Ad Hoc Networks 164 (2024) 103613

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

A digital twin-based traffic light management system using BIRCH algorithm

Haitham Y. Adarbah [a,*], Mehdi Sookhak [a,*], Mohammed Atiquzzaman [b]

[a] Computer Science Department, Texas A&M University-Corpus Christi, Corpus Christi, 78412, TX, USA
[b] School of Computer Science, University of Oklahoma, Norman, 73019, OK, USA

ARTICLE INFO

Keywords:
Digital twin
Traffic management
Signal control
Clustering

ABSTRACT

Urban transportation networks are vital for the economic and environmental well-being of cities and they are faced with the integration of Human-Driven Vehicles (HVs) and Connected and Autonomous Vehicles (CAVs) challenge. Most of the traditional traffic management systems fail to effectively manage the dynamic and complex flows of mixed traffic, mainly because of large computational requirements and the restrictions that control models of traffic lights directly based on extensive and continuous training data. Most of the times, the operational flexibility of CAVs is severely compromised for the safety of HVs, or CAVs are given high priority without taking into account the efficiency of HVs leading to lower performance, especially at low CAV penetration rates. On the other hand, the existing adaptive traffic light approaches were usually partial and could not adapt to the real-time behaviors of the traffic system. Some systems operate with inflexible temporal control plans that cannot react to variations in traffic flow or use adaptive control strategies that are based on a limited set of static traffic conditions. This paper presents a novel traffic light control approach utilizing the BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) clustering algorithm combined with digital twins for a more adaptive and efficient system. The BIRCH is effective in processing large datasets because it clusters data points incrementally and dynamically into a small set of representatives. The suggested method does not only enable better simulation and prediction of traffic patterns but also makes possible the real-time adaptive control of traffic signals at signalized intersections. It also improves traffic flow, reduces congestion, and minimizes vehicle idling time by adjusting the green and red light durations dynamically based on both real-time and historical traffic data. This approach is assessed under different traffic intensities, which include low, moderate, and high, while efficiency, fuel consumption, and the number of stops are being compared with the traditional and the existing adaptive traffic management systems.

1. Introduction

Urban transportation networks are pivotal in shaping the economic and environmental landscape of cities [1]. With the advent of CAVs, alongside HVs, traditional traffic management systems increasingly fall short in handling dynamic and mixed traffic flows efficiently. The burgeoning technology of digital twins offers a novel approach to simulate and manage urban traffic through virtual replicas that reflect real-world conditions, thus enabling adaptive control of traffic systems in real-time [2,3].

Digital twins are virtual representations of physical systems that enable real-time simulation, monitoring, and control. They allow for a comprehensive and evolving model of systems. A digital twin updates its state and behavior in real-time through the data it gets from sensors and other inputs This entails risk assessment, process control, and preventive maintenance hence increasing effectiveness and decreasing expenses [1,4,5]. For instance, in traffic management, digital twins use real and historical traffic data with the goal of, through the use of advanced mathematics, predicting traffic conditions and adapting signal control accordingly. This helps in the efficient flow of traffic flow and reduction of traffic jams while promoting the development of smart cities with the least environmental effects.

The BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm is most efficient for large datasets [6,7]. It provides an incremental and dynamic mechanism for grouping multi-dimensional metric data points into a concise summary that is useful in real-time traffic management applications [8,9].

Adaptive traffic light management methods rely heavily on accurate real-time data, which is typically obtained from sensors, such as induction loops and cameras. These sensors have their own set of limitations, such as poor performance in adverse weather conditions and limited detection capabilities (e.g., induction loops that only detect the presence of a vehicle at a specific point). This can lead to sub-optimal

* Corresponding authors.
E-mail addresses: haitham.adarbah@tamucc.edu (H.Y. Adarbah), mehdi.sookhak@tamucc.edu (M. Sookhak).

https://doi.org/10.1016/j.adhoc.2024.103613
Received 15 May 2024; Received in revised form 21 July 2024; Accepted 28 July 2024
Available online 2 August 2024
1570-8705/Published by Elsevier B.V.

traffic light control decisions in real-world scenarios. Implementing and maintaining adaptive traffic control systems is also both complex and costly. They require infrastructure and continuous maintenance to ensure effective operation. The high cost and complexity can be a barrier to widespread adoption, particularly in regions with limited resources [10]. Coordinating adaptive traffic signals across multiple intersections remains a challenge. Effective synchronization is crucial for optimizing traffic flow across a broader network, but achieving this can be difficult due to the asynchronous nature of traffic patterns and the limitations of current technologies in handling complex scenarios [11]. Integrating new adaptive control technologies with existing traffic management infrastructure can pose significant challenges. Compatibility issues between new and old systems can lead to inefficiencies and may require additional investment to resolve [12].

However, despite advancements, existing traffic light management systems often struggle with dynamic and static scheduling, failing to accommodate the unpredictable variations in traffic flow [13–16]. This leads to exacerbated congestion, increased travel times, and elevated fuel consumption and greenhouse gas emissions. Such inefficiencies highlight the critical need for traffic systems that can dynamically adapt to fluctuating traffic conditions to optimize flow and reduce environmental impacts [17]. Here are the main unresolved limitations in the literature:

- The direct traffic light control models need a considerable amount of computational resources and long-term training data.
- Most of the studies do not consider the operability of both CAVs and HVs in the areas of intersections. The adaptability of CAVs is often greatly limited in the interest of HV safety, or CAVs are preferred without respect to HV efficiency. The imbalance results in retarded system performance, especially at low CAV penetration rates.
- Some models use fixed signal phases, which are unable to adapt to variations in traffic, while others exploit adaptive algorithms, which, however, only consider a limited set of static traffic features.

This paper is an extension of the research work published in [18]. The extension introduces a novel digital twin-based traffic light management system that incorporates the BIRCH clustering algorithm for traffic data analysis [19]. Unlike traditional methods, this system utilizes both real-time and historical traffic data to adaptively manage traffic signals, thereby enhancing traffic flow efficiency and safety at signalized intersections with mixed traffic conditions. The primary objective of this research is to demonstrate how the integration of digital twin technology with real-time data processing can revolutionize traffic management systems. Our methodology is designed to significantly reduce stop times, enhance traffic flow, and curtail fuel consumption by dynamically adapting traffic signals to changing conditions, thereby contributing to a sustainable urban environment. To achieve these objectives, we take into account real-time and historical data, manage different traffic loads, and provide system reliability in different scenarios. The BIRCH algorithm will cluster large data sets and it allows for online adjustments with real-time traffic information. This would decrease the time that vehicles spend at red lights thus decreasing the total time that a vehicle spends on the road and increasing traffic flow. Thus, based on the real-time data received and the possibility to predict traffic situations, the system will be able to adjust the traffic light signal timings to provide smooth traffic flow. Also, short stopping times and the smoother traffic flow result in less stop-and-go driving which in turn results in reduced fuel consumption and greenhouse gas emissions.

The BIRCH algorithm, paired with Digital Twin technology for traffic light management, substantially improves the control of urban traffic systems. The BIRCH algorithm handles the data streams coming from sensors and cameras in the traffic network on a real-time and large scale. It compresses information into a Clustering Feature Tree (CF Tree) that identifies vehicle clusters dynamically based on properties like intensity, type, and speed without regenerating all the data every time. The flexibility dynamics help in the variance of the traffic system conditions. The digital twin model uses the cluster data to carry out a variety of traffic scenarios, including peak hour surges, or lane closures, and predicts the results of different traffic light settings in a virtual environment. BIRCH algorithm enables traffic managers to make timing adjustments that are not reactive, in the sense that these changes are made in response to predicting insights provided by the real-time data. Consequently, BIRCH coupled with digital twins becomes a proactive traffic management system that does not only react to the current traffic state but also predicts future traffic patterns, improving general traffic performance and safety.

The main contributions of the papers are as follows:

1. Conducting a comprehensive study of existing traditional and adaptive traffic control systems.
2. Improving real-time and historical data processing for traffic light control and prediction analysis by integrating BIRCH clustering with digital twins.
3. Implementing and evaluating the suggested method across various traffic intensities (high, moderate, and low traffic conditions).

The rest of the paper is organized as follows: Section 2 reviews related works. Section 3 provides background on the research work. Section 4 describes the methodology employed in developing the digital twin-based traffic light management system combined with the BIRCH algorithm. Sections 5 and 6 present the experimental setup and results. Finally, Section 7 concludes the paper and outlines future research directions.

2. Literature review

This section provides a summary of current intersection management techniques and investigates the research gaps and limitations within previous studies. Digital twins allow for a realistic and dynamic representation of traffic systems, facilitating better decision-making and adaptation to real-time conditions. Del Campo et al. in [20] integrate various technologies, such as IoT and VR to create responsive and interconnected traffic control systems that can adapt to live changes in traffic conditions. Additionally, modeling traffic control systems can involve sophisticated algorithms that optimize traffic flow at intersections, potentially reducing congestion and enhancing traffic management efficiency [21].

Table 1 provides studies that developed intersection control methods for mixed traffic. These studies are classified by intersection type (Unsignalized and Signalized) and by control strategy (Fixed Time and Adaptive). Some studies [14,22] used communications of connected and autonomous vehicles such as V2V and V2I to control signal-free intersections. While these methods are adaptive by making a CAV change its behavior when it is near an intersection, they are usually inappropriate for humans.

Signalized intersection is a more viable option compared to signal-free intersections when mixed traffic control strategies and policies are being implemented. The present mixed traffic operation at signalized intersections is mainly based on signal control strategies and policies in managing CAVs. The strategies are classified as fixed timing control and adaptive control. Fixed linked timing control is a cyclic and round-robin based approach, which is simple to implement but generally leads to inefficiency of traffic due to insufficient real-time traffic condition analysis [15,25–28] and adaptive control [26,30,31]. On the other hand, adaptive signal control changes signal timings – both phase duration and sequence – upon real-time data and, therefore, is more effective in dealing with dynamic traffic conditions.

However, the findings of the literature show that most of the current adaptive signal control approaches are not comprehensive, which may

Table 1
Traffic signal control for mixed traffic related studies.

Paper	Intersection categories	Signal control strategies	Main idea
CHEN2024 [13]	Signal-free	–	Study evaluates vehicle sequencing at smart intersections using Markov process modeling.
SHI2022 [14]	Signal-free	–	Study shows higher AV rates enhance traffic flow and safety. Deep reinforcement learning used with Flow framework and SUMO simulator.
QUA2020[22]	Signal-free	–	Study uses PPO in VRCIS with V2X to enhance urban traffic safety.
BUD2018[23]	Signal-free	–	Study used intelligent roadside units to improve traffic flow and safety by leveraging vehicle-to-infrastructure communications
YAN2023 [24]	Signalized	Fixed timing control	Study proposes three two-step prediction optimization methods that match the traffic arrival in different periods with the corresponding optimal signal scheme.
CHEN2021 [25]	Signalized	Fixed timing control	Optimal control of mixed vehicle platoons at signalized intersections enhances dynamics and efficiency.
YAO2020 [15]	Signalized	Fixed timing control	Safety, Decentralized control model for connected automated vehicle trajectories at signalized intersections optimizes efficiency without significant system loss.
ZHA2018 [26]	Signalized	Fixed timing control	Cooperative eco-driving model for mixed AVs and HVs reduces fuel consumption at signalized intersections, maintaining efficiency and comfort.
SHA2017 [27]	Signalized	Fixed timing control	Paper introduces H-AIM for mixed traffic, reducing delays at low AV penetration.
DRE2008 [28]	Signalized	Fixed timing control	Study presents multiagent approach, replacing human coordination with autonomous systems for safer, more efficient traffic management.
MOH2022 [29]	Signalized	Adaptive control	Study presents an optimal signal control algorithm that significantly reduces vehicle delays at oversaturated intersections.
PAR2022 [30]	Signalized	Adaptive control	Enhanced H-AIM protocol integrates with traffic signals, effectively reducing delays.
MA2022 [31]	Signalized	Adaptive control	SPDL traffic control model minimizes delays, enhances capacity at mixed-traffic intersections.
LIU2018 [32]	Signalized	Adaptive control	Safe intersection management for mixed human-driven and autonomous vehicle systems.

result in reduced operational efficiency. For example, the Liu et al. [32] approach utilizes cyclical changes to signal timing based on permissions assigned to CAVs but fails to integrate traffic data from HVs which could affect the efficiency of the system. Likewise, Ma et al. [31] proposed a three-tier model that focuses on reducing traffic delays by optimizing signal ordering and timing, however, this model predominantly utilizes the vehicle arrival rate as its input that may not comprehensively represent the complex traffic dynamics.

In addition, intensive research has been carried out to optimize the operation of CAVs for taking advantage of their accurate, and predictable driving characteristics, postulating that these are not to affect the behavior of HV within a safe traffic environment [33]. As shown in Table 1, previous studies have established various types of control policies for CAV. For example, CAVs have to obey traffic signal rules as HVs do [15,25,26,31,34]. These kinds of studies optimize CAV trajectories to make the mixed traffic perform efficiently, stable, and green at the individual level or network level.

Previous research in the management of mixed traffic at intersections used different signal control strategies, but they carried several shortcomings. To work properly, direct traffic light control models need a lot of computational resources and a large amount of training data. Many of these studies do not consider the efficiency of both CAVs and HVs, leading to the loss of CAV flexibility in favor of HV safety or the advantage of CAVs at the cost of HV efficiency, particularly at low CAV levels. Also, the current approaches often do not respond to dynamic alterations in traffic, with some systems being rigid signal plans and others considering only a few traffic attributes.

To the best of our knowledge, there are no studies on the application of data-driven insights for pre-programmed adjustments in traffic light management. Utilizing the BIRCH algorithm for this purpose alone offers a simple and efficient method that is particularly useful in situations where simplicity and resource efficiency are more important than more complex direct traffic light adjustment models, which require substantial computational resources and long training data.

3. Background

This section extends the research discussed in our previous paper presented at conference [18], and defines the key elements, and the requirements that are necessary for a 5G V2X digital twin architecture. To gain an in-depth view, refer to the conference paper. Fig. 1 outlines the architectural blueprint for the 5G V2X Digital Twin.

3.1. Sensing and data acquisition

The digital twin framework starts with the implementation of multi-heterogeneous sensors that record data live from the physical environment. Such sensors include LIDAR, radar, cameras, and other IoT devices, positioned in urban settings and vehicles. They gather multidimensional data on vehicle locations, velocities, environmental parameters, and infrastructural dynamics.

3.2. Data ingestion and processing

The data acquisition is the first stage, and the second stage consists of relevant data ingestion pipelines that unify and synchronize data from various sources. Through edge computing systems data processors near the data source carry out initial data analysis. This decreases latency and bandwidth consumption, also critical for real-time applications. Processed at the edge data include aggregation, normalization, and preliminary anomaly detection that simplifies the datasets before they move to centralized servers for advanced processing.

3.3. The 5G network

The 5G network infrastructure plays a critical role in its ultra-reliable low latency communication (URLLC) capabilities that enable real-time attributes of digital twins. Such 5G features, such as massive MIMO and beamforming improve channels and spectrum to make the

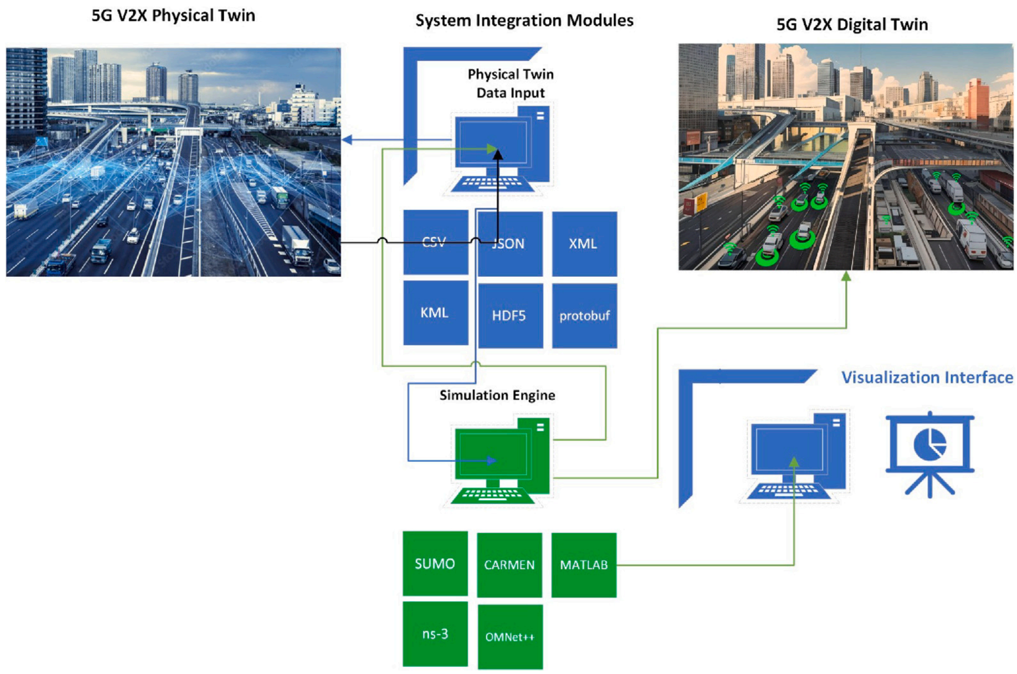


Fig. 1. C-V2X Digital Twin architecture [18].

data paths and bandwidth optimal which in turn increases robustness and response of the communication network. Network slicing enables the formation of dedicated virtual networks with optimized resources and security settings for particular applications in a digital twin ecosystem.

3.4. A model and simulate engine

The ontology-based design of the data model structurally represents the digital and physical entities, where relationships and entities are defined in a scheme that mirrors the real-world configuration of the V2X environment. These models are interpreted by the simulation engine, which is typically implemented using simulation software like MATLAB/Simulink or specialized simulation frameworks like AnyLogic, to simulate different scenarios. This engine incorporates AI algorithms, for example, neural networks for pattern recognition or reinforcement learning for optimization tasks to forecast outcomes and recommendations for proactive actions.

3.5. Data storage and management

Data management solutions are in line with the demanding nature of high-velocity and high-volume V2X data streams. Data lakes when implemented on platforms are a single store for all raw and processed data and can support the processing of massive data for complex analytics and machine learning.

3.6. User interface and visualization

UI and visualization tools are established to follow principles of human–computer interaction to make sure that users can naturally work with the digital twin system. Utilizing modern visualization tools enables the stakeholders to observe and interact with the simulation results in a lifelike and involving setting. Moreover, personalized dashboards give live data analysis in the form of interactive graphs and charts.

3.7. Security, privacy, and interoperability

Security architectures consist of multi-layered approaches among which end-to-end encryption, secure boot, and hardware-based security modules (HSM) ensure data integrity and confidentiality. Privacy by Design principle ensures that the development process is regulated by some of the laws such as GDPR. Interoperability is established by using common communication protocols such as MQTT or CoAP and standard data formats such as JSON or Protocol Buffers, which ensures that IoT platforms can be integrated with other systems and components.

3.8. Analytics and decision support

Statistical models, machine learning algorithms, and data mining techniques are used by the analytics engine to find patterns and insights from this data. This sub-system plays an important role in converting raw data into actionable information, aiding decision-making processes. Decision support systems with optimization algorithms provide advice through real-time analysis which improves operational efficiency and predictive powers of the digital twin.

The digital twin architecture, which concentrates on these technical aspects, not only replicates the physical world precisely but also improves the decision-making and operational processes within the 5G V2X ecosystem, thus creating dynamic and efficient transportation systems.

4. Proposed methodology

Fig. 2 and algorithm 1 highlight the main process of the proposed methods for adjusting pre-programmed traffic light systems, followed by a function-based explanation of the method in more detail.

The necessary data structures are created to store real-time and historical traffic data, the results of the intermediate analysis, clustering results and control traffic decisions. The algorithm runs cyclically, on a specific frequency, with the traffic data being gathered from different sources like the V2I communication, sensors, and cameras (Algorithm, line 3). The data is then cleaned to make it standard and relevant by removing irrelevant information, scaling, transforming, and detecting anomalies (Algorithm, line 4). The normalized data is then subjected to

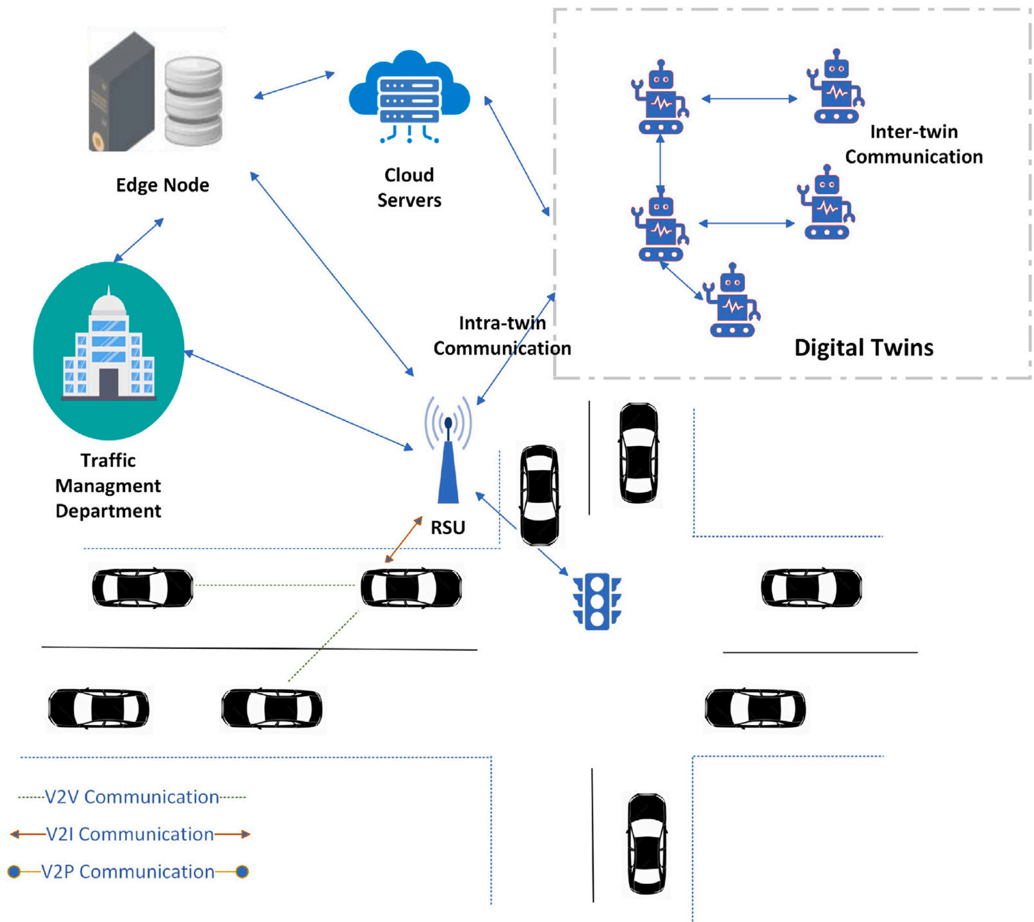


Fig. 2. The suggested system model for traffic light management.

Algorithm 1 The suggested process for adjusting pre-programmed traffic light system

1: InitializeDataStructures()
2: **loop** every predefined_interval
3: TrafficData ← CollectData()
4: NormalizedTrafficData ← PreprocessData(TrafficData)
5: (T, B) ← DefineBIRCHParameters()
6: BIRCHTree ← BIRCHClustering(NormalizedTrafficData, T, B)
7: ClusterProperties ← ClusterAnalysis(BIRCHTree)
8: TrafficForecasts ← TrafficPrediction(ClusterProperties)
9: Anomalies ← MisbehaviorDetection(ClusterProperties)
10: SignalAdjustments ← SignalAdjustmentLogic(TrafficForecasts, Anomalies)
11: ImplementControlAdjustments(SignalAdjustments)
12: FeedbackLoop()
13: SystemMetrics ← PerformanceEvaluation()
14: ModelUpdate(SystemMetrics)
15: **end loop**

the BIRCH algorithm with parameters such as the threshold value and the branching factor to generate the CF Tree (Algorithm 1, lines 5–6). The characteristics of these clusters are then examined to obtain useful knowledge that can be applied to future traffic situations and anomaly identification (Algorithm 1, line 7). According to these predictions and anomalies, the required changes to the traffic signal systems are decided and applied at the actual time (Algorithm 1, lines 10–11). A control loop is created for the adaptive enhancement of the control logic based on the performance assessments, including the traffic flow rate, stop duration, and fuel usage (Algorithm 1, lines 12–14). These performance metrics are incorporated in the digital twin model to enhance future traffic control and simulation, thus, making the system efficient in traffic regulation and handling. This cyclical process guarantees the constant supervision and the real-time modification of the traffic signals' timings, which in turn optimizes the traffic flow control in cities.

4.1. Initialize data structures

Initiation of the data structures is one of the key steps of the suggested traffic management system. This component is crucial in allowing the search of the memory structures where real-time and past traffic data, intermediate analysis findings, clustering outputs, and traffic control decisions will be stored (Algorithm 1, line 1). The value of these structures is dynamic and coordination of the real traffic environment with the digital twin, so traffic is controlled in real-time and decisions are made on time.

Data structures like VehicleData, SensorData, TrafficData, etc. are intended to make the system work effectively with the expected traffic data volume and variety. VehicleData is a key part of information like vehicle ID, location coordinates, speed, acceleration, direction, and timestamps, and it is critical for accurate tracking of vehicle behaviors and movements. This data is very important not only for the present traffic analytics but also to enhance the Digital Twin scenario modeling functionalities [35–37]. SensorData allows gathering data from a traffic sensor, for example, sensor IDs and vehicle IDs, and is important for associating the true sensor data with the digital twin sensor data, thus enhancing the accuracy of traffic condition simulation. TrafficData aggregation includes the incorporation of VehicleData along with SensorData as a full dataset for the BIRCH clustering algorithm and the Digital Twin simulations aimed at optimizing fluid traffic movement and interflow analysis.

In this step, we define the parameters for each data structure, select the appropriate data type, and allocate memory for easier access and effective processing. This method uses local edge devices for the initial data selection and pre-processing to reduce the latencies and the amount of data sent to central systems. BIRCH clustering is an intensive processing activity that is centralized in server or cloud storage repositories, where the high computational powers provide for sophisticated data processing and broad integration with the Digital Twin [19]. The powerful infrastructure of the system allows the system to support real-time data processing, which in turn improves the performance and accuracy of the Digital Twin.

The loop time is modified based on the needs of the traffic management system. Such flexibility is crucial for giving the system responsiveness and balancing computing resources. The smaller intervals that are imposed within the loop make the system more sensitive, and hence it reacts faster to the traffic control and management policies. Still, the intervals below 50 ms require greater computing power and resource allocation. Alternatively, a long interval may save some of the computational resources while the system loses its reactivity. This feature of adjustable loop timing is crucial to the overall performance of a traffic management system, providing the system with the flexibility to adapt to the varying traffic conditions and operational requirements for efficient fine-tuning.

This looping procedure ensures the Digital Twin and BIRCH clustering algorithms are continuously supplied with new data, which increases the capabilities of the system to offer precise and timely traffic predictions and management decisions. The continuous data processing loop makes the traffic management system adaptive and proactive, allowing it to react to the urban traffic flow complexities, immediately.

4.2. Collect data

In this step, the process of data collection takes base traffic data from many sources. They include traffic sensors, cameras, and V2I communications (Algorithm 1, line 3). This is when these tools help us by providing a diversity of data types, such as vehicle counts, speeds, directions, and types. Every kind of data is important to obtain a full picture of the situation with traffic at a given point in time. That is, traffic sensors are commonly placed at intersections and on highway routes to collect quantitative traffic flow information and traffic intensity. Cameras generate visual feedback that is utilized to facilitate more advanced requirements like vehicle classification and tracking, V2I communication provides real-time information from the vehicles to the traffic management systems to perform advanced strategies such as dynamic signaling that is based on the current condition of traffic. Data generated from traffic sensors, cameras, and V2I communications are first pre-processed and organized through edge computing devices, in which raw data is cleansed, aggregated/normalized, anomalous readings are detected, and data transformation is done before the streamlined data is sent to the central server for further analysis and clustering with the help of the BIRCH algorithm.

The way data is processed, and the acquisition of data are maximized to ensure both speed and quality. Sometimes the system employs edge computing devices to handle the large inflow of data streams. The sensors that are located close to the data generators (e.g., near the traffic lights, by the roadside, etc.) perform the preliminary part of data processing. Edge data processing significantly reduces the latency usually associated with sending raw data to central servers for processing. Latency reduction is a crucial part of a traffic management system based on data arriving just in time to make real-time decisions. To add to that, the edge pre-processes the data. During the edge pre-processing stage, several important operations are carried out to prepare the data for the next levels of analysis. First, data cleaning is applied to eliminate unnecessary information and remove noise thereby increasing the quality of data. Then, data aggregation gathers data from multiple sensors regarding vehicles' counts, speed, and direction, and creates a unified data set. After this, data normalization is done to make sure that all the values are on a comparable level, usually, the mean is zero and the standard deviation is one; this is crucial when performing analysis. Anomaly detection is then performed to detect the outliers or any abnormal traffic behavior which may be due to some errors or important traffic events to ensure data quality. Data transformation is the last step that makes the raw sensor data suitable for application to clustering algorithms by restructuring it. Thus, the pre-processing that is done at the edge side helps to minimize the amount of data to be sent to the central servers, to manage the networks more effectively, and provide the traffic management system to work efficiently and accurately.

4.3. Normalize traffic data

This step involves several key practices: data scaling and transformation, automated processing, and how outliers and errors are handled (Algorithm 1, line 4). Every module is adapted to polish the gathered raw traffic data, suited for detailed analysis and modeling within the system.

The Z-score standardization scaling technique transforms the data to have a mean of zero and a standard deviation of one [38]. This transformation is important for addressing variables such as vehicle speeds or counts that can otherwise bias the analysis because of differences in magnitude scales. Z-score standardization is one of the most popular data preprocessing methods which map the data into a normal distribution with the mean of zero and the standard deviation of one. This way, the general questions of different sizes in datasets are addressed while improving the performance of many algorithms.[39]

Error checking and outlier detection are important steps in the realization of normalization. This stage aims at enhancing the quality of the data by detecting and correcting anomalies or errors that are likely to create false results in the analysis. Devices like setting acceptable data ranges allow the identification of data points outside of the expected frames. Moreover, statistical techniques can be employed to detect outliers, which are data points that significantly differ from the other observations. How the outliers are treated—whether to exclude, adjust, or separately analyze them is of critical importance since proper treatment of outliers is what preserves the integrity and accuracy of traffic data analysis.

Outliers in the traffic data are identified through statistical analysis like the Z-score method, which determines points that are far away from the mean. A measure of the spread of the data is the range, which is defined as the difference between the first quartile and the third quartile. These outliers are then dealt with in one of the following ways; if they are thought to be errors then they are removed, if they are not errors then they are altered in a way that reduces their effect, or they are examined individually to gain a better understanding of the traffic conditions in the data set. The clustering in the proposed framework is performed using the BIRCH algorithm and in this step, outlier detection is incorporated with edge computing to filter the data to eliminate the noise before it is passed on to the next stage [39].

4.4. BIRCH clustering and VANET Digital Twins

The main use of the BIRCH algorithm in traffic light management in combination with Digital Twin technology is that of the clustering abilities of BIRCH and the simulation and predictive analytics of Digital Twins. Here's a detailed explanation of how this integration technically manages traffic lights:

Fundamentally, the BIRCH algorithm processes and analyses live traffic data obtained from sensors and cameras located at different

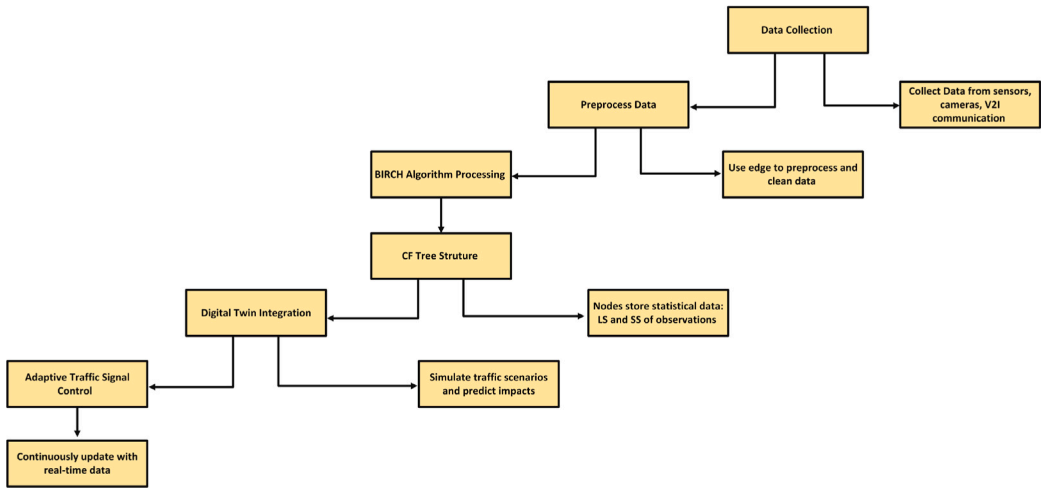


Fig. 3. Phase 1: Illustration of the process.

nodes within the traffic network. This algorithm is well-equipped to deal with large data streams by creating a feature-based summary of the incoming data points as a CF Tree (Algorithm 1, line 6). The CF Tree identifies clusters in terms of attributes such as intensity and type of vehicle, time intervals, and speed. Each node in this tree represents the statistical data of observations (vehicles) that are located within this cluster, such as the linear sum and squared sum of the data points, which are important in efficient computation of cluster centroids and radii. This feature provides the opportunity to dynamically modify the clusters as new data arrive without reprocessing all the data; therefore, it is useful for dynamic environments such as traffic systems, where the conditions are continuously changing [40].

After the clusters are created, this classified data drives the digital twin model of the traffic system, virtual visualization of the physical intersections, and traffic flows. This model takes advantage of the cluster information to simulate various traffic situations, including the effect of a traffic surge during the peak hours or a lane closure. The Digital twin can forecast the consequences of different traffic light modifications, like varying green light times for specific lanes, based on the clusters generated by BIRCH. This model enables traffic managers to see the possible bottlenecks and see the solutions without even implementing them in reality in a virtual environment, therefore optimizing the traffic flow and reducing congestion.

The real-time communication between the BIRCH algorithm and the Digital Twin model results in an adaptive traffic light management system that acts not only on past but also future traffic issues. Such proactive action is productive and can substantially boost the efficiency of urban traffic management systems.

The BIRCH algorithm has three main phases, as follows:

Phase 1: Contextual CF Tree Building with Traffic Data As it can be seen in Fig. 3, the historical vehicle information that is obtained by VANET collected from traffic sensors, cameras as well as V2I communication. The edge devices then clean up this data and then the processing of the data is done by the centralized servers using the BIRCH algorithm. The data collected in clusters is then used in digital twin simulations to estimate and control the traffic situation. The CF Tree is built from the historical vehicle information gathered. The data of each vehicle is a d-dimensional vector, with its position and velocity forming a cluster feature for each cluster of the CF Tree [41].

If the d-dimensional number of data points (vehicles) is N, then $CF = (N, LS, SS)$ is defined as follows:

$$LS = \sum_{i=1}^{N} X_i \tag{1}$$

Sums vectors representing the position and velocity of vehicles and gives a direction and speed aggregate for the cluster.

$$SS = \sum_{i=1}^{N} (X_i)^2 \tag{2}$$

A sum of squares of the vehicle vectors represents the spread of vehicle movement within the cluster.

Phase 2: Hierarchical Clustering for Traffic Control The second stage is aimed at the optimization of the clustering process in the context of the centroids of clusters, i.e. the essence of the traffic flow in each cluster.

The centroid is the core part of the cluster since it determines the average flow direction and speed in the cluster known as $C = \frac{LS}{N}$.

This average provides insight into the overall traffic flow that may not be apparent from individual data points.

Spread Clustering Process: The spread clustering process is an evaluation of the current data grouping to check whether the existing clusters need some adjustments, in the form of dividing large clusters into smaller, more homogenized groups, or merging smaller clusters into larger ones.

Homogeneity within clusters: If a cluster includes data points such that their distances are rather large (i.e., the speeds and directions of the vehicles are quite different) it might be split to have the clustering more precisely reflect similar traffic behaviors.

Traffic flow optimization: Alternatively, if two or more clusters have nearly the same centroids, their fusion may simplify the traffic model but preserve a good part of the information, which would facilitate the monitoring and management of these areas.

Strategic Clustering for Traffic Control: In this phase, the hierarchical clustering is smart, making adjustments to the clusters based on both real-time and historical traffic data in order to optimize the traffic flow. The analysis of the centroids and the spread within clusters will guide traffic system managers who can then make appropriate decisions in respect of traffic light timings, lane usage, and other control measures that kill congestion and improve traffic flow. By way of example, consistent heavy congestion in one direction within a particular cluster during peak hours could be combated with specific measures, such as signal timing adjustments or temporary lane reversals, to better distribute this flow.

In this stage of hierarchy clustering, the traffic is dynamic control with predictive analytical. Continuous improvement of the cluster analysis by the BIRCH algorithm in the VANET Digital Twins framework ensures real-time adaptive control system that manages both normal traffic patterns as well as unforeseen situations effectively.

Table 2
Experiments environment setup parameters.

Parameter.	Desc.
Co-simulation	SUMO-CARLA [43]
Area	A grid of 4 × 4 blocks (with each block measuring 100 m on each side.)
Lanes	2 lanes per direction
Intersections	5
Simulation time	2 h (fast-forwarded at 4x speed)
Low Traffic Scenario	1800 vehicles/h
Moderate Traffic Scenario	3600 vehicles/h
High Traffic Scenario	5400 vehicles/h
Traffic Light_Cycle Times	90 s
Traffic Light_Green Phase	30 s for main roads, and 20 s for secondary roads
Adaptive Signal Control	±10 s
T	0.5 (Threshold value for BIRCH clustering algorithm, representing the maximum radius of a cluster before a new cluster is formed
B	50 (Branching factor for BIRCH clustering algorithm, representing the maximum number of child nodes per cluster node)

Fig. 4. A traffic simulation scenario image.

Phase 3: Traffic Clustering Refinement

Distances that exist between clusters CF_1 and CF_2 may indicate which clusters to join or how to shift so that the system becomes better (Algorithm 1, line 8) [42].

The Euclidean distance between the centroids of the two clusters is computed, the terms are expanded, and then the terms are related to the cluster features LS and SS through their definitions. D is the direct use of the distance formula in terms of the cluster features that have been summarized:

$$D = \sqrt{\frac{N_1 \cdot SS_2 + N_2 - 2 \cdot LS_1 \cdot LS_2}{N_1 \cdot N_2}} \tag{3}$$

Optimal traffic control and safety in urban environments using Digital Twins and VANETs require the perfect set of parameters. A lower threshold (T) value is recommended for better capture of variable urban traffic patterns. The value of the threshold has to be adjusted by the system's performance. An intermediate branching factor (B) allows for effective data processing without the loss of granularity, a requirement for thorough traffic analysis.

The parameters are readjusted continuously by the system due to its constant performance evaluation process to guarantee that the traffic clustering remains relevant and representative of the present urban traffic state. Nevertheless, the refinement stage is relevant since it continuously improves traffic control and safety measures through the dynamic abilities of Digital Twins and VANETs to adapt to the urban traffic situation. Through the changes of the analysis and clustering parameters, the system becomes more efficient in urban traffic management; it becomes more orderly, and congestion and traffic dangers are eliminated throughout the network.

5. Experiment setup

In our traffic management system experimental setup, we utilize an integrated simulation framework that encompasses SUMO [44], CARLA [43], and Digital Twin technologies, thus permitting a comprehensive simulation of traffic [45].

SUMO-CARLA [43] improves the simulation by highlighting the behavioral analysis of individual vehicles. CARLA's advanced graphics and physics engines can be used to simulate the complex interactions between AV and dynamic urban elements including pedestrian flows, unexpected road obstructions, and various weather conditions. These interactions are essential for rigorous verification of the sensory and navigational algorithms of AVs under real-world and difficult conditions.

The main responsibility of the Digital Twins in such an integrated simulation environment is critical. It functions as a link between SUMO [44] traffic simulations and the vehicular interactions that CARLA models [43]. The Digital Twin is constantly updating with real-time data from both simulation platforms and becomes a live, dynamic model, which does not only replicate current traffic situation but also predicts future states by the events and changes in the environment. This predictive facility makes it possible for the proactive decisions of traffic management and therefore, the optimization strategies to be very effective.

Table 2 lists the main simulation parameters. We installed the detectors and the sensor setup for our traffic experiments 50 m before each junction on all approaches. The data-gathering process is scheduled to run every 30 s to increase the accuracy of the observations. To provide a visual understanding of the simulation testbed, refer to Fig. 4.

Our approach to this experiment was to develop and assess three relevant traffic scenarios: low, moderate, and high traffic volumes given different traffic management strategies using SUMO [44], CARLA [43], and digital twins traffic tools. Each scenario is carefully designed to simulate different levels of congestion and traffic flow by establishing concrete rates and intervals for various kinds of vehicles, like sedans, SUVs, and buses. This approach helps us to realize the effects of various traffic intensities on road utilization and to determine which management tactics are the best for improving traffic flow and alleviating congestion. Deriving testing scenarios is beneficial as they allow to model various conditions, from light traffic, with an understanding that a system might perform suitably in non-peak times, up to highly congested situations that challenge the resiliency and efficiency of traffic systems at peak times.

The low traffic scenario works at the overall capacity of 1,800 vehicles per hour which is 50% of a road's full capacity with the vehicle generation interval set as 2.86 s for sedans, 8 s for SUVs, and 40 s for buses. Such longer periods between the vehicle generations, in effect, simulate a less congested environment, which is typical for testing of traffic management strategies in light traffic conditions which is seen at the beginning or end of peak periods.

The paper considers full road capacity to be 3,600 vehicles per hour, which is applied to the moderate traffic situation. The parameters of the system are defined as follows where the low traffic flow is assumed to be 1800veh/hr, which is half of the full capacity. This benchmark can be useful in assessing the traffic management measures when the traffic is not dense. The high traffic scenario with 5400 vehicles per hour is used to examine the system in the worst conditions severe traffic congestion. Thus, 50% of a road's full capacity is obtained from these predefined cases.

Under the moderate traffic scenario, the simulation utilizes the full capacity of the road which is 3,600 vehicles per hour. The vehicle generation intervals are 1.43 s for sedans, 4 s for SUVs, and 20 s for buses, reflecting typical peak traffic flows. This situation presents a well-proportioned layout that represents regular traffic jams, making it possible to evaluate traffic control solutions in usual peak-time conditions with flows as well as with light jams.

A high-traffic scenario assumes a road overcapacity with 5,400 vehicles per hour and shorter cars' generation intervals of 0.95 s, SUVs' intervals – 2.67 s, and buses' intervals – 13.33 s. This situation describes a case of very high traffic congestion, which tests the capacity of the road and traffic management systems. It provides a worst-case scenario for the assessment of the operability of adaptive control participants during severe congestion and lower-speed traffic.

These values were selected based on the running of the simulations several times in a bid to find the most appropriate parameters. This way the process helped to achieve the high similarity of simulations to real traffic intensity in peak hours and over capacity.

6. Simulation results

The proposed system in three different scenarios of low, moderate, and high traffic always demonstrated its ability to constantly dynamically adapt to real-time traffic condition in comparison to both YAO [15] and traditional methods. The method by Yao and Li [15] is based on CAV trajectory optimization fixed timing, and signal control.

6.1. Comparative analysis

This section provides a detailed comparison between Yao and Li [15] and the suggested approach.

Yao and Li [15] investigated the decentralized management of the CAV trajectories in the mixed traffic stream at an isolated intersection with traffic signals. It is based on the approach of finding the best possible path of each vehicle in terms of time, consumption, and security. This is different from other models of control that dictate the paths of all the vehicles as a whole. The control architecture proposed by Yao and Li works on the premise that every CAV tries to find the optimal path on their own with the use of local information and goals. The basic techniques that the authors use are trajectory optimization, decentralized control, and the conversion of the continuous problem into a discrete one to enhance the solvability of the problem; the solution is found with the help of the DIRECT algorithm.

Nevertheless, our approach stands in contrast to Yao and Li's in several important ways owing to the fact that both HVs and CAVs are to co-exist on the road and the common goal of enhancing traffic flow. Yao and Li's model is a distributed model; this is because the system does not have a central control unit and all the vehicles make their decisions on their own. This method is rather efficient in terms of computation but it can lead to poor global solutions in some cases because it does not have the complete picture. On the other hand, our approach is based on the centralized strategy, which means that we take into account the overall view of the traffic system and design the control flow of traffic lights and vehicles' movements. This approach enables the coordination of the optimization thus leading to higher overall system optimization than in the decentralized approach. However, the work done by Yao and Li is confined to isolated intersections and this reduces the range of application of the study. Here is our proposal that covers even more elaborate urban traffic scenarios, and we integrate a Digital Twin concept for efficient data fusion and traffic signal adaptation. This generalization helps our method to coordinate the traffic at different intersections and even under different traffic conditions.

The second major difference can be observed in the technological application. Despite the fact that Yao and Li employ a discrete optimization model that is best for individual intersections, our approach is based on the BIRCH clustering algorithm as well as digital twins. This integration enables it to cluster traffic data in real time and also allow traffic signal adaption to the traffic flow hence making our system very sensitive to traffic conditions.

Yao and Li show better system performance with decentralized control but there is a possibility of not being able to attain global system optimum. On the other hand, the proposed method demonstrates better versatility and performance in regulating traffic flow, especially in areas of congestion. Thus, by keeping the stop rates lower and applying optimal fuel economy technologies with the help of adaptive controls our system shows its efficiency in managing urban traffic in a sustainable manner.

6.2. Average stop rate

The stop rate considers the traffic light cycle frequency, the green phase duration, and the intensity. The parameters can be defined as follows: N: Total number of vehicles approaching the intersection during a given period. S: number of vehicles that stop at least once during this period. T: total traffic light cycle time (sum of all phases including green, yellow, and red). G: duration of the green phase. L: average vehicle arrival rate (vehicles per second).

$$\text{Average Stop Rate (\%)} = \left(\frac{S}{N}\right) \times 100 \tag{4}$$

As the vehicles will stop during the non-green phase:

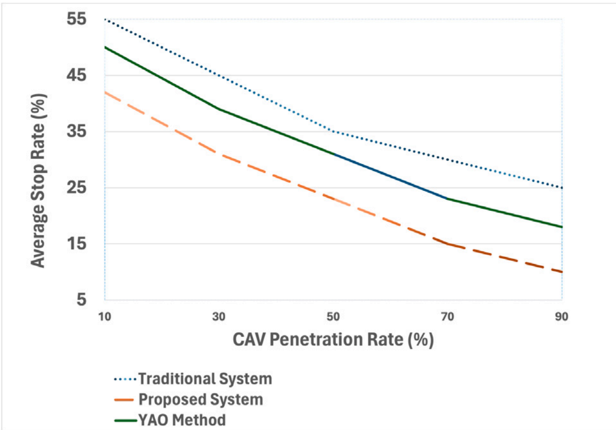
$$S = N - \left(L \times G \times \frac{T}{\text{period}}\right) \tag{5}$$

Where $\frac{T}{\text{period}}$ is the duration of the observation period in seconds.

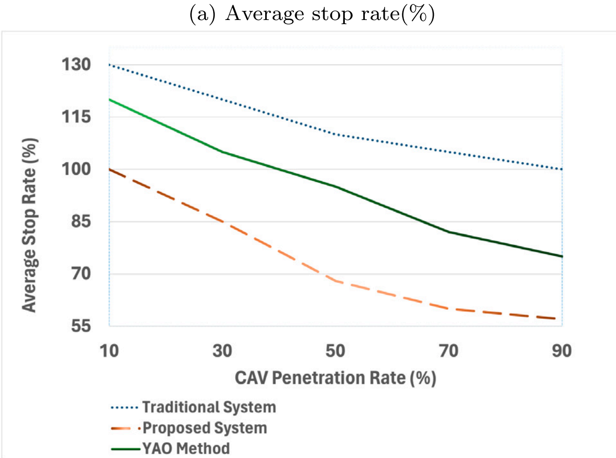
We assume uniform arrival rates for a simple estimation:

$$S \approx L \times (T - G) \times \text{period} \tag{6}$$

$$\text{Average Stop Rate (\%)} \approx \left(1 - \frac{G}{T}\right) \times 100 \tag{7}$$

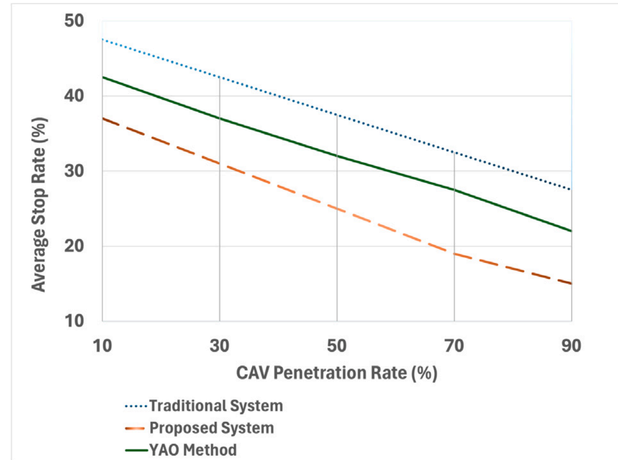


(a) Average stop rate(%)

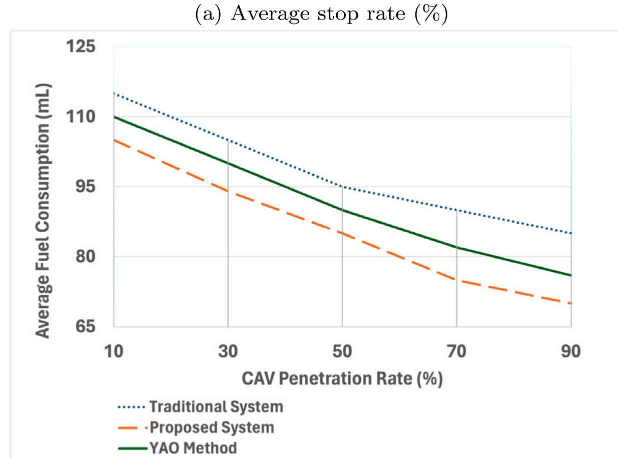


(b) Average fuel consumption (mL)

Fig. 5. Performance evaluation in low traffic scenario.



(a) Average stop rate (%)



(b) Average fuel consumption (mL)

Fig. 6. Performance evaluation in moderate traffic scenario.

6.3. Average fuel consumption (mL)

Fuel consumption depends on idling and stop-and-go conditions. We consider the time spent idling and the fuel rate during idling; R: fuel consumption rate while idling (mL per second) [46]; I: total idling time for a vehicle (seconds).

$$\text{Fuel Consumption per vehicle (mL)} = R \times I \tag{8}$$

For the total fuel consumption over N vehicles:

$$\text{Total Fuel Consumption (mL)} = N \times (R \times I) \tag{9}$$

The frequency and duration of stops directly affect Idling time I, which considers traffic light cycles.

$$I = \text{Number of Stops} \times (\text{Red} + \text{Yellow Phase Duration}) \tag{10}$$

$$\text{Average Fuel Consumption (mL)} = \frac{N \times R \times \text{Average Idling Time}}{N} \tag{11}$$

This equation simplifies the understanding of fuel consumption by focusing on two critical factors: The rate of fuel consumption when the engine is running without any load and the time when this occurs. It contributes to defining the effect of idling on the total fuel consumption in traffic control conditions.

6.4. Low traffic scenario

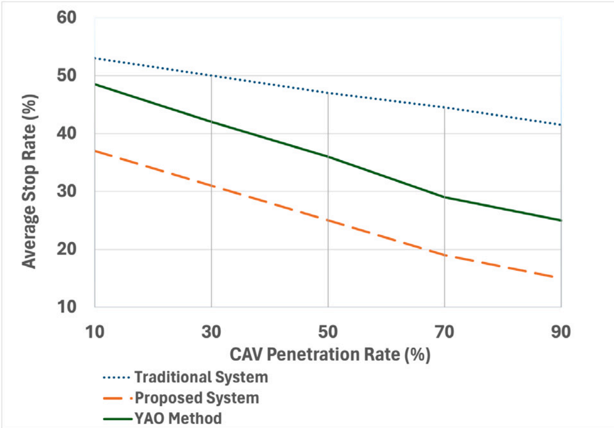
As can be seen in Fig. 5 the traditional method shows inefficiencies due to static signal timings which do not adapt to real-time changes, resulting in higher stop rates and increased fuel consumption.

The traditional method applies fixed timings allocated to specific signals without consideration of the congestion state of the road. This works on fixed timings of green, yellow, and red lights depending on the information gathered from previous experience and is recurrent in a never-ending cycle, irrespective of the present traffic conditions. In this case, it results in some disadvantages that include; making extra stops, consuming more fuel, and congestion in case of flow changes during rush hours or at certain times of the day.

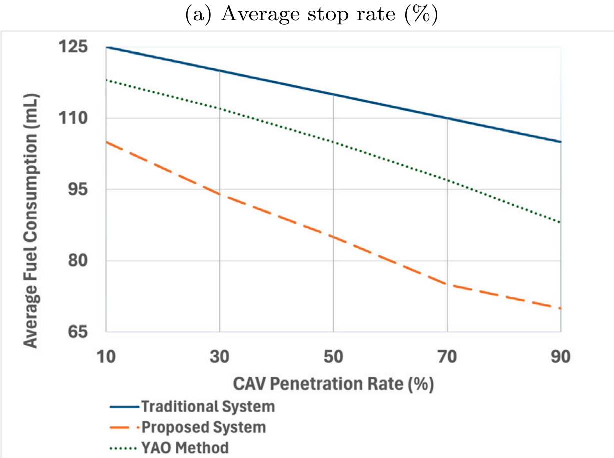
YAO method provides moderate improvements over the traditional method with gradual decreases in stop rates as CAV penetration increases, yet still lacks full optimization for dynamic traffic conditions. The proposed method that utilizes digital twins technology significantly outperforms both the YAO method and the traditional method by dynamically adjusting traffic signals using real-time data. This results in substantial reductions in stop rates – from 42% at 10% CAV penetration to 12% at 90% – and marked decreases in fuel consumption, showcasing superior adaptability and efficiency in traffic management.

6.5. Moderate traffic scenario

In moderate traffic scenarios, the traditional method struggles with peak traffic flows, leading to inefficient traffic management and higher resource use due to its non-adaptive nature (see Fig. 6). However, the YAO Method shows some adaptability with progressive improvements in stop rates, yet remains limited in fully optimizing traffic signal adjustments in response to real-time traffic dynamics. Meanwhile, the proposed system demonstrates exceptional adaptability, with notable



(a) Average stop rate (%)



(b) Average fuel consumption (mL)

Fig. 7. Performance evaluation in high traffic scenario.

improvements in both stop rates and fuel consumption across all CAV penetration levels.

Higher CAV penetration improves the performance of the traffic control networks. Owing to the efficient driving and real-time data, CAVs help in improving signal optimization which helps in reducing fuel consumption and the number of stops. The digital twin model and the BIRCH algorithm are key in this regard to help design this intelligent control system that would enable the traffic to flow smoothly with fewer stops.

The method adjusts to fluctuating intensities and regular traffic jams effectively, showcasing its capability to handle moderate traffic conditions far better than the YAO and Traditional methods.

6.6. High traffic scenario

In terms of traffic intensity, the traditional method fails to manage extreme congestion effectively due to its reliance on static signal timing, resulting in high stop rates and excessive fuel consumption (see Fig. 7). Although the YAO Method reduces stop rates with increasing CAV penetration, it starts from a high base and does not decrease sufficiently to manage high congestion effectively. Our proposed method excels in severe congestion scenarios by maintaining lower stop rates and optimizing fuel usage through advanced adaptive controls. It starts from a lower base stop rate of 37% at 10% CAV penetration and reduces it impressively to 15% at 90%, alongside significant reductions in fuel consumption from 102 mL to 70 mL at the same penetrations. This method's dynamic adjustments based on real-time conditions prove crucial for maintaining operational efficiency in high-intensity traffic.

7. Conclusion

The paper introduces a novel digital twin-based traffic light control solution that makes use of the BIRCH clustering algorithm to optimize traffic signal management in urban areas with the coexistence of CAVs and HVs. Our method shows concern for increased traffic flow efficiency, lower congestion, reduced vehicle idling time, and environmental impacts, that come from the constant changes that our method makes to the traffic signals, using both real-time and historical data. The suggested technique is highly superior to traditional and existing adaptive traffic management systems for different traffic intensities. For future research, more detailed data concerning pedestrians including the census of pedestrians, their movement patterns, and the incidents related to their safety can be used to bring a more realistic approach to the management of traffic in urban areas. This will help in the creation of a better traffic control system that can effectively manage traffic flow, decrease traffic jams, and at the same time increase pedestrian safety as well as boost the quality of life in urban areas.

CRediT authorship contribution statement

Haitham Y. Adarbah: Writing – original draft. **Mehdi Sookhak:** Writing – review & editing. **Mohammed Atiquzzaman:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The work of Mehdi Sookhak was supported by the National Science Foundation (NSF) under grant number CNS-2318725 and the US Department of Transportation (USDOT) Tier-1 University Transportation Center (UTC) Transportation Cybersecurity Center for Advanced Research and Education (CYBER-CARE) (Grant No. 69A3552348332).

References

[1] C. Schwarz, Z. Wang, The role of digital twins in connected and automated vehicles, IEEE Intell. Transp. Syst. Mag. 14 (6) (2022) 41–51.

[2] X. Liao, X. Zhao, Z. Wang, Z. Zhao, K. Han, R. Gupta, M.J. Barth, G. Wu, Driver digital twin for online prediction of personalized lane change behavior, IEEE Internet Things J. (2023).

[3] M.U. Shoukat, S. Yu, S. Shi, Y. Li, J. Yu, Evaluate the connected autonomous vehicles infrastructure using digital twin model based on cyber-physical combination of intelligent network, in: 2021 5th CAA International Conference on Vehicular Control and Intelligence, CVCI, IEEE, 2021, pp. 1–6.

[4] J. Guo, M. Bilal, Y. Qiu, C. Qian, X. Xu, K.-K.R. Choo, Survey on digital twins for internet of vehicles: Fundamentals, challenges, and opportunities, Digit. Commun. Netw. 10 (2) (2024) 237–247.

[5] Z. Lv, Y. Li, H. Feng, H. Lv, Deep learning for security in digital twins of cooperative intelligent transportation systems, IEEE Trans. Intell. Transp. Syst. 23 (9) (2021) 16666–16675.

[6] D. Xu, Y. Tian, A comprehensive survey of clustering algorithms, Ann. Data Sci. 2 (2015) 165–193.

[7] J. Harrington, M. Salibián-Barrera, Finding approximate solutions to combinatorial problems with very large data sets using BIRCH, Comput. Stat. Data Anal. 54 (3) (2010) 655–667.

[8] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O.P. Patel, A. Tiwari, M.J. Er, W. Ding, C.-T. Lin, A review of clustering techniques and developments, Neurocomputing 267 (2017) 664–681.

[9] Noel, Balanced iterative reducing and clustering using hierarchies (BIRCH), 2021, https://medium.com/@noel.cs21/balanced-iterative-reducing-and-clustering-using-heirachies-birch-5680adffaa58. (Accessed 15 June 2024).

[10] Y. Wang, X. Yang, H. Liang, Y. Liu, et al., A review of the self-adaptive traffic signal control system based on future traffic environment, J. Adv. Transp. 2018 (2018).

[11] M. Tubaishat, Y. Shang, H. Shi, Adaptive traffic light control with wireless sensor networks, in: 2007 4th IEEE Consumer Communications and Networking Conference, IEEE, 2007, pp. 187–191.

[12] B. Zhou, J. Cao, X. Zeng, H. Wu, Adaptive traffic light control in wireless sensor network-based intelligent transportation system, in: 2010 IEEE 72nd Vehicular Technology Conference-Fall, IEEE, 2010, pp. 1–5.

[13] X. Cheng, W. Tang, M. Yang, L. Jin, Vehicle sequencing at signal-free intersections: Analytical performance guarantees based on PDMP formulation, IEEE Trans. Control Syst. Technol. (2024).

[14] Y. Shi, Y. Liu, Y. Qi, Q. Han, A control method with reinforcement learning for urban un-signalized intersection in hybrid traffic environment, Sensors 22 (3) (2022) 779.

[15] H. Yao, X. Li, Decentralized control of connected automated vehicle trajectories in mixed traffic at an isolated signalized intersection, Transp. Res. C 121 (2020) 102846.

[16] D. Li, F. Zhu, J. Wu, Y.D. Wong, T. Chen, Managing mixed traffic at signalized intersections: An adaptive signal control and CAV coordination system based on deep reinforcement learning, Expert Syst. Appl. 238 (2024) 121959.

[17] D.R. Aleko, S. Djahel, An efficient adaptive traffic light control system for urban road traffic congestion reduction in smart cities, Information 11 (2) (2020) 119.

[18] H.Y. Adarbah, M. Sookhak, M. Atiquzzaman, A digital twin environment for 5G vehicle-to-everything: Architecture and open issues, in: Proceedings of the Int'L ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks, 2023, pp. 115–122.

[19] B. Lorbeer, A. Kosareva, B. Deva, D. Softić, P. Ruppel, A. Küpper, Variations on the clustering algorithm BIRCH, Big Data Res. 11 (2018) 44–53.

[20] G. Del Campo, L. Piovano, F.P.L. Oostrom, E. Saavedra, G. Zissis, A. Santamaria, Digital twins for street lighting: Challenges for a virtual reality solution based on internet-of-things devices and photometry rendering, in: 2023 IEEE Sustainable Smart Lighting World Conference & Expo, LS18, IEEE, 2023, pp. 1–6.

[21] Z. Rezaei, M.H. Vahidnia, H. Aghamohammadi, Z. Azizi, S. Behzadi, Digital twins and 3D information modeling in a smart city for traffic controlling: A review, J. Geogr. Cartogr. 6 (1) (2023) 1865.

[22] D. Quang Tran, S.-H. Bae, Proximal policy optimization through a deep reinforcement learning framework for multiple autonomous vehicles at a non-signalized intersection, Appl. Sci. 10 (16) (2020) 5722.

[23] G. Budan, K. Hayatleh, D. Morrey, P. Ball, P. Shadbolt, An analysis of vehicle-to-infrastructure communications for non-signalised intersection control under mixed driving behaviour, Analog Integr. Circuits Signal Process. 95 (2018) 415–422.

[24] C. Yang, S. Jin, J.A. Alagbé, C. Bai, A semi-"smart predict, then optimize" method for traffic signal control, IEEE Intell. Transp. Syst. Mag. (2023).

[25] C. Chen, J. Wang, Q. Xu, J. Wang, K. Li, Mixed platoon control of automated and human-driven vehicles at a signalized intersection: dynamical analysis and optimal control, Transp. Res. C 127 (2021) 103138.

[26] W. Zhao, D. Ngoduy, S. Shepherd, R. Liu, M. Papageorgiou, A platoon based cooperative eco-driving model for mixed automated and human-driven vehicles at a signalised intersection, Transp. Res. C 95 (2018) 802–821.

[27] G. Sharon, P. Stone, A protocol for mixed autonomous and human-operated vehicles at intersections, in: Autonomous Agents and Multiagent Systems: AAMAS 2017 Workshops, Best Papers, SãO Paulo, Brazil, May 8-12, 2017, Revised Selected Papers 16, Springer, 2017, pp. 151–167.

[28] K. Dresner, P. Stone, A multiagent approach to autonomous intersection management, J. Artif. Intell. Res. 31 (2008) 591–656.

[29] R. Mohajerpoor, C. Cai, M. Ramezani, Optimal traffic signal control of isolated oversaturated intersections using predicted demand, IEEE Trans. Intell. Transp. Syst. 24 (1) (2022) 815–826.

[30] A. Parks-Young, G. Sharon, Intersection management protocol for mixed autonomous and human-operated vehicles, IEEE Trans. Intell. Transp. Syst. 23 (10) (2022) 18315–18325.

[31] W. Ma, J. Li, C. Yu, Shared-phase-dedicated-lane based intersection control with mixed traffic of human-driven vehicles and connected and automated vehicles, Transp. Res. C 135 (2022) 103509.

[32] X. Liu, P.-C. Hsieh, P. Kumar, Safe intersection management for mixed transportation systems with human-driven and autonomous vehicles, in: 2018 56th Annual Allerton Conference on Communication, Control, and Computing, Allerton, IEEE, 2018, pp. 834–841.

[33] M. Al-Turki, N.T. Ratrout, S.M. Rahman, K.J. Assi, Signalized intersection control in mixed autonomous and regular vehicles traffic environment—A critical review focusing on future control, IEEE Access 10 (2022) 16942–16951.

[34] M. Pourmehrab, L. Elefteriadou, S. Ranka, M. Martin-Gasulla, Optimizing signalized intersections performance under conventional and automated vehicles traffic, IEEE Trans. Intell. Transp. Syst. 21 (7) (2019) 2864–2873.

[35] Y. Li, W. Zhang, Traffic flow digital twin generation for highway scenario based on radar-camera paired fusion, Sci. Rep. 13 (1) (2023) 642.

[36] V.K. Kumarasamy, A.J. Saroj, Y. Liang, D. Wu, M.P. Hunter, A. Guin, M. Sartipi, Integration of decentralized graph-based multi-agent reinforcement learning with digital twin for traffic signal optimization, Symmetry 16 (4) (2024) 448.

[37] X. Ji, W. Yue, C. Li, Y. Chen, N. Xue, Z. Sha, Digital twin empowered model free prediction of accident-induced congestion in urban road networks, in: 2022 IEEE 95th Vehicular Technology Conference, VTC2022-Spring, IEEE, 2022, pp. 1–6.

[38] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognit. 38 (12) (2005) 2270–2285.

[39] D. Singh, B. Singh, Investigating the impact of data normalization on classification performance, Appl. Soft Comput. 97 (2020) 105524.

[40] H.-C. Ryu, S. Jung, S. Pramanik, An effective clustering method over CF^{++} tree using multiple range queries, IEEE Trans. Knowl. Data Eng. 32 (9) (2019) 1694–1706.

[41] A. Lang, E. Schubert, BETULA: Fast clustering of large data with improved BIRCH CF-trees, Inf. Syst. 108 (2022) 101918.

[42] A.D. Fontanini, J. Abreu, A data-driven BIRCH clustering method for extracting typical load profiles for big data, in: 2018 IEEE Power & Energy Society General Meeting, PESGM, IEEE, 2018, pp. 1–5.

[43] CARLA Simulator Team, Advanced SUMO, 2024, URL https://carla.readthedocs.io/en/latest/adv_sumo/. (Accessed 6 May 2024).

[44] German Aerospace Center (DLR), SUMO - simulation of urban mobility, 2024, https://sumo.dlr.de/docs/index.html. (Accessed 2 May 2024).

[45] M. Behrisch, L. Bieker, J. Erdmann, D. Krajzewicz, SUMO–simulation of urban mobility: an overview, in: Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation, ThinkMind, 2011.

[46] M. Zhou, H. Jin, W. Wang, A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing, Transp. Res. D 49 (2016) 203–218.

Dr.Haitham Y. Adarbah is a postdoctoral researcher at Texas A&M University, Corpus Christi, TX, USA. Before joining TAMU-CC, Dr. Adarbah worked as an IT lecturer at Gulf College in Muscat, Oman, from April 2009 to August 2024. He received the bachelor's degree in computer science from the Al-Zaytoonah University of Jordan, Amman, Jordan, in 2004, the master's degree in computer science from Amman Arab University, Amman, in 2009, and the Ph.D. degree in wireless networks from De Montfort University, Leicester, U.K., in 2015. He is very interested in 5G, V2X, Digital Twin, and mobile ad hoc networks (MANET): route discovery schemes, routing techniques, bandwidth utilization, power consumption, delay, security wireless networks, and network simulation and modeling. He has authored several research papers, which were all published in various international journals and conferences.

Dr. Mehdi Sookhak (M'12, SM'19) is an Assistant Professor of Computer Science at Texas A&M University – Corpus Christi (TAMU-CC), where he leads the Intelligent Edge and Cellular Network Laboratory (IECNL) and co Lead of Cybersecurity Center for Advanced Research and Education (CYBER-CAR). He received a Ph.D. in Computer Science from the University of Malaya (U.M.), Malaysia in 2015 where he contributed as a research assistant at the Center for Mobile Cloud Computing Research (C4MCCR). Following this, from 2015 to 2017, he served as a Postdoctoral Fellow in the Department of Systems and Computer Engineering at Carleton University, Canada. Before joining TAMU-CC, Dr. Sookhak was an assistant professor in Arizona State University between 2017 and 2019 and was an assistant professor in Illinois State University between 2019 and 2021. His research expertise spans multiple areas, including vehicular networks, Wireless sensors and actuators, mobile communication, cloud and edge computing, blockchain technology, and computer security. He possesses an impressive publication record, with over 70 articles to his credit in renowned journals and conferences such as IEEE Globecom, IEEE TPDS, IEEE TVT, IEEE CST, IEEE System, ACM Communication, ACM TOSN, and IEEE Communication. Furthermore, Dr. Sookhak contributes significantly to the academic community through his editorial roles on the boards of several ISI-indexed journals, including Vehicular Communications, IEE CEM, IEEE Access, and Electronics. He has also taken on leadership roles by chairing conferences like IEEE WTS and IEEE ICC. Dr. Sookhak's dedication to research extends beyond publications, as he serves as the Principal Investigator (PI) for numerous funded projects by organizations like NSF, NIH, NIFA, and DOT. He is a senior member of Institute of Electrical and Electronics Engineers (IEEE).

Mohammed Atiquzzaman obtained his M.S. and Ph.D. in Electrical Engineering and Electronics from the University of Manchester (UK), and BS from Bangladesh University of Engineering and Technology. He holds the Edith Kinney Gaylord Presidential Professorship and the Hitachi Chair Professorship in the School of Computer Science at the University of Oklahoma and is a senior member of IEEE. Dr. Atiquzzaman is the editor-in-chief of the Journal of Networks and Computer Applications, founding Editor-in-Chief of Vehicular Communications, former co-editor-in-chief of Computer Communication, and has served/serving on the editorial boards of many highly-ranked journals, such as IEEE Transactions on Mobile Computing, IEEE JSAC, IEEE Communications Magazine, and International Journal on Wireless and Optical Communications. He also guest-edited 12 special issues in various journals. He has served as symposium co-chair for many IEEE Globecom and IEEE ICC conferences and co-chair of IEEE High Performance Switching and Routing Symposium. He was the Workshop Chair of Globecom 2022, panels co-chair of INFOCOM'05, and is/has been on the program committee of numerous conferences such as INFOCOM, ICCCN, and Local Computer Networks. He serves on the review panels of funding agencies such as the National Science Foundation, National Research Council (Canada) and Australian Research Council (Australia). In recognition of his contribution to NASA research, he received the NASA Group Achievement Award for "outstanding work to further NASA Glenn Research Center's effort in the area of Advanced Communications/Air Traffic Management's Fiber Optic Signal Distribution for Aeronautical Communications" project. He received the IEEE 2023 Internet of Things, Ad Hoc and Sensor Networks Technical Achievement Award for valuable contributions to the Internet of Things, Ad Hoc and Sensor Networks scientific community, the IEEE 2018 Satellite and Space Communications Technical Recognition Award for valuable contributions to the Satellite and Space Communications scientific community, the IEEE 2017 Distinguished Technical Achievement Award in recognition of outstanding technical contributions and services in the area of communications switching and routing, and the IEEE 2020 Distinguished Service Award in recognition of outstanding service for the benefit and advancement of the IEEE Communications Society. Dr. Atiquzzaman received the IEEE Fred W. Ellersick Prize for his paper entitled "Evaluation of SCTP for Space Networks". He has been invited to deliver keynote talks on over 40 international conferences around the globe, including USA, Brazil, Italy, Poland, Portugal, Australia, China, Taiwan, Malaysia, and Korea. His research interests are in satellite communications, wireless and mobile networks, vehicular networks, ad hoc networks, Quality of Service, and optical communications. His research has been funded by the National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), U.S. Air Force, Cisco, Honeywell, Oklahoma Department of Transportation, and Oklahoma Highway Safety Office. He has over 450 refereed technical publications, most of which can be accessed at www.cs.ou.edu/~atiq.