Data Science and Security in Digital Governance Aspects and an Elastic Bus Transportation Scheme

Movses Musaelian[1], Md Zakirul Alam Bhuiyan[1*], Gary Weiss[1], Tian Wang[2*], Aliuz Zaman[1], and Thaier Hayajneh[1]

[1]*Department of Computer and Information Science, Fordham University, New York, USA 10458*

[2]*Department of Computer Science and Technology, Huaqiao University, Xiamen, Fujian, China, 361021*

*Corresponding Author

***Abstract*—E-Government systems hold immense potential for revolutionizing the relationships between citizens and government services. Our work aims to firstly lay out and analyze the components necessary for that digital ecosystem and also we propose how a bus transportation system, an important city service, can be enhanced with these methods and technologies. In recent years, significant research has been done in data security, smart city technologies, and digital identity, however, research for how such an ecosystem can optimize bus transportation has been lacking; mainly limited to private ride-sharing technologies. Our introspection into the critical components of digital governance leads us to a framework based on smart card data collection allowing rudimentary algorithm application for bus scheduling optimization. From this, we achieve a testable and tangible methodology to make bus scheduling more data driven; thus, contributing to a multi-faceted e-government society.**

I. Introduction

As data science is revolutionizing decision making and data collection in our society, the same technologies have taken a natural step towards revamping government functionalities. With all the moving parts required for the success of a digitized bureaucracy, there is a need to properly ascertain what specific components need to contribute to this ecosystem. A tangible framework for how e-government ecosystems can exist brings this vision closer to fruition. In this context, we highlight how public bus scheduling can be optimized using the data analytic and collection techniques that are propelling the development of digital governance systems. While recent research has highlighted different aspects of e-governance, there is a lack of research that encompasses and ties how these components contribute to one whole system. Furthermore, while data science has really propelled private ride-sharing programs, it hasn't really been adequately applied to more city-wide public transportation systems, which holds a great impact for citizens on a daily basis.

Our findings concentrate on data security, digital identity, and smart cities as critical components in an e-government system for which we bring forth the latest research and examples. In conceiving a framework to optimize a bus system, we carefully detail three important aspects: data collection, data analysis, and information relay. We emphasize the importance of data collection via smart card technology in order to apply algorithms that can make bus scheduling and allocation more efficient and cognizant. This sets up a tangible and testable framework that can be further developed to improve a key service in the smart city ecosystem. In all, we see this research as two-pronged: adding to the understanding of the big picture of multi-component e-governance and with that, proposing a specific improvement, which demonstrates a powerful fusion between data science and public transportation.

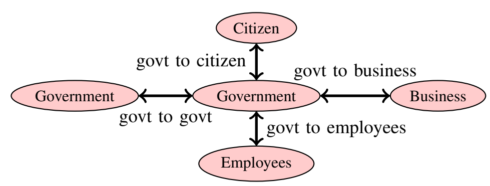

Fig. 1. The various relationships in an e-Government ecosystem.

II. Basic Data Ecosystem of Digital Governance

Here we lay out some of the basic elements and terminology of E-Governance systems and per various research works, lay out the data integration steps that are present in digital governance systems. Then, we introduce the concept of citizen-centered data, which is a key term in the context of this work. Finally, in Table I, we lay out a concise taxonomy of different models reflecting aspects of Digital Governance [7] [8] [10] [11] [12] [13].

A. E-Government & Data Science

It should be noted that there are two distinct terms. E-Government refers to how this technology improves the different activities and processes of government, while E-Governance refers to governing with the help of such technologies, especially data science. Serban et. al. and also Smitha et al. talk about the implications of database security to e-governance and highlight the different models that exist in e-governance: Government to Government, Government to Citizens, and Government to Business [1] [2]. With the help of data science, a data-driven approach to this model brings about solutions to the other interfaces and models. In other words, having a strong foundation in the electronic governance system for ordinary citizens, serves as a direct platform for how you can interact with corporations, small businesses, and within government agencies. Yet, without a concrete foundation, a synergistic approach to all the types of e-government models cannot be achieved.

One helpful diagram seen in Figure 1 illustrates how E-Governance should be viewed as and is used by Navdeep et al. [3].

B. E-Government Maturity & Success

Jin Sangki in his work about a new E-Government maturity model refers to three organization methods of E-Government. The first one is based on the method of service delivery, the second based on citizen participation, and the third based on government innovation or services [6]. It is with this third stage that citizenry participation is accelerated and you achieve both horizontal and vertical integration. This is the natural evolutionary step for E-Government implementations. Researchers such as J. Ramon Gil-Garcia et al. posit the determinants of success of government inter-organizational information sharing and within that research talk about the importance of technical infrastructure to make such sharing viable. Such an infrastructure should be seen in the context of the wider E-Government ecosystem and how that can enable not only citizen-government interaction, but more seamless inter-government interaction [5].

C. Citizen-centered Data

One key concept that needs to be elaborated upon is the concept of citizen centered data; this is crucial for understanding the implication and potential for E-Governance systems. Research about e-government systems has evolved from the higher level view of how a simple, governmental transactional system is to work with citizens to a much more personalized "data-centric" view connecting the notions of e-government and smart cities. This conjunction can be seen in the work by Jingrui Ju et al who write about citizen-centered big data as a "governance intelligence Scheme" for smart cities [16]. Firstly, they define a smart city as an ecosystem that is comprised of different societal actors in which there is a necessary amount of collaboration. In short, they don't view citizens as just "customers" in a transactional ecosystem, but rather individuals who are continuously producing "intelligence".

III. Data Security Methods & Lessons for Digital Governance Data

In order to make all the components of e-government viable, adequate data security is essential, exemplified by the latest research. [9] [15]. It is easy to see how failures in security measures, leading to data compromises and leakages, can severely degrade citizen trust in such digital platforms and with that dissuade them from using such systems actively. In this section, we introduce an anonymizing graph model and fog computing approach as potential methods for securing sensitive e-government data. We also discuss two concrete examples regarding personal data security in Russia and the data breach of Equifax as our main idea is to relay potential solutions and existing problems of securing data in digital governance.

TABLE I
THE DIFFERENT E-GOVERNMENT DEVELOPMENT MODELS

Authors	Taxonomy Model	Brief Description
Sharma et al.	Service Quality Model	Overall Quality leads to Citizen Satisfaction and Citizen trust
Kim et al.	Digital Government Evolution	Three largest clusters that have been developing since 2000 (e-government enactment Scheme, e-government, and online trust). Includes other clusters that have also seen some development in recent years.
Linders, D., et al.	Proactive e-Governance	Citizen-centric model that combines set of elements such as unified web portal, call center, roaming civil servants, multi-channel communication into one integrated system.
Dash et al.	Six Pillars of e-Governance Scheme	Author in his strategy for E-Government implementation gives the following pillars: Common State Wide Projects, Core e-Governance Common infrastructure, Institutional Scheme, Funding Scheme, Capacity Building Scheme, and Legal Scheme.
Albert Meijer	Barriers to E-Government development	Structural, Cultural, Citizen barriers that need to be addressed for successful E-Government development.
A.Ranerup	Human and technological actors in e-governance development	Demonstrates the multi-level political relationship that affect E-Government development and how the networks of human and technological processes are in continuous flux in this process.

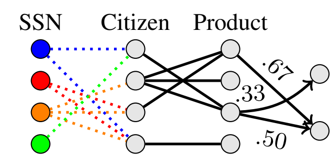

Fig. 2. Possible graph-based multifold anonymization on government data

A. Anonymizing Graph Structures

A work by Li-E. Wang and Xianxian Li talks about a novel graph-based multifold model for anonymizing data; an approach very relevant to how government data could be treated [17]. The main issue these authors tackle is how one can anonymize data that contains multiple types of attributes. In essence their approach calls for a graph based approach in which they "fuzz" sensitive attributes and convert "associations among items into an uncertain form". They also develop a model in the graph to prevent multi-objective attacks. To achieve this, they utilize the fact that in a given graph certain nodes are going to be sensitive while other nodes insensitive (one can imagine social security number as being a sensitive node). With that in mind, a set of association rules are developed and the sensitive rules are converted into an uncertain form. A "fuzzy" technique is thus designed by the authors to anonymize sensitive relational attributes.

Figure 2 shows how this respective anonymization technique could be a applied to government data. The dotted lines ("fuzz") denote "sensitive relationships" between the social security numbers and citizen files; thus, in reaching a final node state for a given product for a citizen the added noise will prevent the sensitive nodes (SSN) from being deduced. Probability estimates for the final nodes exemplify that uncertainty. Expectedly, such a model would be used to strictly prevent success of any type of intrusion, while accurate information retrieval would require additional authenticating steps.

With all this, the research provides a view of a very robust model that is able, with very minimal loss, to protect a whole host of sensitive attributes and their relations. The α-

safety of an anonymized graph and the clustering-based fuzzy algorithm protects the multifold privacy. Their work shows the potential that such a robust graph model can have for sensitive government and citizen data, as it is a much welcome departure from a vulnerable transactional database approach, which in many respects, has severe limitations.

Another example that exemplifies the research being done in graph anonymization in recent years is in the work by Yidong Li et al. regarding models to protect private weighted graphs [25]. Their research addresses the issue of preventing identity disclosure when a given adversary has background information; such as, information about the weight-related properties. All these techniques should be seen in the context of the wider move to cloud computing. Recent years have seen a large drive of organization computing needs to cloud services. While for security reasons, it would be expected that a Government would provide its own cloud service, Kh. E. Ali et al. for example talk about how cloud computing can overcome some of the difficulties of E-Government implementation [27]. They mainly cite reasons of cost-saving and scalability which are the added benefits we are seeing today from cloud computing.

B. Fog Computing Approach

A similar privacy preserving approach to user data comes from the proposal put forth by Piao et al. regarding a "fog-computing-based differential privacy approach" [19]. The researchers specifically address the issue of protecting government data, especially in the context of sensitive data of citizens. They focus on the issue that when a cloud platform is attacked and the attacker has prior "background knowledge", a lot of information can be compromised. In essence, their method calls for a differential method with which Laplace noises are added to the original data, which helps obscure the sensitive data from a potential hacker, even if they have strong background knowledge.

C. Subject Anonymization in Russia

Going forward, Anna Zharova and Vladimir Elin discuss how this issue of personal data security applies in Russia and offer some useful insight on the needs for data security in a concrete example [20]. They talk about the glaring flaw of how even anonymized personal data can be combined and refined to infer a person's identity. They state how the Russian model for personal data protection is based on "depersonalisation of subjects", yet with the current tools and information available to us today, it is possible to reconstruct such data and re-identify the given subjects. They attribute the impossibility of complete anonymization to such "proliferation tools".

D. Equifax Data Security Lesson

In the context of information security, the data breach of Equifax Company where hundreds of thousands of social security numbers were leaked demonstrated a huge failure of personal data security. In this context, Luke Briner writes about how information security should be approached in the light of such a high profile breach [24]. One interesting suggestion that comes from his article is that we should remedy the "over-sharing problem". The current landscape allows for a great amount of user data being sold and sent to companies of which the user many times is not aware of.

IV. Data Science Technologies in Smart Cities

The smart city is an emerging concept of how data science technology and information flow can be applied to the workings of city infrastructures [4] [18]. This intersection seeks to improve the everyday lives of citizens and also make local governance more informed and streamlined. We highlight the utility of geo-crowdsouring as a means to collect data and data-driven transportation, while we also overview how big data networks can be applied to such cities and how to assess smart city potential.

A. Geo-crowdsourcing for Data Driven Solutions

An important modern method for data collection, Witanto et al. talk about how geo-crowdsourcing data can be used to detect different types of anomalies and events in a given city location [29]. With the use of today's cellular devices, it is very easy to send location data; thus, we are at a point where such geo-crowdsourcing is feasible and efficient. The Scheme that the authors are proposing is one which utilizes and encourages citizen engagement and the combination of location data and also topic extraction from social media. In essence, the authors combine a few critical sources of information in today's society to develop a useful data Scheme for a smart government.

B. Data Driven Public Transportation Systems

One key aspect of the smart city is its ability to enhance public transportation as the issue of transportation relies on carefully planned logistics, something that data driven approaches can enhance. One very relevant problem is assessing bus punctuality as accurate evaluation of this can lead to better transportation planning. Research done by Yuyang Zhou et al. seeks to tackle that very problem as they develop a bus arrival time calculation model [30]. The researchers recognize the flaw in assessing bus arrival time simply based on the card swipes when entering the bus. Their calculation model takes into account the time lag between consecutive card swiping and the last card swiping time. Using data that they had also manually collected, they are able to come up with a set of heuristics to help in calculating arrival times more accurately. With this, they have developed a rather robust and efficient model that can help in bus network planning.

Similarly in another recent study, Attila Nagy and Vilmos Simon use machine learning to make traffic predictions as accurate as ever all in the context of a functional smart city [31]. In their extensive work, the researchers use data from sensors, both moving and stationary, and different machine learning models such as KNN and neural networks. They found that their most promising models take advantage of the spatio-temporal property of traffic flows. Another interesting application of such smart cities is using sensor data to enhance

TABLE II
DIFFERENT SMART CITY DATA NETWORKS

Big Data Network	Example Use Case
Preventive local administration	pro-active preventative action by local government such as for crime or congestion
Local operations management	Smart trash pickup or traffic control
Local network development	Wi-Fi hotspot optimization
Local information diffusion	Intelligent navigation, weather monitoring

door-to-door garbage collection, which was researched by Yin Chen et al [32]. In short, they employ the notion of "automotive sensing" in which sensors are fixed on several moving vehicles such as buses and taxis, whose data helps inform urban sensing tasks. In the particular research from this work, they are able to cover a large part of the city with their proposed sensing platform. The implications of their success are great for the potential of such smart cities as this sensor technology infrastructure helps enable many important tasks, such as sanitary tasks performed by local city administrations.

C. Big Data Networks in Smart Cities

This formulation in Table II shows how the data networks of a smart city affect different actors in differing approaches. One point that they also touch upon is the challenge of information delivery. Regarding this very topic Sarah Barns talks about how to design interfaces for smart governance, which touches upon the very same challenge of effective information delivery. Barns talks about the general evolution from 'smart cities 1.0' to 'smart cities 2.0', where open governance leads to a 'government as platform' understanding [37].

Aguilera et al. builds upon this concept in their research where they propose a mechanism that can provide common access to different data sources from the city and in turn bring citizens to a more proactive role [34]. In other words, making the citizen more of a stakeholder in a smart city ecosystem via the implementation of citizen-centric data services. Ahmed Osman and also Malik et al. in their research, look at the development of a big data analytics Scheme for smart cities using a layered design approach like seen in previous literature [35] [36].

D. Assessing Smart City Potential

How can the potential of Smart Cities be assessed? What factors drive a city to be ready to be a smart city? This very question is tackled by Miguel Zotano and Hugues Bersini in their research about assessing the potential of Smart Cities in a data-driven manner [38]. Zotano et al. prescribe certain elements that need to be present for a smart city's feasibility: data availability, eligibility (data can be re-used), completeness. Table III aims to summarize the holistic approach towards smart city data:

TABLE III
NECESSARY STEPS TO ASSESS "SMART CITY POTENTIAL"

Process Step	Explanation
Raw Data Sources	Data collected from citizens and sensors is stored.
Identification	Find relevant data from the raw data that can be put to action.
Filtering	Make proper encoding for the data in order to make it eligible for your analytics.
Matching & Linking	Link your analytics to the smart city ontology; in other words, transmission of intelligence back to the smart city.

V. DIGITAL IDENTITY IN ESTONIA'S E-GOVERNMENT

A key component in the e-governance infrastructure is the digital identity of a given citizen. Here, we look at how digital identity has been implemented in Estonia's rather developed e-Government ecosystem.

A. E-Residency in Estonia

One of the most famous examples of the implementation of digital identity is the Estonian e-Residency program. In a nutshell, the e-Residency program gives people around the world the ability to establish "e-Residency" in Estonia and with that have access to a whole host of government services, especially when it comes to registering a business and opening a bank account. All this is done without ever stepping foot in Estonia because services such as electronic signatures and blockchain technology enable an e-Resident to remotely and securely access and use a whole host of e-government and private sector services. In all, the program has been very successful, bringing added revenue to Estonia and proving the worth of this electronic infrastructure.

We come back to a recent news article written about the goals and aspirations that Marten Kaevats, Estonia's lead digital advisor, has for his country's very advanced e-Government ecosystem [40]. One interesting concept that Kaevats talks about is the notion of "invisible government", in other words, having instantaneous and automatic access to anything that you are entitled to get from the government. This two-way very direct responsiveness leads to a natural feeling of the absence of the burden of bureaucracy. Another principle that Kaevats mentions, which is crucial for e-Government success, is the "once only principle", which stipulates that government cannot ask data from citizens that is already held by a national public body.

VI. A PROPOSED DATA DRIVEN SCHEME FOR ELASTIC BUS TRANSPORTATION

Given the synergistic aspects of digital governance, we offer a potential scheme for the optimization of a public bus transportation, which we believe is a very tangible citizen-government relationship within the broader e-Government sphere. More specifically, this is a fundamental local government service (govt to citizen relationship), which can really benefit from a data-driven approach. Very conspicuously, ride-share applications have revolutionized transportation, especially taxis, where cars are dispatched based on geo-location algorithms optimizing the logistics of the ride. These sophisticated algorithms have also extended to ”pool rides” where different customers board the car who are going in a similar direction. How can a similar revolution be brought to bus transportation, which operates on a much larger scale for cities? Our main goal is to propose how such a system could potentially work, starting from how data can be collected and organized; then, what kind of algorithms could be applied to extract route optimizations, and finally how all this information could be effectively relayed between the actors.

Bus networks operate on fixed schedules, which vary based on part of the day, weekday or weekend, and so forth. However, the real revolution in the advent of smart city technologies is the transformation from a static, top-down administered city to a reactive city, power by bottom-up data inflows (citizen/sensor data to city government). For this problem, however, the main limitations stem from the difficulty of real-world testing and also the complex factors both infrastructural and logistical for its implementation. Certain inspirations and lessons can be taken from how health monitoring is being greatly improved in the medical sector with certain technological solutions [23].

A. Data Collection

Consistent data collection is imperative for effective data science. City wide sensors in general don't provide the granularity needed to power the given optimization that we are proposing. For this, we need a citizen level data collection mechanism for which we propose the use of 'smart cards'. Such a smart card, which can already be seen in practice in many cities, is a contactless method of payment (tap to pay), with seamless refilling for fares. Our main objective is to collect data as efficiently as possible, which can in turn fuel a whole host of useful metrics. For this we need:

- Location of Station & Time of when user X begins to wait at station
- Time when user X boards bus & bus number
- Time and location when user X exits the bus

How are these data points going to be collected? While there are few approaches for this collection, one approach we propose is the following:

1) User taps card at station upon arrival → collects **user_id**, **station**, **time**
2) User enters bus (no tap necessary) → collects **bus id**, **station**, **bus arrival time**
3) User exits bus (tap to exit) → collects **user_id**, **station**, **time**, **bus id**, & collects payment based on starting station.

These data points can be used to calculate relevant metrics; here are some examples:

- **user trip**: user_id is connected to a specific bus route and trip (station A to station B) by connecting user_id from first tap to user_id in the exit tap.
- **wait_time**: bus_arrival_time - user_station_arrival_time (exit tap records user's bus number) to get a user's wait time.

TABLE IV
STATION CARD TAP DATA

user_id	station	time_stamp
732948	River Rd. 25	2019-04-01T18:07:10
642123	Main St. 12	2019-04-01T18:09:45

TABLE V
BUS ARRIVAL DATA

bus_id	station	time_stamp
BX3578	River Rd. 25	2019-04-01T18:10:40
BL2075	Main St. 12	2019-04-01T18:15:20

TABLE VI
USER EXIT TAP DATA

user_id	station	bus_id	time_stamp
732948	Main St. 12	BX3578	2019-04-01T18:37:14
642123	Market St. 3	BL2075	2019-04-01T19:03:24

- **trip_time**: user_exit_time - bus_arrival_time (per user or per station)
- **station_queue_length**: select count(distinct user_id) from db where time in time_range group by station
- **avg_trip_length_from_stationA**: take average of user_exit_time - bus_arrival_time times filtered for user_id with stationA in their initial tap.

These are just some of the metrics that can be mined from the proposed collected data and we see how these metrics can already on a rudimentary level provide helpful metrics. Of course, in its practice there will be many nuances to deal with (e.g. multiple initial taps in different stations, fare evasion, etc.), but with this we aim to propose a ”naïve” Scheme, that will have to be refined in a practical implementation. It should be noted that the initial tap upon arrival to the station is crucial in order assess wait times and by that tap they will have ”initialized” their fare payment. We assume a fare structure which is based on the distance between your entry/exit points (zone based); thus, it requires respective taps at your starting and ending point. You can see an example data entry in table IV, which is sent to the database after user taps card at station, then when the bus arrives at the station it sends the data seen in table V, and finally when the user taps their card to exit table VI example entry is sent to the database.

B. Algorithm Concepts

We lay out some potential algorithm concepts in order to make optimizations to the bus transport network, mainly: reduce wait time, reduce trip time, minimize overcrowding. Fortunately, given the data we have at our disposal, many of these optimizations reduce from a much more complex problem to a more straightforward solution.

1) ”Artery Reduce”: We conceive the term ”Artery Reduce” to refer to the reduction of the ”weight” of transportation

Fig. 3. Trip Time Edges between Stations

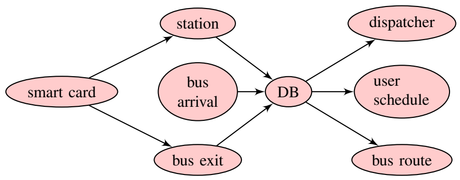

Fig. 4. Information Relay Overall Schema

paths or in other words the trip time. It is expected that trip times are at their highest during traffic which in turn is at the highest during the rush hour times. Given there are no physically separated bus lanes and refraining from the very logistically difficult task of re-routing buses, we ”reduce the artery” by skipping stations. For a given time frame, if we imagine our bus's path in a graph model like in figure 3, we signify the weight of the arrows as the trip times between the respective stations. We thus develop a heuristic that when trip times are a certain standard deviation above the mean trip time for that time frame and day, the station in between (B) is skipped. This change then informs the upcoming bus to skip A but stop at B. We are keeping in mind that these changes are calculated *in advance* of the given day.

2) Bus Allocation: Smart bus allocation is essential for reducing wait time and also minimizing bus overcrowding. Based on the data we have, we can know in advance that certain stations have large amount of waiting users and furthermore what are the most frequent bus routes taken from that station. If we think of this problem in the context of a graph model, we essentially don't want any of our nodes (stations) ”growing too large” at any given time if the size of the node represents the amount of people waiting. Thus, at strategic times, more buses in frequency are dispatched in advance.

3) Ride Profiles: From this data, ride profiles can be easily extracted (origin_station,arrival_station,time) which in turn can yield groups of popular routes taken by people. This can in turn inform route planning; for example, if you prioritize a route that is able to carry the popular ride profiles. Ride profiles can also inform improvements needed at stations, bus capacity, and other such elements. A bottom-up approach can also be taken by clustering. If we have the longitude/latitude coordinates of our respective stations and a timestamp, we can apply a simply k-means algorithm to uncover groupings of what can be labeled as ride profiles, done for a certain time window (e.g. 8.00-9.00). With this, the decision makers have an essential snapshot of how rides cluster, with large clusters signifying very popular rides that are similar to each other. Such clustering can also be done to riders given user_id is available, though caution needs to be taken given people replace cards, lose cards, and so forth.

C. Information Relay

Figure 4 shows the overall information relay schema and it should be noted that bus arrival sends its data upon bus's arrival to the station (not dependent on smart card tap). In all, this is a simplified schema, but conveys the scheme of information flow that we want to establish during the transportation ride cycle. An equally important element of this mechanism is how the data driven solutions are conveyed to the riders, buses, and dispatchers/organizers. There are a few points that need to be established about this:

- Possibly calculate a week's schedule ahead of time
- Maintain consistency $\rightarrow$ no volatile changes, no radical route changes
- if substantial route change made:
  give more time & forewarning

These points convey the point that while data can power dynamic scheduling, there needs to be a consistency maintained. We believe the calculation of a week's schedule ahead of time provides enough forewarning for riders and also enough granularity to take advantage of the analytics. Public bus transportation is not a taxi service after all, so the flexibility is much less. Furthermore, frequent, drastic changes to routes can confuse riders and create dissatisfaction. These dynamic changes need to firstly happen in terms of the bus time scheduling and the dispatching. Skipping stations for certain buses, while a bit more drastic, can be seen to be within the limits. Complete route changes however will require much more forewarning and time, and cannot be done very frequently. Of course the easiest way to relay this information to the rider is through mobile applications that can for their specified direction tell them the day's bus schedule and respective stations. We believe that the efficiency gained from such data driven scheduling will outweigh the inconvenience of not having a very fixed schedule. Stations should also be equipped with dashboards that can inform riders of the schedule for that station. Relaying this changing information, can be a research topic all to itself; effective data communication will be critical for the success of such a bus system.

VII. Our Future Work for the Elastic Bus Transportation Concept

While we lay out a general Scheme and concept for such a elastic transportation systems, there is still much to be done in order to make such an idea both viable and implementable. Here we touch upon two of those avenues: testing on both existing and synthetic data and a comprehensive analysis of bus transportation deficiencies.

A. Testing

In order to make this concept more concrete, testing should be done to further develop the proposed algorithms and adjust as needed. Both synthetic and real world bus data can be used and experiments can be run to see if the given data structure and heuristics would yield better performance.

B. Bus Transportation Deficiency Analysis

A more qualitative and comprehensive analysis needs to be done on existing bus transportation systems to see how well such a proposed concept could solve those deficiencies. While such deficiencies are complex and stemming from often localized problems, it needs to be seen if such an improvement can be worth the investment of the cost of implementation.

VIII. Conclusion

In this paper, we have concisely laid out the most fundamental and emerging aspects of E-Government Systems and how data science is revolutionizing that progress. We have shown the careful interplay between the different components such as data security, digital identity, smart cities that all need to co-exist successfully in order to ensure the success of the ecosystem. Our proposed elastic bus transportation system seeks to make an important facet of city infrastructure more data driven and dynamic, which can exemplify the goodness that data science can bring to everyday life. Finally, we have pointed out our future work for this proposal in order to bring it closer to fruition.

References

[1] Mariuta Serbana, Raluca-Mariana Stefan, Eduard-Ionel Ionescu. *Information Protection Security, Clustering and E-governance*. Procedia Economics and Finance, 2014.

[2] Smitha K K, Dr. Tony Thomas, Chitharanjan K. *Cloud Based E-Governance System: A Survey*. Procedia Engineering, 2012.

[3] Preet Navdeep, Dr. Manish Arora, Dr. Neeraj Sharma. *Role of Big Data Analytics in Analyzing e-Governance Projects*. GIAN JYOTI E-JOURNAL, 2016.

[4] Md Zakirul Alam Bhuiyan, Jie Wu, Guojun Wang, Zhigang Chen, Jianer Chen, and Tian Wang. *Quality-Guaranteed and Event-Senstive Data Collection and Monitoring in Wireless Vibration Sensor Networks*. IEEE Transactions on Industrial Informatics, 13(2): 572-583, April 2017.

[5] J. Ramon Gil-Garcia, Djoko Sigit Sayogo. *Government inter-organizational information sharing initiatives: Understanding the main determinants of success*. Government Information Quarterly, 2016.

[6] Jin Sangki *Vision of future e-government via new e-government maturity model: Based on Korea's e-government practices*. Telecommunications Policy, 2017.

[7] Pratyush Nidhi Sharma, Forrest V. Morgeson III, Sunil Mithas, Salman Aljazzaf. *An empirical and comparative analysis of E-government performance measurement models: Model selection via explanation, prediction, and parsimony*. Government Information Quarterly, 2019.

[8] Yushim Kim, Jing Zhang. *Digital government and wicked problems*. Government Information Quarterly, 2016.

[9] H. Tao, Md Zakirul Alam Bhuiyan, A. Abdalla, M. Hassan, J. Jain, and T. Hayajneh. *Secured Data Collection with Hardware-based Ciphers for IoT-based Healthcare*. IEEE Internet of Things Journal, 6(1): 410-420, Feb. 2019.

[10] Dennis Linders, Calvin Zhou-Peng Liao, Cheng-Ming Wang. *Proactive e-Governance: Flipping the service delivery model from pull to push in Taiwan*. Government Information Quarterly, 2015.

[11] Satyabrata Dash, Subhendu Kumar Pani. *E-Governance Paradigm Using Cloud Infrastructure: Benefits and Challenges*. Procedia Computer Science, 2016.

[12] Albert Meijer. *E-governance innovation: Barriers and strategies*. Government Information Quarterly, 2015.

[13] Agneta Ranerup. *The socio-material pragmatics of e-governance mobilization*. Government Information Quarterly, 2012.

[14] Aizhan Tursunbayeva, Massimo Franco, Claudia Pagliari. *Use of social media for e-Government in the public health sector: A systematic review of published studies*. Government Information Quarterly, 2017.

[15] Entao Luo, Md Zakirul Alam Bhuiyan*, Guojun Wang, Md Arafatur Rahman, Jie Wu, and Mohammed Atiquzzaman. *PrivacyProtector: Privacy-Protected Patient Data Collection in IoT-based Healthcare Systems*. IEEE Communication Magazine (COMMAG), 56(2): 163-168, 2018.

[16] Jingrui Ju, Luning Liu, Yuqiang Feng. *Citizen-centered big data analysis-driven governance intelligence Scheme for smart cities*. Telecommunications Policy, 2017.

[17] Li-E. Wang, Xianxian Li. *A graph-based multifold model for anonymizing data with attributes of multiple types*. Computers & Security, 2018.

[18] Tian Wang, Md Zakirul Alam Bhuiyan, Guojun Wang, Md. Arafatur Rahman, Jie Wu, and Jiannong Cao. *Big Data Reduction for Smart City's Critical Infrastructural Health Monitoring*. IEEE Communication Magazine (COMMAG), 56(3): 128-133, 15 March 2018.

[19] Chunhui Piao, Yajuan Shi, Jiaqi Yan, Changyou Zhang, Liping Liu. *Privacy-preserving governmental data publishing: A fog-computing-based differential privacy approach*. Future Generation Computer Systems, 2019.

[20] Anna Konstantinovna Zharova, Vladimir Mikhailovich Elin. *The use of Big Data: A Russian perspective of personal data security*. Computer Law & Security Review, 2017.

[21] Colin Tankard. *Encryption as the cornerstone of big data security*. Network Security, 2017.

[22] Yuehua Wu. *Protecting personal data in E-government: A cross-country study*. Government Information Quarterly, 2014.

[23] Md. Zakirul Alam Bhuiyan, Guojun Wang, Jie Wu, Jiannong Cao, Xuefeng Liu, Tian Wang. *Dependable Structural Health Monitoring Using Wireless Sensor Networks*. IEEE Trans. Dependable Sec. Comput. 14(4): 363-376, 2017.

[24] Luke Briner. *Making information security easier*. Computer Fraud & Security, 2018.

[25] Yidong Li, Hong Shen, Congyan Lang, Hairong Dong. *Practical anonymity models on protecting private weighted graphs*. Neurocomputing, 2016.

[26] Lifeng Nai, Yinglong Xia, Ilie G. Tanase, Hyesoon Kim. *Exploring big graph computing An empirical study from architectural perspective*. J. Parallel Distrib. Comput., 2017.

[27] Kh. E. Ali, Sh. A. Mazen, E.E. Hassanein. *A Proposed Hybrid Model For Adopting Cloud Computing In E-Government*. Future Computing and Informatics Journal, 2018.

[28] Lishan Qiao, Limei Zhang, Songcan Chen, Dinggang Shen. *Data-driven graph construction and graph learning: A review*. Neurocomputing, 2018.

[29] Joseph Nathanael Witanto, Hyotaek Lim, Mohammed Atiquzzaman. *Smart government Scheme with geo-crowdsourcing and social media analysis*. Future Generation Computer Systems, 2018.

[30] Yuyang Zhou, Lin Yao, Yanyan Chen, Yi Gong, Jianhui Lai. *Bus arrival time calculation model based on smart card data*. Transportation Research Part C, 2017.

[31] Attila M. Nagy, Vilmos Simon. *Survey on traffic prediction in smart cities*. Pervasive and Mobile Computing, 2018.

[32] Yin Chen, Jin Nakazawa, Takuro Yonezawa, Hideyuki Tokuda. *Cruisers: An automotive sensing platform for smart cities using door-to-door garbage collecting trucks*. Ad Hoc Networks, 2019.

[33] Chiehyeon Lim, Kwang-Jae Kim, Paul P. Maglio. *Smart cities with big data: Reference models, challenges, and considerations*. Cities, 2019.

[34] Unai Aguilera, Oscar Pea, Oscar Belmonte, Diego Lpez-de-Ipia. *Citizen-centric data services for smarter cities*. Future Generation Computer Systems, 2017.

[35] Ahmed M. Shahat Osman. *A novel big data analytics Scheme for smart cities*. Future Generation Computer Systems, 2018.

[36] Kaleem Razzaq Malik, Yacine Sam, Majid Hussain, Abdelrahman Abuarqoub. *A methodology for real-time data sustainability in smart city: Towards inferencing and analytics for big-data*. Sustainable Cities and Society, 2018.

[37] Sarah Barns. *Smart cities and urban data platforms: Designing interfaces for smart governance*. City, Culture and Society, 2018.

[38] Miguel Angel Gomez Zotano, Hugues Bersini. *A data-driven approach to assess the potential of Smart Cities: the case of open data for Brussels Capital Region*. Energy Procedia, 2017.

[39] Luyao Chen, Md Momin Aziz, Noman Mohammed, Xiaoqian Jiang. *Secure large-scale genome data storage and query*. Computer Methods and Programs in Biomedicine, 2018.

[40] David Whitehouse. *Enabling e-government: Estonias National Digital Advisor Marten Kaevats*. Global Government Forum, 2019.