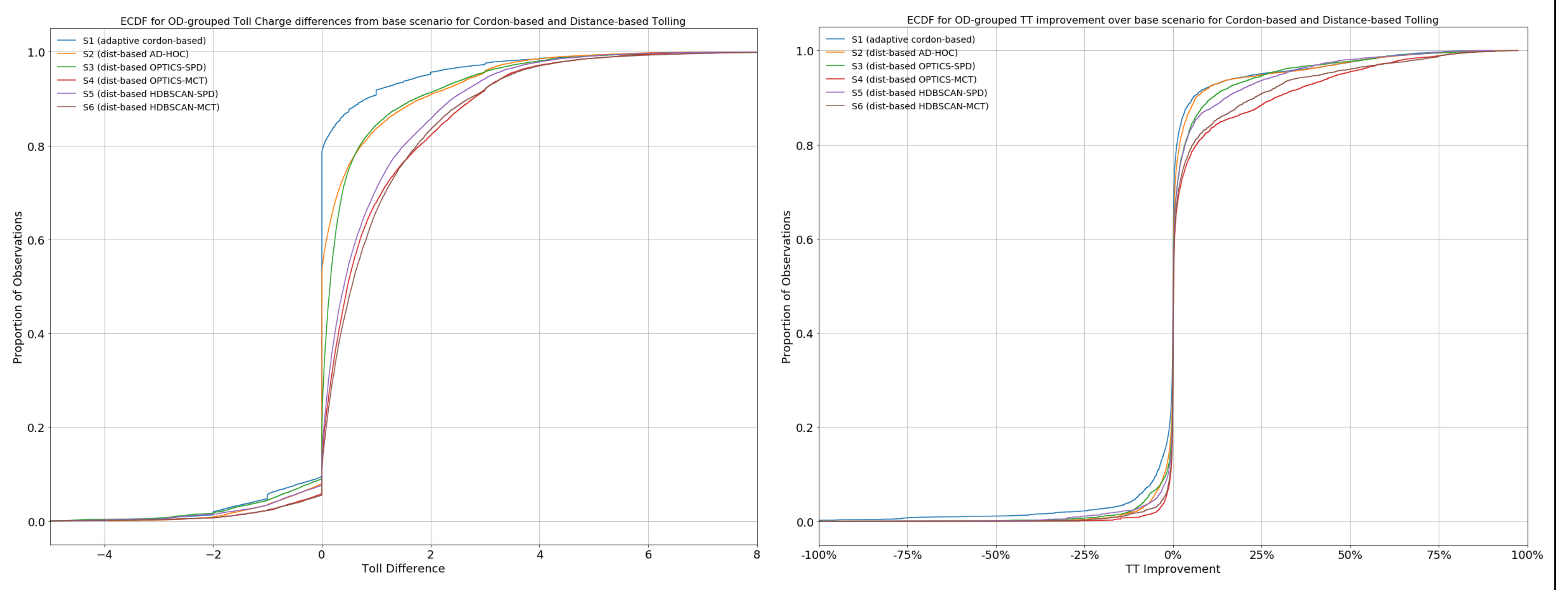


Fig. 3 ECDF FOR OD-GROUPED TOLL CHARGE DIFFERENCES FROM S0 AND OD-GROUPED TT IMPROVEMENT OVER S0

CONCLUSIONS & FUTURE WORK

- Distance-based tolling schemes result in higher tolls per traveler, and a more efficient tolling based on network utilization, thus leading to significant improvements in social welfare and network performance.
- It is also evident that the largest proportion of the traveler population subset that benefits from lower travel times corresponds to S4, followed closely by S6.
- Despite the higher toll charges in case of distance-based tolling, the significant improvements in travel time and schedule delays lead to improved consumer surplus, with tolling zone definitions derived from Marginal Cost Toll (both with OPTICS and HDBSCAN*) resulting in more efficient tolling, and hence, the best performance in terms of SW, CS, TT.
- While clustering performance evaluation results (**SC**, **DB**) are a good predictor of tolling optimization performance improvement, other factors, such as type of feature used in deriving tolling zone definitions may play a deciding role.
- Network topology and spatio-temporal patterns of demand and congestion are likely to play an important role, and more tests are required in different contexts to further examine the robustness of the methodology.
- We aim to investigate the possibility of implementing hierarchical density-based clustering algorithms, as well as other appropriate clustering algorithms, to define sets of tolling zones on data sets from denser urban traffic networks, with additional features (density, speed, marginal cost toll or any combination thereof), and evaluate their impact on network performance when used as part of the distance-based tolling optimization framework.

ACKNOWLEDGEMENT

This research is supported by the National Research Foundation, Prime Minister's Office. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the National Research Foundation, Prime Minister's Office, Singapore.