

Research Article

Scalable System for Smart Urban Transport Management

**Nauman Ahmad Khan, Jean-Christophe Nebel , Souheil Khaddaj,
and Vesna Brujic-Okretic**

School of Computer Science & Mathematics, Kingston University London, Kingston KT1 2EE, UK

Correspondence should be addressed to Jean-Christophe Nebel; j.nebel@kingston.ac.uk

Received 3 March 2020; Revised 3 September 2020; Accepted 8 September 2020; Published 16 September 2020

Academic Editor: Stefano De Luca

Copyright © 2020 Nauman Ahmad Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Efficient management of smart transport systems requires the integration of various sensing technologies, as well as fast processing of a high volume of heterogeneous data, in order to perform smart analytics of urban networks in real time. However, dynamic response that relies on intelligent demand-side transport management is particularly challenging due to the increasing flow of transmitted sensor data. In this work, a novel smart service-driven, adaptable middleware architecture is proposed to acquire, store, manipulate, and integrate information from heterogeneous data sources in order to deliver smart analytics aimed at supporting strategic decision-making. The architecture offers adaptive and scalable data integration services for acquiring and processing dynamic data, delivering fast response time, and offering data mining and machine learning models for real-time prediction, combined with advanced visualisation techniques. The proposed solution has been implemented and validated, demonstrating its ability to provide real-time performance on the existing, operational, and large-scale bus network of a European capital city.

1. Introduction

Advancements in smart devices and their interoperability have led many industries to develop sensor-based solutions that collect and exploit data not only from individual devices but also from their operating environment [1–3]. Modern public transport vehicles, such as buses and trains, are now equipped with many sensors that gather a huge amount of information including their speed, direction, location, fuel consumption, driving behaviour, departure and arrival times, and even on-board videos [4, 5]. While this already assists those vehicles' individual navigation and enhances their passengers' security [6], real-time combination and analysis of all the data collected within an entire transport network have the potential to provide extremely valuable information to transport companies and their consumers. Consequently, the demand for the development of intelligent and context-aware transport systems [7] has been driven by both increasingly demanding users, which require smartphone apps informing them on the network status, and operators, which need to monitor and predict the behaviour

of the network to provide a better service. Indeed, context-awareness can be used for in-route planning to resolve conflicts between different vehicles, accurate route selection if there is congestion, and incident monitoring [8, 9], suggesting alternate routes based on commuters' preferences [4, 10] and helping bus drivers to find the nearest locations for emergency help and refuelling [11, 12].

Provision of intelligence in such context-aware public transport systems is implemented by channelling the contextual data being collected from the transport infrastructure into prediction systems so that behaviour of the transport system can be forecast. Yet, this can only be delivered if highly efficient information management architecture is available to handle an unprecedented volume of data in real time [13–17]. That challenge is further exacerbated by the fact that the data to be combined and analysed are heterogeneous and come from a variety of sources and sensors through diverse infrastructures [5]. This requires a high level of coordination in the transport system: data captured in the vehicles need to be transferred in real time to a common processing platform where they are transformed in order to

be usable by potential recipients. As data validity is short in a dynamic transport system, real-time processing and delivery of such information to the target audience require a high performance and scalable platform. Ease of integration, adaptability, and sustainability are other important quality factors, as new devices and their corresponding services have to be accommodated without affecting the system. In addition, high accuracy and usability, particularly in terms of prediction and data visualisation, are also very important requirements.

Traditional system architectures have been unable to meet many of those quality requirements of modern transport systems [18–20]. Thus, a unified adaptable and scalable software architecture for smart transport systems using service orientation principles, memory caching, and advanced data mining and machine learning models is proposed herein. Engineering such a component-based architecture not only has to meet the identified quality requirements but also has to consider implementation and integration of all the components and tiers in the system. The system architecture needs to be able to acquire, store, manipulate, and integrate information from heterogeneous data sources in order to produce eventually reliable predictions. This will help transport managers and planners to optimise, in real time, scheduling, stops usage, and passengers' time. The system architecture has to offer efficient data structures for quick storage and retrieval, on the fly analysis capabilities and preparation of use-ready information, to reduce processing at the time of request. Such system will be able to conduct effective analytics on the data as it comes, while keeping processed data ready to be transferred to requesting devices or vehicles.

After discussing advances in transport management systems together with data and architectural considerations, the core components of the proposed architecture are presented with a focus on integration services, data structure, and analytics components. Then, experimental results using real-time data fed from an actual transport network are presented, validating both the proposed architecture and the value of smart data analytics. Finally, the proposed solution is discussed and compared with other systems.

2. Related Work

2.1. Intelligent Transport Management Systems. The concept of intelligent transport system (ITS) emerged in late 1970s well before the growth of digital devices and availability of platforms allowing sharing the data produced or consumed by these devices [21]. ITSs—defined as a group of technologies, systems, and services for efficient and secure transport services—cover a wide domain that includes both private and commercial transportation systems. Benefitting from intelligent technology integration, ITSs have already provided many new opportunities for building safe, reliable, and scalable service infrastructures for transport [22–24]. Consequently, ITSs have become a key element of smart cities and connected digital ecosystems [25].

The complexity of ITS architectures rises from the combination of many different software and hardware

components and their connectivity [21]. Thus, it is important to identify the interfaces between those components so that communication between them can be standardised [26]. They provide the core of the architectural framework that can be used for planning, defining, deploying, and integrating intelligent transport systems. Such framework usually defines the functional entities, the workflows to connect them, and the services offered to the user [27]. These services typically include information about traveling, management of different traffic modes, collection of data for performance improvement, monitoring driving behaviour, incident response handling, and managing vehicle fleets in public and private settings [28]. These user services are delivered with the support of many different entities that are connected to each of the services offered by the ITS [25]. In the context of a bus network, they comprise direct users of the system such as passengers, bus management operators, transport companies, and external entities that the ITS interacts with. They include data providers (e.g., weather and roadworks) and the different sensors (e.g., GPS, speed, direction, temperature, and light detectors) that are installed on the vehicles [13, 29].

Therefore, the design of a successful ITS architecture should consider all these parameters, using information and data flows, to define the system's functional details [30, 31]. Each user service offered by the ITS should be represented as a workflow where each step is bound with the required process inputs and outputs. Moreover, as an ITS is a distributed system where different entities may be quite remote from each other, design workflows for its services have to be formally followed to represent how different functions are formed, so that their compliance with standards can be verified. Like any large software, ITSs are very complex systems that require formal specification in order to ensure that the relationships between the different components are well defined and their behaviours are predictable [32]. In summary, the ITS architecture plays a key role in the identification of the system complexities and the ability to implement solutions.

Depending on the environment or region in which an ITS is implemented, different stages of deployment should be planned [24], so that the system is developed according to the knowledge of the context it belongs to. As the demand for services changes and new functionality needs to be introduced to meet them, an ITS should be able to continuously improve. This ongoing process involves evolution of existing services and integration with new features while maintaining existing functionalities [24]. Consequently, in common with other large-scale systems that have the capacity to evolve and expand their services, ITSs must offer many different quality characteristics including compatibility, extendibility, interoperability, reliability, integration, and standardisation [29, 30]. In addition, they should deliver system performance and accurate predictions based on historical transport data, and scalability, that is, providing good quality of service under increased load.

2.2. Transport Data Analytics. To take full advantage of data collected from public transport networks, smart analytics

tools have to be included in the system to improve quality of service by supporting data-driven decision-making [1, 33–36]. Through integration of data visualisation, data mining, and machine learning techniques in the data acquisition and storage processes, pattern may be detected and predictions can be made. For example, not only could an ITS predict bus arrival times at specific stops, but also it could identify particular behaviours based on information available from different stops either on a given route or across different routes. This may lead to the redesign of routes or part of the network to improve the transport system.

There has been a considerable amount of research into designing effective approaches so that accurate and adaptive prediction systems can be offered to both bus companies and passengers [37–39]. Historical data, such as arrival times against planned times for given stops over an extended period, are essential to provide a deep insight into the behaviour of a transport network. Combined with machine learning techniques, such data allow forecasting about the network's status, even when there is no information available about external factors [23, 40–43]. Existing systems have exploited neural networks and regression and clustering techniques to predict bus arrival times [43–47], while others have performed route prediction using GPS data observation [48]. Overall, conversion of data into knowledge by the application of smart analytics techniques supports strategic decision-making and system automation, delivering improved operations and services.

2.3. Architectural Consideration. Many ITS architectures [49–53] have been developed according to the developed architectural concepts previously described and the standards set by the working group on ITS architecture of the International Organization for Standardization (ISO) [28, 42, 54]. They include a transport-knowledge-based architecture able to optimise the routes of a network. Using a hierarchical strategy of bus nodes and connection points at different level, static, such as planned arrivals, and dynamic schedule information are combined for effective predictions [52]. Alternatively, a cloud-based design of an ITS was proposed taking advantage of cloud platform and Internet of things technologies [53]. The choice of a cloud platform was motivated by the fact that as the flow of heterogeneous data produced by a growing number of sensors increases in speed and volume, a traditional client server architecture would not be able to handle it. However, since those ITS architectures are of high level, they do not include the required implementations to support actual public transport systems.

Nevertheless, there have been some attempts at implementing small-scale smart systems, such as the University of Wollongong's shuttle service [21]. That service (UniShuttle) runs buses between different campuses and other key destinations including the city centre and the beach. However, that system's functionality and performance are limited: as scalability was not among its requirements, quality of service has degraded with the increase of the number of buses in the network. Moreover, since its architecture was not designed to be interoperable

and compatible with external systems, it cannot integrate with any external data source. Finally, its prediction model only considers real-time data from the segment the bus is driving through, making it blind to information from previous stops such as existing delays.

In summary, there has been an increasing demand for smart context-aware public transport systems in order to deliver a better service. This relies on the development of architectures that can provide a basis for intelligent information processing and decision-making support. So far, as their implementations have been of quite high level, they have not been able to provide the essential services expected by transport systems. In particular, such architectures have to consider the integration of all data relevant to a specific network. In addition, they need to be flexible in order to adapt and evolve to rapidly changing requirements; otherwise, those systems cannot remain competitive when novel technologies become available. Attaching quality attributes to such software architecture can help in their evaluation, optimisation, and adaptation of evolving needs [55]. All those issues, especially those associated with the architectural design and the prediction models, will be addressed in the proposed architecture that is introduced in the following section.

3. Service-Based Intelligent Transport Architecture

3.1. Key Principles. As the state-of-the-art review has revealed, there is a need for a new and practical architecture for transport systems, which is designed according to both its intrinsic requirements and its expected interactions with other entities in a smart environment. In addition, it is suitable that this novel service-based intelligent transport architecture is able to allow not only the creation of new smart transport systems but also the transformation and support of existing systems. The principles on which the proposed architecture is based aim at delivering the best quality of service. More specifically, it delivers the following five key features.

Plug and Play architecture: Any component in the architecture can be replaced by a more advanced one without any additional change being required to the system. This feature is particularly critical with analytics algorithms, since they belong to an extremely dynamic research field. Consequently, any service-based intelligent architecture must be able to be updated regularly to support competitive decision-making.

(1) The Real-Time Cache Manager. While most existing transport systems process information through expensive spatial queries from the database [56, 57], usage of a real-time cache manager enables the system to work on a novel data structure based on hierarchical regions so that the effort of performing complex spatial queries can be drastically reduced, contributing to the system performance. Consequently, this facilitates integration of the system with any other internal or external system without affecting the required functionality.

(2) *Automation*. Absence of automation may lead different processes to rely on each other, which can conduce to deadlocks and component failures when they are run in parallel [58]. Therefore, by automating the whole flow of functionality, processes from data acquisition to data visualisation can be autonomous, which ensures system robustness when they are run concurrently.

(3) *Genericity*. Although the proposed architecture is primarily designed to be used on public transport systems, it can be easily customised to any other data-oriented domain for the production of predictive analytics. Thus, the implemented analytics algorithms are generic and independent from the application and the data warehouse. In principle, database type or data format should not affect the production of predictive analysis and visualisation for the delivery of strategic decision-making and operational excellence.

(4) *Flexibility*. Since many systems do not have integration or expanding capabilities, they encounter performance and scalability issues [21]. The proposed architecture is easily applicable to any existing system, not only allowing data integration and offering scalability but also empowering it with predictive analytics and monitoring.

3.2. Architecture Description. The proposed architecture is a generic system, where data processing components are conceived to be scalable and efficient. It is designed not only to deliver the provision of transport services but also to adhere to quality of service requirements and to provide data analytics capability. This architecture relies on six modules. Figure 1 depicts those modules and their interactions. For each of them, their essential features are elaborated in the following paragraphs.

(1) *The Web Services Module*. It is based on a dynamic WCF web service management system with asynchronous access allowing requests, which do not need immediate data response, not to have to wait for data to be stored and processed [59]. The data collected from the vehicles' different sensors are dropped at designated web services that add the data requests to their processing pool. Those different web service components serve different types of data requests coming into the system such as data coming from internal networks and those generated by heterogeneous external sources, for example, third-party transport and prediction data providers. For example, data arriving at a vehicle's current location and those produced by the vehicle's speed metre are linked so that both speed and location values can be combined in subsequent analyses. Such an approach allows integration of new data sources into the architecture without affecting the inner working of the system, which delivers elements of pluggability and extendibility.

(2) *The Real-Time Cache Manager Module*. It is essential to enhance performance and support real-time monitoring and

predictions [60]. Its purpose is to reduce database lookup calls when information is needed so that data can be dispatched on the go without any delay in accessing this information. This module is also responsible for providing short-term on-the-fly predictions based on the transport network's current situation such as arrival times. Moreover, it generates alerts when specific areas are identified as being prone to congestion and/or delay issues. To ensure real-time response for both incoming and outgoing data, the module keeps a snapshot of the current situation of the data about the location and predictions for short-term future events. This is performed by maintaining an optimised hierarchical data structure to support storage and retrieval of the information. In particular, as queries for spatial data can be time-expensive once data volumes are important, the module's data structure enables the system to run spatial queries on filtered and context-specific data, for example, transport network's areas.

(3) *The Analytics Component Module*. It provides the system's intelligence. It processes data coming from different internal and external sources so that they can be converted into usable information for strategic decision-making. Once data retrieval is completed, they are fed into the data clean-up component that removes any anomalies and transforms the data's structure to make it compatible with the data structure required for visualisation, data mining, and predictions. The visualisation component contributes to the visual representation of the data, as well as predicted probabilities. A key specificity is that, instead of only plotting the data, it also indicates accuracy associated with prediction processes. The data analytics component is comprised of many data mining and machine learning techniques so that accurate statistics and prediction models can be generated to support decision-making. Its implementation includes a variety of algorithms including Clustering, Linear Regression, Logistic Regression, Decision Trees, and Neural Networks [61–64].

(4) *The Transformation Module*. It is responsible for interoperability with other internal or external systems when data exchange is required. As data are available in many different formats, like JSON and XML, the transformation module offers integration between different data types and structures. Since its components implement a layer for each type, the system implementation is independent of the format of the data that are either imported from or exported to other systems through the web services. Thus, the transformation module ensures data interoperability by acting as a layer either between different components or for intersystem communication.

(5) *The Components Factory Module*. It equips the system with dynamic flexibility for usage and integration of new components without the need for new deployment. It has the reference of the different components available in the system to conduct different functional activities such as data retrieval, clean-up, and application of prediction components. Although the current system integrates data

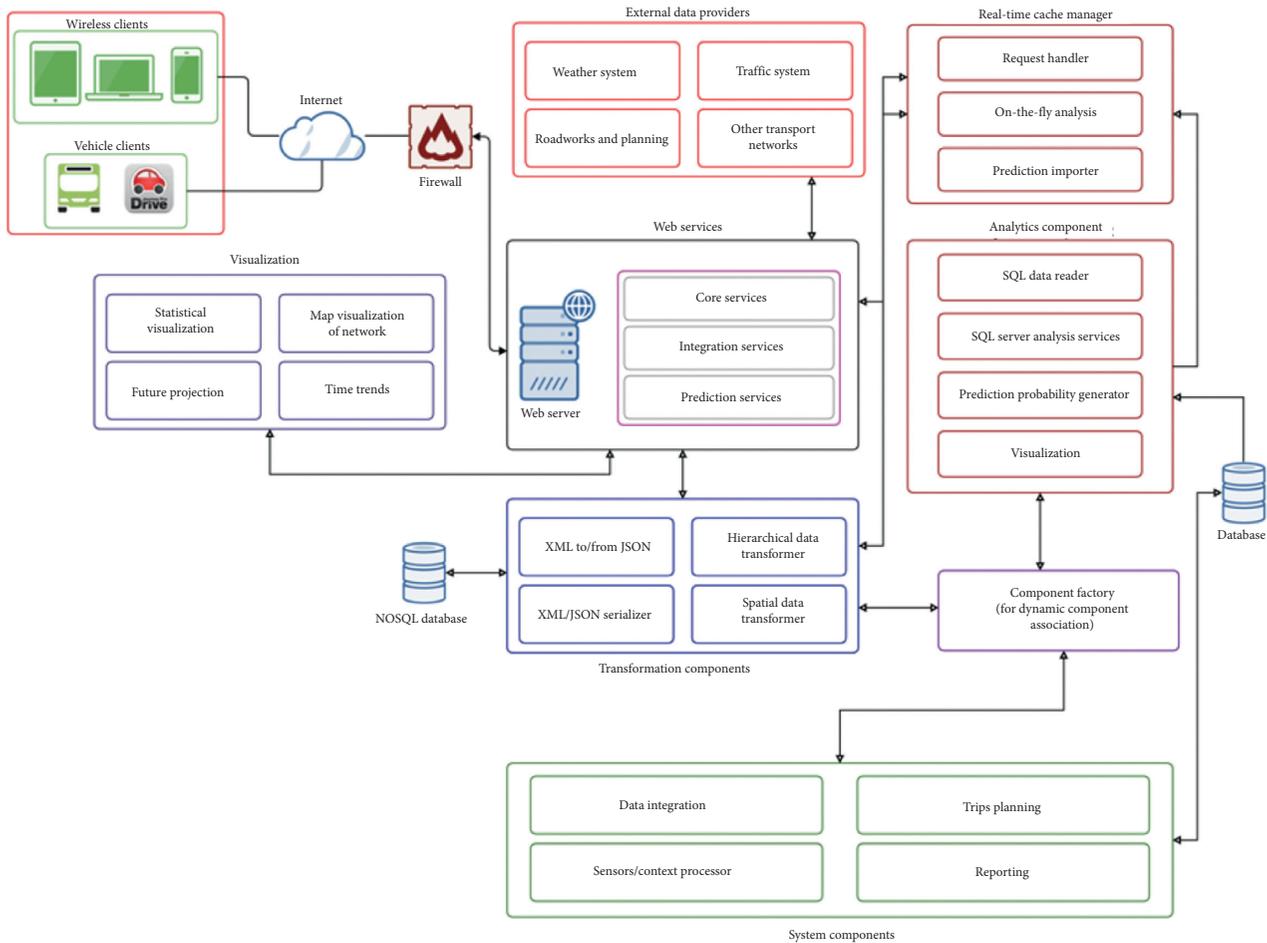


FIGURE 1: Modules of the proposed system architecture and their interactions.

from an SQL server database, the system can equally work with components built to work with data coming from any other database system. By using the relevant transformation components, those data can be fed into the prediction system with the expected structure regardless of which database platform they came from. The module connects with all major components of the system like the real-time cache manager, therefore adding to the system the capacity of using different real-time prediction systems.

(6) *The System Components Module.* It plays an important role in processing and archiving the data coming from different sources and converting them to compatible structure so that the analytics component can easily fetch data. It provides the basis of the prediction process and enables the real-time cache manager to make an on-the-fly estimation of arrival times and detection of traffic situations like congestion. Its integration component has endpoints for data coming from different external data providers about, for examples, planned trips, roadworks, and diversions. It is responsible for interconnection of different formats and establishing criteria to join information from more than one source to create a single standard schema of information that is used for analysis and prediction.

4. Evaluation

As the proposed system has been designed to collect, deliver, and process data in real time while being scalable, its meaningful evaluation requires an implementation on an existing, operational, and large-scale transport system. This was performed through a partnership with Mermaid Technology [65], a company that runs an urban transport system with 1200 buses in the city of Copenhagen, Denmark. This capital city is particularly appropriate to evaluate the system since it is a typical European city with urban environment, modern establishment, heavy footfall, and streets and roads insufficient to handle the traffic requirements of the city. Although bicycle usage is high, public transport is the preferred option to save time in commuting. Despite the presence of an underground train system connecting various destinations in the city, buses remain the main means of transport.

Mermaid's buses are equipped with on-board sensors collecting in real-time bus operations, bus position (GPS based), speed and direction, status of local signs and traffic lights, and images from CCTV cameras placed on the buses. This corresponds to a data volume of over 1 KB per second per bus. As those sensors are connected to a central hub, the data are transmitted in real time to a server where

information is stored, processed, and shared with other buses to inform them of the activity of neighbouring buses allowing some journey optimisation.

In this study, the main emphasis is on the most challenging routes: those transporting the most passengers and those involved in complex scenarios, that is, routes running through the busy city centre and routes offering many connections to other routes. As a result, out of the 1298 routes operated by Mermaid's 1200 buses, the experiments focused on the data produced by 128 buses and their associated 188 routes. The rich context information captured by the on-board sensors enabled the server system to support provision of real-time versatile and customer-driven visualisation, data mining, and predictions. A sample of data stored in the system is displayed in Table 1. For each stop of a completed journey that was expected to start at 23:53:00 on 31st January 2018, the table provides the bus's predicted and actual arrival and departure times. Although the system is able to handle in real time without error (see Section 2.5.1) the flow of sensor data transmitted by the active buses, data associated with a given journey may occasionally be missing. For example, some sensors were not active or collected invalid data; in addition, some buses were rerouted or did not complete their journey. Consequently, data were pre-processed to flag such journeys so that their associated data were not considered for analysis tasks such as visualisation and predictions.

In the following sections, the proposed system is rigorously evaluated using the data produced by the Mermaid's buses. Most importantly, performance, scalability, and robustness of web services and the associated real-time cache components are assessed, including under challenging conditions. Following validation of the core of the system, the analytical values of the implemented analytic components are illustrated. First, examples of the versatility and customisation of real-time visualisation are provided. Second, benefits of real-time data mining and prediction applications are demonstrated.

4.1. Performance, Scalability, and Robustness of Web Services and Real-Time Cache Components. The web services must provide the performance and scalability required for real-time data processing. In particular, prediction criteria have to be updated with the latest data being added along with the real-time data feed. Evaluation was conducted by connecting the web services to real client components loaded on the buses to evaluate the readiness, availability, and automatic scaling of the services provision. In addition to measuring response time according to the number of users, the presence of potential errors was monitored if any call failed. While initially between 500 and 1000 buses were connected to the system, eventually 25000 unique buses were considered in a staging environment.

As shown in Table 2, despite significant increases in the number of calls that were made, the throughput, that is, number of calls processed per minute, delivered by the web services components was only reduced moderately. This indicates that the system is highly scalable and can serve

concurrent calls very efficiently. The system is able to handle more than 25,000 unique clients per minute. The response time of less than 395 milliseconds (Table 2), even when it is serving more than 70,000 requests per minute, shows that the system can handle thousands of concurrent requests without any impact on its performance. Moreover, they do not cause any interference on the response time for other requests being made to the system. It is also important to note that data are not static; they continuously change and need to be read from the caching-based database. During those experiments, none of the calls failed as each one was served with requested data, demonstrating the robustness of the system. That shows the uptime and availability of the service under stress time as requests are served within the threshold time of data availability, that is, one minute. As the server used for these experiments did not have high-end specifications—an Intel Xerox 2-core CPU with only 2 GB RAM installed was used—it is expected that usage of a machine with higher specifications would deliver significantly better response time and throughput without generating any error.

4.2. Real-Time Visualisation. As the core of the system was validated, it can be exploited for data analytics. Real-time visualisation is an important functionality allowing human users to detect patterns that will support their decisions. However, provision of pertinent visualisation is not an easy task. For instance, by only considering 30 routes out of the 1298 existing ones, visualisation of their operational stops on a map is able to provide meaningful information (see Figure 2(a)). However, the additional mapping of live buses on the stops (Figure 2(b)) to indicate how many connections exist between different routes increases its complexity significantly making gaining clear understanding of the bus situation and their routes difficult.

To address this, the monitoring system, built in the system's architecture, creates a variety of visualisation representations at different levels, such as system, zone, and route, so that information can be viewed at the most relevant resolution using the most appropriate depiction. For example, Figure 3 presents in real time updated statuses of all the routes on the 15th of April 2018 between 7:00 and 9:00 am; the colour indicates the type of delay being reported at each stop. Such a map is particularly useful to infer possible relationship between different routes in terms of delays. As the proposed architecture is able to provide simultaneous real-time monitoring of more than 10,000 stops, continuously updating information about delays and expected arrivals, it provides a strategic advantage for decision-making.

4.3. Data Mining and Prediction Components. While visualisations are able to present complex data so that users can gain a better insight into the status of the transport network, usage of data mining and machine learning algorithms permits going much further in terms of pattern recognition and predictions. For example, the system allows extraction of planned and actual arrival and departure times (see data samples in Table 1), so that the bus company can visualise

TABLE 1: Stored data of predicted and actual arrival and departure times for all stops of a bus journey (identical planned times for consecutive stops denote the presence of optional stops).

Stop name	Planned arrival time	Actual arrival time	Planned departure time	Actual departure time
Ringsted st.	1/31/2018 23:53:00	1/31/2018 23:46:10	1/31/2018 23:53:00	NULL
Hovmarksvej	1/31/2018 23:53:00	1/31/2018 23:54:48	1/31/2018 23:53:00	1/31/2018 23:55:00
Søndergade	1/31/2018 23:54:00	1/31/2018 23:56:24	1/31/2018 23:54:00	1/31/2018 23:56:35
Ørnevej	1/31/2018 23:55:00	1/31/2018 23:56:59	1/31/2018 23:55:00	1/31/2018 23:57:10
Kildemarken	1/31/2018 23:56:00	1/31/2018 23:58:00	1/31/2018 23:56:00	1/31/2018 23:58:09
Kildeskovvej	1/31/2018 23:56:00	1/31/2018 23:58:35	1/31/2018 23:56:00	1/31/2018 23:58:43
Kærehave	1/31/2018 23:58:00	1/31/2018 23:59:44	1/31/2018 23:58:00	1/31/2018 23:59:48
Adamshøjvej	1/31/2018 23:59:00	2/1/2018 0:00:56	1/31/2018 23:59:00	2/1/2018 0:01:03
Kværkebyvej	2/1/2018 0:00:00	2/1/2018 0:01:53	2/1/2018 0:00:00	2/1/2018 0:02:00
Bedstedvej	2/1/2018 0:00:00	2/1/2018 0:02:24	2/1/2018 0:00:00	2/1/2018 0:02:30
Fjællebro	2/1/2018 0:00:00	2/1/2018 0:02:53	2/1/2018 0:00:00	2/1/2018 0:02:59
Kværkeby Stationsvej	2/1/2018 0:01:00	2/1/2018 0:03:48	2/1/2018 0:01:00	2/1/2018 0:03:54
Fjællebro Huse	2/1/2018 0:02:00	2/1/2018 0:04:28	2/1/2018 0:02:00	2/1/2018 0:04:33
Slimminge, Egelundsvej	2/1/2018 0:04:00	2/1/2018 0:07:01	2/1/2018 0:04:00	2/1/2018 0:07:10
Slimminge, Slimmingevej	2/1/2018 0:04:00	2/1/2018 0:07:45	2/1/2018 0:04:00	2/1/2018 0:07:55
Skovbo Efterskole	2/1/2018 0:05:00	2/1/2018 0:08:35	2/1/2018 0:05:00	2/1/2018 0:08:39
Lundegårdsvej	2/1/2018 0:06:00	2/1/2018 0:09:10	2/1/2018 0:06:00	2/1/2018 0:09:14
Gummersmarkvej	2/1/2018 0:08:00	2/1/2018 0:10:05	2/1/2018 0:08:00	2/1/2018 0:10:09
Tingbjerggård	2/1/2018 0:09:00	2/1/2018 0:10:36	2/1/2018 0:09:00	2/1/2018 0:10:40
Vollerslevvej	2/1/2018 0:10:00	2/1/2018 0:11:04	2/1/2018 0:10:00	2/1/2018 0:11:11
Bjæverskov, Lidemarksvej	2/1/2018 0:11:00	2/1/2018 0:11:54	2/1/2018 0:11:00	2/1/2018 0:12:44
Sandet	2/1/2018 0:12:00	2/1/2018 0:13:52	2/1/2018 0:12:00	2/1/2018 0:14:00
Vindegårdsvej	2/1/2018 0:14:00	2/1/2018 0:15:34	2/1/2018 0:14:00	2/1/2018 0:15:43
Ud for Fyrrevej	2/1/2018 0:15:00	2/1/2018 0:16:07	2/1/2018 0:15:00	2/1/2018 0:16:14
Vemmedrup	2/1/2018 0:16:00	2/1/2018 0:16:44	2/1/2018 0:16:00	2/1/2018 0:16:55
Lellinge Kirke	2/1/2018 0:20:00	2/1/2018 0:19:55	2/1/2018 0:20:00	2/1/2018 0:20:04
Østergård	2/1/2018 0:22:00	2/1/2018 0:21:41	2/1/2018 0:22:00	2/1/2018 0:21:47
Krageskovbakken	2/1/2018 0:24:00	2/1/2018 0:22:31	2/1/2018 0:24:00	2/1/2018 0:22:38
Køge Gymnasium	2/1/2018 0:27:00	2/1/2018 0:24:00	2/1/2018 0:27:00	2/1/2018 0:24:09
Torvebyen	2/1/2018 0:28:00	2/1/2018 0:24:53	2/1/2018 0:28:00	2/1/2018 0:25:04
Fændediget	2/1/2018 0:29:00	2/1/2018 0:25:58	2/1/2018 0:29:00	2/1/2018 0:26:09
Køge st.	2/1/2018 0:31:00	2/1/2018 0:27:00	2/1/2018 0:31:00	NULL

TABLE 2: Performance of the web service in terms of throughput and response time according to the numbers of unique users and calls.

Unique users	Iterations	Throughput	Average (milliseconds)	Median (milliseconds)	Deviation (milliseconds)	Total calls made
50	20 each	49,147 calls/min	69	54	109	10,000
500	50 each	70,501 calls/min	137	84	357	25,000
1000	25 each	46,143 calls/min	383	105	1242	25,000
500	No limit for 3 mins	73,715 calls/min	395	108	1483	238,345
25,000	5	29,400 calls/min	895	523	2243	125,000

and identify delay patterns for given times, days, routes, and/or drivers. This is illustrated in Figure 4, where the route followed by bus “5676” starting at “18:56” is analysed for a given working week (Monday to Friday). Plotting of delays and early arrivals at all stops for five days reveals that, along that planned 39-minute journey, the bus reached certain stops up to 5 minutes before or after the scheduled times. Moreover, it shows that, at the start of its journey, the bus systematically accumulates delays, up to 5 minutes. Then,

after stop 8, when the bus leaves the busy city centre, the drivers are able to reduce that delay reaching the final stop either on time or even early. Investigation of delays associated with that journey at different times of the day confirms this pattern (data not shown). Such knowledge suggests that not only should the bus company review that route’s timetable to better reflect the time needed to arrive at the first stops, but also those unrealistic arrival times put drivers under pressure to recover that delay. Eventually, this

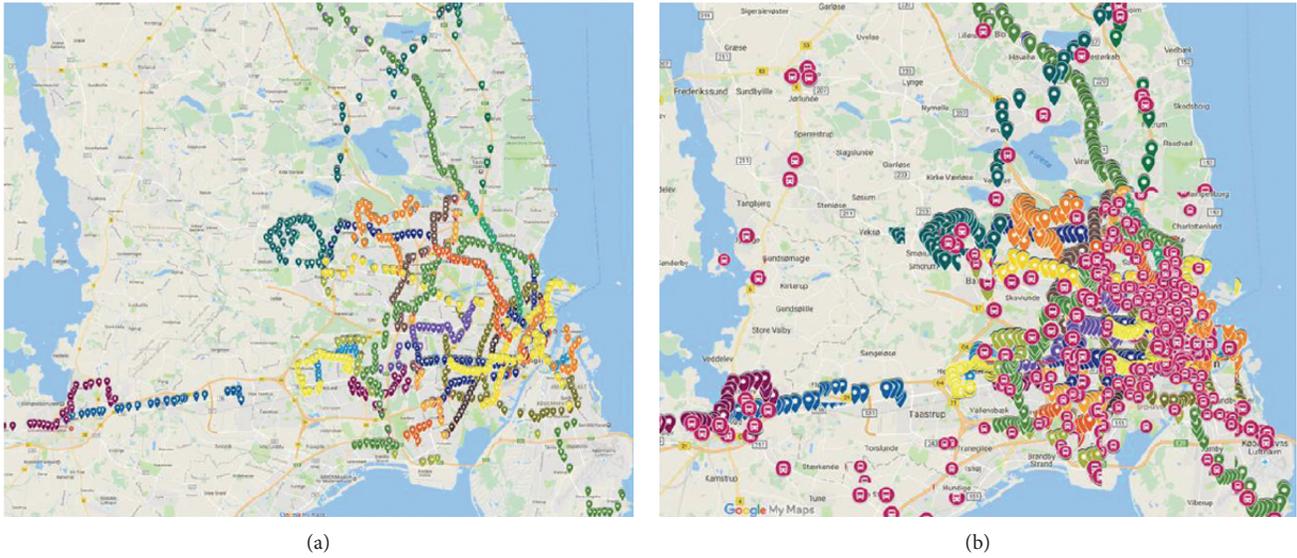


FIGURE 2: (a) Operational stops on 30 bus routes; (b) with additional mapping of live buses on the stops.

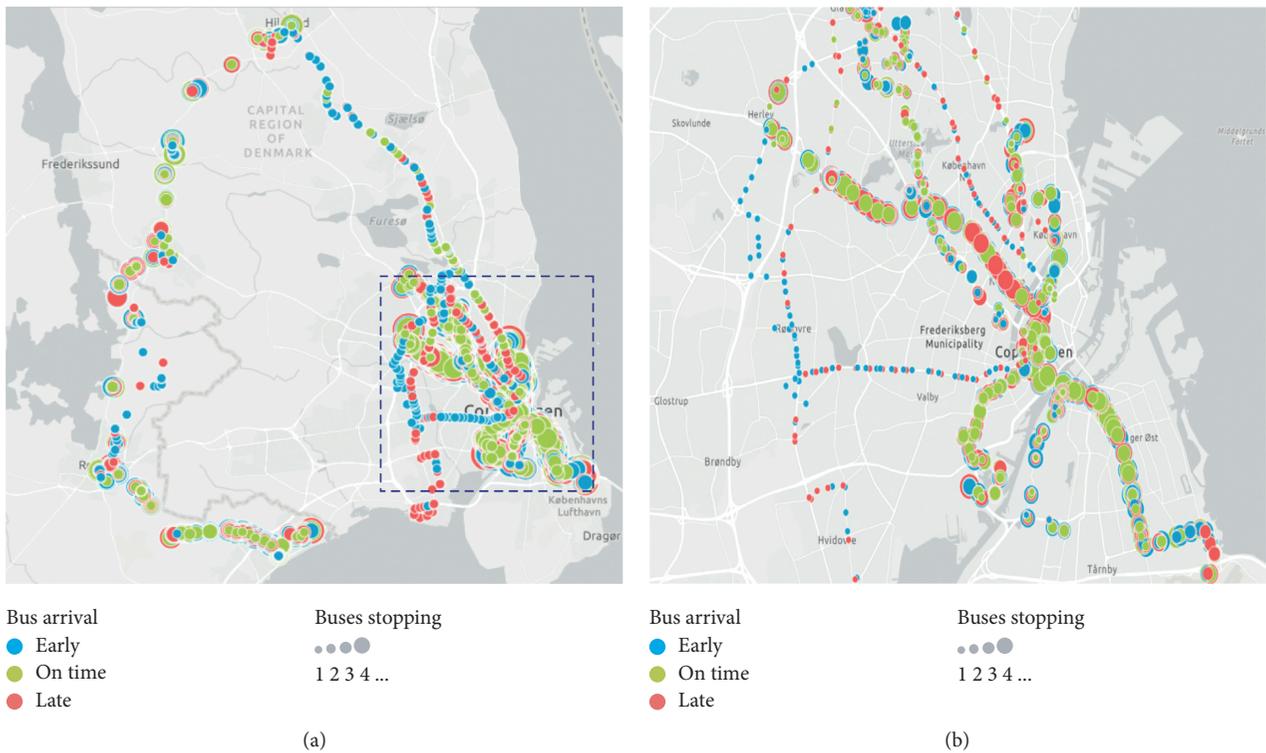


FIGURE 3: Real-time visualisation of route statuses, where the colour indicates the type of delay being reported; (a) status at all bus stops of the city of Copenhagen; (b) status at bus stops around the city centre.

too often leads to early arrival at the last bus stops, providing a suboptimal service to the users.

The ability to predict if a bus will arrive on time is critical to improve the service provided to users. Since the system gives access to both a large amount of historic data and current real-time data, machine-learning-based classifiers can be trained to inform users of future delays. As Neural Networks (NNs) have made remarkable

progress in the last few years [63, 64] and they have already been used with success in similar tasks [44], they are the technology of choice for building such classifier. Thus, the analytic component includes an implementation of a relatively simple NN, that is, a multilayer perceptron network with an input layer passing the input variables to two hidden layers and an output layer predicting the delay status [62].

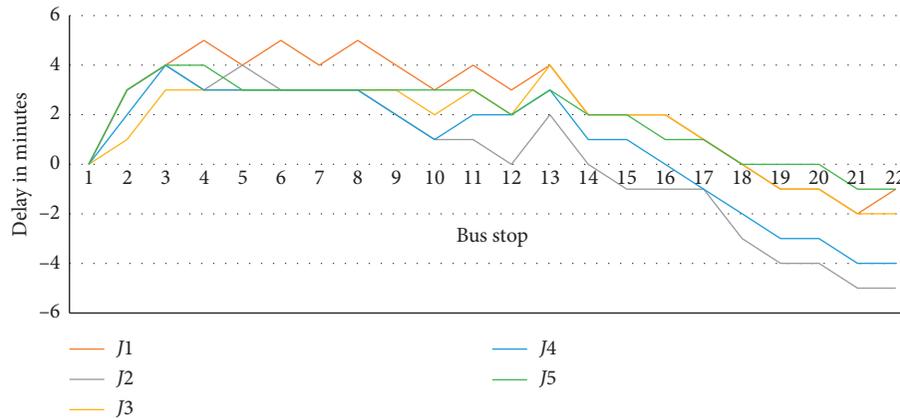


FIGURE 4: Delay at each bus stop associated with a journey performed by bus “5676” during a 5-day working week.

Since, intuitively, the time of the day, the delay history of the stop of interest, and the delay at the previous stop are expected to provide key information to predict the delay status at a given stop, only four input variables are considered: hour, minute, delay history, and delay duration at the last stop of the current journey. Focusing initially on a single route, 218,027 data points of past journeys completed at different times of the day were used to train the NN. More specifically, whereas the model was trained using all data captured between the 1st of January and the 31st of March 2018 (both working days and weekends), the model was evaluated using all data generated from the 1st of April to the 30th of April 2018. The experiment using those unseen data revealed that the NN-based classifier is able to perform accurate prediction—99.21% prediction probability—if a bus will be on time at the next bus stop.

Although such a classifier produces excellent results that provide value to users, it suffers from the fact that it operates like a “black box”: the logic used for decision-making cannot be accessed and, therefore, no new knowledge can be gained from such predictions [66]. To address this, an alternative machine-learning-based predictor was also implemented using the Expectation Maximization (EM) algorithm [67]. As it relies on producing consistent clusters, their composition and nature can provide useful insight into the particular features that lead to the integration of a sample to a group.

Using the same dataset and features as previously, 10 clusters were produced. While delivering lower accuracy than the NN-based classifier—91.73% prediction probability—results can still be used practically. In addition, as Figure 5 shows, profiles of the generated clusters are accessible, making them available for further analysis. First, if a new instance is classified as belonging to one of 6 clusters (2, 5, 6, 7, 8, and 9), this indicates unambiguously that a delay is expected. Second, while membership of clusters 1 and 10 suggests delays, that of clusters 3 and 4 implies that the bus will be on time. Interestingly, while information about delays is a major component of cluster 10, knowledge of the time of day has priority in all the other clusters. Remarkably, cluster 4 considers the minute more important than the hour. As it

is associated with the absence of delay, this may indicate that its membership is composed of stop times that are away from hourly peak traffic times. Obviously, additional data mining tools are required to extract more practical knowledge from those profiles.

Additional experiments using data extracted from 227 different routes while keeping the same experimental setup indicate that both approaches, that is, NN and EM, remain accurate in this more challenging setting. Indeed, prediction probabilities only decrease slightly: 98.89% and 88.77% for the NN- and EM-based predictors, respectively. Performances of those models are illustrated in Figure 6, where a lift chart [68]—a variation on the receiver operating characteristic (ROC) curve—plots the percentage of correct predictions according to a percentage of the overall stops. This chart confirms the excellent accuracy of the NN-based predictor.

5. Discussion and Comparison

The primary aim of this work is to develop a real-time predictive monitoring system for urban transport, which is based on a highly adaptable architectural framework for intelligent transport management, therefore addressing many of the issues associated with transport systems presented in Section 2. Transport is a core and vital service for modern economies and hence it is a natural progression to adopt service orientation philosophy for its management. Thus, one of the novel aspects of the developed architectural framework is integrating the service-oriented paradigm shift in information technology, together with the associated Quality of Service (QoS) factors that are necessary for the success of such complex systems. This is combined with latest advancement in data mining and machine learning, novel performance enhancing techniques, sensor and storage technologies including data gathering and transformation, and advanced visualisation techniques.

As presented in Section 2, a number of transport management systems, architectures, and approaches have been described in the literature. They can be classified into three categories: firstly, high-level frameworks/models and approaches, secondly, practical and special purpose systems,



FIGURE 5: Profiles of the 10 clusters generated by EM showing if they are associated with delay or not, as well as the attributes that are the most predominant for each of them.

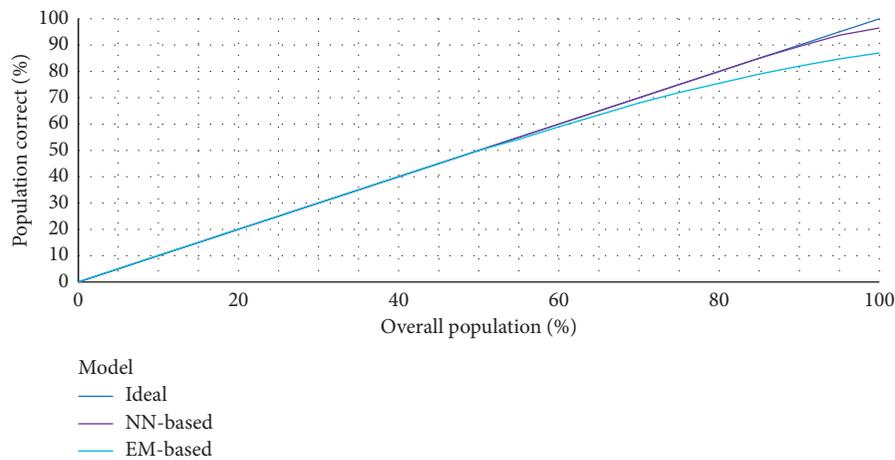


FIGURE 6: Lift chart comparing accuracies of the NN- and EM-based predictors against the ideal model.

and finally studies that are mainly focused on the use of machine learning algorithms and techniques.

Overall, there have been a number of high-level frameworks/architectures and approaches proposed in the literature. Some of those are highly comprehensive yet personalised using standards for a set of user services based on their individual requirements [51], while others are presented as a guide for the development of future systems standards [54]. However, as their main focus is on standards and requirement policies, they do not cover aspects of practical implementation and construction of a public transport system, which are fully addressed in the proposed framework. Another approach that is adopted in [28] is kept relatively simple so that it can be used as a base for developing more complex systems. Indeed, the absence of constraints-based architecture implementation makes it suitable to other applications. However, that approach is still at relatively high level as it does not present any recommendation or details regarding actual implementation or evaluation. On the other hand, those aspects are fully covered in our system as described in Sections 2.4.2 and 2.5.

Moreover, although a few practical architectures and systems have been proposed, they do not offer full solutions. An interesting piece of work was presented in [21] for a bus

management system that is based on the IETF presence model, which is a subscription and observer model. Although it reported a reasonable level of prediction, when using the historical model, there was no focus on the QoS attributes of the system, which are very important, particularly in terms of performance and scalability, when the number of buses increases in the network. In contrast, these issues have been evaluated extensively in the proposed system: the obtained results demonstrate very good performance and scalability factors (Table 2) and higher prediction accuracy (Figure 6). In addition, whereas the limited integration of data mining on the fly is not included in the architecture design of that previous work, this has also been addressed in the proposed system. Other works include the following: (i) the exploitation of a pillar-based architecture, which does not reflect the complexity of the implementation and construction of large transport management systems [18], and (ii) the use of mainly three layers, that is, sensors grouping for input, cloud storage, and client interface, following broadly a three-tier architectural approach [53] and, therefore, suffering from all its limitations particularly in terms of extensibility, reusability, and scalability. Alternatively, some researches have focused on back-end data-driven decision support systems rather than full implementation and evaluation [66]. Overall, unlike other proposed

systems, ours offers a fully integrated solution that is general, ensuring portability and adaptability, and is designed considering QoS factors.

Although a large number of papers have focused on the exploitation of data analytics and machine learning algorithms, ranging from clustering to the use of neural networks, they offer prediction results that are either comparable [43, 47] or inferior [39, 40] to the ones obtained by our fully integrated system (Figure 6). Indeed, many of the listed algorithms and models have been integrated in our system to form one of the core components of the proposed architecture. This allows different algorithms to be evaluated and eventually adopted within the self-contained analytics component, without affecting the rest of the system. Ultimately, the proposed approach covers the complete data-driven transport management application lifecycle, starting with data gathering and transformation, followed by going through analytics and visualisation in a modular adaptable architecture. Such holistic approach is extremely important as new data sources are continuously emerging, and their integration requires a new generic modular architecture.

6. Conclusions

This paper presents a data integration generic architecture that provides an implementation blueprint for developing intelligent public transport systems. The architecture follows service-oriented principles and meets the quality requirements for transport management systems with real-time monitoring and forecasting. The implementation and evaluation of the proposed system architecture were conducted on the large-scale bus network of a European capital city. The web services component of the system was benchmarked for performance for different numbers of concurrent users requesting the information through web services or pushing information into the system again through web services. Then, real-time data visualisation of the data stream was also presented to illustrate the performance achieved through implementation of smart analytics. Finally, results produced by the implemented data mining and machine learning models for real-time prediction were analysed in terms of their value for decision-making and accuracy.

The experimental evaluation of the system has demonstrated the possibility of developing a decision support system for complex urban transport management. The system is able to handle the heterogeneous data generated by thousands of buses going through tens of thousands of bus stops. Moreover, the production of informative visualisations and accurate predictions in real time allow improving the quality of service along scheduled routes. With the continuous development of sensor, smart analytics technologies, and the growing vehicle infrastructures, the plug-and-play and scalability characteristics of the architecture are essential to permit the integration of new components, while still delivering real-time performance. In particular, the adoption and customisation of additional domain-specific data mining and machine learning

techniques will deliver enhanced intelligence to support further decision-making.

Data Availability

Data are available upon request.

Disclosure

Some of the materials contained in this paper were submitted to the School of Computer Science & Mathematics, Kingston University, London, by Nauman Ahmad Khan in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science, August 2017 [69].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Mermaid Technology for giving them access to real-time data and relevant infrastructure. This work was made possible by a Ph.D. scholarship from Kingston University, London.

References

- [1] O. Amft and P. Lukowicz, "From backpacks to smartphones: past, present, and future of wearable computers," *IEEE Pervasive Computing*, vol. 8, no. 3, pp. 8–13, 2009.
- [2] R. George, A. J. Marsh, and S. Maglavera, "Enabling pervasive computing with smart phones," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 20–27, 2005.
- [3] O. Nuria and A. P. Pentland, "Graphical models for driver behavior recognition in a smartcar," in *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No.00TH8511)*, pp. 7–12, Dearborn, MI, USA, October 2000.
- [4] Y. Jaybal, C. Ramanathan, and A. Rajagopalan, "HDSAnalytics: a data analytics framework for heterogeneous data sources," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data—CoDS-COMAD '18*, pp. 11–19, Goa, India, January 2018.
- [5] U. Shafi, A. Safi, A. R. Shahid, S. Ziauddin, and M. Q. Saleem, "Vehicle remote health monitoring and prognostic maintenance system," *Journal of Advanced Transportation*, vol. 2018, Article ID 8061514, 10 pages, 2018.
- [6] M. Kamargianni, M. Matyas, W. Li, and J. Muscat, "Londoners' attitudes towards car-ownership and mobility-as-a-service: impact assessment and opportunities that lie ahead," MaaS Lab—UCL Energy Institute Report, University College London (UCL), London, UK, 2018.
- [7] K. Henriksen, I. Jadwiga, and A. Rakotonirainy, "Modeling context information in pervasive computing systems," in *Proceedings of the International Conference On Pervasive Computing*, Heidelberg, Germany, August 2002.
- [8] Y. Kumagai, T. Sakaguchi, O. Masayuki, and A. Motoyuki, "Prediction of driving behaviour through probabilistic inference," in *Proceedings of the 8th Conference on Engineering Applications of Neural Networks (EANN'03)*, Malaga, Spain, September 2003.

- [9] N. Risto and H. Summala, *Road-User Behaviour and Traffic Accidents*, North-Holland Publishing Company, Amsterdam, Netherlands, 1976.
- [10] Olsson and C. Magnus, "Taking the next step in context-aware applications," in *Proceedings of the IRIS26*, Helsinki, Finland, August 2003.
- [11] A. Rakotonirainy, T. Cutmore, T. Steele, and D. A. James, "An investigation into peripheral physiological markers that predict monotony," in *Proceedings of the Road Safety Research, Policing and Education Conference*, pp. 14–16, Perth, Australia, 2004.
- [12] S. Buthpitiya, F. Luqman, M. Griss, B. Xing, and A. K. Dey, "Hermes—a context-aware application development framework and toolkit for the mobile environment," in *Proceedings of the 2012 26th International Conference on Advanced Information Networking and Applications Workshops*, Fukuoka, Japan, March 2012.
- [13] K. N. Qureshi and H. Abdullah, "A survey on intelligent transportation systems," *Middle-East Journal of Scientific Research*, vol. 15, no. 5, pp. 629–642, 2013.
- [14] M. Garofalakis, J. Gehrke, and R. Rastogi, *Data Stream Management: Processing High-Speed Data Streams*, Springer, Berlin, Germany, 2016.
- [15] T. Li, C. Zeng, Y. Jiang et al., "Data-driven techniques in disaster information management," *ACM Computing Surveys*, vol. 50, no. 3, pp. 1–43, 2017.
- [16] J. Baek, Q. H. Vu, J. K. Liu, X. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid," *IEEE Transactions on Cloud Computing*, vol. 3, no. 2, pp. 233–244, 2015.
- [17] A. Gandami and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [18] H. Wieker, J. Vogt, and M. Fuenfroeken, "Intelligent transportation system infrastruct and software challenges," in *Automotive Systems and Software Engineering*, Y. Dajsuren and M. Van Den Brand, Eds., Springer, Cham, Switzerland, pp. 295–319, 2019.
- [19] S. E. Shladover, "Connected and automated vehicle systems: introduction and overview," *Journal of Intelligent Transportation Systems*, vol. 22, no. 3, pp. 190–200, 2018.
- [20] F. Camacho, C. Cárdenas, and D. Muñoz, "Emerging technologies and research challenges for intelligent transportation systems: 5G, HetNets, and SDN," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 12, no. 1, pp. 327–335, 2018.
- [21] D. E. Yong, *UniShuttle—a small-scale intelligent transport system in the connected mobility digital ecosystem*, Ph.D. thesis, University of Wollongong, Wollongong, Australia, 2015.
- [22] S. Maiti, A. Pal, A. Pal, T. Chattopadhyay, and A. Mukherjee, "Historical data based real time prediction of vehicle arrival time," in *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China, October 2014.
- [23] Q. Shi, L. Qixin, and Z. Weizhong, "Architecture analysis of common information platform for intelligent transportation systems (ITS) and its construction means," *Journal of Transportation Engineering and Information*, vol. 1, no. 1, pp. 41–47, 2003.
- [24] J. Crowcroft, I. Wakeman, Z. Wang, and D. Sirovica, "Is layering, harmful? (remote procedure call)," *IEEE Network*, vol. 6, no. 1, pp. 20–24, 1992.
- [25] N. Lathia, J. Froehlich, and L. Capra, "Mining public transport usage for personalised intelligent transport systems," in *Proceedings of the 2010 IEEE International Conference on Data Mining*, Sydney, Australia, December 2010.
- [26] A. Moreno, E. Onieva, A. Perallos, G. Iovino, and P. Fernández, "Cooperative decision-making its architecture based on distributed RSUs," in *Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 84–90, Cham, Switzerland, December 2015.
- [27] Z. Hass, "A protocol structure for high-speed communication over broadband ISDN," *IEEE Network*, vol. 5, no. 1, pp. 64–70, 1991.
- [28] AASHTO/ITE/ NEMA, *National Transportation Communications Protocol for ITS [NTCIP] 9001*, AASHTO/ITE/ NEMA, Washington, DC, USA, 2009.
- [29] N. A. Khan, V. Brujic-Okretic, and S. Khaddaj, "Intelligent integration framework for smart transport system," in *Proceedings of the 12th International Conference on Intelligent Environments (IE)*, pp. 76–79, London, UK, September 2016.
- [30] W. Alsaqaf, M. Daneva, and R. Wieringa, "Quality requirements in large-scale distributed agile projects—a systematic literature review," in *Proceedings of the International Working Conference on Requirements Engineering*, pp. 219–234, Cham, Switzerland, February 2017.
- [31] J. P. G. Sterbenz and G. M. Parulkar, "AXON: application-oriented lightweight transport protocol design," in *Proceedings 10th International Conference on Distributed Computing Systems*, Paris, France, June 1990.
- [32] ITS SA, *ITS Organizational Issues—The Road Ahead*, ITS SA, In press, 2000.
- [33] D. Advani, *How Melbourne Is Fast-Tracking Smart Transit Using Big Data*, HuffPost, New York, NY, USA, 2019, <http://www.huffingtonpost.com/>.
- [34] A. Abazorius, *MIT Big Data Initiative Launches Transportation Challenge, Privacy Working Group*, MIT Press, Cambridge, MA, USA, 2019, <http://web.mit.edu/newsoffice/>.
- [35] E. Deakin, "Intelligent transport system," *ACCESS Magazine*, vol. 1, no. 34, pp. 29–34, 2009.
- [36] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: from big data to big insights," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215–225, 2016.
- [37] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Information Systems*, vol. 64, pp. 266–280, 2017.
- [38] B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, "Bus travel time prediction using a time-space discretization approach," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 308–332, 2017.
- [39] B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, "Pattern-based time-discretized method for bus travel time prediction," *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 6, Article ID 0401701, 2017.
- [40] J. Ranhee and R. Rilett, "Bus arrival time prediction using artificial neural network model," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, Washington, WA, USA, October 2004.
- [41] S. Pongnumkul, T. Pechprasarn, N. Kunaseth, and K. Chaipah, "Improving arrival time prediction of Thailand's passenger trains using historical travel times," in *Proceedings of the 2014 11th International Joint Conference on IEEE Computer Science and Software Engineering (JCSSE)*, Chon Buri, Thailand, May 2014.

- [42] I. Bošnjak, *Intelligent Transportation Systems*, Faculty of Traffic Science, Zagreb, Croatia, 2005.
- [43] A. Agafonov and V. Myasnikov, "An adaptive algorithm for public transport arrival time prediction based on hierarchical regression," in *Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, Canary Islands, Spain, September 2015.
- [44] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [45] S. Maiti, A. Pal, A. Pal, T. Chattopadhyay, and A. Mukherjee, "Historical data based real time prediction of vehicle arrival time," in *Proceedings of the IEEE 17th International Conference on Intelligent Transportation Systems (ISTC)*, Qingdao, China, October 2014.
- [46] J. Liu, C. Gong, J. Li, J. Li, and X. Cui, "Social sensing enhanced time ruler for real-time bus service," in *Proceedings of the 2013 Ninth International Conference on Semantics, Knowledge and Grids*, Beijing, China, October 2013.
- [47] T. Liu, J. Ma, W. Guan, Y. Song, and H. Niu, "Bus arrival time prediction based on the k -nearest neighbor method," in *Proceedings of the IEEE 5th International Joint Conference on Computational Sciences and Optimization*, Harbin, China, June 2012.
- [48] J. Froehlich and J. Krumm, "Route prediction from trip observations," SAE International, Warrendale, PA, USA, SAE Technical Paper, no. 2008-01-0201, 2008.
- [49] Z. Bělinová, P. Bureš, and P. Jesty, "Intelligent transport system architecture different approaches and future trends," in *Advances in Intelligent and Soft Computing*, pp. 115–125, Springer, Berlin, Heidelberg, 2010.
- [50] M. Mochizuki, A. Suzuki, and T. Tajima, "UTMS System Architecture," in *Proceedings of the 6th World Congress on Intelligent Transport Systems*, Toronto, Canada, November 1999.
- [51] Frame (2004), European ITS Framework Architecture, <http://www.frame-online.net>, 2019.
- [52] S. Ahmed and S. S. Kanere, "SKVR: scalable knowledge-based routing architecture for public transport networks," in *Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks*, pp. 92–93, New York, NY, USA, 2006.
- [53] K. Ashokkumar, B. Sam, R. Arshadprabhu, and Britto, "Cloud based intelligent transport system," *Procedia Computer Science*, vol. 50, pp. 58–63, 2015.
- [54] ITS America, *ITS Architecture Documentation*, ITS America, Washington, DC, USA, 1996.
- [55] A. Patidar and U. Suman, "A survey on software architecture evaluation methods," in *Proceedings of the IEEE 2nd International Conference on Computing for Sustainable Global Development*, pp. 967–972, New Delhi, India, March 2015.
- [56] M. A. Javed, S. Zeadally, and E. B. Hamida, "Data analytics for cooperative intelligent transport systems," *Vehicular Communications*, vol. 15, pp. 63–72, 2019.
- [57] M. Kaur, S. Singh, and N. Aggarwal, "Study of NoSQL databases for storing and analyzing traffic data," in *Proceedings of the 2nd International Conference on Communication, Computing and Networking*, C. Krishna, M. Dutta, and R. Kumar, Eds., Springer, Berlin, Germany, pp. 577–587, 2019.
- [58] T. Neele, M. Rol, and J. F. Groote, "Verifying system-ide-properties of industrial componentased software," in *Fundamentals of Software Engineering. FSEN 2019*, H. Hojjat and M. Massink, Eds., Springer, Berlin, Germany, pp. 158–175, 2019..
- [59] M. Vargas-Santiago, L. Morales-Rosales, S. E. Pomares Hernández, and K. Drira, "Autonomic web services based on asynchronous checkpointing," *IEEE Access*, vol. 6, pp. 5538–5547, 2018.
- [60] D. Li, M. Dong, Y. Yuan, J. Chen, K. Ota, and Y. Tang, "SEER-MCache: a prefetchable memory object caching system for IoT real-time data processing," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3648–3660, 2018.
- [61] M. Samp, "A survey of machine learning algorithms and their application in information security," in *Guide to Vulnerability Analysis for Computer Networks and Systems*, S. Parkinson, A. Crampton, and R. Hill, Eds., Springer, Berlin, Germany, 2018.
- [62] J. Pang, J. Huang, Y. Du, H. Yu, Q. Huang, and B. Yin, "Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network," *IEEE Transaction on Intelligent Transportation*, vol. 20, no. 9, pp. 3283–3293, 2018.
- [63] M. Jiber, I. Lamouik, Y. Ali, and M. A. Sabri, "Traffic flow prediction using neural network," in *Proceedings of the International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Morocco, April 2018.
- [64] H. Nguyen, L.-M. Kieu, T. Wen, and C. Cai, "Deep learning methods in transportation domain: a review," *IET Intelligent Transport Systems*, vol. 12, no. 9, pp. 998–1004, 2018.
- [65] Mermaid Technology, A Leading Provider of Public Transport Solutions, <http://www.mermaid.dk>, 2019.
- [66] F. Sun, A. Dubey, J. White, and A. Gokhale, "Transit-hub: a smart public transportation decision support system with multi-timescale analytical services," *Cluster Computing*, vol. 22, no. S1, pp. 2239–2254, 2019.
- [67] Y. Y. Kim, W. Lee, and I. C. Moon, "Black-box expectation-maximization algorithm for estimating latent states of high-speed vehicles," in *Proceedings of the 2018 18th International Conference on Control, Automation and Systems (ICCAS)*, Daejeon, Republic of Korea, 2018.
- [68] I. H. Witten and E. Frank, *Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, Burlington, MA, USA, 2000.
- [69] N. A. Khan, *Real time predictive monitoring systems for urban transport*, Ph.D. thesis, Kingston University, London, UK, 2017.