Integrating Citizens' Avatars in Urban Digital Twins

Rafael García-Luque*, Lorenzo Toro-Gálvez, Nathalie Moreno, Javier Troya, Carlos Canal and Ernesto Pimentel

ITIS Software, Universidad de Málaga, Spain
E-mail: rafagarcialuque@uma.es; lorenzotoro@uma.es; nmv@uma.es; jtroya@uma.es; carloscanal@uma.es; epimentel@uma.es
**Corresponding Author*

Received 10 August 2023; Accepted 02 October 2023; Publication 23 December 2023

Abstract

Urban Digital Twins (UDT) represent a powerful tool to effectively make cities smart. Generally, UDTs are linked with other Digital Twins to build ecosystems where the citizen is at the heart of the ecosystem. This is why citizens should be considered first-class entities in the UDTs. At the same time, citizens' privacy cannot be compromised. In this paper, we propose to integrate citizens' digital twin in UDTs through their digital avatars (DAs). DAs allow exploiting citizens' information, behavioral habits and personal preferences, while allowing them to have full control of their own data. We present our architecture that makes use of the Cloud-to-Thing Continuum to optimize available processing resources. We focus on a case study of the public transport service of the city of Malaga (Spain) and describe how we are approaching its implementation. Finally, we validate such UDT implementation through tests that evaluate its accuracy and the factors affecting its performance.

Keywords: Urban digital twin, digital avatar and cloud-to-thing continuum.

Journal of Web Engineering, Vol. 22_6, 913–938.
doi: 10.13052/jwe1540-9589.2264

1 Introduction

Smart cities use technology to offer solutions to global problems such as climate change or resource depletion, and aim to provide useful services to their citizens and to solve specific urban problems, such as transportation and accessibility. The concept of Digital Twin City [6], also known as Urban Digital Twin (UDT), has been defined as a way to effectively make cities smart. A UDT is a digital twin capable of modeling specific city aspects such as transportation, heat maps of population density or environmental factors. Similar to other digital twins, UDTs facilitate two-way feedback between the model and the physical entities represented in it. These digital twins cannot only enable real-time remote monitoring for cities, but also suggest adaptation policies and enhance decision-making processes in areas such as improved urban governance, smart healthcare or smart transportation.

Over the last few years, the interest in the social aspects of smart cities is growing fast. For this reason, citizens must be considered as first-class entities of the UDT, since they are the fundamental key to this ecosystem. However, the role of people has often been ignored, treating them as a crowd and not as individuals. And if citizens are to be considered as individuals, then their privacy would be compromised [2]. In previous works [3,4,15,16], we presented the concept of *Digital Avatar (DA)*, which represents a virtual representation of its user and through which the latter can decide which personal information to share. The DA resides in the smartphone of its user.

In this paper we propose to integrate citizens through their DAs into UDTs. The idea is to exploit citizens' information, behavioral habits and personal preferences, at the same time allowing them to have full control of their own data, including access control and storage location. Citizens who share specific information can benefit from better services from the organizations to which they entrust their data.

In addition to addressing the integration of citizens into UDTs, our proposal also aims to face the challenge of managing the large number of devices involved in these systems, that can require high computing capacity. To this end, we have developed a framework which takes advantage of the Cloud-to-Thing Continuum [11–13] for optimizing the available processing resources. For this, we propose a distributed architecture based on four layers, namely *mist*, *edge*, *fog* and *cloud*, and define the devices and services to be placed in each layer by focusing on a case study of the public transportation service of the city of Malaga (Spain). We also present the technology stack with which we plan to implement our framework, and give a glimpse of the implementation centered on digital twins.

This work is an extension of [17]. The remainder of this paper is structured as follows. Section 2 gives an insight into the basics needed to understand our approach. Then, Section 3 presents a case study of a UDT dealing with public transportation that motivates our proposal. Sections 4 and 5 present the architecture of our framework and the implementation centered on digital twins that we are carrying out, respectively. The performance and accuracy of our UDT is discussed in Section 6. Next, Section 7 analyzes how the citizen relates to the UDT through its own digital twin. Finally, Section 8 concludes the paper with an insight to future work.

2 Background

2.1 Urban Digital Twins

A Digital Twin (DT) is a comprehensive digital representation of a real system, service or product (the Physical Twin, PT), synchronized at a specified frequency and fidelity [1]. The DT includes the properties, conditions, and behavior of the physical entity through models and data, and is continuously updated with real-time data about the PT's performance, maintenance, and health status throughout its entire lifetime [5, 9].

One specific type of DT is the Urban Digital Twin (UDT), which is a virtual representation of a city's physical assets that utilizes data, analytics, and machine learning to create real-time, adaptable simulation models. The digital twin describes the reality (and the history of it), while it is the additional applications that bring the real intelligence and help create the common picture of reality that is the added value of an urban digital twin. UDTs enable more informed decision-making, participatory governance, and improved urban planning by offering a risk-free testing environment for long-term predictions and impact assessments.

2.2 Digital Avatars

A Digital Avatar (DA) [4, 15] is a virtual entity residing in an individual's smartphone or tablet that records information about its owner and offers different services for interacting with their surroundings and other users' DAs. This concept arises from the People as a Service (PeaaS) [10] paradigm, which provides a conceptual model for application development focused on the smartphone as a representative and interface to its owner. The main purpose of a DA is to serve as a digital representation of a person, facilitating their participation in collaborative social computing applications. In this context, the DA of a user is designed to collect data about the user and their habits, integrate this information with external sensor data or open data sources, and interact with both the environment and other users' DAs.

Additionally, a DA is responsible for representing the user in responding to requests coming from other avatars or social computing applications, as well as adjusting its records and anticipated behavior as needed. Consequently, DAs empower users by enabling them to (i) take control of the information and contents they create, (ii) manage how all that information is accessed and exploited in a secure manner by third parties, and (iii) be in control of their changes and adaptations. Thus, privacy and security can be ensured [14], since users can have the control over their own data and decide whether or not they want to share it, and with whom. However, there are limitations to the capabilities of DAs, such as lack of support for global decision making and comprehensive planning, which are key elements of any smart city application.

3 Case Study

The case study we are dealing with is urban transportation in smart cities, specifically in the city of Malaga (Spain). We have access to real-time information about the use of buses in the city: transport lines, stops and schedules, GPS position of the buses, traffic status, among others, provided by the City Council as open data.[1]

Our main goal is to develop an urban digital twin with the ability to use all the information about the transportation network to make predictions about peak occupancy hours, recognize usage patterns and allocate the resources for optimal use. The digital twin can also monitor the current state of transport and react or adapt to unexpected peaks or incidents.

Furthermore, the UDT would be improved if it was connected to the citizens' DAs. Thus, one of our goals is also to infer information from them: how individuals use the transportation system, their habits and routines in relation to it, or their behavioral patterns for commuting to work or moving around the city. This will allow for more accurate predictions made on individuals. In this manner, citizens may get improved and personalized services.

Another benefit of DAs is that personal preferences could be taken into consideration to suggest alternative transportation methods more suited to

[1] https://datosabiertos.malaga.eu/group/transporte

citizens' likings and needs. Knowing their preferences could be very useful in deciding about the overall transport solutions to be implemented.

4 Architecture

To carry out the above case study, we rely on the Cloud-to-Thing Continuum [11, 18]. This concept refers to a highly distributed, decentralized, and dynamic environment spanning from IoT devices to the cloud. The continuum helps us to solve the problem of computing capacity since it is designed to address the challenges of next-generation IoT systems that involve a large number of heterogeneous devices generating massive amounts of data and require real-time processing, low service response times, and enhanced reliability and security.

Following the continuum model, we propose a four-tier architecture where each layer offloads the upper layer by taking over some of its functionalities. Next, we present the four layers starting from the lowest layer of the architecture: mist, edge, fog and cloud computing [13]. A snapshot of our architecture is shown in Figure 1.

Mist represents the layer closest to the citizens, and we place here the citizens' smartphones that include their DAs. Therefore, all citizens' data

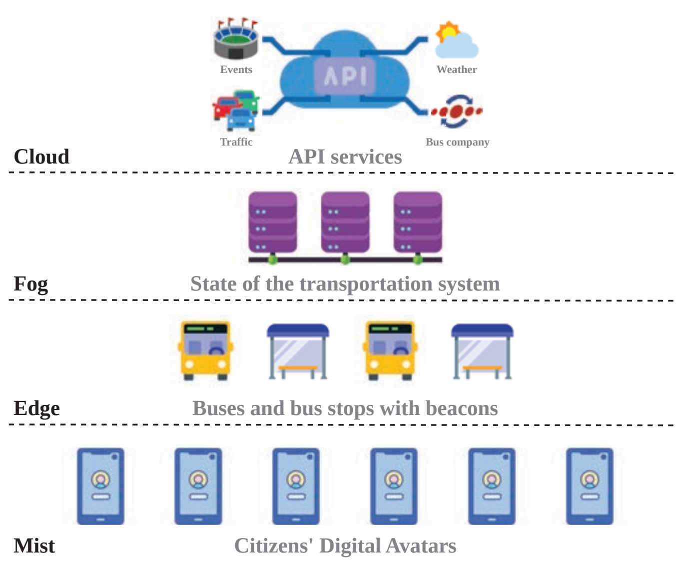


Figure 1 Proposed architecture for the transportation system based on the continuum.

coming from their DAs is collected at this level: positioning, information on their habits and routines, personal preferences, etc., always ensuring their privacy and data anonymization and allowing users to decide with whom they share them. Recall that citizens' information resides in their smartphones, from where they decide what to share and with whom [15]. *Mist* is also the layer through which citizens receive individualized and personalized information, coming from the processing in the upper layers, based on their virtual profiles.

The next layer is *edge* computing. This layer processes data directly coming from the mist as well as data gathered in this layer. Here, we consider that the bus stops and buses can gather and process information depending on the environment and different situations. Thus, this architecture makes it possible to run specific applications in a fixed location providing direct data transmission. Specifically, the inclusion of tracking and processing devices such as Bluetooth beacons in the different buses or bus stops makes it possible to send notifications to citizens and provide significant information, e.g., about the bus being late, overcrowded, or skipping a particular stop.

The third layer is *fog* computing, which enables ubiquitous access to a shared range of scalable computing resources. In the context of the proposed scenario, the *fog* layer will include the state of the transportation system including relevant factors organized in city areas. By using fog computing to process and analyze real-time data from the transportation system and citizens, we can gain valuable insights into factors such as traffic congestion, passenger load and route optimization, all of which can help improve the overall efficiency and reliability of the transportation system. Furthermore, this layer is essential for making predictions. By analyzing real-time and historical data, we can predict transportation costs, identify potential traffic bottlenecks and crowded areas, forecast peak demand times, and recognize popular routes and transit hubs.

The *cloud* layer, as the top tier of our proposed architecture, provides access to a range of services and real-time data that can have a direct effect in both the current and future state of the transportation system. This information is directly translated to the fog layer for analysis. Examples of services that can have an effect in the state of the transportation system include weather information and forecast, events taking place in the city (sport games, concerts or festivals) and real-time traffic information – for instance, the TomTom API offers information on the current traffic status. All these services offer information that can be crucial for optimizing transportation routes and reducing travel time for commuters.

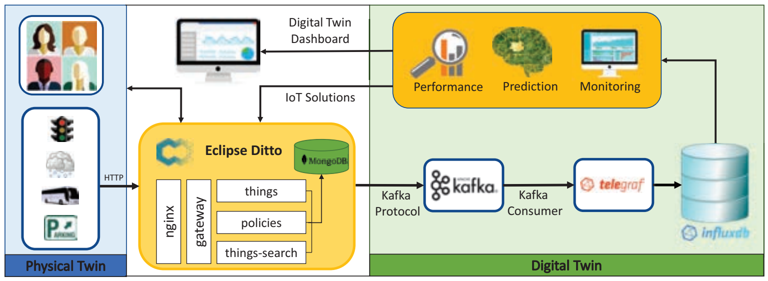


Figure 2 An overview of the Digital Twin Architecture composed of microservices.

Note that some information on the urban public transportation system is also accessible from the cloud layer, such as the route or position of the different buses. Information coming from this layer can be used to alert citizens about general incidents in the transportation system, such as accidents or cancellations, non-operational lines on special days, etc. Recall that other information related to transportation comes from the mist layer (specifically from the location of citizens' digital avatars) or from the edge layer (thanks to the beacons mentioned above).

5 Implementation

For materializing the architecture described before, we employ Eclipse Ditto[2], an open-source domain-agnostic framework that helps build digital twins. Figure 2 shows our architecture from a digital twins point of view. It is composed of independent microservices that communicate with one another in real time, using protocols such as AMQP or Apache Kafka. Each of these services is encapsulated in containers managed by Docker, ensuring isolation for both, which in turn enables portability and guarantees correct execution.[3]

The physical twin in our system is comprised of real-world entities providing information to the digital twin, with the Eclipse Ditto platform situated between them. The physical twin is scalable, since more entities can be added, and so is the digital twin. We also include the citizens, through their digital avatars, as part of the physical twin, since they are an integral part in any smart city.

Information is gathered continuously from the physical twin. For this, we have implemented a Java application, using Maven, that makes requests

[2]https://www.eclipse.org/ditto/intro-overview.html

[3]The source code of our UDT is available for donwload from our github repository [8]

to the different APIs and collects the information through Eclipse Ditto. However, Ditto is only responsible for storing and updating the most recent state of each entity in its database, implemented in MongoDB, which prevents having historical records. For this reason, we use, as part of the digital twin, InfluxDB,[4] an open-source database that enables fast storage and retrieval of time series data, with high availability. Since Ditto requires an intermediary in the connection to InfluxDB, we use Telegraf,[5] which plays the role of a server-based agent responsible for collecting and sending metrics to databases, IoT systems, etc.

The information stored as time series data can be used for different purposes, such as obtaining performance insights, monitoring and doing predictions. All these can be used to improve or correct the service offered by the different entities of the physical twin. Some of this information can also be used to improve and create easy-to-understand recommendations to citizens through their digital avatars, such as suggesting alternative routes when commuting to work or avoiding traffic congestion in rush hours.

As for the data we collect from the services of the physical entities, we are making periodic requests to the APIs of (i) Malaga City Council, (ii) OpenWeatherMap, and (iii) Tomtom. This means we are storing information about real-time buses location, parking occupancy, urban traffic and meteorological parameters of the city of Malaga. For instance, regarding information from public buses, our framework has currently collected around 79k documents containing the geolocation of all buses in each time interval (requests are made every minute). This data-base knowledge allows to set the location of each bus with respect to its established stop and, thus, estimate the duration of all the routes.

The combination of the times taken by the buses, together with the climatic aspects, parking occupancy and traffic congestion, will pave the way for establishing patterns that infer the performance of buses and, based on this, adapt it to the preferences and needs of the citizens.

6 Validation

Validation is fundamental in the construction of digital twins. Through this process, the quality, accuracy and consistency of the replicated system is verified in comparison with the original system. A validated digital twin

[4]https://www.influxdata.com/products/influxdb-overview/
[5]https://www.influxdata.com/time-series-platform/telegraf/

can perform simulations whose results are reliable and consistent with the behavior of the real system.

In the specific case of the digital twin presented in the paper, different kinds of tests were carried out. Some of them consisted in verifying the accuracy of the arrival predictions of the digital twin and their agreement with actual bus arrival data. In addition, comparative analyses of bus service performance on weekdays and holidays, as well as in dry and rainy weather conditions, were carried out. These tests allowed us to evaluate the performance and consistency of the UDT in different scenarios and conditions, not only ensuring that the results obtained were reliable and useful in average, but also demonstrating the potential of such a tool to analyse and optimise the physical system under particular conditions.

6.1 Accuracy of Arrival Predictions

The accuracy of our digital twin has been validated by analyzing its performance for different bus lines and destination stops. Both bus lines and destination stops were chosen randomly and prior to the data analysis in order not to influence the results.

The graphs shown in Figures 3 to 6 describe aspects related to the predictive behavior of the UDT that help us to better understand the real

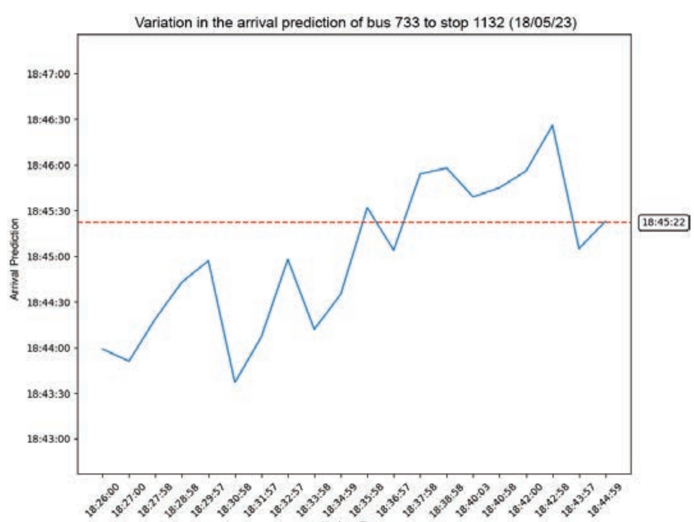


Figure 3 Bus 733 arriving at bus stop 1132.

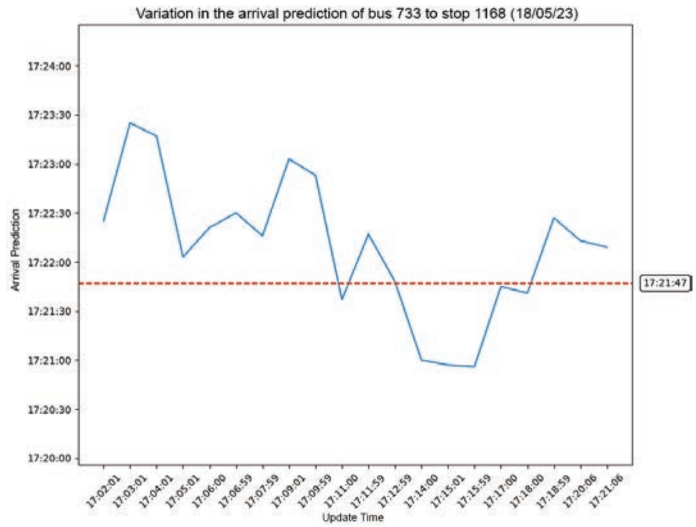


Figure 4 Bus 733 arriving at bus stop 1168.

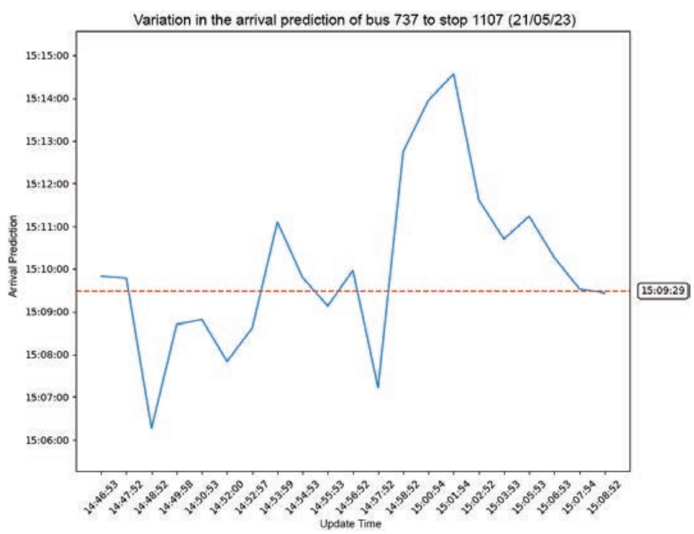


Figure 5 Bus 737 arriving at bus stop 1107.

system. In particular, we studied how the predictions fluctuate along the timeline. The data represented are updated every minute along with the time the UDT predicts that a given bus will arrive at stop 'X'. These predictions are shown on the graph as a blue line on the Y-axis. In red, on the other hand, we

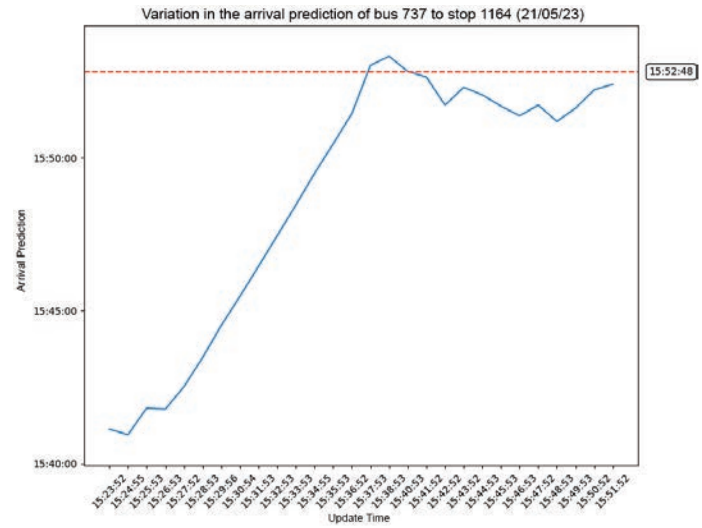


Figure 6 Bus 737 arriving at bus stop 1164.

can observe the actual arrival time at the stop according to the data published openly by the city council of Malaga. When the blue line is over the red line, it means that the prediction expects the bus to take longer than it finally did; when it is below the red line it means that the prediction expects the bus to arrive faster than it finally did.

As shown in Figures 3 and 4, the predictions start approximately twenty minutes before the arrival time of the bus at the target stop. With two minutes difference (above or below the actual time), the UDT begins its sequence of calibrations.

The arrival time predictions of the digital twin fluctuate until they converge to the actual arrival time with a margin of error of less than half a minute. The changes shown in the last 9 minutes in both graphs suggest a correction by the driver in the way of driving, trying to attain scheduled arrival times for the bus line: (a) accelerating in the case of Figure 3 and (b) moving slowler in the case of Figure 4.

Another significant example can be found in Figure 5, where the UDT predictions shoot up significantly with respect to the actual value for a small time interval. This type of behavior can be justified by the presence of traffic jams in the city during the bus route at peak hours.

Arrival to the last stop of the line and changes in the direction of travel of the bus, or exchanging bus drivers also have an important impact on the behavior of our digital twin. In particular, they affect the predictions made for stops located downstream of such a change. Figure 6 shows an example in which we can see how for a period of about 10 minutes the bus has been stopped. It is around 15:35 when it resumes and the predictions again converge accurately to the actual value.

In summary, although the speed of buses within the city is usually relatively constant in average, it is not easy to predict when a given bus will arrive at a particular stop. In addition to the speed, other factors such as traffic flow, the presence of traffic lights, the number of citizens using the service that day, or changes in direction have an influence. Our digital twin currently detects these anomalies although it is not yet able to use this information for future predictions.

6.2 Rainfall Variations

It seems reasonable to think that rain will significantly affect the average time it takes for a bus to travel through the stops – especially in cities like Malaga, where there is no rain most of the year. We have tried to verify this hypothesis with the data collected during the time period analyzed. Although it does not often rain in Malaga, at least on Thursday, May 18, it did heavily, especially between 2 PM and 3 PM. This allowed us to compare the measurements of that day with the average data obtained on other Thursdays of previous weeks when it did not rain during the same time period.

In order to calculate these average arrival times, we collected the actual GPS positions of the buses of line no. 11 every minute, over a period of four months, from February to June 2023. Line 11 serves the whole city from East to West and back spanning over 12 kms and more than 50 bus stops.

The red line in Figure 7 represents the mean time to arrival on Thursdays between 2 PM and 3 PM over this four-month period, in percentage (i.e. it reads always 100% no matter the distance and travel time between bus stops), while the blue line represents the mean time to arrival during the raining time (as a percentage of the mean time to arrival for the same day of the week and hour over the four months studied). Again, when the blue line is over the red line, it means that the bus took longer than average to arrive at the stop; while when it is below the red line it means that the bus arrived faster than the mean time.

Figure 7 directly relates the mean time between stops for Thursdays without rain (in red), versus the values obtained on May 18 (in blue) when it rained. Using a percentage scale, where 100% corresponds to the times

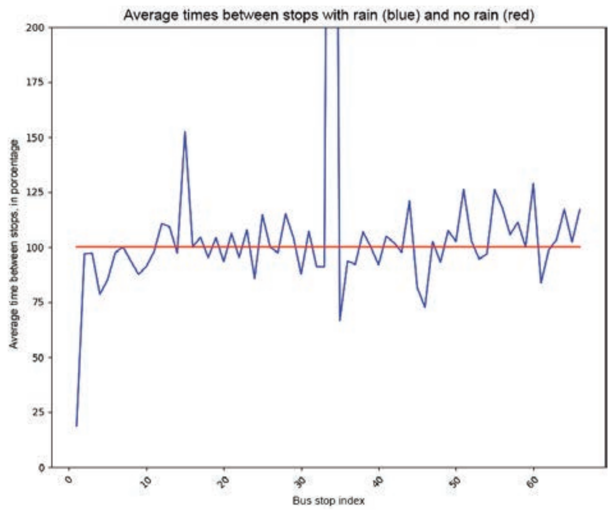


Figure 7 Mean times for dry and rainy times, as a percentage over dry time.

obtained in dry conditions, it can be seen how the average times in rainy conditions fluctuate between 20% and 25% more than the times in dry conditions. These variations are not as significant as might be expected in principle. This leads us to think that the bus lines contemplate the presence of possible climatic anomalies within their time schedules.

Two relevant aspects should also be highlighted in this graph. On the one hand, at the first stop there is a large significant variation with respect to the values collected on days without rain. This notable discrepancy is due to two factors: (a) the absence of previous data to help the system to calibrate and (b) the impact of the change of direction in which the bus travels through the stops on the estimated times. On the other hand, the average times collected at stops with indexes 15 and 35 are also significant. In both cases these times increase by more than 50%. Both indexes correspond to stops that are either very central (in the first case) or take place after the bus has changed direction (second case).

Based on the above results, we cannot conclude that rain has such a significant impact as expected in our initial hypothesis. However, we have also analysed the standard deviations as a measure to evaluate the "stability" of the averages obtained. In this graph (see Figure 8), the influence of rain on bus travel times is more evident. Although at some stops the standard deviation does not differ so much from the values recorded during dry days, in some cases it is more than 70% higher than the values without rain. This

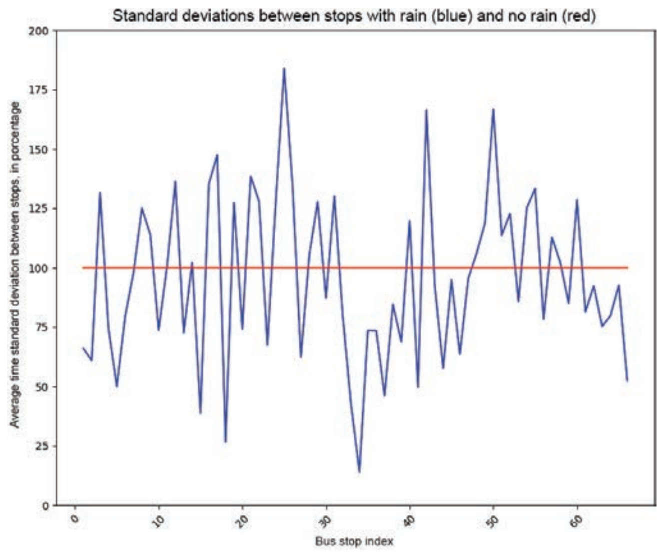


Figure 8 Standard deviations of mean dry and rainy times as a percentage of dry weather.

occurs, in particular, at the points which correspond to enclaves within the city where traffic congestion is frequent (consequently even more on rainy days).

In conclusion, it is possible that we did not find more significant variations in the data because the congestion generated by traffic during the rain did not affect the bus route as much. Obviously, for the conclusions to be more robust, it would be necessary to conduct a more thorough comparison of multiple rainy days with multiple dry days. This would allow us to more accurately analyze the impact of weather conditions on traffic and assess whether there are recurring patterns on different rainy days. However, we believe our digital twin implementation provides the means to perform such kinds of analyses.

6.3 Variations Between Working Days and Weekends

Traffic and the way of life in cities differ considerably on weekends compared to weekdays. We have performed a comparative analysis of average stop times between these two scenarios. The data collected corresponds to April 2023. Specifically, from April 11 until the end of that month. The days prior to April 11 took place in Malaga the Easter celebrations and the analysed bus line suffered significant alterations in its route through the city centre, so we have discarded these data.

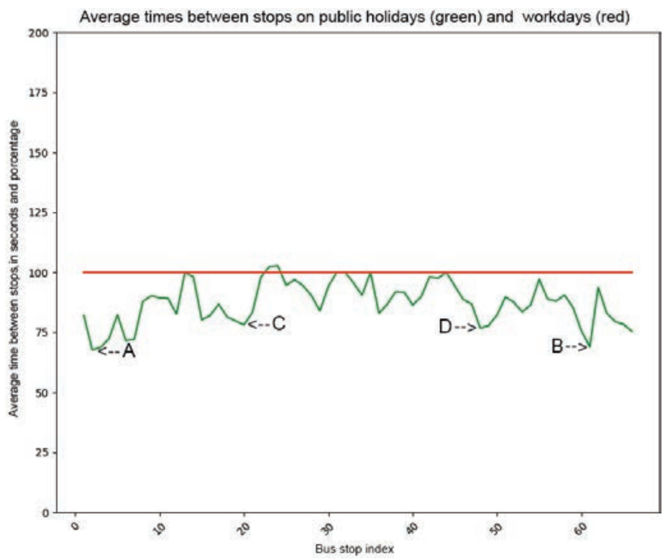


Figure 9 Average times on weekends and working days.

Figure 9 shows the average times between stops for both types of scenarios. The red line represents weekdays, while the green line represents holidays and weekends. The values are scaled with respect to working days, with the latter always having a value of 100%. According to the graph, the mean travel times between stops on holidays are consistently lower than the times on weekdays. This difference can be attributed to the higher traffic flow generated by people commuting to and from their workplaces. This hypothesis is further reinforced by the following observations. The stops with the greatest variation between holidays and working days are mostly found in the University area, since there are no lessons on weekends. These are points A and B on the graph. Similarly, points C and D correspond to stops where the number of offices is relatively high, so that many people congregate on weekdays.

As in the previous case, we also analysed the standard deviations of the mean time between stops (see Figure 10). In the graph, we can see how the lower index stops leading into the city show a smaller deviation on holidays, which translates into greater consistency in bus schedules. However, more centrally located stops (e.g., 15 or 20) show very significant deviations on holidays. This suggests that there may be specific factors to consider at those locations that affect bus punctuality during holidays.

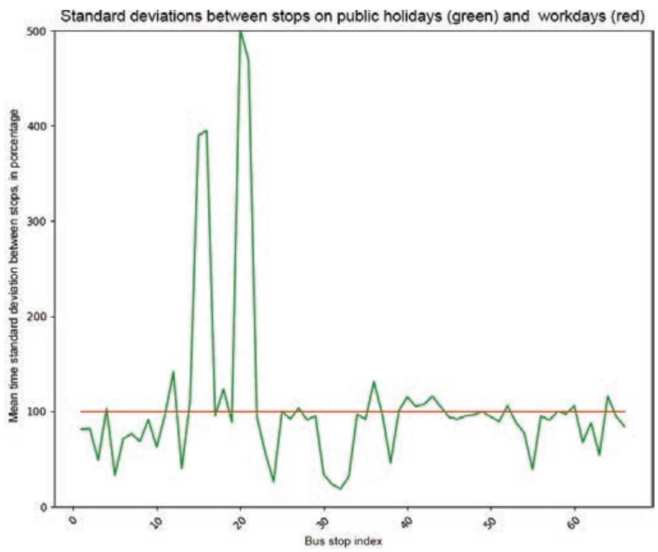


Figure 10 Standard deviations of average times on holidays and weekdays.

Finally, although there are other stops with variations in deviations, these differences are not as significant as those mentioned above. These results suggest that the influence of the flow of people commuting to work on weekdays has a general effect on bus arrival times. It would be interesting to carry out a study detecting these peak hours on weekdays, which would help to improve the efficiency of the service by avoiding crowds of users.

6.4 Patterns Identified

Based on the analyses performed, certain behavioral patterns can be extracted that help us to better understand both the functioning of the digital twin and the physical system to which it refers. Examining the graphs, it can be seen that predictions about the arrival time of a bus at a bus stop, made more than 10 minutes before the final time, tend to generate values with lower accuracy. In contrast, predictions made when the bus is less than 10 minutes away from the stop tend to converge to the actual values. In particular, such predictions tend to be very accurate in the last 5 minutes.

Directional changes in the route taken by the bus also affect the accuracy of the digital twin predictions. When traveling stops in reverse direction, buses remain stopped at the last stop for a longer time than usual, and the prediction algorithm does not take this variable into account, requiring some time to recalibrate itself. During this adjustment period, the predictions may be delayed, which is reflected in the variations of the times between stops.

The digital twin also allows to infer patterns about the behavior or driving mode of bus drivers. Analyzing the data collected, it is observed that drivers usually drive more slowly at the beginning of a trip. This leisurely driving usually changes to a more accelerated mode when approximately halfway through the trip is reached. It is then that drivers have to accelerate to fit their schedule, thus generating sudden changes that affect the predictions.

The aforementioned behavioral patterns provide further insight into the variations that take place in the times between stops. These times are additionally affected by other factors such as the specific time of day, the current stretch, discrepancies in forecasts and weather conditions. While further analysis of possible variations between the behavior of the digital twin and the actual physical system is required, these tests validate the effectiveness of the digital twin in predicting and understanding bus arrival times.

7 Citizens' Digital Twins

Although digital twins are an emerging technology, their deployment in different domains is occurring at an accelerating pace. Their exploitation in the context of smart cities is opening up new challenges among which the creation of 'Digital Twin Ecosystems" (DTEs) stands out. This is the creation of multiple digital twins interconnected with each other in a natural way. To extract value from these interconnections, digital twins need to combine/share resources and information. This results in insight and analytics that go beyond the capabilities of the individual components.

Adopting an ecosystem approach in the case study at hand implies modeling and analyzing the behavior of the citizens who regularly use the bus lines of the city transportation system. In other words, building digital twins associated with each of the users and feeding both kinds of twins – transportation system and citizens – with data generated by their counterparts.

For this purpose, an Android application was implemented whose central core is the citizen's digital avatar. The latter is stored in the user smartphone in a small local database using Couchbase Lite technology. The basic structure of the digital avatar is composed of both common personal information (user name, age, phone number, email, etc.) and data related to how the citizen uses the transportation system, in particular the most frequented bus stops, as well as the trips made (initial and final stop, start and end time, day of the week, bus line, set of GPS positions in the minutes after the bus trip). All this is

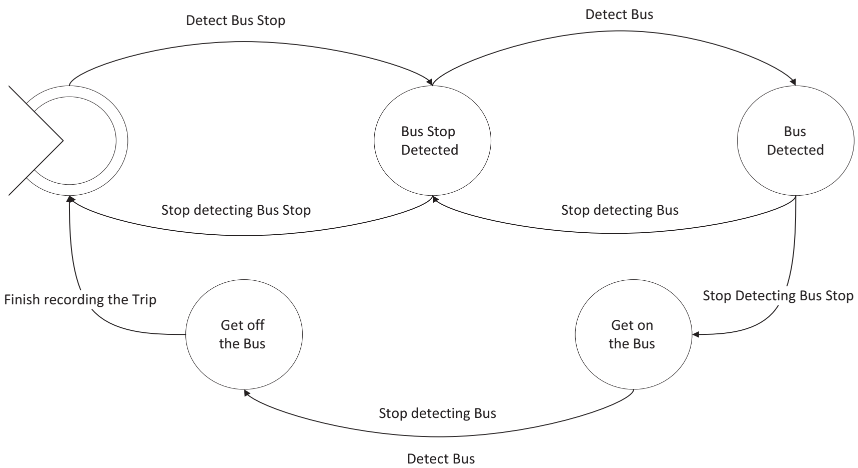


Figure 11 State diagram for user detection at bus stops and buses.

stored in a JSON document in which the privacy level for reading and writing of the each indicated field is recorded.

In order for the application to run successfully, the mobile device must support GPS location and Bluetooth connectivity. In particular, the application makes use of the Bluetooth signal to detect the beacons deployed by the city council both at bus stops and inside the buses. Thus, when a citizen approaches a bus stop, the mobile application implemented detects the Bluetooth signal identifying the stop, and it generates a new record in the digital avatar, marking a potential bus trip. This record contains the geographical coordinates of the stop as the starting point of the journey. Similarly, the beacon located inside the bus allows the system to determine when a user has actually taken the bus. This occurs when the application no longer detects the bus stop beacon but the one inside the bus. Finally, when a new bus stop beacon is again detected, but not the bus beacon anymore, the system infers that the user has gotten off the bus, ending the trip, and it also records this information in the digital avatar.

Figure 11 summarises the implemented algorithm. Note how the stop at the end of one trip can be the starting point of a new trip, if the user changes from a bus line to another one. This way, the citizen's use of the transportation system is recorded, from which usage patterns (usual times, days of the week, etc.) can be inferred.

The application, called uMuv, has an interactive map that reflects in real time the location of buses and bus stops respectively (see Figure 12).

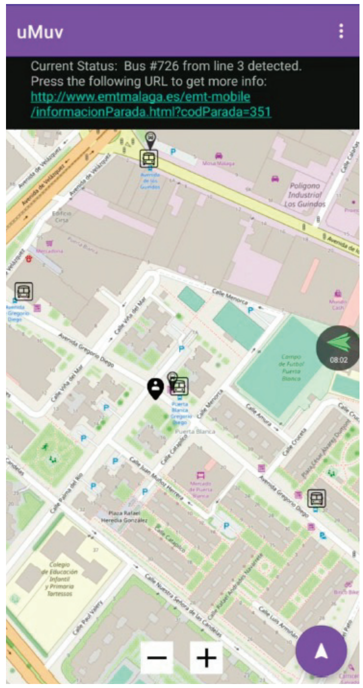


Figure 12 uMuv Interactive map.

The data represented in this map comes from the open information published by the City Council of Malaga. Each of the stops made by the bus constitutes an intermediate point of the route that is finally registered in the digital avatar. Moreover, uMuv has an additional functionality. When a user's trip ends, during a few minutes the GPS position where the citizen is located is saved. This utility can be used by the responsible entity concerned (i.e. the local municipality) to carry out studies and/or statistics on the routes taken by users *after* they leave the bus. As a result of such a study, the need for an additional stop could be inferred given the frequency and affluence of users walking to that point.[6]

Additionally, a notification system was implemented through OneSignal where the citizen decides whether she wants to share her data with external entities. These external entities can range from digital twins of other citizens, our UDT, or the City Council itself, among others. The information contained

[6]The source code of uMuv is available for donwload from the following github repository [7]

in the digital avatar can be of great interest to other close contacts of the user as it provides important data that can be used to plan future bus trips.

But undoubtedly, this information is even more valuable for our UDT. As we have mentioned in previous sections, the UDT is fed by the open data published by the City Council on bus lines and bus stops. This data is updated every minute, but during that minute it is not possible to know the exact location of the bus. However, if we interconnect the UDT with the DT of any user on the bus (one is enough), we have accurate and detailed information of what is happening at each moment. We thus have finer grained information about the behavior of the bus and we can solve questions such as why a bus takes so long between two stops or two time intervals, for instance due to traffic lights or specific traffic lights. Using the Digital Avatars framework, citizens' DTs can be put at the service of the UDT to improve its performance and the information it handles. We are currently working on the implementation of the connections between the UDT and the citizens' digital twins.

8 Conclusions and Future Work

This work describes the architecture and implementation of a framework for integrating citizens through their DAs into UDTs, making use of the Cloud-to-Thing Continuum to arrange different components in each of the four proposed layers: *mist*, *edge*, *fog* and *cloud*.

As a proof of concept of our approach, we have developed an initial version of the UDT based on Eclipse Ditto, that takes into consideration real-time data of services such as the bus transportation system and weather information. Deploying cheap and simple elements like Bluetooth beacons on both buses and bus stops permits to detect when a given person uses the bus transportation system, by means of her smartphone. This way, information coming from open data sources publicly available – like real-time GPS positions of buses – has been integrated into the UDT with citizens' custom information to know how people use the transportation system for moving around the city, their daily routines and their preferences. All this information permits a more comprehensive analysis of the transportation system.

This work can be continued in several directions. In a first step, we plan to incorporate data on traffic conditions and specific events happening in the city, such as festivals or sports events, which may have an impact on the transportation system. Next, we also plan to explore the integration of other techniques like ML and AI to further improve the system's performance and accuracy in its predictions.

Overall, the proposed framework has the potential to significantly improve the transportation system of the city of Malaga by providing real-time data and analysis that can optimize routes, reduce travel time, and enhance the efficiency and reliability of the system.

Acknowledgments

This work was partially funded by the Spanish Government (FEDER/Ministerio de Ciencia e Innovación) under projects PID2021-125527NB-I00 and TED2021-130523B-I00. We would also like to thank the graduate students Daniel Roura and Tomás García for their contribution developing software artefacts for this work.

References

[1] Digital Twin Consortium. 2021. Glossary of Digital Twins.

[2] Claudio Badii, Pierfrancesco Bellini, Angelo Difino, and Paolo Nesi. Smart City IoT Platform Respecting GDPR Privacy and Security Aspects. *IEEE Access*, 8:23601–23623, 2020.

[3] David Bandera, Nicolás Pozas, Manuel F. Bertoa, José María Álvarez, Carlos Canal, and Ernesto Pimentel. Extensión de Digital Avatars para crowdsensing distribuido. In *Sistedes JCIS*, 2021.

[4] Manuel F. Bertoa, Nathalie Moreno, Alejandro Pérez-Vereda, David Bandera, José M. Álvarez-Palomo, and Carlos Canal. Digital Avatars: Promoting Independent Living for Older Adults. *Wirel. Commun. Mob. Comput.*, 2020:1–11, 2020.

[5] Francis Bordeleau, Benoît Combemale, Romina Eramo, Mark van den Brand, and Manuel Wimmer. Towards Model-Driven Digital Twin Engineering: Current Opportunities and Future Challenges. In *ICSMM Proceedings*, volume 1262, pages 43–54, 2020.

[6] Tianhu Deng, Keren Zhang, and Zuo-Jun (Max) Shen. A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *MSC Journal*, 6(2):125–134, 2021.

[7] Rafael García-Luque, Lorenzo Toro-Gálvez, Nathalie Moreno, Javier Troya, Carlos Canal, and Ernerto Pimentel. Citizen digital twin - Git repository. https://github.com/atenearesearchgroup/umuv, 2023.

[8] Rafael García-Luque, Lorenzo Toro-Gálvez, Nathalie Moreno, Javier Troya, Carlos Canal, and Ernerto Pimentel. Transport service digital twin - Git repository. https://github.com/atenearesearchgroup/malaga-bus-dt, 2023.

[9] Michael Grieves. Digital Twin: Manufacturing excellence through virtual factory replication. 1:1–7, 2014.

[10] Joaquín Guillén, Javier Miranda, Javier Berrocal, José García-Alonso, Juan Manuel Murillo, and Carlos Canal. People as a Service: A Mobile-centric Model for Providing Collective Sociological Profiles. *IEEE Softw.*, 31(2):48–53, 2014.

[11] Juan Luis Herrera, Jaime Galán-Jiménez, Jose Garcia-Alonso, Javier Berrocal, and Juan Manuel Murillo. Joint Optimization of Response Time and Deployment Cost in Next-Gen IoT Applications. *IEEE IoT Journal*, 10(5):3968–3981, 2023.

[12] Sergio Laso, Javier Berrocal, Pablo Fernandez, José María García, Jose Garcia-Alonso, Juan M. Murillo, Antonio Ruiz-Cortés, and Schahram Dustdar. Elastic data analytics for the cloud-to-things continuum. *IEEE IC*, 26(6):42–49, 2022.

[13] Theo Lynn, John G Mooney, Brian Lee, and Patricia Takako Endo. *The Cloud-to-Thing Continuum: Opportunities and Challenges in Cloud, Fog and Edge Computing*. 2020.

[14] Paula Muñoz, Alejandro Pérez-Vereda, Nathalie Moreno, Javier Troya, and Antonio Vallecillo. Incorporating Trust into Collaborative Social Computing Applications. In *Proc. of EDOC'21*, October 2021.

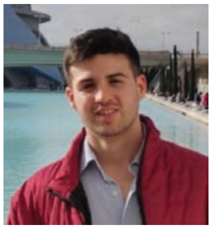
[15] Alejandro Pérez-Vereda, Carlos Canal, and Ernesto Pimentel. Modelling digital avatars: A tuple space approach. *Science of Computer Programming*, 203:102583, 2021.

[16] Alejandro Pérez-Vereda, Juan M. Murillo, and Carlos Canal. Dynamically programmable virtual profiles as a service. In *2019 IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, pages 1789–1794, 2019.

[17] Lorenzo Toro-Gálvez, Rafael García-Luque, Javier Troya, Carlos Canal, and Ernesto Pimentel. Towards the integration of digital avatars in urban digital twins on the cloud-to-thing continuum. In *3rd International Workshop on Big data driven Edge Could Services (BECS @ ICWE 2023)*, 2023.

[18] Ivan Čilić, Ivana Podnar Žarko, and Mario Kušek. Towards Service Orchestration for the Cloud-to-Thing Continuum. In *2021 6th SpliTech*, pages 01–07, 2021.

Biographies



Rafael García-Luque received his B.S. degree in Biomedical Engineering from the University of Málaga (Spain) in 2020, and his M.Sc. degree in Software Engineering and Artificial Intelligence from the same University in 2022. Currently, he is a Ph.D. Student at the Department of Computer Science of the University of Málaga. His main research areas are Bioinformatics, Mobile Computing and Web Development.

Lorenzo Toro-Gálvez graduated in Software Engineering from the University of Malaga (Spain) in 2022. Currently, he is pursuing a Master's degree in Computer Science at the Open University of Catalonia (Spain). His research areas encompass mobile computing and web development.

Nathalie Moreno received the M.Sc. and Ph.D. degrees in computer science from the Department of Computer Science, University of Málaga, Spain, in 2012, where she is currently an Assistant Professor. Her research interest is mainly oriented toward model-driven development. In particular, she focuses on conceptual modeling of digital twins as well as in the representation and propagation of trust and reputation in applications for the Internet of Things.

Javier Troya is Associate Professor of Software Engineering at the University of Malaga, Spain. Previously, he was a post-doctoral researcher as well as Assistant and Associate Professor at the University of Seville, Spain (2016 to 2021), and a post-doctoral researcher in the TU Wien, Austria (2013 to 2015). He obtained his International PhD with honors from the University of Malaga, Spain (2013), and was awarded in 2020 with the I3 Certificate by the Spanish Ministry of Science, Innovation and Universities. His current research interests include model transformation testing, uncertainty modeling and digital twins.

Carlos Canal received his Ph.D. degree in Computer Science from the University of Malaga (Spain) in 2001. He is currently a Full Professor with the University of Malaga. His research interests include digital twins, mobile computing, and cloud web development.

Ernesto Pimentel received his Ph.D. degree in Computer Science from the University of Malaga (Spain) in 1993. He is currently a Full Professor at the University of Malaga. His research interests are focused on Service-Oriented Computing, Formal Methods for Software Engineering, Cloud Computing, and Internet of Things.