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

Movses Musaelian¹, Md Zakirul Alam Bhuiyan¹*, Gary Weiss¹, Tian Wang²*, Guojun Wang³*, Thaier Hayajneh¹

¹Department of Computer and Information Science, Fordham University, New York, USA 10458
²Department of Computer Science and Technology, Huaqiao University, Xiamen, Fujian, China, 361021
³School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China

*Corresponding Author

July 4, 2019
E-Government systems hold immense potential for revolutionizing the relationships between citizens and government services.

We lay out the ecosystem for digital governance and emphasize the components:

- General e-government layout
- Citizen-centered data
- Data Security for digital governance data
- Smart Cities
- Digital Identity

This 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.
Introduction: E-Government Ecosystem Overview

The various relationships in an e-Government ecosystem.
Motivation

A digitized and data driven government system has immense potential for society:

- Transparency & Trust
- Pro-active relationships between citizens and government
- More intelligence and effective public service solutions
- Cost efficiency of government resource allocation

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.
Citizen-centered Data

- 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.
- Jingrui Ju et al. don’t view citizens as just "customers” in a transactional ecosystem, but rather individuals who are continuously producing "intelligence"\(^1\).

---

\(^1\) Jingrui Ju, Luning Liu, Yuqiang Feng. *Citizen-centered big data analysis-driven governance intelligence Scheme for smart cities.* Telecommunications Policy, 2017.
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.  

Possible graph-based multifold anonymization on government data

---

Data Security for Digital Governance Data

Equifax Data Security Lesson

- Hundreds of thousands of social security numbers were leaked demonstrating a huge failure of personal data security
- We should remedy the "over-sharing problem". Marten Kaevats, Estonia’s lead digital advisor, emphasizes the "once only principle", which stipulates that government cannot ask data from citizens that is already held by a national public body.
## Different Smart City Data Networks

<table>
<thead>
<tr>
<th>Big Data Network</th>
<th>Example Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preventive local administration</td>
<td>pro-active preventative action by local government such as for crime or congestion</td>
</tr>
<tr>
<td>Local operations management</td>
<td>Smart trash pickup or traffic control</td>
</tr>
<tr>
<td>Local network development</td>
<td>Wi-Fi hotspot optimization</td>
</tr>
<tr>
<td>Local information diffusion</td>
<td>Intelligent navigation, weather monitoring</td>
</tr>
</tbody>
</table>
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, we seek to use carefully collected rider data to optimize bus scheduling and allocation.
Data Driven Bus Transportation Framework: Data Collection

Our main objective is to collect data regarding the amount of people at given stations at certain times. 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

Chronology:

1. User taps card at station upon arrival $\rightarrow$ collects user\_id, station, time
2. User enters bus (no tap necessary) $\rightarrow$ collects bus id, station, bus arrival time
3. User exits bus (tap to exit) $\rightarrow$ collects user\_id, station, time, bus id, & collects fare based on starting station.
Data Driven Bus Transportation Framework: Data Collection

### Station Card Tap Data

<table>
<thead>
<tr>
<th>user_id</th>
<th>station</th>
<th>time_stamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>732948</td>
<td>River Rd. 25</td>
<td>2019-04-01T18:07:10</td>
</tr>
<tr>
<td>642123</td>
<td>Main St. 12</td>
<td>2019-04-01T18:09:45</td>
</tr>
</tbody>
</table>

### Bus Arrival Data

<table>
<thead>
<tr>
<th>bus_id</th>
<th>station</th>
<th>time_stamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>BX3578</td>
<td>River Rd. 25</td>
<td>2019-04-01T18:10:40</td>
</tr>
<tr>
<td>BL2075</td>
<td>Main St. 12</td>
<td>2019-04-01T18:15:20</td>
</tr>
</tbody>
</table>
Data Driven Bus Transportation Framework: Algorithm Concepts

"Artery Reduce"

- With our bus’s path in a graph model we signify the weight of the arrows as the trip times between the respective stations. We consider station skipping for very "heavy" sequences thus "reducing the artery".

![Trip Time Edges between Stations](image)

Bus Allocation

- We 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.
- We can know that in advance that certain stations have large amount of waiting users and furthermore what are the most frequent bus routes taken from that station. This can inform us of data-driven bus
Ride Profiles

- These profiles can inform about improvements needed at stations, bus capacity, and popular routes.
- 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.
- The decision makers have an essential snapshot of how rides cluster, with large clusters signifying very popular rides that are similar to each other.
Data Driven Bus Transportation Framework: Information Relay

Information Relay Overall Schema
Data Driven Bus Transportation Framework: Future Work

Testing

• Further develop the proposed algorithms and adjust as needed.
• Synthetic and real world bus data can be applied to see if the given data structure and heuristics may yield such better performance.

Bus Transportation Deficiency Analysis

• What are the current deficiencies and problems in existing bus transportation systems?
• Would such an improvement be worth the investment of the cost of implementation?
Conclusion

- 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 together 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.