

FINAL REPORT

**Crowdsourced Social Media Monitoring System
Development**



**OFFICE OF RESEARCH
15 KENNEDY DRIVE
FOREST PARK, GA 30297-2534**

TECHNICAL REPORT STANDARD TITLE PAGE

| | | | | | |
|---|---|------------------------------|---|-----------------------------|--|
| 1. Report No.: FHWA-GA-17-1614 | | 2. Government Accession No.: | | 3. Recipient's Catalog No.: | |
| 4. Title and Subtitle: Crowdsourced Social Media Monitoring System Development | | | 5. Report Date: August 2017 | | |
| | | | 6. Performing Organization Code: | | |
| 7. Author(s): Dr. Amit Kumar, Dr. Catherine L. Ross, Dr. Alex A Karner, and Rohan Katyal | | | 8. Performing Organ. Report No.: 16-14 | | |
| 9. Performing Organization Name and Address: Georgia Tech Research Corporation College of Design Center for Quality Growth and Regional Development (CQGRD) 760 Spring Street, Suite 213 Atlanta, GA 30332-0790 | | | 10. Work Unit No.: | | |
| | | | 11. Contract or Grant No.: 2014-32, RES PROJ 16-14 | | |
| 12. Sponsoring Agency Name and Address: Georgia Department of Transportation Office of Research 15 Kennedy Drive Forest Park, GA 30297-2534 | | | 13. Type of Report and Period Covered: Draft Final; May 2016 – August 2017 | | |
| | | | 14. Sponsoring Agency Code: | | |
| 15. Supplementary Notes: Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration. | | | | | |
| 16. Abstract: Crowdsourcing is a relatively new area of research, but it is already generating an enormous amount of interest among both researchers and practitioners, and is finding applications in multiple domains. It is particularly useful for efficient traffic management and increasing public participation. Many state DOTs are already using crowdsourced technologies and others are exploring its applications for traffic management. Researchers are using sensor-rich mobile phones and online social networks for fetching data from network users. Despite recent advancements, there remain gaps between the state of the art and practice that need to be bridged. Programs like the Waze Connected Citizens Program and Strava Metro Data Program are success stories in practice. This study explores the implementation of crowdsourced traffic management by Georgia DOT (GDOT) and the challenges specific to them. The reliability of data and filtering high volumes of information were found to be the two primary concerns. The team proposed a system which can potentially tackle those challenges. The system consists of a mobile application and a text mining application that together leverage the existing Twitter technology stack. Based on interviews with traffic management professionals and a visit to GDOT, the report contains recommendations that would improve the workflow at the TMC. Computer vision, data management and social media analytics would be particularly beneficial to decrease operator burden. A system with multiple sources of information integrated into one would be particularly beneficial. We are on the cusp of a revolution with respect to big data and crowdsourcing. This is the ideal time for GDOT to invest in crowdsourcing technologies to reap the benefits in the future. | | | | | |
| 17. Key Words: Crowdsourced Traffic Management, Social Media Monitoring, Incident Management | | | 18. Distribution Statement: | | |
| 19. Security Classification (of this report): Unclassified | 20. Security Classification (of this page): Unclassified | 21. Number of Pages: 844 | 22. Price: | | |

GDOT Research Project No. 16-14

Draft Final Report

CROWDSOURCED SOCIAL MEDIA MONITORING SYSTEM DEVELOPMENT

By

Dr. Amit Kumar (Principal Investigator)
Dr. Catherine Ross (Co-Principal Investigator)
Dr. Alex Karner (Co-Principal Investigator)

College of Design
Center for Quality Growth and Regional Development (CQGRD)
Georgia Tech Research Corporation

Contract with

Georgia Department of Transportation

In cooperation with

U.S. Department of Transportation
Federal Highway Administration

August, 2017

The contents of this report reflect the views of the author(s) who is (are) responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Georgia Department of Transportation or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Crowdsourced Social Media Monitoring System Development

Table of Contents

| | |
|--|-----|
| Abbreviations | vi |
| EXECUTIVE SUMMARY | vii |
| CHAPTER I. INTRODUCTION | 1 |
| Organization of the report | 2 |
| CHAPTER II. LITERATURE REVIEW | 4 |
| Introduction | 4 |
| Types of Crowdsourcing for Traffic Management | 5 |
| Advantages and Disadvantages of Crowdsourcing Types | 6 |
| Benefits of Crowdsourcing | 6 |
| Disadvantages of Crowdsourcing | 7 |
| Active Crowdsourcing | 8 |
| Passive Crowdsourcing | 8 |
| Combined Crowdsourcing | 9 |
| State of the Art | 9 |
| Sensor Based Crowdsourcing Systems | 9 |
| Human Mobility Based Crowdsourcing Systems | 10 |
| Gamification Techniques to Encourage Participation..... | 11 |
| Extracting Relevant Data and Filtering..... | 12 |
| State of Practice | 12 |
| Advanced Crowdsourced Data Applications | 14 |
| Strava Metro Data Service | 14 |
| Waze Connected Citizens Program..... | 15 |
| Adoption of Crowdsourced Technologies in a TMC Workflow..... | 15 |
| Commercially Available Traffic Management Technologies..... | 16 |
| Summary | 17 |

About the Center for Quality Growth and Regional Development

The Center for Quality Growth and Regional Development (CQGRD) is an applied research center of the Georgia Institute of Technology. The Center serves communities by producing, disseminating, and helping to implement new ideas and technologies that improve the theory and practice of quality growth.

For more information, visit www.cqgrd.gatech.edu

| | |
|--|-----------|
| CHAPTER III. CHALLENGES AND LESSONS LEARNED FROM PRACTICE | 19 |
| Introduction..... | 19 |
| Questionnaire Design..... | 19 |
| Selection of Respondents and Data Collection | 21 |
| Analysis of the State of Practice | 23 |
| ATMS and ITS..... | 23 |
| 511 System..... | 23 |
| Online Social Networks | 24 |
| Waze Connected Citizens Program..... | 28 |
| Dedicated Platforms Developed by DOTs..... | 28 |
| Challenges faced by TMCs | 35 |
| Cost of Systems..... | 35 |
| Multiple Data Sources..... | 36 |
| Workflow after an Incident | 36 |
| Variable Volume of Incidents | 37 |
| Network Coverage | 37 |
| Challenges for Crowdsourced TMC | 37 |
| Large Volume of Data Generated | 37 |
| Confidence in Data | 38 |
| Gaining User Engagement in Crowdsourcing Systems | 39 |
| Lessons Learned from Practice | 39 |
| One Stop Solution | 40 |
| Historical Data | 40 |
| Automating the TMC Workflow..... | 40 |
| Training the Staff | 41 |
| Quality Assurance | 41 |
| Mobile and Offline Access | 41 |
| Dissemination of Information | 42 |
| Summary | 42 |
| CHAPTER IV. CROWDSOURCED TMC FOR GEORGIA: NEEDS ANALYSIS..... | 43 |
| Introduction..... | 43 |
| Assessing the Georgia DOT TMC Operations..... | 45 |
| Assessing the Challenges | 45 |

| | |
|--|----|
| Needs and Usefulness of Crowdsourced Traffic and Incident Management (TIM) | 49 |
| SWOT Analysis | 51 |
| Proposed System for Georgia | 52 |
| Technical Architecture | 52 |
| Mobile Application | 53 |
| Text Mining Algorithm | 53 |
| Gamification to Increase Users | 53 |
| Advantages and Disadvantages of Proposed System | 54 |
| Advantages of the System | 55 |
| Weaknesses of the System | 55 |
| Prioritizing Action Items for the TMC | 56 |
| Summary | 57 |
| CHAPTER V. SUMMARY AND CONCLUSIONS | 59 |
| Summary | 59 |
| Recommendations | 62 |
| Limitations and Future Study | 63 |
| REFERENCES | 65 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: States with TMCs that Were Targeted for an Interview | 22 |
| Figure 2: Interviews Conducted with TMCs Located in these States..... | 22 |
| Figure 3: Usage of 511 Systems by DOTs..... | 24 |
| Figure 4: Reported Number of Twitter Followers by Interviewed DOTs | 26 |
| Figure 5: Reported Number of Likes on Facebook for Interviewed DOTs | 26 |
| Figure 6: Adoption of Crowdsourced Tools by DOTs | 27 |
| Figure 7: OSN Usage by DOTs | 27 |
| Figure 8: Number of DOTs Using Mobile Platforms, Web Platforms or Both..... | 29 |
| Figure 9: Usage of Various Platforms Based on Interview Data..... | 30 |
| Figure 10: Hits on Mi Drive by Platform Type | 30 |
| Figure 11: Variation in Incident Reporting by Day (Source: UDOT (2014)) | 31 |
| Figure 12: TMC Interaction with Citizens..... | 32 |
| Figure 13: Number of Users for DOT Mobile Applications on All Platforms | 33 |
| Figure 14: Number of Users for Web Platforms..... | 34 |
| Figure 15: Percentage of DOTs with Dedicated Staff for Social Media | 35 |
| Figure 16: Concerns about Volume of Data Generated..... | 38 |
| Figure 17: Panoramic View of the GDOT TMC | 43 |
| Figure 18: Workspaces of GDOT TMC Operators..... | 44 |
| Figure 19: Live Feed of High Definition Cameras at GDOT TMC | 46 |
| Figure 20: Live Feed of Cameras near the March, 2017 I-85 Bridge Collapse Site..... | 47 |
| Figure 21: Live Feed of Cameras at Strategic Locations..... | 47 |
| Figure 22: Select Twitter Handles that the Social Media Manager Monitors and Manages | 49 |
| Figure 23: SWOT Analysis..... | 52 |
| Figure 24: Architecture of Proposed System | 54 |
| Figure 25: Average Retention and Churn Rate after the First, Second and Third Months..... | 56 |

LIST OF TABLES

| | |
|---|----|
| Table 1: State of Practice at Different Agencies or Jurisdictions | 13 |
|---|----|

Abbreviations

ATMS – Advanced Traffic Management System

DOT – Department of Transportation

GPS – Global Positioning System

HERO - Highway Emergency Response Operators

ITS – Intelligent Transportation System

OSN – Online Social Network

SCATS – Sydney Coordinated Adaptive Traffic System

TIM – Traffic and Incident Management

TMC – Transportation Management Center

Georgia DOT – Georgia Department of Transportation

Florida DOT – Florida Department of Transportation

Utah DOT – Utah Department of Transportation

EXECUTIVE SUMMARY

Recent advances in mobile and ubiquitous computing have led to a massive increase in the amount of data generated through the use of social media and personal portable devices. These “crowdsourced” data can be used in many different application areas and are particularly useful to Departments of Transportation for traffic and incident management (TIM). Crowdsourcing is a relatively new area of research which is generating an enormous amount of interest among both practitioners and the research community.

The allure of crowdsourcing is clear—state DOTs have limited personnel resources and cannot constantly monitor all links and intersections in their jurisdictions. Crowdsourced data can overcome those constraints by engaging network users as sensors. Even though the state of the practice in incident reporting and management has moved beyond traditional technologies such as CCTV and loop detectors, there is still a huge gap between the crowdsourcing state of the art and the state of the practice. A majority of DOTs have a strong social media presence and engage with citizens online. Others are experimenting with systems that allow crowdsourced citizen reporting. Programs like Waze “Connected Citizens Program” and Strava Metro Data have been successful in providing information for low coverage areas. These programs can also supplement DOT coverage in areas where cameras are sparse. Georgia DOT’s ability to handle major events and large volumes of data needs to evolve as massive amounts of data are increasingly generated online. Once GDOT develops and deploys systems that can automatically fetch, filter and prioritize reported incidents from crowdsourced data, the use of these systems will benefit GDOT’s TIM program.

Many state DOTs, including Utah, Florida, Michigan and Washington D.C., have already employed crowdsourced technologies for TIM. Other state DOTs are also looking into the

possibility of using such systems. Crowdsourced data can be used to overcome data gaps and deficiencies and also increase public participation. The objective of this research is to study different options for collecting and utilizing crowdsourced data, and apply those findings to recommend a crowdsourcing solution for the Georgia DOT (GDOT).

The research methods employed in this study include identifying different options to obtain crowdsourced data for traffic management, and delineating the advantages and disadvantages of each method. The research team then studied the challenges and lessons learned from the implementation of crowdsourced TIM in other states, to determine how Georgia DOT (GDOT) can best make use of crowdsourced information.

This study also clarifies how crowdsourcing can reduce gaps in information involving incidents and level of service loss in a transportation network. It provides a foundation of knowledge that GDOT can draw from to implement a process of automated incident detection and confirmation using multiple data sources and the rapid dispatch of response teams.

The team proposes a low-cost crowdsourced TIM system for GDOT. The system consists of a mobile application which allows citizens to report incidents. The incidents are then automatically tweeted. A text mining application running at GDOT mines these reports. This system leverages Twitter's infrastructure to minimize development effort, cost and maintenance. Based on practitioner interviews and a visit to the TMC operated by GDOT, the study includes several recommendations to enhance the efficiency of the TMC. Automatic dissemination of information and computer vision technology to automatically detect incidents from camera feeds will reduce operator burden. Improved data management and social media analytics will assist in swiftly filtering through the high volume of data generated. An integrated TIM architecture where

information from multiple sources feeds into a single system leading to automated incident detection and validation should be developed to realize the greatest positive impact on traffic management. With technologies like machine learning and artificial intelligence allowing machines to make independent decisions and make sense of large volumes of data, the future of traffic management lies in crowdsourcing.

ACKNOWLEDGEMENTS

The research presented in the following report was sponsored by Georgia Department of Transportation through Research Project Number 16-14. The authors would like to acknowledge Sarah Lamothe, Office of Research, Georgia DOT for her assistance in completing this project.

The authors are indebted to the managers, social media coordinators and other TMC personnel across the US who participated in interviews conducted by the research team and helped us to understand the challenges of TMCs operations. Our special appreciation is extended to TMC personnel in Georgia, in particular, Mark Demidovich for his kind cooperation during this study.

The authors would also like to acknowledge the assistance from researchers and graduate students at the Center for Quality Growth and Regional Development (CQGRD) at Georgia Tech.

The opinions and conclusions expressed herein are those of authors and do not represent the opinions, conclusions, policies, standards or specifications of the Georgia Department of Transportation or of the cooperating organizations.

CHAPTER I. INTRODUCTION

Recent trends in ubiquitous computing have given rise to new avenues of collaboration. Data are being produced and consumed at a colossal scale. Online social networks (OSNs) such as Twitter, Facebook, Waze, Strava and Instagram are generating textual, image-based and geotagged data. These data are being fetched, processed and analyzed across many issue domains to learn from the wisdom of the masses. This process of making sense of these data and applying them to solve a particular problem is called “crowdsourcing.”

Planners are increasingly using crowdsourced data as either their primary source of information or to supplement traditional technologies. Certain Departments of Transportation (DOTs) have developed dedicated applications to fetch crowdsourced data that feed into existing Intelligent Transportation Systems (ITS) in real-time. Crowdsourced systems also have the potential to increase and improve public participation.

A Traffic Management Center (TMC) typically has to deploy a large number of sensors and cameras to monitor and detect incidents in a highway network. Maintenance of these sensors is challenging; they are distributed across the network and must be repaired in person. In addition, simultaneously monitoring all the video feeds generated by a large number of cameras is simply not possible with available manpower.

Crowdsourcing has the potential to satisfy TMC data needs and improve their operational efficiency, but these new approaches must be evaluated before they are adopted in practice. There are multiple forms of crowdsourcing, each with inherent strengths, weaknesses, opportunities, and threats. A TMC can only select the most appropriate technology/method after a comprehensive evaluation. SWOT (strengths, weaknesses, opportunities and threats) analysis is one available option to identify optimal strategies for a particular organization through the analysis of internal

organizational factors (strengths and weaknesses) and external factors (opportunities and threats) faced by the organization.

The objective of this research is to study different options for collecting and utilizing crowdsourced data, and apply those findings to recommend a crowdsourcing solution for the Georgia DOT (GDOT).

Organization of the report

This report has five chapters, including this introduction. The remaining chapters are organized as follows.

- *Chapter II: Literature review*

The research team conducted an extensive literature search to understand both the state of the art and the potential future of crowdsourcing techniques. The search included refereed journal articles, conference proceedings, blogs and project reports from other DOTs/TMCs.

- *Chapter III: Challenges and lessons learned*

The team interviewed TMCs operating in multiple states that have implemented or are in the process of implementing traffic management systems using crowdsourced data. The goal of the interviews was to understand the advantages and disadvantages of different crowdsourcing models and methods, and apply the lessons learned to the research team's recommendations for GDOT.

- *Chapter IV: Evaluate the potential for crowdsourcing applications at GDOT*

In order to understand GDOT's workflow and Georgia-specific traffic problems, the team conducted in-house interviews with GDOT TMC staff to understand the strengths, weaknesses, and the needs of the statewide TMC located in Atlanta. A SWOT analysis was also performed

to better understand the possible utility of a crowdsourced traffic management system for the GDOT TMC. The recommendations include a potential architecture (system strategy) for the crowdsourced solution for the TMC operated by GDOT.

- *Chapter V: Summary and conclusions*

The concluding section summarizes the recommendations and discusses the limitations of the work as well as opportunities for future study.

CHAPTER II. LITERATURE REVIEW

Introduction

The increasingly widespread use of social media applications, where users create and upload content consumable either by their own social network or the broader public, including Twitter, Facebook, Waze and Instagram is creating a staggering amount of data that are potentially useful for a wide range of applications in transportation infrastructure planning and operations. These applications include, among others, real-time coordination of traffic signals, cycling infrastructure location siting, disaster relief, and incident management and public engagement (Aubry, Silverston, Lahmadi, & Festor, 2014; Barron, Manso, Alcarria, & Gomez, 2014; Molina, 2014; Schweitzer, 2014; Steinfeld, Zimmerman, Tomasic, Yoo, & Aziz, 2011).

Additionally, crowdsourced information on incident location, weather, congestion, or roadway conditions can be used to inform real-time traffic management systems (Boulos et al., 2011; Myr, 2002; Pinto, 2007). Crowdsourcing such information in real-time through mobile phones and other personal electronic devices is increasingly attracting the interest of public agencies.

For example, the number of critical intersections that need to be monitored by TMCs often exceed the CCTV cameras that an agency can afford to deploy. Moreover, staffing limitations restrict the number of video feeds that can be monitored simultaneously. Because of the widespread use of OSNs, existing infrastructure can be leveraged to accumulate crowdsourced incident data.

Many states (e.g. Iowa, Florida, and DC) have developed, and additional states are developing, mechanisms for traffic and incident management using crowdsourced data. In principle, these data can be combined with more traditional traffic data sourced from sensors, detectors, and cameras to aid with real-time traffic management (El Faouzi, Leung, & Kurian, 2011; Van Lint & Hoogendoorn, 2010).

In addition to these practical applications, the academic literature on crowdsourced data has rapidly been expanding. Earlier literature was focused on the concept and definition of crowdsourcing. Bozzon et al. (2013) define it as an emerging way of involving humans in performing information seeking and computation tasks. Doan et al. (2011), and Hossain and Kauranen (2015) associate terms like peer production, user-powered systems, user-generated content, collaborative systems, community systems and peer production with crowdsourcing.

Having realized the potential, the literature now is more focused on applications. Crowdsourcing is considered a suitable model for crowds to participate in public planning projects (Brabham, 2009; Hilgers & Ihl, 2010).

Types of Crowdsourcing for Traffic Management

“Crowdsourced” data generated from sensor-rich smart phones connected to social media services (including Twitter, Facebook, Waze and Instagram, among others) can supplement existing traffic management systems to develop a comprehensive reporting platform. Crowdsourcing real-time information through these devices has been attracting great interest not only from the general public but also from public agencies. Many states (including Iowa, Florida, and DC) have developed and are developing mechanisms for traffic and incident management using crowdsourced data. There are generally three types of crowdsourcing:

1. **Active:** requiring some action from participants. For example, Twitter, Facebook, and Instagram posts all require users to create content and upload it to the web.
2. **Passive:** requiring no action from participants. For example, TomTom and Google Maps collect data passively, and do not require action from participants.

3. **Combined:** where active input from the user is not required but can be sought to supplement passively collected data. Waze is an example of a combined crowdsourced dataset.

Advantages and Disadvantages of Crowdsourcing Types

This section is an examination of the advantages and disadvantages of crowdsourcing in general and that of the various crowdsourcing types in particular.

Benefits of Crowdsourcing

Conventional traffic management technologies are constrained by both the cost of installing sensors and the volume of data that needs to be processed to obtain and utilize the enhanced road network coverage. Although the cost associated with crowdsourcing systems depends on the type of system being considered, generally the cost is lower than conventional means of gathering information. The existing infrastructure technology such as loop detectors and cameras have a high maintenance and running cost. Crowdsourced data can be used to augment the data obtained through these other means.

Gathering data through crowdsourcing can also result in representation from a greater variety of demographic groups and geographic regions. Moreover, it allows for much quicker access to local data and costs less than conventional data collection methods. For example, Waze can fetch traffic information without the constraints of a fixed location (such as a camera). Waze has the reach and accessibility in the form of mobile devices that no conventional technology can provide.

Public participation in urban planning is vital to the development of successful policies. Crowdsourcing techniques allow a large number of private citizens to provide transportation data as well as a systematic platform for the public to share feedback and suggestions. Urban planners

can also engage in interactive policy development with the community through the use of social media.

Disadvantages of Crowdsourcing

Crowdsourced data has several limitations. Accurate crowdsourced information depends on the system, contributors, and type of information collected. Research suggests that 90% of crowdsourcing efforts fail due to lack of interest from the public (Dahlander & Piezunka, 2014). The authors monitored 23,809 organizations that had been using crowdsourcing software to solicit feedback from the public. Only 1% of those organizations achieved the level of one suggestion per day while 90% received fewer than 30 suggestions per year.

Data reliability is another primary concern. Internal DOT verification of the validity of the data requires manual intervention, which creates the same staffing and resource limitations as conventional data collection methods. It also limits the scalability of the crowdsourced data process. After reliability, the next biggest concern is quality and repetition. The same incident might be reported differently by different citizens. Some users may intentionally report inaccurate information, which then must be culled. Quality assurance and assessment mechanisms need to be standardized and implemented to increase confidence in the data.

There are privacy concerns associated with crowdsourcing, particularly on OSNs such as Facebook and Twitter. Often a user's understanding of the system's privacy policy and the actual policy are far apart. When fetching personally identifiable information, the user's consent needs to be given and the user should be made aware of what data is being collected. Not having a clear policy and not informing the user how their data is being used can decrease the confidence of the user in the system and gradually discourage the user from participation.

Recruiting and retaining participants is another major challenge. Users need to be motivated or incentivized to participate (De Vreede et al., 2013; Komarov, Reinecke, & Gajos, 2013; J. Yang, Adamic, & Ackerman, 2008). The strategies to do so depend on the type of crowdsourcing and format of data being collected. For example, if a large number of contributors are required, social media is a useful medium to recruit participants. Like any other data collection method, crowdsourcing efforts suffer from bias. Without a desired number of contributors, it's difficult to neutralize the effect of the bias (Smith & Fehr & Peers, 2015).

Active Crowdsourcing

Active crowdsourcing requires active participation from users. For such crowdsourcing to be effective, transportation network users need to be motivated to report incidents. This form of crowdsourcing also suffers from self-reporting error which can result in poor data quality. A user might not know the correct location or cause of an incident if reporting while passing by in a moving vehicle. In spite of these drawbacks, active crowdsourcing can be effective since data can be fetched through multiple formats and is not limited by the sensor being used. Automatic filtering on unstructured data is also critical, as users might report the same incident in different ways (e.g., pictures, video or different text descriptions of the same incident). Some states also have hands free laws that requires a driver to engage both hands with the steering wheel to avoid distracted driving. Participation in active crowdsourcing is therefore not possible in single occupancy vehicles due to this law.

Passive Crowdsourcing

Passive crowdsourcing techniques do not require any action from the participants beyond the initial installation of the application. This form of crowdsourcing is already used by public agencies (Charalabidis, N. Loukis, Androutsopoulou, Karkaletsis, & Triantafillou, 2014; Loukis &

Charalabidis, 2015). The likelihood of users participating in this form of crowdsourcing is higher as data are collected automatically with minimal involvement from the user. The data obtained through such methods have a limited amount of detail as compared to active approaches, since they are mined using sensors. But the data generated in this manner are structured, appearing in a standard format, which can improve quality control and the ability to filter.

Combined Crowdsourcing

While enjoying the benefits of active and passive crowdsourcing techniques, combined crowdsourcing may suffer from the disadvantages of both. For example, network users may not be interested in participating without an incentive, or if they have privacy concerns. Despite these limitations, this form of crowdsourcing can be more robust as the data being reported by users are supplemented with automatically generated passive data.

State of the Art

While the state of practice has clearly moved beyond its prior emphasis on CCTV and phone calls, the literature review suggests that many of the practices in traffic management are still lagging behind the most innovative, state of the art techniques available. This section presents the state of the art of crowdsourcing techniques relevant to TMCs.

Sensor Based Crowdsourcing Systems

Smartphones can create mobile sensor networks capable of collecting rich information. Several crowdsourcing solutions such as NoiseTube (Stevens & D'Hondt, 2010), Pothole (Eriksson, Girod, et al., 2008) and CityExplorer (Matyas et al., 2008) make use of mobile sensors. Mohan et al. (2008) proposes Nericell, a system for monitoring road and traffic conditions in a city using smartphones. They focus specifically on sensing, using the accelerometer, microphone, GSM radio, and GPS sensors to detect bumps, braking and honking. Campolo et al. (2012) developed

SMaRTCaR, a smartphone based platform that leverages dedicated hardware in vehicles to collect data. SMaRTCaR reports not only vehicle speed, but uses external sensors to monitor specific physical parameters such as pollution, humidity and temperature. This collected data can be used for predictive analysis and defining the strategy of the DOT among other uses. VTrack (Thiagarajan et al., 2009) is another system for travel time estimation using the fetched sensor data. It can work around the limitations of sensors such as GPS which are unreliable under some conditions (e.g. heavy tree canopy, cloudy skies, and urban high-rise buildings) and have high power consumption. VTrack can use WiFi to estimate the travel time along the route. Cabernet (Eriksson, Balakrishnan, & Madden, 2008) is another system which uses WiFi. It sends data to and from moving vehicles using open WiFi access points opportunistically during travel.

Human Mobility Based Crowdsourcing Systems

Researchers have conducted pilot projects to estimate travel speed and driving patterns based on location data. Mobile Millennium was one such pilot project launched on the University of California, Berkeley campus in 2008 for a duration of one year (Herrera et al., 2010). During that time, more than 5,000 users downloaded the application and shared their GPS location to a server for aggregation and analysis. Similar to Mobile Millennium, iCarTel2 is an iPhone application developed as a part of the CarTel project by researchers at Massachusetts Institute of Technology (MIT). It is still in active use in Boston (Balakrishnan & Madden, 2014).

Unpredictable travel times and congestion make traffic management in developing regions a more complex process. To deal with the challenges of traffic management in developing regions Sen et al. (2009) developed a technique to estimate vehicle speed based on the Doppler shift of frequencies for vehicular honks. Frank et al. (2014) have developed mobile applications that monitor traffic conditions in real-time. The application mines the location and velocity data and

uses that to identify alternate routes and identify congestion. A similar application is LuxTraffic, a traffic sensing system deployed in the country of Luxembourg (Kovacheva, Frank, & Engel, 2013).

Many researchers have proposed solutions using the GPS-equipped public transportation networks such as taxis and busses to detect traffic conditions. Pan et al. (2013) deployed a system in China to monitor traffic anomalies using GPS trajectory and social media data. The system was tested on 30,000 cabs and content from WeiBo (a Chinese social networking site; www.weibo.com) was used to detect reasons for anomalies.

Integrated hardware and software solutions have also been developed to monitor traffic movement. A piece of hardware developed by the CarTel team can be installed in a vehicle to monitor movement using GPS. The hardware opportunistically sends data to the server. This information is then used for better route planning. Traffic monitoring can be done in a more efficient manner using dynamic data from social media. This data is used to perform a congestion analysis and clustering of congested areas (Chatterjee, Mridha, Bhattacharyya, Shakhari, & Bhattacharyya, 2016).

Gamification Techniques to Encourage Participation

The success of crowdsourced systems is dependent on the number of contributors. In order to increase adoption of such technologies, researchers have designed gamification and incentivization techniques. Quinn and Bederson (2011) study incentives that can be used to engage the crowd, which can include pay, altruism, enjoyment, reputation, and so on. *RasteyRishtey* is a social incentive based system which assists users in meeting each other at specific places and times (Sen, 2014). McCall and Koenig (2012) propose the design of a game which provides financial as well as social incentives for users to alter their driving patterns. The user can join leagues to compete

against each other. The application also makes suggestions to alter individual behavior (e.g., leave at 10 am instead of 9 am) and provides incentives for doing so, including discounts and free parking.

Extracting Relevant Data and Filtering

One promising area where advanced research could aid practice is in filtering signal from noise within vast quantities of crowdsourced data. The sheer volume of information created through crowdsourcing can be overwhelming. Extracting, storing and sorting the right data is an extensively studied problem in modern research. Takeshi et al (2010) proposed a solution for real-time event detection by social sensors. They consider each user to be a sensor. The proposed solution detects target events on features such as keywords in a tweet and the number of words. In the case of traffic management systems, this information needs to be complemented with additional information such as location. Often identifying reliable data is a challenge even after filtering. Artikis et al. (2014) built a mechanism to tackle data veracity and resolve the disagreement of information fetched from various sensors (fixed sensors at intersections and mobile sensors embedded on buses) via crowdsourcing participants.

State of Practice

Traditional practices for traffic management include the use of 511 systems, 911 systems, CCTV cameras, mailing lists, police radio channel, patrol teams, loop detectors, video detection, and microwave radars. The literature review of DOT practices reveals that multiple agencies in the US are currently employing crowdsourced data of some type instead of or in addition to traditional systems. These are summarized in Table 1. As demonstrated in the table, one of the most prevalent use cases is reporting incidents to agencies using some type of geo-referenced data. Submission of this type of information enables authorities to take quick action. The table contains a column

categorizing the crowdsourcing system into one of three categories: Active, Passive and Combined.

Table 1: State of Practice at Different Agencies or Jurisdictions

| Agency Name or Jurisdiction | Brief Description | Type of System | References/ Link |
|---------------------------------------|--|--|---|
| Washington DOT | 511 traveler system for citizens to report incidents. | Active | http://wsdot.wa.gov/traffic/511 |
| Iowa DOT | Prominent social media presence via 12 Twitter and six Facebook accounts. The staff actively engages with people on OSNs during office hours. | Active | https://iowadot.gov/stayingconnected#/stayingconnected |
| Utah DOT | Deployed three mobile applications for increasing public participation. They have prominent presence on social networks and have an active blog online. | Active | https://www.utah.gov/connect/mobile.html |
| Florida DOT | In order to provide data on interstate highways and arterial roadways FDOT collaborated with Waze in 2014. This has enabled them to increase the quality and quantity of data. FDOT allows for two-way communication on their social network accounts as well. FDOT also leverages the capabilities of the 511 system. | Combined (Waze and Strava Metro) Active (511) Active (Online Social Media) | http://www.fdot.gov/traffic/Newsletters/2014/2014-Aug.pdf http://www.fdot.gov/info/newsroom.shtm |
| Michigan DOT | MiDrive is a mobile application available on the Android and iOS store that allows for two way communication between motorists and the DOT | Active | http://www.michigan.gov/mdot/0,4616,7-151-9620-341121--,00.html |
| Minnesota DOT | Citizens can post reports on the 511 traveler system | Active | http://hb.511mn.org/#login?timeFrame=TODAY&layers=allReports%2CroadReports%2CwinterDriving%2CvoxReports%2CweatherWarnings%2Cflooding%2CotherStates |
| Tennessee DOT | Collaboration with Waze as a part of the connected citizens program. The collaboration is designed as a two-way data sharing of publicly available traffic information | Combined | https://www.tn.gov/tdot/news/tdot-joins-waze-connected-citizen-program |
| Oregon DOT | Using Strava Metro data to decide where to optimally place bike counters to capture maximum cycling behavior | Combined | http://metro.strava.com/ |
| Seattle DOT | Using Strava Metro data to gain insights into preferred cyclists route and spot the dangerous intersections | Combined | http://metro.strava.com/ |
| Queensland Transportation (Australia) | Using Strava Metro data to develop more bike routes | Combined | http://metro.strava.com/ |

| Agency Name or Jurisdiction | Brief Description | Type of System | References/ Link |
|---|--|----------------|---|
| San Francisco County Transportation Authority | Developed a mobile application called CycleTracks that collects bike ride data. This data is used to develop a route choice model. Initially, it was deployed to learn the usage patterns of bike lanes. The rider can also provide additional data such as purpose of ride. | Passive | http://www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android |
| Vermont Transportation | Using Strava Metro as their key data layer in the state-wide VTrans On-Road Bicycle Plan | Combined | http://metro.strava.com/ |
| City of Austin, Texas | Developed an application which pulls GPS data from motorists to change the signal in real-time | Passive | http://www.govtech.com/applications/Austin-Turns-to-Crowdsourcing-to-Improve-Traffic-Congestion.html |
| City of Ottawa | Signed up for 2 years of Strava data that assists the planners to get an idea of how cyclists and pedestrians interact with the urban environment | Passive | http://www.cbc.ca/news/canada/ottawa/strava-app-ottawa-1.3546513 |

Advanced Crowdsourced Data Applications

Modern technologies such as machine learning are being used to derive value from the large volumes of data generated by crowdsourcing (Boulos et al., 2011; Gao, Barbier, & Goolsby, 2011; Kittur et al., 2013). MIT Media Lab created an algorithm to determine which places in a city seemed safer (Smith, 2015). *Divvy*, a Chicago based bike share system is using crowdsourcing to decide where to place future bike stations (Divvy Bikes, 2017). This mechanism involves the users in the planning process. It uses modern machine learning techniques to create a relationship between human perception and urban planning. All the data used for the program were collected from surveys. Open Street Map (openstreetmap.org) is an open source map platform that leverages crowdsourced data (Smith, 2015).

Strava Metro Data Service

Strava is a social network for athletes (Strava, 2017). It uses GPS-enabled personal devices to fetch riding, running and walking data. In 2014, Strava launched a data service called Strava Metro. This

service provides anonymized movement data of cyclists and pedestrians to urban planners. These data are being used to optimize the bicycle counter placement and understand the usage patterns of cycle lanes. This dataset has information from over half a million cyclists and pedestrians. Over a million data points are also added every week. Over 70 organizations are currently using Strava Metro data for non-motorized infrastructure planning (Strava, 2014).

Waze Connected Citizens Program

Waze started the “Waze Connected Citizens Program” (CCP) in October 2014, which is a two-way data sharing agreement between DOTs and Waze. The program now has more than 100 partners including city, state and country government agencies, non-profits and first responders.

According to Waze internal data, 70% of the time Waze users report a crash on average 4.5 minutes before it is reported to emergency response centers through 9-1-1 or equivalent channels (Waze, 2014). The City of Boston experienced an 18% month-over-month reduction in congestion at key intersections due to the collaboration with Waze (Waze, n.d.).

Adoption of Crowdsourced Technologies in a TMC Workflow

The operations staff at Florida DOT monitor roadway conditions through Closed Circuit Television (CCTV) cameras and roadway detectors to disseminate traffic information through Dynamic Message Signs (DMS) and the 511 Traveler Information System. In order to provide data on interstate highways and arterial roadways, FDOT collaborated with Waze. This collaboration has enabled them to increase the quality and quantity of data. In March 2014, Florida DOT signed an agreement with Waze for two-way information sharing. Waze obtains data from the 511 feed, reported incident data and traffic conditions data are received as a single feed. Florida DOT obtains crowd sourced traffic alerts from the Waze feed and displays the data on its own crowdsourced map. Utah DOT has social media accounts on Twitter, Facebook, Flickr and

YouTube. They disseminate and receive information through multiple social media accounts for different regions and programs. They have an active blog as well. Utah DOT has deployed three mobile applications on android and iOS:

- **Click ‘n Fix:** Allows all citizens to report an issue by dropping a pin on a map at the location of the problem.
- **Citizen reporter:** Enlists volunteers to report on current road conditions along specific roadway segments.
- **General purpose traffic application:** Allows access to road conditions and traffic information on mobile devices.

Commercially Available Traffic Management Technologies

Several commercial tools for intelligent traffic monitoring have also been developed, the most noteworthy ones are INRIX, SCATS and Crowdsense+. INRIX provides analytics tools, connected car and predictive technologies. They have developed two mobile applications, namely INRIX Traffic and ParkMe. INRIX Traffic predicts traffic based on historical data and suggests the best route. It automatically learns the user’s driving habits and personalizes routes to avoid traffic congestion. ParkMe, as the name suggests, guides the driver to parking spots. Sydney Coordinated Adaptive Traffic System (SCATS) is an adaptive urban traffic management system that synchronizes traffic signals to optimize traffic flow across a city, region or corridor. SCATS allows the departments to implement complex, objective-oriented, traffic management strategies. In order to deploy the system, the department needs the following: a SCATS-compatible traffic signal controller, a centralized computer system to manage all traffic signal controllers, reliable communications network for the centralized computer system to exchange data with all traffic

signal controllers in the city, and vehicle detectors at each intersection, usually in the form of loops in the road pavement. Crowdsourc+ is a tool developed by Fehr & Peers which enables citizens to quickly report geocoded feedback for improvement of their community. The data are aggregated, supplemented by traditional in person meetings, and then reported. This reported data can be used to determine safe routes to school, campus plans, bicycle storage etc.

Summary

Crowdsourcing is a relatively new area of research, but it is already generating an enormous amount of interest among both researchers and practitioners and is finding applications in multiple domains. Crowdsourcing can reduce the gaps in information about important incidents and loss in level of service in the state wide transportation network.

Data can be collected from OSN consumers in three ways: active, passive and combined. The data collected can be used to predict travel time, suggest best routes, predictive analysis and report incidents to the authorities.

With advances in mobile computing and OSNs, the use of crowdsourcing has increased. The new applications offer functionalities such as crowdsourced incident detection, traffic monitoring and public participation.

The sheer volume of the information online can be overwhelming, especially in cases which are time critical. Several of the proposed solutions aggregate data from mobile sensors such as GPS, accelerometer and microphones to generate crowdsourced maps of real-time traffic (Mobile Millennium, iCartel2 and TrafficSense). GPS equipped public transportation networks such as taxis and busses have also been used to detect traffic conditions. The large volume of data being posted online is noisy, complicating the extraction of relevant data. The only way to verify the validity of the data is by manual intervention. Sometimes the location of the post is not consistent

with the actual location. Moreover, the initial learning curve for the staff is steep. For the successful implementation of crowdsourced applications, the user needs to be incentivized to download the application and regularly engage with the application. A social incentive based system which allows the users to meet using their phone location has been successfully tested in India (using a system termed “RasteyRishtey”). The user’s driving pattern can be altered by providing them financial and social incentives to compete against each other.

The social media presence of DOTs has slowly evolved. The staff actively engages with citizens on OSNs during office hours (e.g. Iowa DOT). In the recent past dedicated software has been developed to interact with citizens. They can report incidents using mobile applications developed by the DOTs (e.g. Utah and Florida DOT). DOTs have collaborated with commercial crowdsourcing platforms such as Waze. This collaboration is based on mutual data exchange and benefits (e.g. Florida DOT). Despite the increase in using crowdsourced data by the DOTs, there are several barriers that need to be overcome.

CHAPTER III. CHALLENGES AND LESSONS LEARNED FROM PRACTICE

Introduction

The literature review suggests that although the TMCs are increasingly becoming aware of crowdsourcing techniques and are supplementing data collected through their ITS architecture with social media monitoring systems, a distinct gap between the state of the art and the state of the practice exists. In order to bridge the gap between the two, it is absolutely essential to study current systems that use crowdsourced data, as well as the architecture and challenges associated with these systems. Through this process, the state of the practice could be brought closer to the state of the art.

In order to better understand the crowdsourced data collection practices currently in place at DOTs around the country, the project team conducted a series of telephone interviews. The interview schedule was finalized with 11 DOTs/TMCs via email. In preparation for the interviews, the project team designed a questionnaire based on the findings from the literature review and preliminary analysis. The majority of the questions included in the questionnaire were open-ended in nature, which allowed the flexibility to discuss related topics during the interview.

In order to conduct the data analysis, the project team quantified the collective interview responses. The analysis provided insights into current crowdsourced technologies used by the DOTs as well as possibilities for the future.

Questionnaire Design

The questionnaire was designed to deepen the research team's understanding of both successes and challenges encountered by traffic management personnel from other states. It was revised in consultation with the project advisory team from Georgia DOT. Questions were designed to gather

information on existing systems, implementation details, and future plans for crowdsourcing and social media at the DOT. If a DOT was not yet utilizing crowdsourcing for traffic management operations, that was also noted. The interview responses were incorporated in this report to provide GDOT with lessons learned from the experiences of TMCs in other states.

For the DOTs which had mobile phone based applications, questions to quantify the number of active users, number of downloads (tabulated separately for android and iOS) and the growth of the users over time, were also included. This was done to understand what type of applications appeal to users and, from this, how higher download rates could be achieved. Several different techniques to incentivize application users were outlined in the literature review. The interviews also included questions regarding techniques to successfully incentivize users.

Focusing on technical details, the team included specific questions on system architecture, volume and type of data being collected. This information helped with the analysis of the complexity of the system and was then correlated with the adoption of the system at the TMC.

To understand the agency's perspective, questions were included inquiring about the most popular feature and how the collected data are being used. This section also focused on the response and acceptance by DOT officials. The team also collected information about how the workflow at the DOT had changed after the implementation of crowdsourced traffic management and how the changes in workflow could be subsequently designed for the greatest benefit.

The next questions were aimed at understanding the primary challenges faced by the TMC in real life scenarios. These questions involved challenges in filtering the data, judging the credibility of the information, and training agency employees on the use of the system. The questions also

focused on how these problems are mitigated. Responses to these questions assisted the research team in determining the best practices to address common challenges faced by TMCs.

The final part of the questionnaire focused on the future plans and any interview specific questions, including any problems with the current system and what was anticipated for the future of crowdsourcing at the DOT.

Selection of Respondents and Data Collection

The team selected 21 TMCs managed by state DOTs across the US as potential respondents. Some of the DOTs already have established crowdsourced data systems, while others are in the process of setting up such systems. The selected DOTs had different types of crowdsourcing mechanisms in place. DOTs from both cold and warm climates were included, as challenges faced by DOTs in different climates are drastically different. Interviewing DOTs with varied crowdsourcing experience helped the research team to gain an in-depth understanding of both the real world challenges faced by agencies that have implemented crowdsourcing and the expectations of agencies that have not. Figure 1 shows the 21 states with TMCs that were invited for an interview. Not all invited TMCs were able to participate due to time constraints; ultimately, 11 TMC interviews were scheduled (Figure 2).

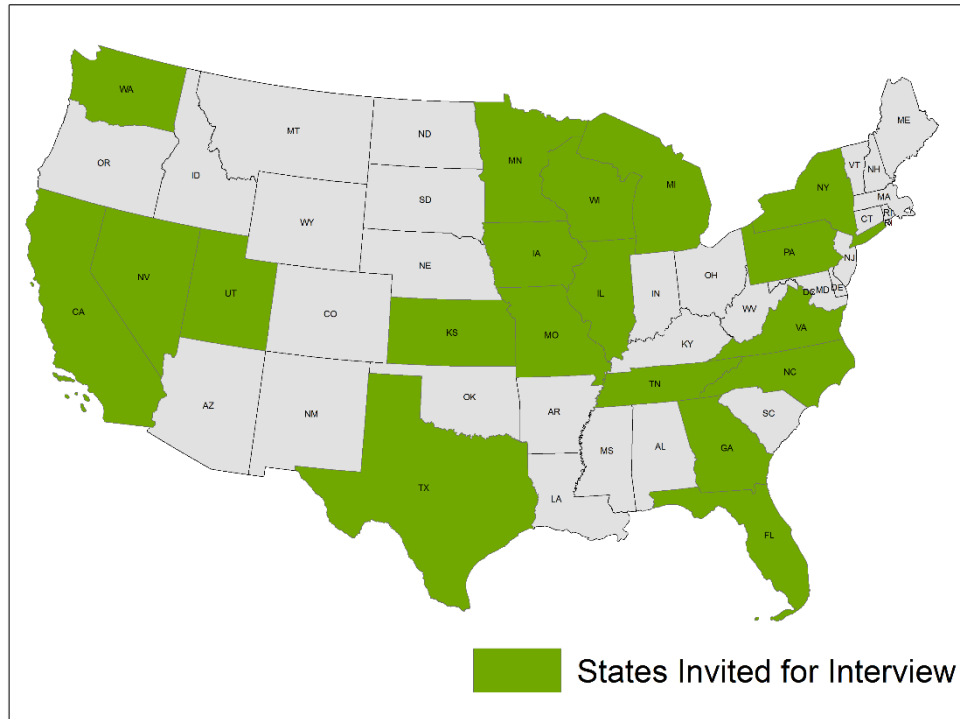


Figure 1: States with TMCs that Were Targeted for an Interview

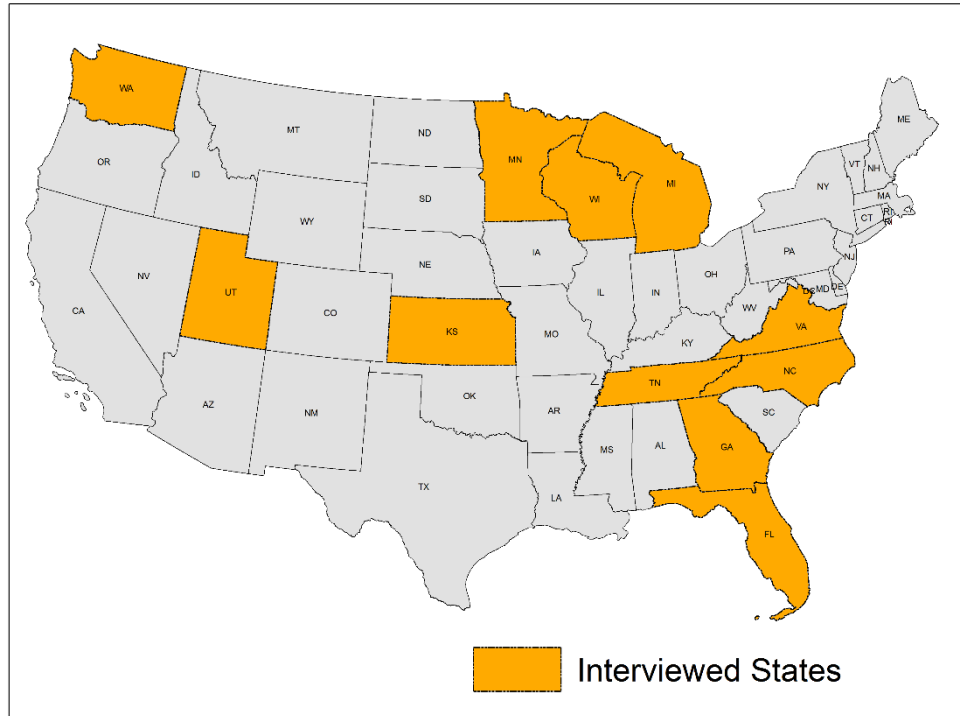


Figure 2: Interviews Conducted with TMCs Located in these States

Analysis of the State of Practice

Responses from the interviews reveal that different traffic-related problems are experienced by different states. DOTs are using or are in the process of using modern technologies to tackle these problems. This section details the use of these measures and technologies.

ATMS and ITS

All of the DOTs/TMCs interviewed have deployed an Advanced Traffic Management System (ATMS) and have a well-built Intelligent Traffic System (ITS) architecture to improve highway travel within the state. These systems typically fetch information regarding incidents using traditional technologies such as cameras, sensors and loop detectors. In general, cameras are used to monitor traffic and verify reported incidents. Georgia DOT has deployed a pilot project in which a video detection system automatically detects incidents in the video footage. However, it is not fully adopted yet, since it has raised multiple false alarms. Sensors such as microwave radars, video detectors and loop detectors are used to gauge the traffic flow. Dynamic message boards or variable message signs (VMS) are used to quickly disseminate information to the highway network users.

511 System

As shown in Figure 3, 10 out of 11 DOTs are currently using a 511 system. The 511 system is differently configured for different states. For Wisconsin DOT, the 511 system is a type of brand name that comprises multiple systems including the IVR, Twitter, phone system and mobile application. For the majority of DOTs, the 511 system only includes a website and phone system. Three DOTs out of those interviewed allow for two-way dissemination of information through the 511 system. The Georgia DOT website has features such as changeable message signs, breaking news, accidents, Highway Emergency Response Operators (HERO) information and construction

updates. The 511 email and text subscriptions are popular among Georgia residents. Georgia DOT has 13,000 registered users and receives 20,000 calls on average per week. Washington DOT received 786,000 calls in 2016; this volume has since declined with the increase in use of OSNs. The volume of calls significantly varies with weather and typically surges during times of adverse weather and natural disasters. Tennessee DOT experiences a 50% increase in call volume per day during the winter. The calls increase from approximately 1,200/day to 1,800/day. North Carolina DOT received around 120,000 calls during hurricane Matthew.

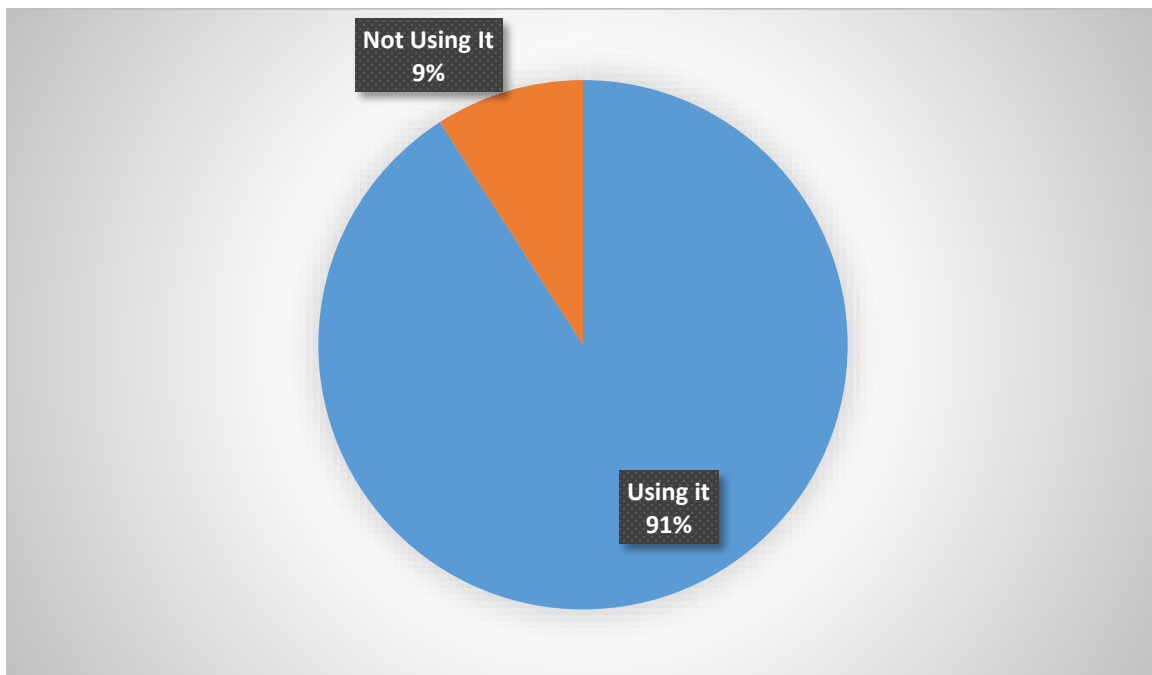


Figure 3: Usage of 511 Systems by DOTs

Online Social Networks

Interviewed agencies have also been crowdsourcing data using platforms like Twitter, Facebook and Waze. However, only five out of the 11 TMCs interviewed reported having followers on Twitter. Figure 4 shows the number of Twitter followers as reported by five DOTs. The average

number of followers of these five TMCs' Twitter accounts was found to be 60,920 (based on historical data provided by interview respondents), and the average number of "likes" on Facebook reported by the TMCs is 26,233. Figure 5 shows the number of Facebook page likes as reported by TMCs/DOTs.

These social media platforms are being used for not only information dissemination, but also for increasing public participation. Figure 6 shows the adoption of OSNs and other crowdsourced platforms among the 11 interviewed TMCs. Figure 7 shows the split between DOTs using OSNs for one-way dissemination of information versus those actively using them for two-way interaction. DOTs create separate accounts for different regions and special projects. Minnesota DOT's Twitter feed is automatically populated with all incident information, and it is not manually monitored. On the other hand, Michigan DOT's media representative reviews OSN posts on a one-on-one basis. Twitter usage is significantly higher than Facebook because of the real-time nature of the platform. Twitter allows for quick dissemination of information during time critical events. Kansas DOT encourages citizens to post online with certain hashtags like #NEKansasTraffic, this allows the social media managers to quickly filter the data. DOTs make use of third party software to quickly filter and analyze the content posted online. Kansas DOT and Tennessee DOT use Tweetdeck and Hootsuite, respectively, for this purpose.

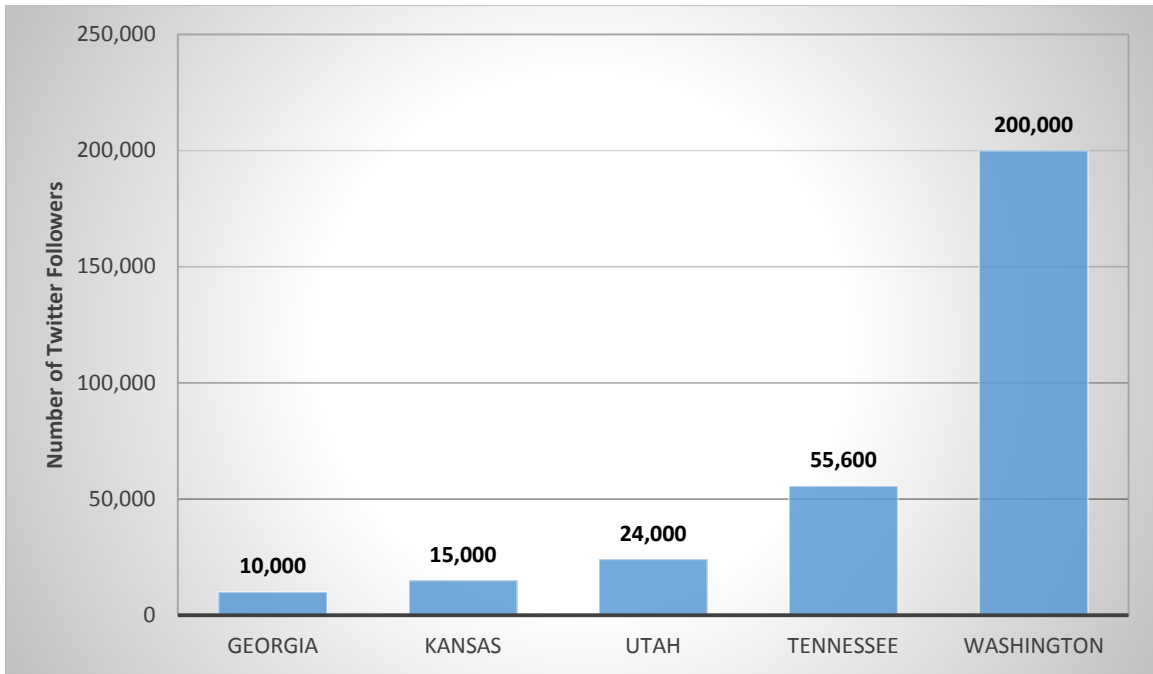


Figure 4: Reported Number of Twitter Followers by Interviewed DOTs

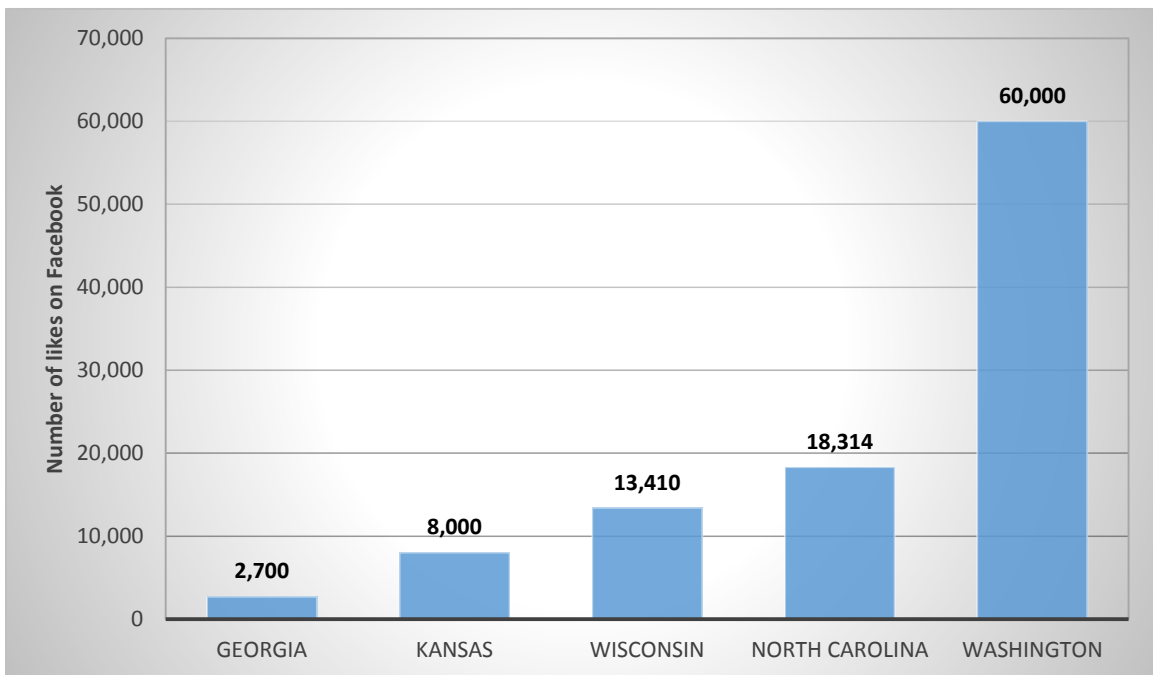


Figure 5: Reported Number of Likes on Facebook for Interviewed DOTs

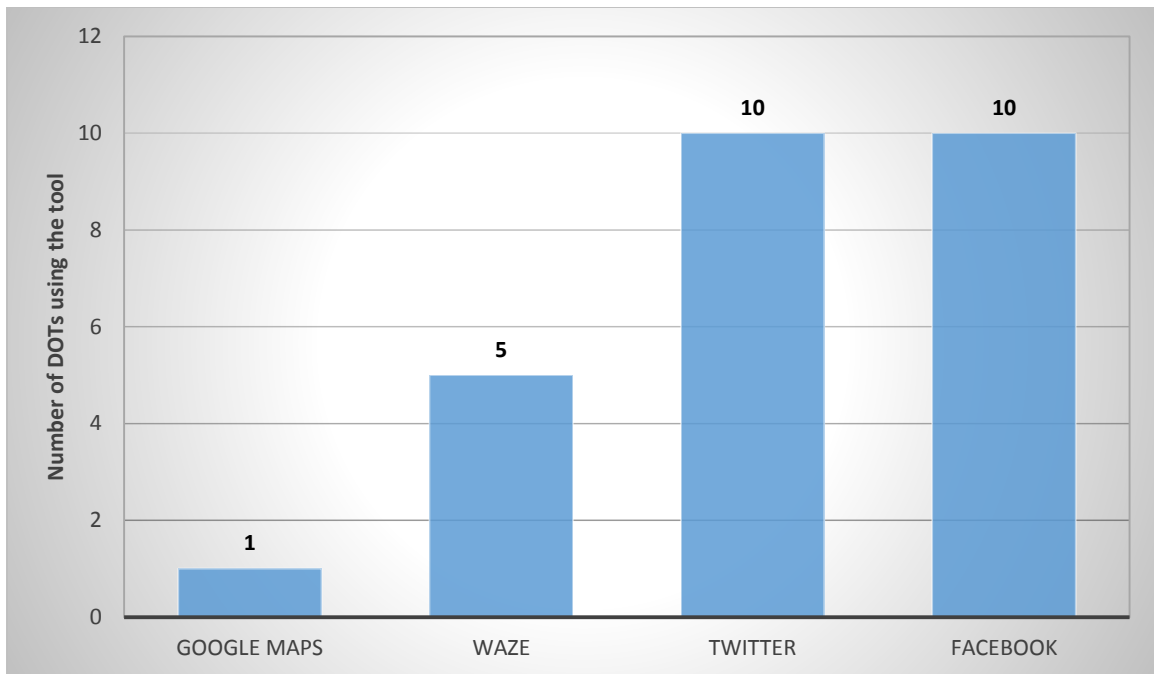


Figure 6: Adoption of Crowdsourced Tools by DOTs

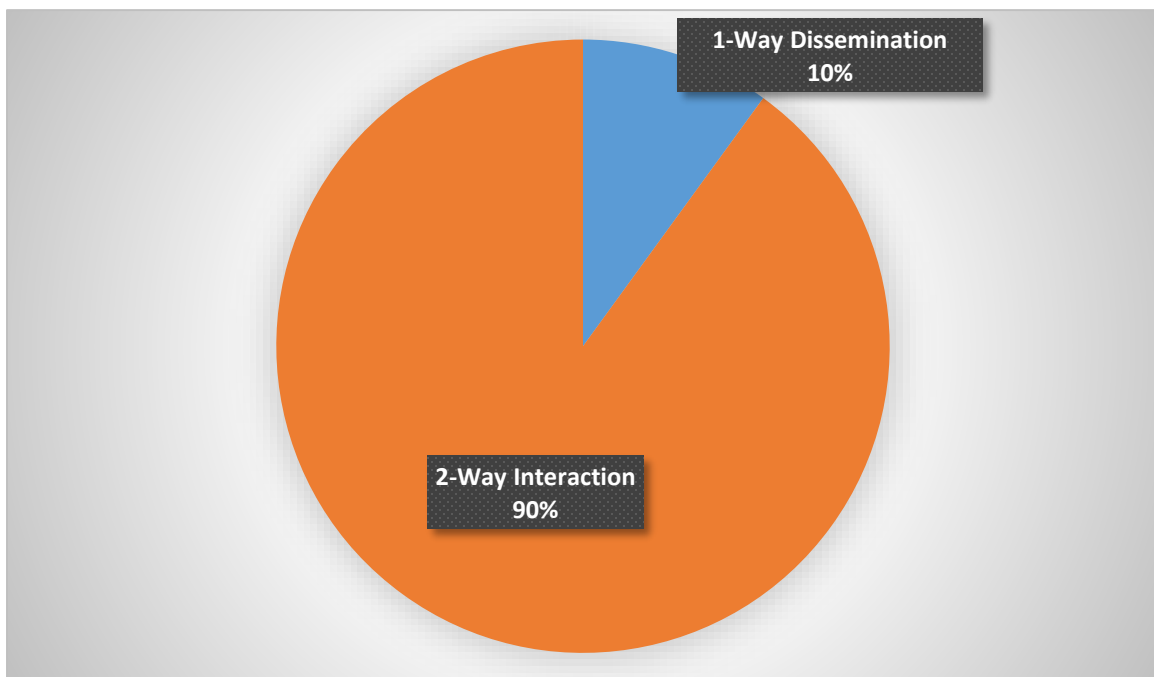


Figure 7: OSN Usage by DOTs

Waze Connected Citizens Program

Since its launch, the Waze “Connected Citizens Program” has expanded to more than 80 partners including city, state, and federal agencies, non-profits and first responders (Waze, n.d.). There are several advantages to the Waze program. Waze provides access to a large set of users (known as “Wazers”) who help to determine the credibility of a piece of information. Additionally, the Waze algorithm to filter duplicates and approximate the exact location of an incident is particularly useful. Usage of Waze varied across the DOTs interviewed. For example, the Tennessee DOT only posts information from Wazers that exceed a certain reporting threshold. Michigan DOT uses Google Maps instead of Waze because they feel that Waze has a lot of noise. Georgia DOT is using Waze, and finds it helpful to fetch information from rural areas, while Washington DOT does not use Waze because they feel that it is only beneficial for areas with high density or urban areas. Kansas DOT actively monitors Waze but is not a part of the connected citizens program.

Dedicated Platforms Developed by DOTs

Several DOTs have developed their own tools in the form of a mobile application or an interactive website (see Figure 8). Developing a platform unique to individual DOTs has pros and cons. On the one hand, it allows the DOT to have greater flexibility and customizability of the data being collected. All relevant data can be collected, in the exact format desired, including geotags and images. The data can then be easily funneled into existing systems. On the other hand, there are associated development and maintenance costs.

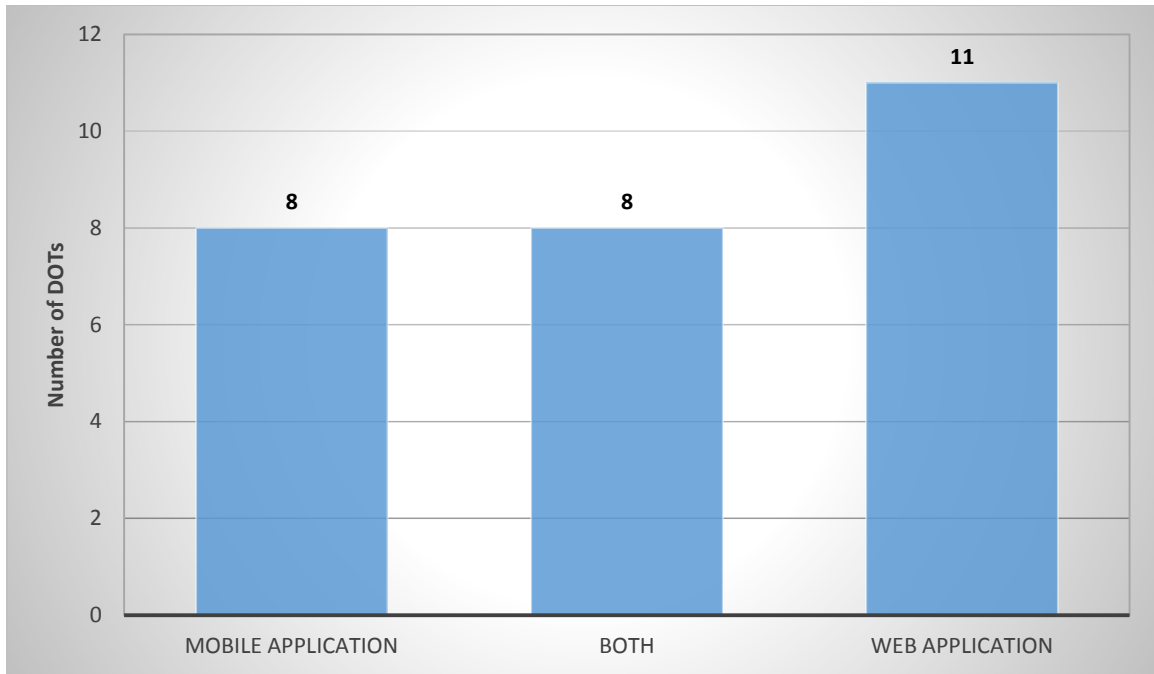


Figure 8: Number of DOTs Using Mobile Platforms, Web Platforms or Both

Figure 9 presents the most popular platforms in different states. Phone-based systems are most commonly used, while mobile applications are the least. However, DOTs reported that as the use of OSNs and mobile applications increases, the number of calls they receive declines. One example of the distribution is shown in Figure 10. It illustrates the hits on the Mi Drive system (Michigan DOT's traffic information and reporting system) in 2016 for the three different platforms namely, the mobile website, web application, and the mobile application. Citizens use the mobile website the most, and the number of hits during the winter season is significantly higher.

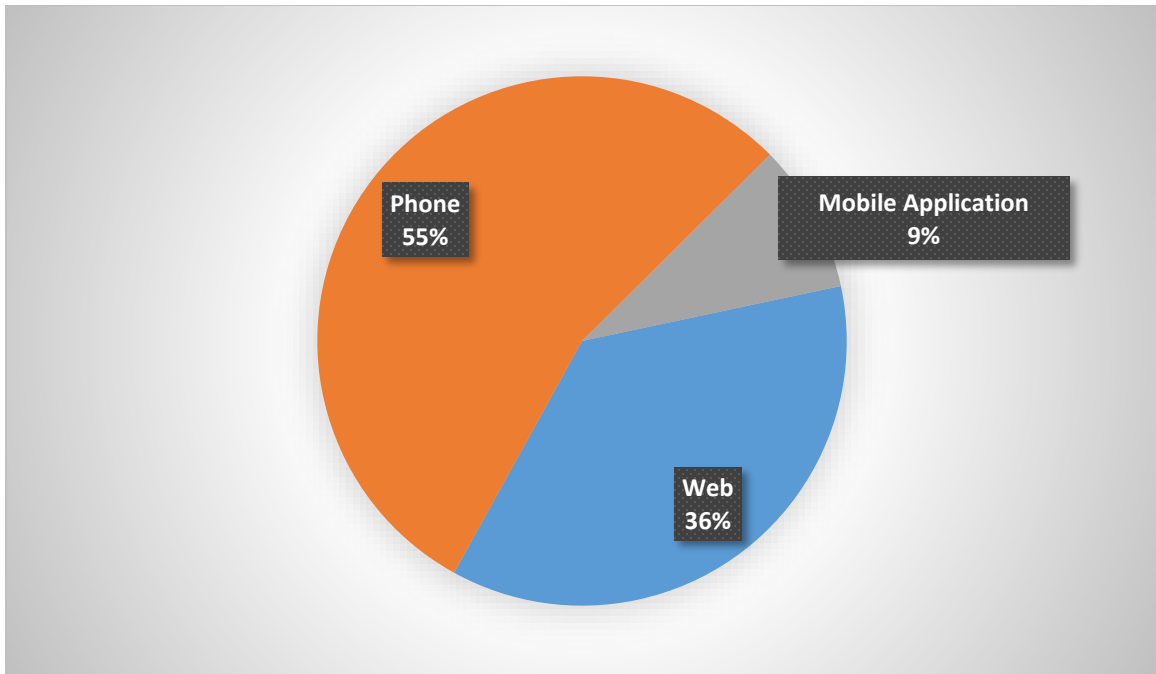


Figure 9: Usage of Various Platforms Based on Interview Data

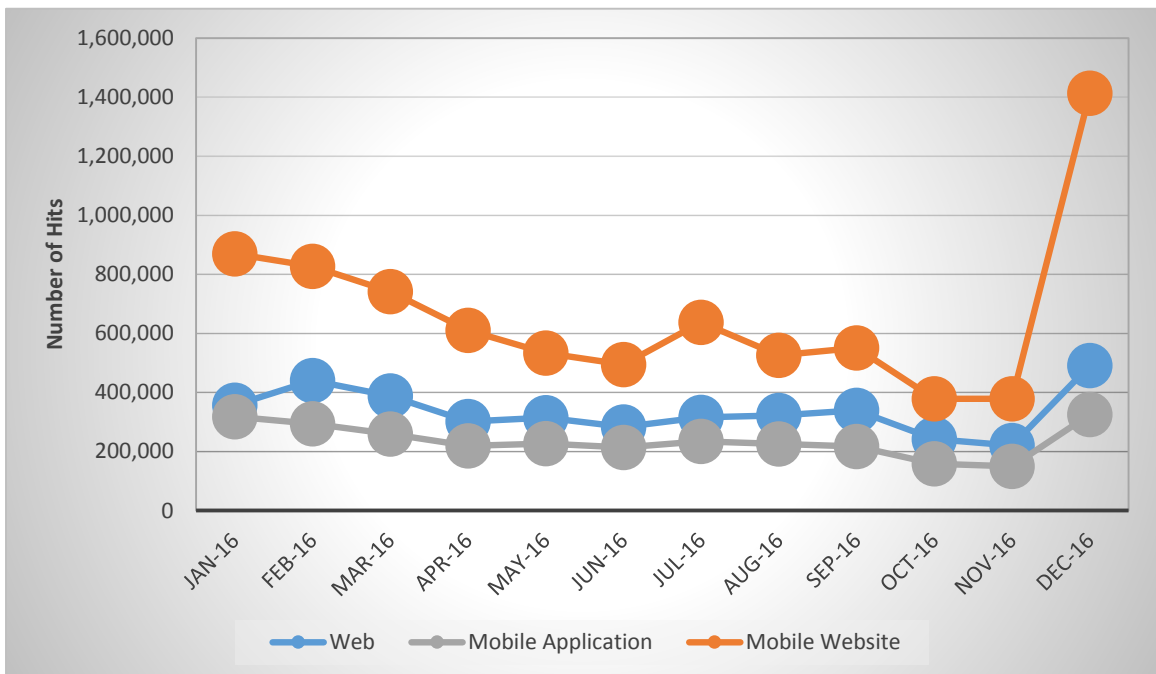


Figure 10: Hits on Mi Drive by Platform Type

As another example, the Utah DOT has developed two mobile applications, one for user reporting and the other for traffic updates. User reporting has been very effective in Utah. As seen in Figure 11, Utah DOT received over 1,800 road condition reports during November 2013. During a winter storm in December of that year, over 130 reports were submitted in one day.

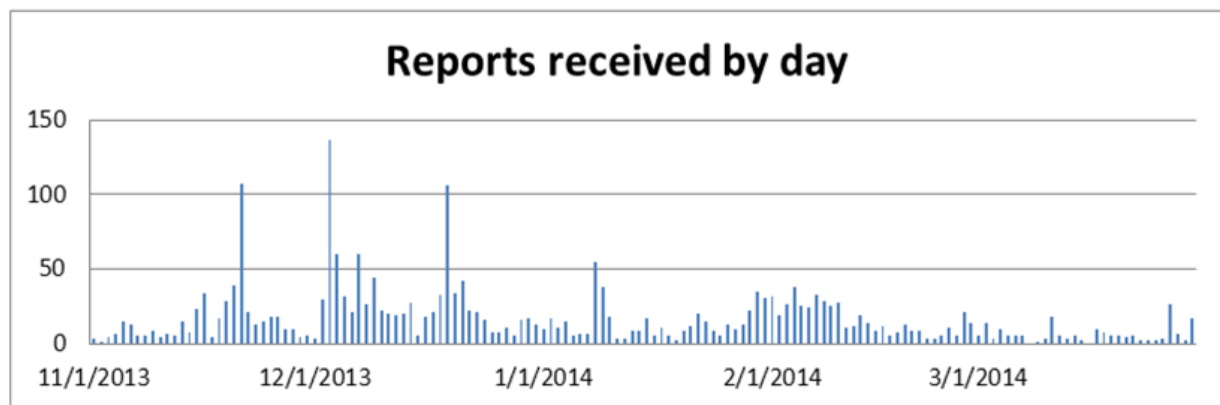


Figure 11: Variation in Incident Reporting by Day (Source: UDOT (2014))

Most DOTs use their mobile applications for one-way communication. Figure 12 shows the percentage of DOTs that use their mobile application for two-way interaction with residents. Georgia DOT primarily uses their mobile application to push out information to its 40,000 active users. Therefore, the custom applications developed by other states provide an example of effective two-way interaction which could be useful for GDOT. Both the Minnesota and Utah DOT's applications include an easy to use reporting feature to specifically gather information about road conditions, weather conditions, and other traffic-related issues. Application simplicity and ease of use is important to improve report quality. For example, Mi Drive has a quick access option to report that you are "Stuck in Traffic." In order to maintain the quality of reports, only users who have undertaken a short training are given access to this feature.

To compare application use, Figure 13 shows the number of users of mobile applications on all mobile operating systems as reported by three state DOTs; the mean value is 290,666. This form of reporting is of a higher quality compared to social media reports, which can vary significantly.

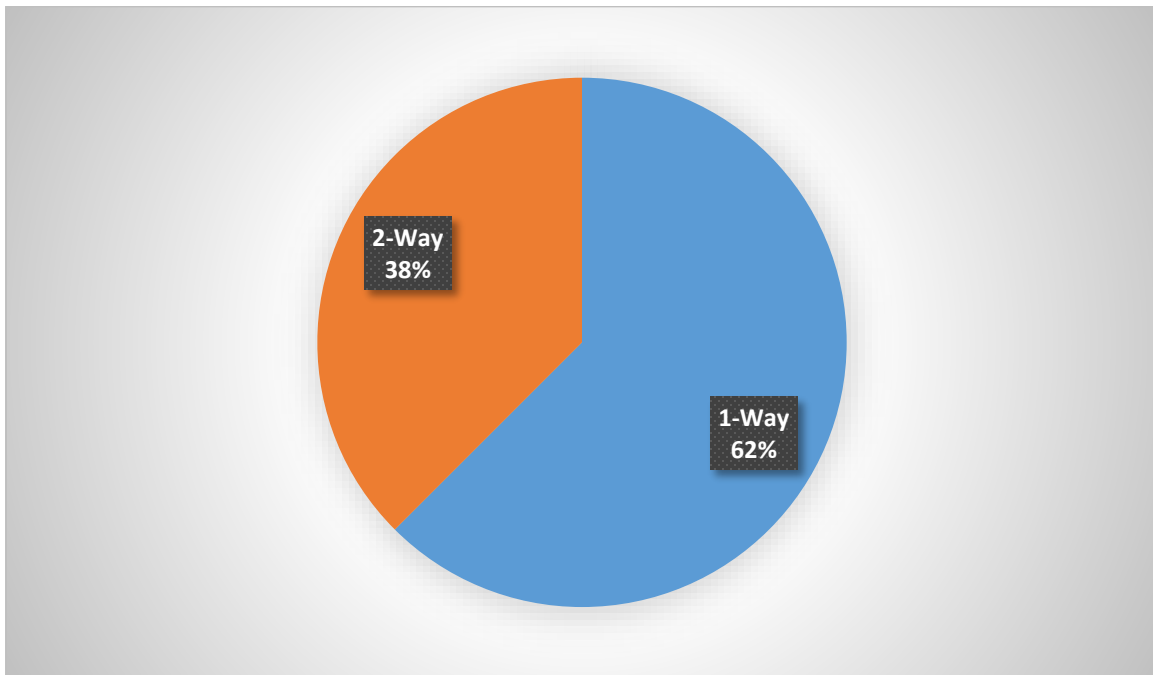


Figure 12: TMC Interaction with Citizens

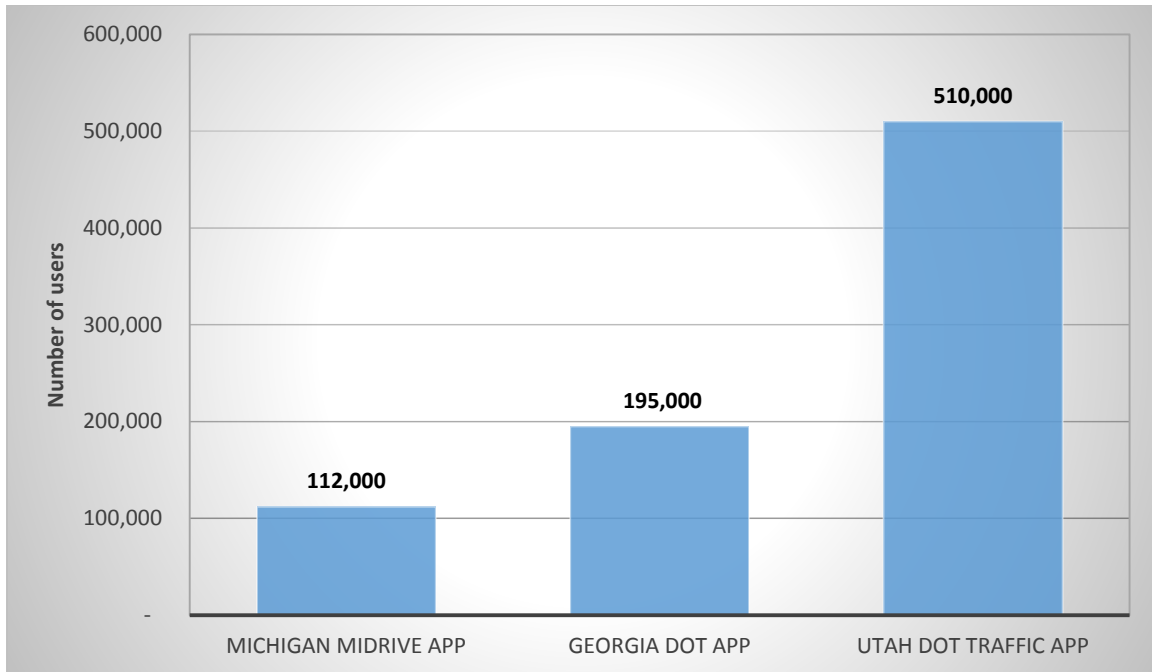


Figure 13: Number of Users for DOT Mobile Applications on All Platforms

As another example, the Kansas DOT uses a customized application, Kandrive.org, as their primary system. The Kansas DOT 511 system feeds into Kandrive.org as well. In Washington and Florida, the web platforms are the most used platforms. Despite the popularity of these web systems, none of the agencies interviewed reported having any incentivization models in place to increase their use.

Interviewees also shared details regarding the data collection process. As soon as an incident is reported or detected, a location is established for the incident and the motorist's information is obtained. Next the TMC staff is notified, including first responders and maintenance workers. The warnings and alerts are put out on the website, 511 system and mobile applications. The DOTs with multiple systems also use the Application Programming Interface (API) to disseminate the information across systems. The local media are then informed. Some DOTs have automated the process of informing the media. For example, Georgia DOT provides live access to all local TV channels. If there are a surge of incidents, priority is given to the ones that affect driver safety (e.g.

on-street crashes). Incidents like construction lane closures and stalls on the shoulder are given the lowest priority during a surge of incidents. Number of users of information disseminated through each platform varied by states. For example, Figure 14 shows the users of web platform for Georgia and Michigan. Some DOTs reported that they have appointed dedicated staff (see Figure 15) to manage crowdsourced channels such as social media, and especially to handle data surges.

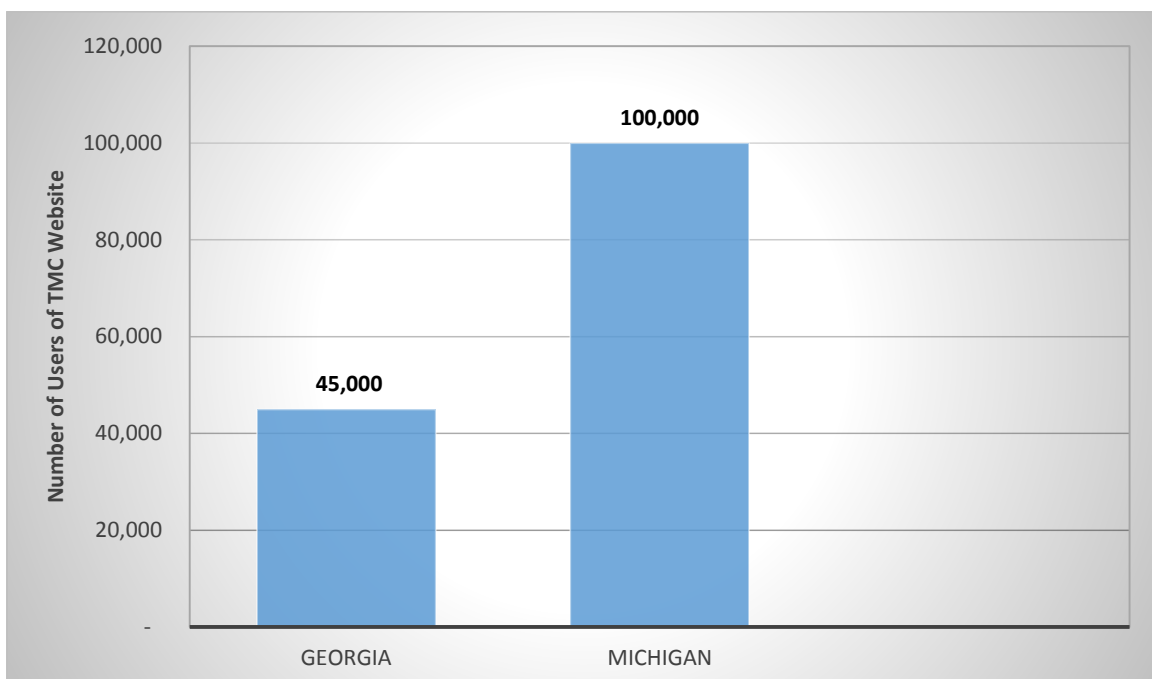


Figure 14: Number of Users for Web Platforms

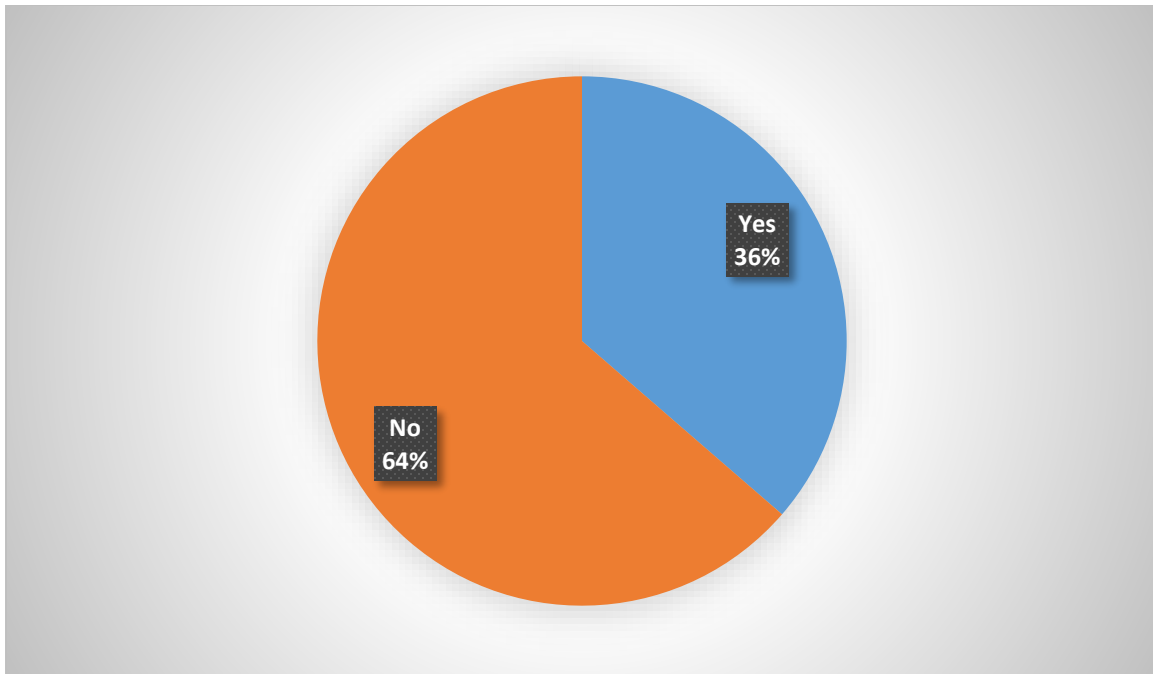


Figure 15: Percentage of DOTs with Dedicated Staff for Social Media

Challenges faced by TMCs

This section details the primary challenges faced by TMCs in performing daily tasks and during critical times. Since TMCs monitor and manage traffic flow and communicate information between different agencies, their preparedness to tackle critical events is vital to the healthy functioning of the state transportation network.

Cost of Systems

Cost is a limiting factor for DOTs when deciding which system to use. Interviewees indicated that ITS deployments are expensive to build, but provide robust data once deployed. The existing ITS in Georgia deployed by Georgia DOT cost over \$1 billion over the course of 20+ years. Moreover, building these systems is time-intensive, with most ITS projects let in Georgia taking 18-24 months to construct. On the other hand, systems to fetch crowdsourced data can be built at a fraction of the cost and time. If executed effectively, crowdsourcing can provide rich data. As one interview

respondent from the Tennessee DOT explained, “We collaborated with Waze because it didn’t cost us anything.”

Multiple Data Sources

Cameras provide another method for state DOTs to monitor traffic conditions and incidents. However, cameras present their own challenges. Monitoring feeds from multiple data sources is extremely difficult. Also, the extent of camera coverage varies by state. Some states have poor camera coverage, whereas other states have much more. Out of the states interviewed, Georgia and Kansas have fairly extensive coverage, with 750 and 456 cameras deployed, respectively. Monitoring all the cameras simultaneously is virtually impossible. Therefore, TMCs typically use some mechanism for prioritizing camera feeds based on the importance of locations and the historical frequency of incidents, while other state DOTs switch camera feeds manually. At the Atlanta TMC, 20 cameras are on the screen at any given time. Interviewees indicated that cameras are mostly used by TMCs to verify reported incidents rather than to identify incidents. Only occasionally has the camera feed led to incident identification, yet some priority measures have been put in place in order to optimally view the video feed. Georgia DOT is currently testing a software in exploratory mode that can potentially use video feeds from CCTV cameras for automated incident detection. Other DOTs have a process in place to open up the video feed for reported incidents, historical data, and real-time speed data.

Workflow after an Incident

When an incident occurs, several tasks need to be performed. The first responders need to be notified, warnings need to be sent out and the media needs to be informed. Interview respondents from the Michigan DOT felt that reaching out to external agencies is the biggest challenge in such scenarios. Additionally, the tasks which are not fully automated cause delays in the response.

Variable Volume of Incidents

As mentioned earlier, the volume of incidents varies significantly throughout the year. During times of natural disaster and bad weather, a higher number of incidents occur. Wisconsin DOT experiences a significant increase in the number of incidents during the winter due to poor road conditions. This wide seasonal variation in the number of incidents creates staffing challenges.

Network Coverage

Interview responses from most TMCs indicated that the extent of camera coverage is poor and that covering all roads in the state is not possible. Verifying or detecting incidents in areas of no/low coverage is a problem. Such incidents need to be manually verified. Moreover, because of hands-free laws, phone-based systems are not a good option while driving, especially with single occupancy vehicles.

Challenges for Crowdsourced TMC

Large Volume of Data Generated

Crowdsourced systems generate large volumes of data. Moreover, crowdsourced data are being generated from multiple sources. Short-staffed DOTs find it difficult to identify and dedicate the necessary resources to analyze these data. Figure 16 shows the percentage of DOTs that expressed concern about the large volume of data being generated. Additional challenges include situations where multiple users report the same incident in different ways, as well as when reported incidents go unacknowledged by DOTs. This second situation could lead citizens to lose confidence in the system and reduce their participation. If DOTs are unable to respond quickly on OSNs such as Twitter and Facebook this could also deter participation. Another challenge is that the data collected on OSNs need to be scrubbed and grouped, which also requires resources. On a related

note, all DOTs focus on incidents which threaten human life. Effectively filtering and prioritizing these types of incidents from a large volume of data is a difficult challenge.

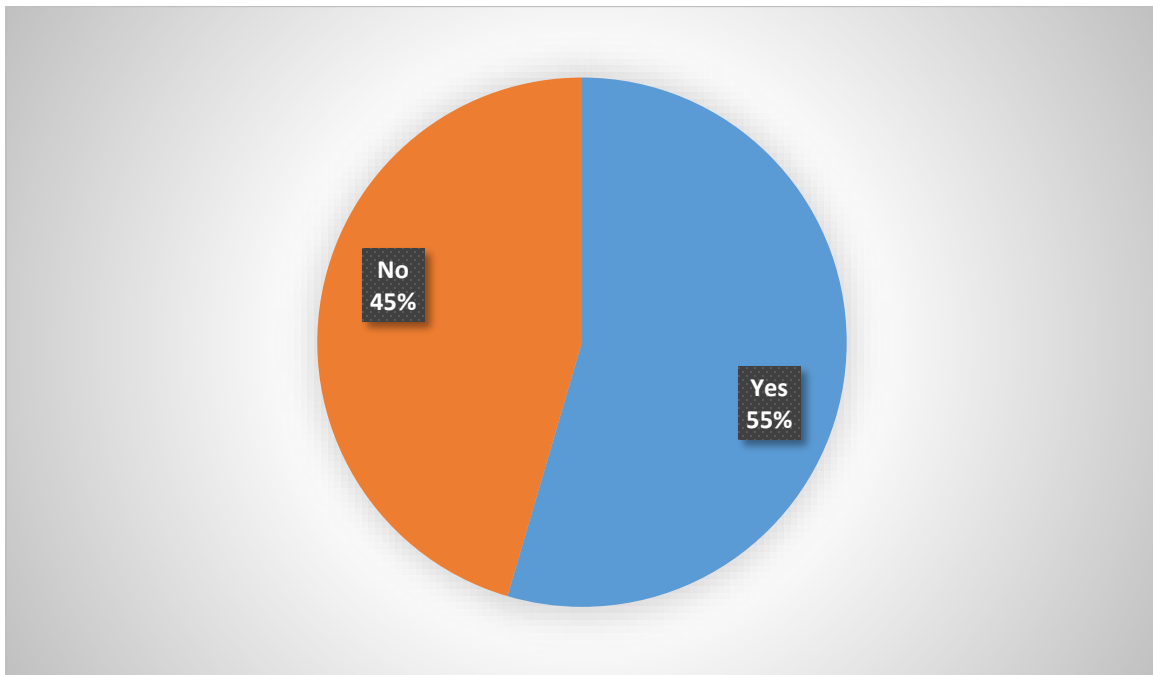


Figure 16: Concerns about Volume of Data Generated

Confidence in Data

Most TMCs expressed some concerns about the credibility of crowdsourced data, although interviewees stated that malicious and/or false reporting had not actually been observed. Often reported incidents do have the wrong location, direction or are missing important details, but this appears to be unintentional. Mobile application users report incidents as they pass by, but by the time the report is generated their location has changed. In response to this challenge, DOTs are developing methods to improve data reliability. The Tennessee DOT uses Waze points for this purpose. Another similar problem results when multiple citizens post about the same incident with different details; in those cases, it is difficult to determine which one is correct and how to combine the related information from multiple reports. A representative from the North Carolina DOT's TMC pointed out that, "Detection is not verification." Because of the issue of confidence in

crowdsourced data, all incidents need to be verified either using cameras or manually by state patrol. This verification process limits the scalability of the system.

However, regardless of the issue of confidence in reported data, some DOTs are turning to crowdsourced systems to obtain information in areas with poor coverage—like rural areas. Conversely, an interview respondent from the Tennessee DOT expressed that systems like Waze are more useful for high density metro regions and the credibility of information generated from rural areas is low.

Interview responses from multiple TMC personnel indicate that the first 30 minutes after an incident occurs is the most difficult time to find credible information and determine what has happened in areas without camera coverage. North Carolina DOT uses “congestion” as a default reason in incidents where no reliable information is available regarding another underlying cause.

Gaining User Engagement in Crowdsourcing Systems

For a crowdsourcing solution to be successful, the client application needs to be useful, non-invasive in terms of privacy, and should require minimal user interaction (Frank et al., 2014). Interviewees indicated that users will only engage with the application if they see a benefit from doing so. The mobile application should have low energy requirements for the user to be able to regularly report incidents. Privacy is also a major concern with regards to crowdsourced data. Personally identifiable information should not be stored. If this type of data is required, it should be encrypted.

Lessons Learned from Practice

This section discusses the lessons the research team learned by studying how the TMCs operate. The most useful ones are as follows. Given the limited number of employees, a system which integrates all the channels of crowdsourced information is required. This system needs to automate

the entire workflow of the TMC. The historical traffic data fetched by the system can play a role in defining the future strategy of DOTs. Quality assurance procedures need to be established to increase the reliability of the data. A mobile solution will help to increase citizen participation. Moreover, a mobile system for TMC employees to access the system would be beneficial.

One Stop Solution

The volume of incident reporting and limited TMC staff means that it is difficult to handle incident detection and verification manually. An integrated single system is necessary. This will allow the DOTs to optimally use their existing resources and personnel. Integration into one system would also assist with detecting duplication.

Historical Data

Archiving and analyzing past data can help determine which crowdsourced system would be most suitable for a state. For example, in a state where a large number of reported incidents have the wrong GPS location, users can be asked to input the location data along with the report. Alternatively, location can be corrected by reviewing historical errors and comparing that to current average speed and direction near the reported location on the road. For example, Minnesota DOT uses past data to define normal traffic volume and speed. The historic data are compared with the real-time traffic data to identify anomalies. Virginia DOT leverages past traffic patterns in a similar way, to detect errors. These data are also used to allocate resources.

Automating the TMC Workflow

Some DOTs (for example Tennessee DOT) do not have an automated mechanism to inform the media about incidents. Once an incident is reported and verified, all the notifications and alerts such as informing the local media, sending out internal notifications, and alerting the first

responders need to be automatic. This workflow will not only minimize unnecessary burden on the operators, but also result in a quicker response from the first responders.

Training the Staff

For crowdsourced systems to be successful, it is essential to train the staff. Training should include both instruction on the use of the system, as well as why it is beneficial to use it. Staff members will be more motivated to use such systems once they understand the benefits.

Quality Assurance

Every DOT interviewee expressed concerns about data credibility. More advanced citizen reporting programs can be established, which involve extensive training before allowing the user access to the reporting channels. DOTs can then prioritize reports from those users who have completed the training.

Mobile and Offline Access

A mobile platform should be deployed to increase citizen participation. As per interview responses, smartphones provide a powerful tool to capture the full potential of crowdsourcing, and allow the public to contribute to complex problem solving. This finding is echoed by researchers (Chatzimilioudis, Konstantinidis, Laoudias, & Zeinalipour-Yazti, 2012). A mobile solution allows citizens to file reports from anywhere, moreover the GPS coordinates from the phone can be used to determine the location of the incident. A mobile solution should be implemented for DOT officials to log into the system as well. This will allow operators to access and verify incidents on the go. SMS reporting should be deployed in areas with low connectivity and poor signal coverage.

Dissemination of Information

Both the retrieval and dissemination of information needs to be efficient. Drivers can make better decisions if they are well informed in real-time. Drivers can alter their behavior to avoid incidents, and therefore not add to the congestion already existing in a location.

Summary

The objective of this study was to understand what methods peer DOTs were using to collect, clean, and analyze crowdsourced data, and to learn from their successes. A total of 11 DOTs participated in the semi-structured telephone interviews conducted by the research team. The questionnaire developed for the interviews focused on understanding various aspects of state TMC operations such as current technologies, systems using crowdsourced information, and workflow. Participant responses highlighted the gap between the state of the art and the state of the practice. The majority of the TMCs interviewed are only using OSNs to obtain crowdsourced information. The cost of these systems, multiple data sources to reconcile, and limited number of employees and resources posed a challenge to the TMCs. Interview respondents raised concerns about the large volume of data generated through crowdsourcing and the reliability of those data. Presently cameras are used to verify the information before acting on it. Waze's connected citizen program has been extremely popular and beneficial for many state DOTs/TMCs, particularly in mining information from areas with poor coverage. The responses suggested the need for an integrated solution, quality assurance procedures and automation of the TMC workflow for a quicker response.

CHAPTER IV. CROWDSOURCED TMC FOR GEORGIA: NEEDS ANALYSIS

Introduction

This chapter analyzes the need for and the utility of crowdsourced systems for traffic and incident management for Georgia DOT's TMC. These recommendations are based on six one-on-one interviews conducted with TMC personnel, contextual inquiry during the Atlanta TMC tour, and observations by the research team. During the tour, the team met with 511 operators, personnel from the Regional Traffic Operations Program (RTOP) and the Social Media Manager. During the one-on-one interviews, the team interviewed an Assistant Traffic Specialist, Traffic Specialist, Operator, Operation Supervisor, TMC Manager, and an Operator II (a part of the HERO program) at the TMC. Figure 17 shows a panoramic view of the TMC from the team's visit. The TMC is well equipped with modern technologies and has been efficiently designed for a smooth workflow. Figure 18 shows the work spaces of TMC operators with access to modern technologies and state of the art software.

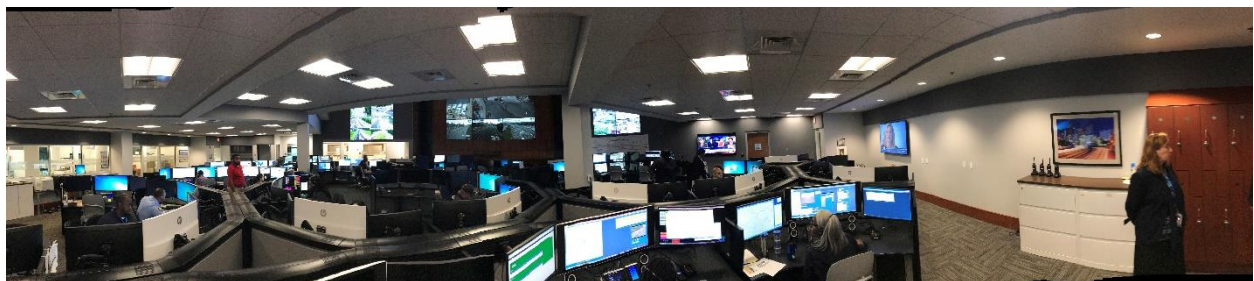


Figure 17: Panoramic View of the GDOT TMC



Figure 18: Workspaces of GDOT TMC Operators

The remainder of this section of the report is organized as follows: first, the research team provides an assessment of the primary challenges faced by Georgia DOT in the TMC operation. Then the results of the SWOT (strengths, weaknesses, opportunities and threats) analysis and a proposed system for Georgia DOT are presented. Next, the benefits and disadvantages of the system are described and probable solutions to overcome the challenges are suggested. Finally, the action items are prioritized to specifically propose a method to manage the high volume of incident reporting that TMCs are experiencing.

Assessing the Georgia DOT TMC Operations

The primary responsibilities of the TMC include managing incidents, controlling reversible lanes, posting messages, and dispatching HEROs. This section details the challenges and requirements of the TMC.

Assessing the Challenges

The TMC faces several challenges in their everyday operation. Cameras are an integral part of their workflow. The GDOT TMC has access to 750 Interstate cameras (plus hundreds more arterial cameras), and ensuring that all the cameras are operational is a challenge. At any point in time, Georgia DOT has 21 camera feeds being projected on the large front screens (Figure 19-21). The operators are only able to view a limited number of cameras at once. The TMC is presently experimenting with a software that automatically detects incidents from the video feed and brings up the images from the camera located at the site of the identified incident. However, this system is still under development, since it currently produces a large number of false incident reports.



Figure 19: Live Feed of High Definition Cameras at GDOT TMC

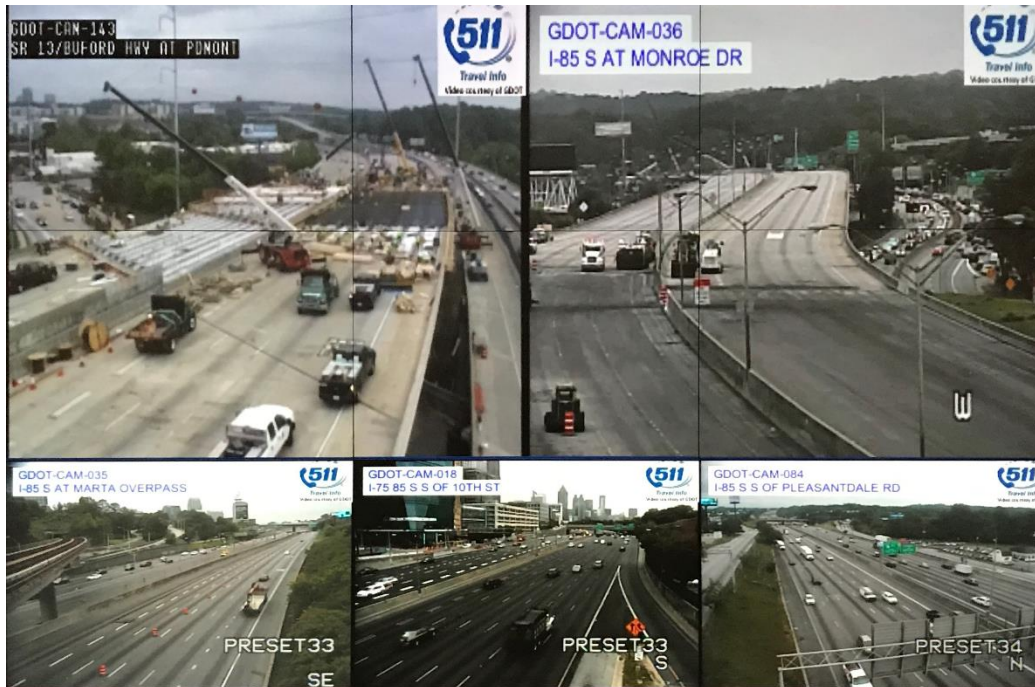


Figure 20: Live Feed of Cameras near the March, 2017 I-85 Bridge Collapse Site



Figure 21: Live Feed of Cameras at Strategic Locations

Operators stated that there is a tremendous increase in the number of calls during major incidents and adverse weather, and the operators at the TMC find it difficult to keep up with the high volume of calls during these surges in data reporting. The data reporting surge continues even several days after an incident, as there are still a large number of calls to the system to obtain status updates. To respond to the surge, the operators have to record messages about the areas to avoid, and manually add this information to the 511 system. These messages also then need to be updated in real-time. With a limited number of operators, every individual operator's workload increases immensely during such times.

There are 500 Million tweets per day on Twitter on all topics (Internet Live Stats, 2017) and 1.15 billion daily active users on Facebook (Zephoria Digital Marketing, 2017). In addition, the manager at the GDOT TMC described Waze as a "firehose" of data. Sifting through these platforms which contain such large volumes of data to find relevant information is a difficult process. In addition to this challenge, there needs to be a mechanism to quickly and effectively visualize the data. For example, having a data dashboard to visualize the entire data influx from OSNs would enable the TMC personnel to quickly spot anomalies. One traffic specialist stated that the ability to query incident/crash data based on certain parameters such as "cars involved" and "property damage only" would be beneficial. At the GDOT TMC, a social media manager manually tweets about the real-time status of incidents and monitors other transportation related handles on Twitter (Figure 22). This manual process is time consuming, and would benefit from a more automated process.

Another bottleneck occurs in the TMC's process when an incident is automatically detected under the current system. This system sometimes generates false incident reports, and the TMC must investigate all reports. The TMC only has 30 HEROs serving the entire network and has to

optimally utilize these resources to verify incidents. Live feed from the cameras is very useful to supplement the HEROs as they try to verify incidents.

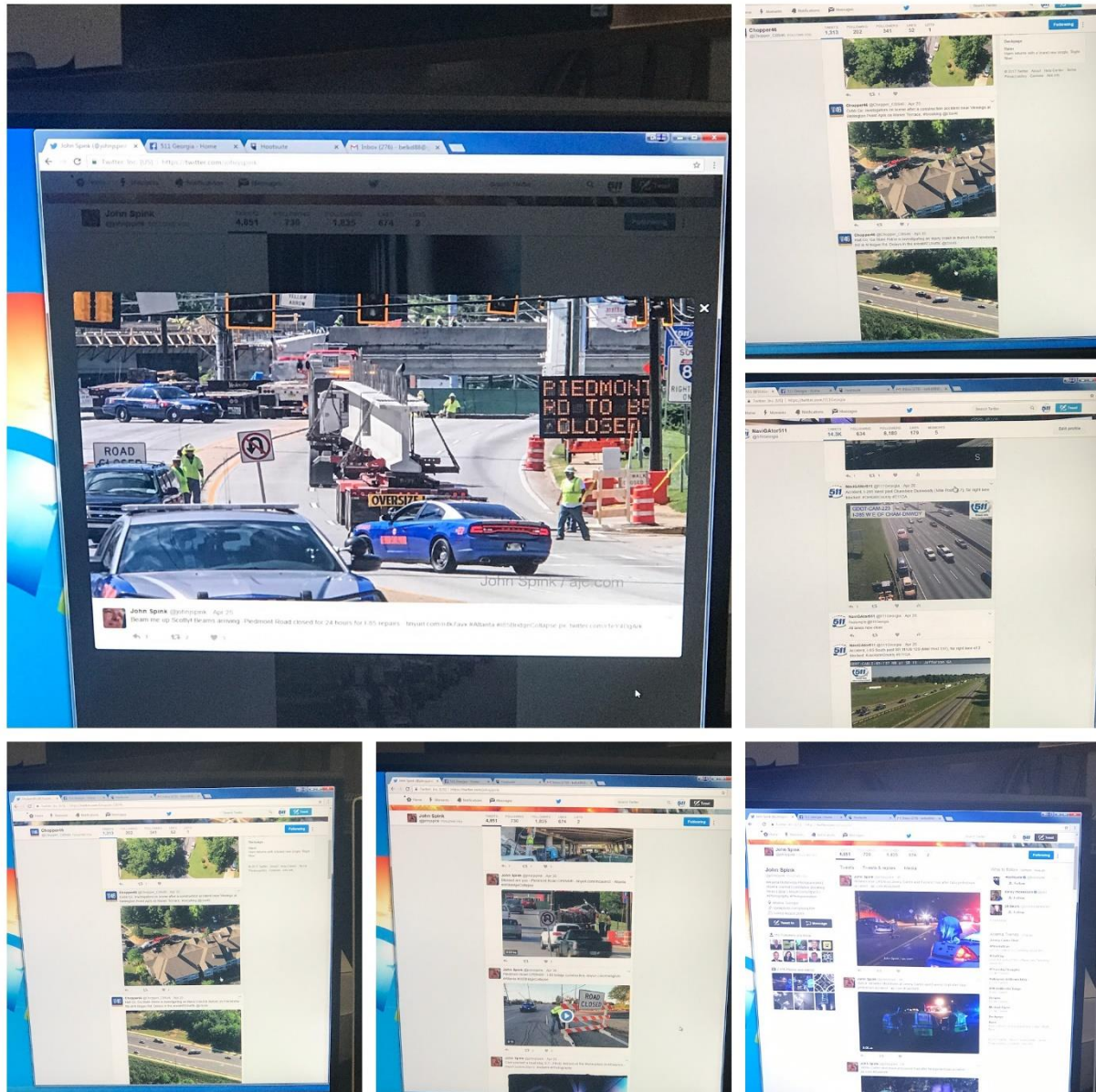


Figure 22: Select Twitter Handles that the Social Media Manager Monitors and Manages

Needs and Usefulness of Crowdsourced Traffic and Incident Management (TIM)

Georgia DOT can benefit from more effective use of crowdsourced data in several ways. Most critically, crowdsourced data could facilitate faster incident detection. Establishing a mechanism

for automatic fetching and prioritization of reported incidents from different sources would assist with this task. With the current workflow, the TMC is well equipped to handle a large volume of calls, but not to deal with a large volume of data flowing from various sources.

Before the collaboration with the Waze “Connected Citizens Program,” the TMC personnel were using Waze and Google Maps to detect possible incidents in areas with congestion. A WAZE representative at an ITS Georgia meeting claimed that incidents are reported via the software an average of 10 minutes before TMC Operators discover the same incident. Operators at the TMC also confirmed that many incidents are detected on Waze before other sources. As the data management program at GDOT evolves, all the sources of data ideally need to be integrated and visualized on a single dashboard for quick and easy access, with major incidents highlighted. When the research team asked about the possibility of developing a customized GDOT application, the GDOT traffic specialist raised concerns about this option, since the user base would be so much smaller than the existing number of Waze users. The success of a crowdsourced application depends on a large user base. Moreover, GDOT could benefit from the data generated by Waze in rural areas and areas with poor camera coverage. Individuals interviewed at the TMC also stated that it would also be beneficial for the HEROs to have access to pictures of the incident before they arrive. This would help reduce the amount of time spent at the site of the incident, as well as help with the ability to decide a plan of action before reaching the incident location, and therefore increase their safety.

The TMC personnel shared other helpful suggestions to improve the function of the current technology. For example, operators are unable to take notes on the TMC system during a phone call. Another finding from the interviews was that the TMC uses NaviGator as the main system. The research team recommends integrating the data from crowdsourced systems into this system

as a comprehensive approach to incident detection and information dissemination. Conducting a needs analysis and contextual inquiry before developing future tools or choosing which tools to use would be helpful.

SWOT Analysis

Figure 23 presents the summary of the SWOT analysis. The TMC in Georgia seems to be well equipped to handle major incidents. Some inefficiencies remain due to the technology stack being used and the burden on the operators. Processing large volumes of data with a limited number of employees is one of the primary challenges faced by the TMC. Data gaps in incident identification pose another challenge, but this problem can be overcome with crowdsourced technologies. Crowdsourcing can not only help in fetching incident data from areas with poor coverage, but also with verifying incidents. In conclusion, the research team recognized multiple opportunities and the enormous potential of crowdsourcing to improve the efficiency of TMC operations.

| INTERNAL FACTORS | |
|--|--|
| STRENGTHS (+) | WEAKNESSES (-) |
| <ul style="list-style-type: none"> • Ability to handle major events • Well equipped to handle large volume calls • Collaboration with Waze • Accidents verification by HEROs | <ul style="list-style-type: none"> • Limited staff • Data management • Limited number of cameras • Difficult on-boarding for existing tools • Limited use of data visualizations in the tools |
| EXTERNAL FACTORS | |
| OPPORTUNITIES (+) | THREATS (-) |
| <ul style="list-style-type: none"> • Automatically put incident details on 511 • Auto generation of customized reports • Enhanced data sharing workflow • Automatically update CMS with details of incident • Enhanced mechanisms to sift through large volumes of data • Integrate different crowdsourcing channels into the existing Navigator system • Better integration and access to media resources • Automated incident detection in camera feed • Single dashboard for all the data sources • More efficient reporting of incidents for HEROs to minimize delay | <ul style="list-style-type: none"> • High volume of data and calls • Traffic in the street delays first responders • Locating affected motorist or site of incident • Users reporting malicious data • Safety of HEROs and first responders |

Figure 23: SWOT Analysis

Proposed System for Georgia

The proposed system recommendations can be useful not only for the Georgia TMC, but also can provide useful guidance for other TMCs across the country. Figure 24 presents an overview of the system proposed by the research team, followed by additional details.

Technical Architecture

Keeping in mind the limited staff and resources, this study proposes an architecture which leverages the existing infrastructure in place at the TMC. The proposed system consists of a mobile application (iOS and Android) which allows the user to tweet details of the incident to Twitter

(www.twitter.com). A text mining application running at the TMC's internal server mines the relevant tweets and stores them in a local database. This system will require minimum development effort and take advantage of the latest open standards in technology and hardware efficiency via Twitter.

Mobile Application

The mobile application will allow public/network users to quickly tweet about the incident. The application will automatically fetch the location of the user and alert the user to self-report the location as well. The self-reported data will supplement the GPS data to compensate for changes in the vehicle's location. The user can select from a list of pre-defined Twitter hashtags to add such as #CarCrash, #MajorAccident and #BrokenDownCar (Twitter, 2017). These hashtags will assist the application at the TMC in mining the tweets. All users will have profiles that will have a rating based on the number of reported incidents and validity of reported incidents. If a reported incident turns out to be false, the user rating will be affected.

Text Mining Algorithm

The text mining algorithm running on the application server at the TMC will fetch the relevant data from the Twitter API. The hashtags will be used by the algorithm to prioritize the incidents. Moreover, this will allow the TMC staff to quickly filter tweets. While prioritizing reported incidents, the algorithm will take into account the rating of the user who tweeted as well.

Gamification to Increase Users

Encouraging people to earn rewards within the application by posting information about incidents can increase user engagement. Users (first few reporters) can be awarded with digital currency that can be redeemed for free parking or other such benefits in the city, since users are more likely to report an incident if they know they will be rewarded. Badges and recognition of achievement are

other means of driving user engagement. These incentives can be shared by GDOT on social media. For example, having a social leader board can encourage citizens to post more incidents to elevate their status and to compare their ratings with other users.

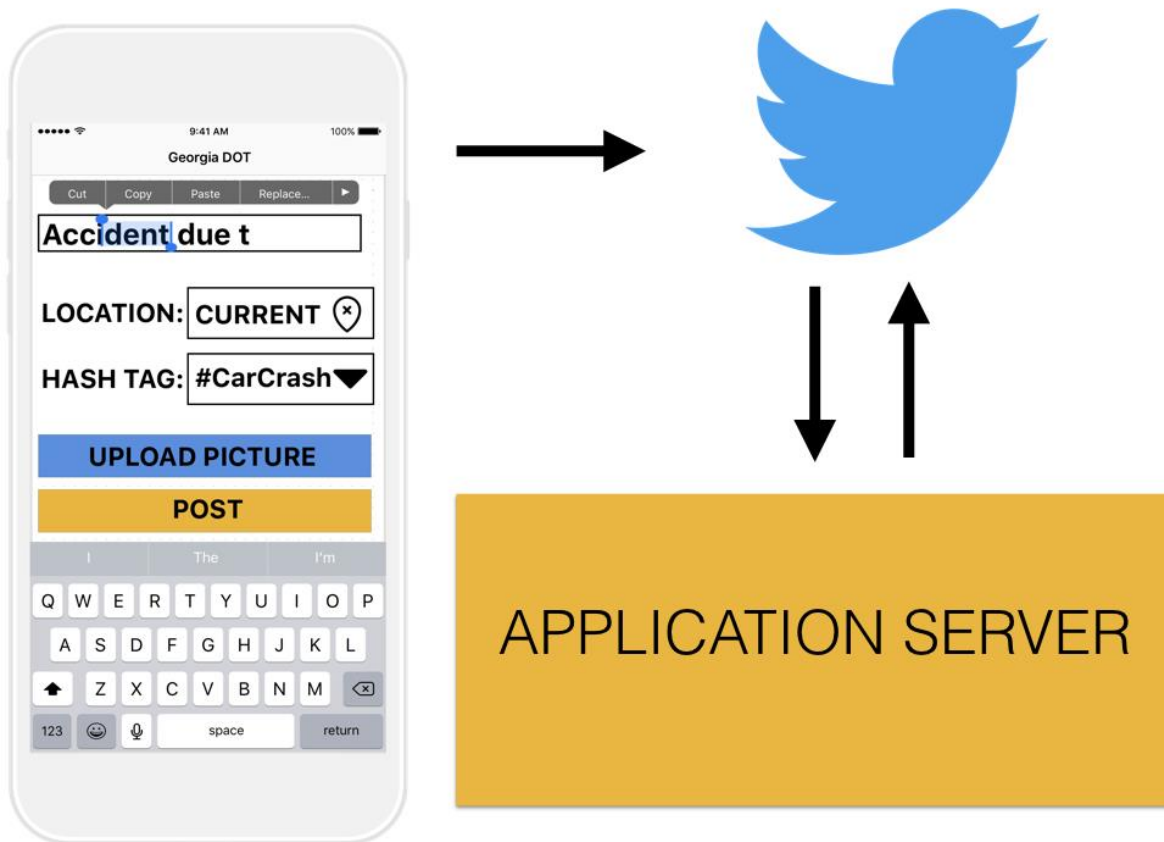


Figure 24: Architecture of Proposed System

Advantages and Disadvantages of Proposed System

This section details the advantages and disadvantages of the system. It includes both the technical and participant induced limitations of the system.

Advantages of the System

This solution minimizes the software development task by utilizing existing Twitter infrastructure. Twitter provides a robust service for users to post the incident details in real-time. Considering Twitter's popularity, it is likely that a large number of Georgia residents /road users already have a Twitter account. One advantage of Twitter is the mandatory use of hashtags, which make it easy to mine and analyze the data. Hashtags can also be used to prioritize the reported incidents. Ratings on user profiles can be used to decide which information from which user should be prioritized. Since the success of a crowdsourced system is based on the volume of users, tracking user engagement and retention is important to determine the effectiveness of the system. The project team proposes several gamification strategies to attract more users and thus increase the number of incident reports.

Weaknesses of the System

Amendments to Title 40 of the Official Code of Georgia, i.e. House Bill 23 and Senate Bill 360 prohibit drivers from using wireless telecommunications devices while driving with some exceptions. These exceptions include reporting a situation in which one's safety is in jeopardy or reporting a traffic accident, a medical emergency, and a serious road hazard. These laws may deter drivers of single occupancy vehicles from using cell phone based mobile applications for reporting minor incidents, road conditions and traffic congestion related issues. Despite the research team's recommended strategies to introduce gamification and social engagement features to increase the user base, the application might not gain traction due to this limitation. Promoting car sharing/pooling can be a possible answer to this problem. Alternatively, some new automobiles are being designed with a hands free Twitter feature.

If a network user does not use a mobile app for a long time, it is likely that he or she loses interest in it. According to Localytics, 80% of all app users churn within 90 days (Figure 25) (Localytics, 2017). Possible strategies to reduce the churn rate could include DOTs sending automated appreciation notes, status updates sharing the usefulness of reporting through the app, or other incentives.

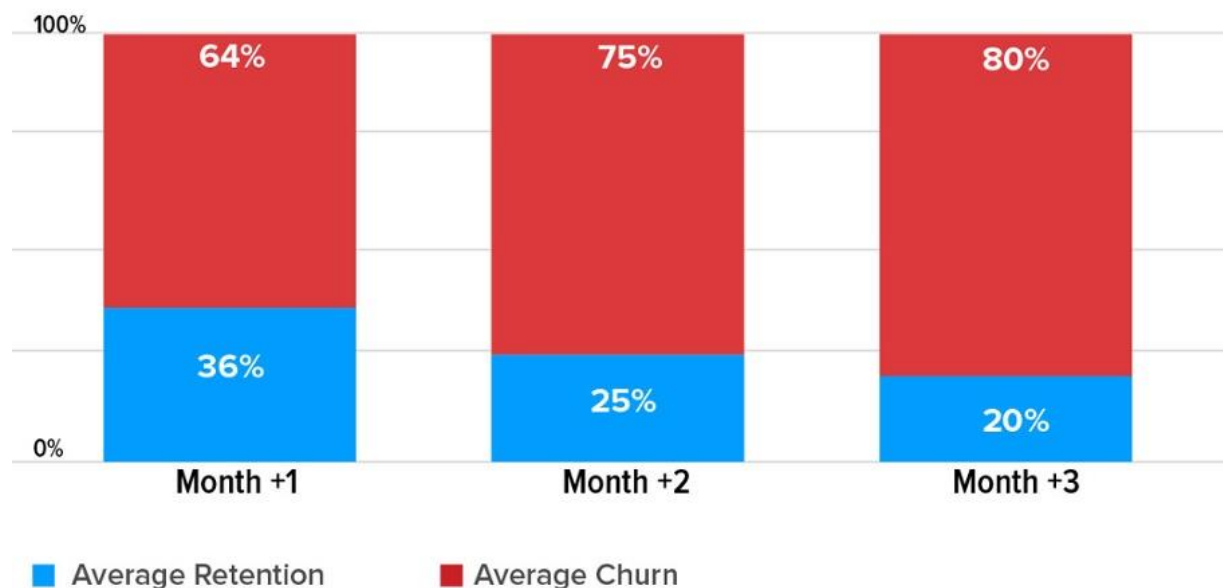


Figure 25: Average Retention and Churn Rate after the First, Second and Third Months

(Source: Web blog on *localytics*¹)

Prioritizing Action Items for the TMC

When a situation occurs that dramatically impacts traffic flow or access, such as a major accident or inclement weather, there is a corresponding surge in the volume of crowdsourced data. In order for DOTs to effectively respond to these situations, there needs to be an automated prioritization

¹ <http://info.localytics.com/hs-fs/hubfs/app%20retention%202017.jpg?t=1493200855873&width=931&name=app%20retention%202017.jpg>

process in place. The data needs to be categorized by type of incident (for example: debris in roadway, stalled vehicles, multi-car crash etc.). The data categorization process needs to include a second layer of data filtering, to determine which incidents pose the greatest threat to life or safety. These incidents can then be addressed first. This prioritization process can change based on staff inputs, for example during adverse weather conditions low impact incidents such as disabled vehicles might pose an additional hazard and would therefore move up on the priority list. This will help to prevent additional accidents from being caused by the disabled vehicle, as well as to reduce the amount of time that the driver is potentially exposed to the elements. As another example, if multiple disabled vehicles are clustered together, this incident location could be assigned a higher priority rank in the system. The system could also include a feature to measure the congestion impact caused by an incident, which could be integrated into the prioritization process. Information regarding flooding, heavy rains, low visibility, high wind and fire should be immediately uploaded on all information dissemination channels. This will allow citizens to plan better, and to use alternate means of transportation or other routes to avoid the affected ones. Such prevention measures will decrease the burden on the already congested areas. The priority ranking of potholes and other structural failures in the road surface would depend on the size of the failure.

Summary

This chapter analyzed the need for a crowdsourced traffic and incident management plan for Georgia. The research team interviewed six TMC personnel with various job titles and responsibilities, and toured the facility to better understand GDOT's capacity to collect and utilize crowdsourced data for traffic management. The research team documented the primary challenges faced by the TMC personnel, and noted them as follows: poor coverage in some regions of the state and limited TMC staff and resources. In addition, the large amount of data produced by

crowdsourcing platforms is overwhelming and cannot be filtered and prioritized efficiently with the existing systems.

Crowdsourced data can be extremely beneficial for Georgia DOT to facilitate prompt incident detection and to gather large amounts of information from the public. However, the reliability of crowdsourced data was voiced as a concern. Currently, the usability of crowdsourced data is somewhat limited since reported incidents are manually verified before being acted upon. This reduces the scalability of the crowdsourced data, and makes it difficult for GDOT to fully integrate it into the TMC's processes.

Keeping in mind the needs and challenges faced by the Georgia DOT, the team proposed a system to fetch crowdsourced data. The system leverages the existing infrastructure of Twitter. It consists of a mobile application which the users can use to tweet their incident report with hashtags. The hashtag helps the text mining algorithm running at the backend of the Twitter software to filter the tweets and prioritize them. The system uses Twitter's infrastructure, so the development time is minimal. Since the success of crowdsourcing systems is based on the volume of incidents being reported, the biggest barrier for the new system will be to gain initial traction.

Prioritization of tasks during times of surge needs to be done based on the impact it has on the traffic. A disabled car on the center of the street as compared to a car stopped on the shoulder has a higher negative impact on congestion levels. Roadway accidents which could cause additional accidents or problems should be given the highest priority in the ranking process.

CHAPTER V. SUMMARY AND CONCLUSIONS

Summary

Traffic management centers often face two significant challenges in traffic and incident management. First, it is difficult to ensure full coverage of the network through sensors and cameras. This is due to the fact that installation and maintenance of a large number of sensors and cameras requires a large outlay of staff and resources. Further, manually monitoring a large number of sensors and camera feeds is difficult. Second, detection of incidents in real-time is often challenging. Sensors and cameras may indicate the potential location of incidents, but incident detection through these two data sources are typically constrained by operational challenges. Trying to detect and monitor incidents through sensors and cameras during congested traffic conditions is difficult and may raise false alarms.

The above stated challenges of TMCs can be alleviated by augmenting the data obtained from the ITS infrastructure of TMCs with real-time crowdsourced data. The increase in popularity of personal devices such as smartphones and ever increasing advances in mobile computational technology provides an enormous opportunity to engage network users and citizens in traffic and incident management. In addition, today's smartphones are often programmable and equipped with a set of embedded sensors, such as a gyroscope, accelerometer, digital compass, GPS, microphone, and camera. Simple mobile apps can turn a smartphone into a powerful sensing device and it can potentially generate extremely useful data passively without any intervention from the user. As per Yang et al. (2012), "One can leverage millions of personal smartphones and a near-pervasive wireless network infrastructure to collect and analyze sensed data far beyond the scale of what was possible before, without the need to deploy thousands of static sensors."

However, there are multiple challenges related to the application of crowdsourced data in traffic and incident management. Most important of these challenges is that it is difficult to get sufficient data through voluntary participation in crowdsourcing. Network users tend to report only when stranded. This challenge can be overcome by implementing incentivizing measures, which has been tried with success in many parts of world. Another challenge related to crowdsourcing is ensuring that the data that is reported is accurate. Although the findings from this study indicate that incidents where citizens intentionally report false data are rare, the possibility of such situations occurring cannot be ruled out.

The challenges posed by crowdsourced data should be thoroughly understood before this data is fully integrated into the TMC's systems and processes. This study attempted to understand the advantages and disadvantages of crowdsourcing technique and its suitability for its potential use by GDOT's TMC operations by surveying the state of art and practice of crowdsourced traffic and incident management systems. For this purpose, two important tasks were performed. First, a thorough literature review was conducted to understand the state of art of this technique around the world. Then, the interviews were conducted with TMC staff members from across the country to understand the state of practice. A targeted set of interviews of TMC personnel was also conducted to understand the strengths and weakness of the TMC in Georgia. Based on the insights gained from the literature review and interviews, a system for crowdsourced traffic management for Georgia has been proposed in this study. Following are some important insights from this study:

- With the advances in mobile computing and OSNs the potential of crowdsourced traffic management has increased.
- Crowdsourced information such as incident location, weather, congestion and roadway conditions are extremely useful for real-time traffic management.

- There are three types of crowdsourcing techniques: active, passive and combined.
- Crowdsourcing increases public participation in data generation which results in more effective traffic management.
- Reliability of crowdsourced data is a major challenge.
- Mobile device sensor are the most prevalent means of data collection.
- Gamification techniques such as a social leader board and a social incentive based systems are used to increase the number of users.
- The majority of DOTs have a strong presence on OSNs.
- A large volume of data is necessary to effectively use crowdsourcing techniques, yet this volume of data is difficult for TMCs to manage.
- Waze Connected Citizens Program has been beneficial for TMCs to obtain data from poor coverage areas.
- A single system that processes information from all the sources of crowdsourced information is required to be effectively utilized by the TMC.
- Automated filtering and prioritization mechanisms need to be developed for the systems being used by the TMC.
- Georgia DOT has an efficient work flow in place with can benefit greatly from the proposed automated crowdsourced system.

Recommendations

1. **Computer Vision Technology (CVT):** CVT can be used to automatically detect incidents in the video feed and alert the personnel monitoring them. This would significantly decrease the chance of missing incidents.
2. **Automatic Dissemination of Information:** This would increase the efficiency of the TMC operations by decreasing the burden on the TMC operators as well as promote citizen engagement.
3. **Social Media Analytics Tools:** Products like Hootsuite (<https://hootsuite.com/>) can help swiftly filter the data and thus verify road network user problems more accurately and in real-time.
4. **Verification Using Crowdsourcing:** Citizens passing through the incident area can be automatically notified and requested to post a picture or tap on “verified” if they see an incident.
5. **Storing Comprehensive Historical Incident Data:** This data is important for developing future strategies and tools. Past data will help product designers understand past usage patterns and design a customized system.
6. **Improved Data Management System:** Sensors, cameras, crowdsourced data, and other data inputs generate huge datasets over time. Therefore, an effective data management system is of paramount importance. The evolving big data management techniques need to be utilized by the TMC for efficient data storage and retrieval.
7. **Improve Usability of Tools:** Conducting usability testing and obtaining feedback from employees regarding the advantages and disadvantages of the tools currently in place at the

TMC could help to improve the efficiency of the TMC personnel, which would result in a smoother workflow. As new processes are integrated to take advantage of crowdsourced data, this also provides an ideal time to review the effectiveness of the overall suite of tools in use at the TMC.

8. **Data Visualizations:** Increased use of data visualizations would make it easier to understand the existing data. Good visualizations are an efficient tool to quickly filter the data.
9. **Crowdsourced System in this Study:** The crowdsourcing system proposed in this study is simple to build and will require a limited investment of resources. This solution will improve the incident detection process and lead to a shorter response time.
10. **System Integration:** The most desirable and overarching recommendation to improve the efficiency of the TMC is to implement a new system with better integration across all functions. The current volume of incident reporting, together with the crowdsourced data input, all primarily captured manually with a fragmented system can result in inefficient TMC operations. A system where all the information feeds into single integrated process would be the most effective. This will allow the DOTs to avoid duplication in incident detection and enable optimal use of existing resources and personnel.

Limitations and Future Study

Due to time constraints, the team was only able to interview personnel from 11 TMCs. If additional TMCs could have been interviewed, the research team might have gained greater insight into the functioning and problems faced by DOTs across the country. In addition, only personnel from TMCs in the United States were interviewed. However, there might be opportunities to learn lessons from crowdsourced traffic management techniques and technology from other countries.

Interviewing transportation departments from developing countries would have enhanced the project team's understanding of the technologies that have been employed under more diverse traffic conditions. Vehicles with GPS installed form a small portion of the overall vehicle fleet in developing regions. Moreover, the camera coverage in the network is low, and resources to increase camera coverage are limited. Therefore, crowdsourced systems are often the only viable option for traffic management. Systems being used in these conditions can offer additional lessons for DOTs in the United States, particularly in rural regions with low coverage.

Based on the findings from this work, the research team has identified several opportunities for additional study. A usability study of the existing suite of tools and technology in place at the TMC could be conducted to determine if there could be improvements in either technology or training. This could happen as a component of the new processes implemented to take advantage of crowdsourced data. Another future extension of this study would be to fine tune the algorithm that mines tweets and prioritizes the reported incidents based on real-time data. The system proposed in this study assumes ideal behavior by the users.

Another area for future study could include designing a data fusion model that combines the data generated by multiple sources to the TMC, including deployed sensors in the network, the camera feeds, and crowdsourced data. The integrated or combined data from such a data fusion model can be harnessed by the TMC for an automated and more robust incident detection and confirmation process than exists today.

REFERENCES

- Artikis, A., Weidlich, M., Schnitzler, F., Boutsis, I., Liebig, T., Piatkowski, N., ... others. (2014). Heterogeneous Stream Processing and Crowdsourcing for Urban Traffic Management. In *EDBT* (Vol. 14, pp. 712–723). Retrieved from <https://pdfs.semanticscholar.org/ce0b/d496c10d0ae55d3e4e9c8c2f356c3a07bb92.pdf>
- Aubry, E., Silverston, T., Lahmadi, A., & Festor, O. (2014). CrowdOut: a mobile crowdsourcing service for road safety in digital cities. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on* (pp. 86–91). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6815170/>
- Balakrishnan, H., & Madden, S. (2014, October 16). CarTel. Retrieved from <http://cartel.csail.mit.edu/doku.php>
- Barron, J. P. G., Manso, M. A., Alcarria, R., & Gomez, R. P. (2014). A mobile crowdsourcing platform for urban infrastructure maintenance. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2014 Eighth International Conference on* (pp. 358–363). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6975489/>
- Boulos, M. N. K., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., ... Chuang, K.-Y. S. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics*, 10(1), 67.
- Bozzon, A., Brambilla, M., Ceri, S., & Mauri, A. (2013). Reactive crowdsourcing. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 153–164). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2488403>

- Brabham, D. C. (2009). Crowdsourcing the public participation process for planning projects. *Planning Theory*, 8(3), 242–262.
- Campolo, C., Iera, A., Molinaro, A., Paratore, S. Y., & Ruggeri, G. (2012). SMaRTCaR: An integrated smartphone-based platform to support traffic management applications. In *Vehicular Traffic Management for Smart Cities (VTM), 2012 First International Workshop on* (pp. 1–6). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6398700/>
- Charalabidis, Y., N. Loukis, E., Androutsopoulou, A., Karkaletsis, V., & Triantafyllou, A. (2014). Passive crowdsourcing in government using social media. *Transforming Government: People, Process and Policy*, 8(2), 283–308.
- Chatterjee, S., Mridha, S. K., Bhattacharyya, S., Shakhari, S., & Bhattacharyya, M. (2016). Dynamic Congestion Analysis for Better Traffic Management Using Social Media. In *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2* (pp. 85–95). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-30927-9_9
- Chatzimilioudis, G., Konstantinidis, A., Laoudias, C., & Zeinalipour-Yazti, D. (2012). Crowdsourcing with smartphones. *IEEE Internet Computing*, 16(5), 36–44.
- Dahlander, L., & Piezunka, H. (2014). Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. *Research Policy*, 43(5), 812–827.
- De Vreede, T., Nguyen, C., De Vreede, G.-J., Boughzala, I., Oh, O., & Reiter-Palmon, R. (2013). A theoretical model of user engagement in crowdsourcing. In *International Conference on Collaboration and Technology* (pp. 94–109). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-41347-6_8

- Divvy Bikes. (2017). Divvy. Retrieved from <https://www.divvybikes.com/how-it-works>
- Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), 86–96.
- El Faouzi, N.-E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges—A survey. *Information Fusion*, 12(1), 4–10.
- Eriksson, J., Balakrishnan, H., & Madden, S. (2008). Cabernet: vehicular content delivery using WiFi. In *Proceedings of the 14th ACM international conference on Mobile computing and networking* (pp. 199–210). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1409968>
- Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S., & Balakrishnan, H. (2008). The pothole patrol: using a mobile sensor network for road surface monitoring. In *Proceedings of the 6th international conference on Mobile systems, applications, and services* (pp. 29–39). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1378605>
- Frank, R., Weitz, H., Castignani, G., & Engel, T. (2014). Collaborative traffic sensing: a case study of a mobile phone based traffic management system. In *Consumer Communications and Networking Conference (CCNC), 2014 IEEE 11th* (pp. 579–584). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/7056314/>
- Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, 26(3), 10–14.
- Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 568–583.

- Hilgers, D., & Ihl, C. (2010). Citizensourcing: Applying the concept of open innovation to the public sector. *The International Journal of Public Participation*, 4(1), 67–88.
- Hossain, M., & Kauranen, I. (2015). Crowdsourcing: a comprehensive literature review. *Strategic Outsourcing: An International Journal*, 8(1), 2–22.
- Internet Live Stats. (2017). Twitter usage statistics. Retrieved from <http://www.internetlivestats.com/twitter-statistics/>
- Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., ... Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1301–1318). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2441923>
- Komarov, S., Reinecke, K., & Gajos, K. Z. (2013). Crowdsourcing performance evaluations of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 207–216). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2470684>
- Kovacheva, A., Frank, R., & Engel, T. (2013). LuxTraffic: A collaborative traffic sensing system. In *Local & Metropolitan Area Networks (LANMAN), 2013 19th IEEE Workshop on* (pp. 1–6). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6528286/>
- Localytics. (2017). Localytics. Retrieved from <https://www.localytics.com/>
- Loukis, E., & Charalabidis, Y. (2015). Active and passive crowdsourcing in government. In *Policy Practice and Digital Science* (pp. 261–289). Springer. Retrieved from http://link.springer.com/10.1007/978-3-319-12784-2_12

- Matyas, S., Matyas, C., Schlieder, C., Kiefer, P., Mitarai, H., & Kamata, M. (2008). Designing location-based mobile games with a purpose: collecting geospatial data with CityExplorer. In *Proceedings of the 2008 international conference on advances in computer entertainment technology* (pp. 244–247). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1501806>
- McCall, R., & Koenig, V. (2012). Gaming concepts and incentives to change driver behaviour. In *Ad Hoc Networking Workshop (Med-Hoc-Net), 2012 The 11th Annual Mediterranean* (pp. 146–151). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6257115/>
- Mohan, P., Padmanabhan, V. N., & Ramjee, R. (2008). Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM conference on Embedded network sensor systems* (pp. 323–336). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1460444>
- Molina, J. (2014). *The Case for Crowdsourcing in Bicycle Planning: An Exploratory Study*. Tufts University. Retrieved from <https://pdfs.semanticscholar.org/d23f/e26e7635f7f56cb2569d555529eacf67ebe3.pdf>
- Myr, D. (2002). *Real time vehicle guidance and forecasting system under traffic jam conditions*. Google Patents. Retrieved from <https://www.google.com/patents/US6480783>
- Pan, B., Zheng, Y., Wilkie, D., & Shahabi, C. (2013). Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 344–353). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2525343>

- Pinto, D. (2007). *Traffic incidents processing system and method for sharing real time traffic information*. Google Patents. Retrieved from <https://www.google.com/patents/US20080255754>
- Quinn, A. J., & Bederson, B. B. (2011). Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1403–1412). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1979148>
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851–860). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1772777>
- Schweitzer, L. (2014). Planning and social media: a case study of public transit and stigma on Twitter. *Journal of the American Planning Association*, 80(3), 218–238.
- Sen, R. (2014). RasteyRishtey: A social incentive system to crowdsource road traffic information in developing regions. In *Mobile Computing and Ubiquitous Networking (ICMU), 2014 Seventh International Conference on* (pp. 171–176). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6799090/>
- Sen, R., Sevani, V., Sharma, P., Koradia, Z., & Raman, B. (2009). Challenges in communication assisted road transportation systems for developing regions. *NSDR '09*. Retrieved from http://dritte.org/nsdr09/files/nsdr09_camera/s4p3_sen09nsdr.pdf
- Smith, A. (2015). Crowdsourcing for Active Transportation. *ITE Journal*, 85(5). Retrieved from <https://trid.trb.org/view.aspx?id=1368024>

- Smith, A., & Fehr & Peers. (2015). *Crowdsourcing pedestrian and cyclist activity data* (White Paper Series No. DTFHGI-11-H-00024). Federal Highway Administration. Retrieved from http://www.pedbikeinfo.org/cms/downloads/PBIC_WhitePaper_Crowdsourcing.pdf
- Steinfeld, A., Zimmerman, J., Tomasic, A., Yoo, D., & Aziz, R. (2011). Mobile transit information from universal design and crowdsourcing. *Transportation Research Record: Journal of the Transportation Research Board*, (2217), 95–102.
- Stevens, M., & D'Hondt, E. (2010). Crowdsourcing of pollution data using smartphones. In *Workshop on Ubiquitous Crowdsourcing*. Retrieved from <http://soft.vub.ac.be/Publications/2010/vub-tr-soft-10-15.pdf>
- Strava. (2014). Stava Metro. Retrieved from <http://metro.strava.com/>
- Strava. (2017). Stava - About us. Retrieved from <https://www.strava.com/about>
- Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S., & Eriksson, J. (2009). VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems* (pp. 85–98). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1644048>
- Twitter. (2017). Using hashtags on Twitter. Retrieved from <https://support.twitter.com/articles/49309>
- UDOT. (2014, July 8). Citizen Reporter reports by day. Retrieved from <http://blog.udot.utah.gov/2014/08/udot-citizen-reporter-program-gathers-volunteer-data/reports-by-day/>

- Van Lint, J. W. C., & Hoogendoorn, S. P. (2010). A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil and Infrastructure Engineering*, 25(8), 596–612.
- Waze. (2014). *Connected citizens program*. Retrieved from <https://docs.google.com/gview?url=https%3A%2F%2Fs3-eu-west-1.amazonaws.com%2Fwaze-partner-assets%2FCCPCaseStudies.pdf>
- Waze. (n.d.). *Connected citizens by Waze*. Retrieved from <https://docs.google.com/gview?url=https%3A%2F%2Fs3-eu-west-1.amazonaws.com%2Fwaze-partner-assets%2FCCPFactSheet.pdf>
- Yang, D., Xue, G., Fang, X., & Tang, J. (2012). Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In *Proceedings of the 18th annual international conference on Mobile computing and networking* (pp. 173–184). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=2348567>
- Yang, J., Adamic, L. A., & Ackerman, M. S. (2008). Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. In *Proceedings of the 9th ACM conference on Electronic commerce* (pp. 246–255). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1386829>
- Zephoria Digital Marketing. (2017). The top 20 valuable Facebook statistics – Updated May 2017. Retrieved from <https://zephoria.com/top-15-valuable-facebook-statistics/>