CHAPTER

1 CHARACTERISTICS OF INTELLIGENT TRANSPORTATION SYSTEMS AND ITS RELATIONSHIP WITH DATA ANALYTICS

Sakib M. Khan, Mizanur Rahman, Amy Apon, and Mashrur Chowdhury
Clemson University, Clemson, SC, United States

1.1 INTELLIGENT TRANSPORTATION SYSTEMS AS DATA-INTENSIVE APPLICATIONS

Intelligent transportation system (ITS) applications are complex, data-intensive applications with characteristics that can be described using the "5Vs of Big Data": (1) volume, (2) variety, (3) velocity, (4) veracity, and (5) value (for the original 3V's, see Ref. [1]). Note that any single one of these characteristics can produce challenges for traditional database management systems, and data with several of these characteristics are untenable for traditional data processing systems. Therefore, data infrastructures and systems that can handle large amounts of historic and real-time data are needed to transform ITS from a conventional technology-driven system to a complex data-driven system.

The first "V" is the volume of ITS data, which is growing exponentially for transportation systems. With the growing number of complex data collection technologies, unprecedented amounts of transportation related data are being generated every second. For example, approximately 480 TB of data was collected by every automotive manufacturer in 2013, which is expected to increase to 11.1 PB/year by 2020 [2]. Similarly, 500 cameras of the closed-circuit television (CCTV) system in the city of London generate 1.2 Gbps [3].

The second "V" of ITS data is the variety of the data, which are collected in various formats and in a number of ways, including numeric data captured from sensors on both vehicles and infrastructure, text data from social media, and image and GIS data loaded from maps. The degree of the organization of this data can vary from semi-structured data (e.g., repair logs, images, videos, and audio files) to structured data (e.g., data from sensor systems and data from within a traffic incident data warehouses) [4]. Social media data is considered to be semi-structured data, containing tags or a common structure with distinct semantic elements. Different datasets have different formats that vary in file size, record length, and encoding schemes, the contents of which can be homogeneous or heterogeneous (i.e., with many data types such as text, discrete numeric data, and continuous numeric data that may or may not be tagged). These heterogeneous data sets, generated by different sources in different formats, impose significant challenges for the ingestion and integration of a data analytics system. However, their fusion enables sophisticated analyses from self-learning algorithms for pattern detection to dimension reduction approaches for complex predictions.

Data Analytics for Intelligent Transportation Systems. DOI: http://dx.doi.org/10.1016/B978-0-12-809715-1.00001-8

The third "V" of ITS data, velocity, varies widely. Data ingest rates and processing requirements vary greatly from batch processing to real-time event processing of online data feeds, inducing high requirements on data infrastructure. Some data are collected continuously, in real-time, whereas other data are collected at regular intervals. For example, most state Departments of Transportation (DOTs) use automated data collectors that feed media outlets with data. One such example is the Commercial/Media Wholesale Web Portal (CWWP) designed by the California DOT (Caltrans) to facilitate the data needs of commercial and media information service providers. The CWWP requests and receives traveler information generated by the data collection devices maintained by Caltrans [5]. Although speed data from traffic is collected continuously, data such as road maps may be updated at less frequent intervals.

The term veracity is the fourth "V" of ITS data and is used to describe the certainty or trustworthiness of ITS data. For example, any decision made from a data stream is predicated upon the integrity of the source and the data stream, that is, the correct calibration of sensors and the correct interpretation of any missing data. Consequently, the goal of collecting reliable and timely transportation related data is a significant challenge for the ITS community.

The final "V" of ITS data, value, can depend on the age of data, their sampling rates, and the intended application. For example, data that are a few minutes old may have no value for a collision avoidance application, but may be useful in a route planning application. The value is a measure of the ability to extract meaningful, actionable business insights from data [6]. The following subsections describe ITS from different data system perspectives, as well as explain different data sources and data collection technologies of ITS.

1.1.1 ITS DATA SYSTEM

The use of a one-dimensional view of ITS will likely simplify some aspects of the system. However, the complexity of ITS requires using multiple perspectives. One way of viewing ITS is as a data-intensive application in which the data are hosted by, and circulate through, an interconnected network of computers, communication infrastructure, and transportation infrastructure. This system is characterized by (1) data producers and consumers, (2) data storage systems, and (3) intelligent decision support components. Communication is supported through both wired and wireless technologies. Through the interconnection network, intelligent decision support applications extract relevant data that are produced by billions of sources, specifically from roadway sensors and ITS devices. The data are then used to provide specific services to road users, transportation planners, and policy makers.

A second way to understand ITS involves considering the various layers of the architecture, similar to the Open Systems Interconnection network model [7]. For this system, the foundation layer contains the physical transportation components, computer networks, computers, and storage devices. These computing components may be commodity off-the-shelf, or may be specifically designed propriety devices that are used by a small community or a single company. The system is also characterized by a series of defined standards that allows networks to connect to computers and storage devices. Above the foundational physical layer is the data link layer, which is characterized by a series of increasingly sophisticated standards that define communication protocols for specific network technologies, such as wireless or wired networks. The internet protocol (IP), which is the standard protocol for connecting different networks together, works above the individual network protocols to allow vehicular communication via cellular phone to a data center that is interconnected with wired network technologies such as 10G ethernet. Transport layer protocols above IP such as transmission control protocol (TCP) and others ensure an end-to-end reliability of communication even when the different sources are moving and changing. The session, presentation, and application layer protocols above the transport layer describe the data formats expected by the applications, then manage the different types of messages communicated between users and systems and between different autonomous systems.

Another view of ITS is that of the "Three Is"—instrumented, interconnected, and intelligent [8]. This is an instrumentation concept that includes advanced devices and sensors that are increasingly varied in the amount and type of data collected. For example, sensors may measure location information, monitor and measure vibration, or capture video using different types of cameras. Probe vehicles on the highway may be deployed to enable the continuous collection of traffic data. Although sensors require a source of power such as a battery or electrical connection, technology advances are enabling the possible widespread deployment of inexpensive sensors onto the transportation infrastructure that can operate without a battery or external power source. Here, sophisticated wired and wireless communication systems transmit the data from sensors to intelligent decision support applications.

1.1.2 ITS DATA SOURCES AND DATA COLLECTION TECHNOLOGIES

Substantial advances in communication and computing technologies have in turn yielded advancements in ITS data collection technologies. The relevant data come from many sources. ITS data sources can be categorized into four broad groups: (1) roadway data, (2) vehicle-based data, (3) traveler-based data, and (4) wide area data. Similarly, data collection technologies are grouped into four categories: (1) roadway data collection technology, (2) vehicle-based data collection technology, (3) traveler-based data collection technology, and (4) wide area data collection technology.

Roadway data collection technologies have been used for decades to collect data from fixed locations along a highway. Sensors used on roadways can be passive in nature, collecting data without disruption to regular traffic operations [9]. One of the most widely deployed roadway data collection technologies is the loop detector. Numerous loop detector-based applications are now in use such as intersection traffic monitoring, incident detection, vehicle classification and vehicle reidentification applications [10,11]. Some types of loop detectors can provide data that include the count or detection of vehicles at a location. Another type of roadway data collector is microwave radar, which can detect vehicle flow, speed, and presence. Infrared sensors can be used to measure the reflected energy from a vehicle, which may be used to infer characteristics about the type or behavior of the vehicle. Ultrasonic sensors can identify vehicle count, presence, and lane occupancy. Another widely used roadway data collection technology is the CCTV camera. Machine learning methods can be applied to the video to detect characteristics of traffic. Once these images are digitized, they are processed and converted into relevant traffic data. Different machine vision algorithms are used to analyze the recorded traffic images for real-time traffic monitoring, incident detection and verification, and vehicle classification.

Vehicle-based data collection technologies, such as vehicles with electronic toll tags and global positioning systems (GPS), when combined with cell phone-based Bluetooth and Wi-Fi radios, are the second data source used in ITS applications. While the roadway data collection technologies are used for data collection at a specific location, the opportunity for collecting data from mobile vehicle sources has motivated the development of such new applications as route choice, origin–destination survey, and travel time estimation. Connected vehicle (CV) technologies, which connect vehicles on a roadway through a dynamic wireless communications network, enable vehicles to share data in real-time with other vehicles and the transportation infrastructure, specifically the roadside units (RSUs). Such seamless real-time connectivity between the vehicles and infrastructure in a CV environment has the potential to enable a new host of the benefits for the existing infrastructure-based ITS applications, which include safety, mobility, and environmental benefits. Thus far, the United States DOT (USDOT) has identified 97 CV applications, and the list is increasing [12].

Motorists using cell phone applications provide a third data collection source for ITS. These widely used communication and cell phone applications and online social media have been used by travelers to voluntarily provide updated traffic information. For example, the Waze cell phone application, now operated by Google, uses location information of travelers to infer traffic slowdown and the potential location of traffic incidents. However, such data from motorists that is derived through online social media platforms is semi-structured and unreliable in that the driver does not provide the specific location information of any traffic event. For example, only 1.6% of Twitter users have their geolocation functionality activated [13].

Wide area data collection technology, which monitors traffic flow via multiple sensor networks, is the fourth data collection source. Photogrammetry and video recording from unmanned aircraft and space-based radar are also available as data collection technologies. Data collected from these technologies include vehicle spacing, speed, and density, which in turn are used for diverse purposes such as traffic monitoring and incident management.

A summary of the different transportation data collection technologies is provided in Table 1.1. Apart from the data collected by the four classical data collection sources, transportation-related data is also generated from such sources as the news media and weather stations. The inclusion of both real-time and archived data collected by both public and private agencies using different technologies in the different transportation decision-making activities has played a remarkable role in the rapid implementation of different ITS applications.

1.2 BIG DATA ANALYTICS AND INFRASTRUCTURE TO SUPPORT ITS

The goal of data analytics is to provide insight and knowledge from the collected data. The ability to analyze data and provide on-demand decision support is critical for ITS, whether the task is to evaluate an existing transportation network or to compare proposed alternatives. Consequently, Big Data analytics methods developed for ITS are based upon the ability to incorporate different types of unstructured, real-time, or archival data sets from diverse data sources. A sample of the key aspects of data analytics for ITS is described here, particularly the fundamental types of data analytics, the role of the time dimension of data, infrastructures for Big Data analytics, and the security of ITS data. More detailed explanations are outlined in the remaining chapters of this book.

Table 1.1 ITS Data Sources and Data Collection Technology

Data Sources	Data Collection Technology	Data Type	User	Advantage	Limitation
Roadway data	Loop detector	Volume, speed, classification, occupancy, presence	Public agency	• Not affected by weather • Most widely used, availability of skilled manpower	• Limited coverage • Extended lifecycle cost • Damage-prone due to truck weight
	Vision-based technology (CCTV camera)	Volume, speed, classification, occupancy, presence	Public agency	• Larger coverage than loop detectors • Not affected by traffic load • Continuous data collection	• Extended lifecycle cost • Highly affected by weather
Vehicle-based data	Floating car data (with GPS and cellular network)	Vehicle position, travel time, speed, lateral and longitudinal acceleration/ deceleration, obstacle detection	Public and private agencies	• Larger coverage than loop detectors and cameras • No special hardware device is necessary in cars • No particular infrastructure is to be built along the road • Continuous data collection • Not affected by weather	• Sophisticated algorithm is required to extract the data • Low location precision for GPS
	Connected vehicle	Vehicle position, travel time, speed, lateral and longitudinal acceleration/ deceleration, obstacle detection	Public and private agencies	• Larger coverage than loop detectors and cameras • Continuous data collection • Not affected by weather	• Sophisticated algorithm is required to extract the data • Dedicated short range communication (DSRC) or other communication devices are necessary
Traveler-based data	Twitter, Waze	Real-time alerts, incident detection	Public and private agencies	Larger coverage due to presence of the travelers	• Low location precision • Semi-structured data
Wide area data	Photogrammetry	Traffic monitoring, incident management, transportation planning and design	Public agency	Can collect data from locations where accessibility is difficult from the ground	• Affected by weather, vegetation, and shadows • Accuracy affected by camera quality and flying height

Source*: [14] S. Bregman, Uses of social media in public transportation, Transportation Research Board, (99) (2012) 18–28. [15] CDOT, Survey Manual, Chapter 4, Aerial Surveys, Colorado Department of Transportation. ⟨https://www.codot.gov/business/manuals/survey/chapter-4/chapter4.pdf⟩, 2015 (accessed 17.07.16); [16] S.M. Khan, Real-time traffic condition assessment with connected vehicles, M.S. Thesis, Clemson University, Clemson, SC, 2015.*

As described in Chapter 2, Data Analytics: Fundamentals, data analytics can be descriptive, diagnostic, predictive, and prescriptive, each of which is used in ITS data analytics. Descriptive analysis uses statistical methods to describe characteristics and patterns in the data. Given observational data about vehicles on a roadway, it is possible to calculate (1) the average number of vehicles along stretches of road at certain times during the day, (2) the average, minimum, and maximum velocity of the vehicles, and (3) the average weight and size of the vehicles. Various visualization tools, described in detail in Chapter 7, Interactive Data Visualization, may help to describe the characteristics of the data. For example, the average count of the daily long-haul trucking traffic data was collected for the US national highway system in 2011, which is shown in Fig. 1.1 [17]. Descriptive analytics has to take into account variations in the source and context of the data. For example, the weekend traffic may be very different than the weekday traffic, and the traffic may vary seasonally. Many organizations publish guidelines on calculating the annual average daily traffic, such as the American Association of State Highway Transportation Officials.

Data analytics seeks to find anomalies or trends in data, which are then used to diagnose problems or to make predictions about the future. Statistical and spatiotemporal analysis tools (e.g., ArcGIS, R), image processing tools (e.g., Matlab as a tool for traffic camera recording), natural language processing tools (e.g., Python as a tool for social media data) are software that are widely used in ITS data analytics. An extensive set of examples using the R language is provided in Chapter 3 of this book.

Referring to the previous example, Fig. 1.2 shows how such predictions are a vital component of ITS data analytics, in terms of the predicted average count of daily long-haul trucking traffic [18]. This figure shows the data from 2040 for the U.S. national highway system [18]. This comparison of the two figures illustrates the highest predicted growth of traffic, which is useful for predictive data analytics.

As described in Chapter 4, The Centrality of Data: Data Lifecycle and Data Pipelines, the data lifecycle and data pipeline used in data analytics entail knowing what data to use, how to compare historical data with current data, and how to use these data for accurate predictions. Not all important data have been recently acquired either. As a framework for integrating CV technologies and for identifying interfaces for standardization, the USDOT developed the Connected Vehicle Reference Implementation Architecture (CVRIA). The requirements for CVRIA are derived from a series of concepts of operations developed by the USDOT through 2012 [19]. Significant information is also available about the spatial and temporal context of the collected data. For example, the highway police collect incident information with the location reference that includes the mile marker along the highway, along with the incident start time and duration. These data, with other incident detection and verification sources such as traffic cameras, 911 emergency call and private company data, are stored in a server in the traffic management center (TMC). These data are stored and merged with respect to time and location of specific incidents. The case studies in Chapter 4, The Centrality of Data: Data Lifecycle and Data Pipelines further illustrate the importance of understanding the context of the collected data and how to value data of varying age.

As described in Chapter 5, Data Infrastructure for Intelligent Transportation Systems, a scalable infrastructure is required that can support the interactions between vehicles, the transportation

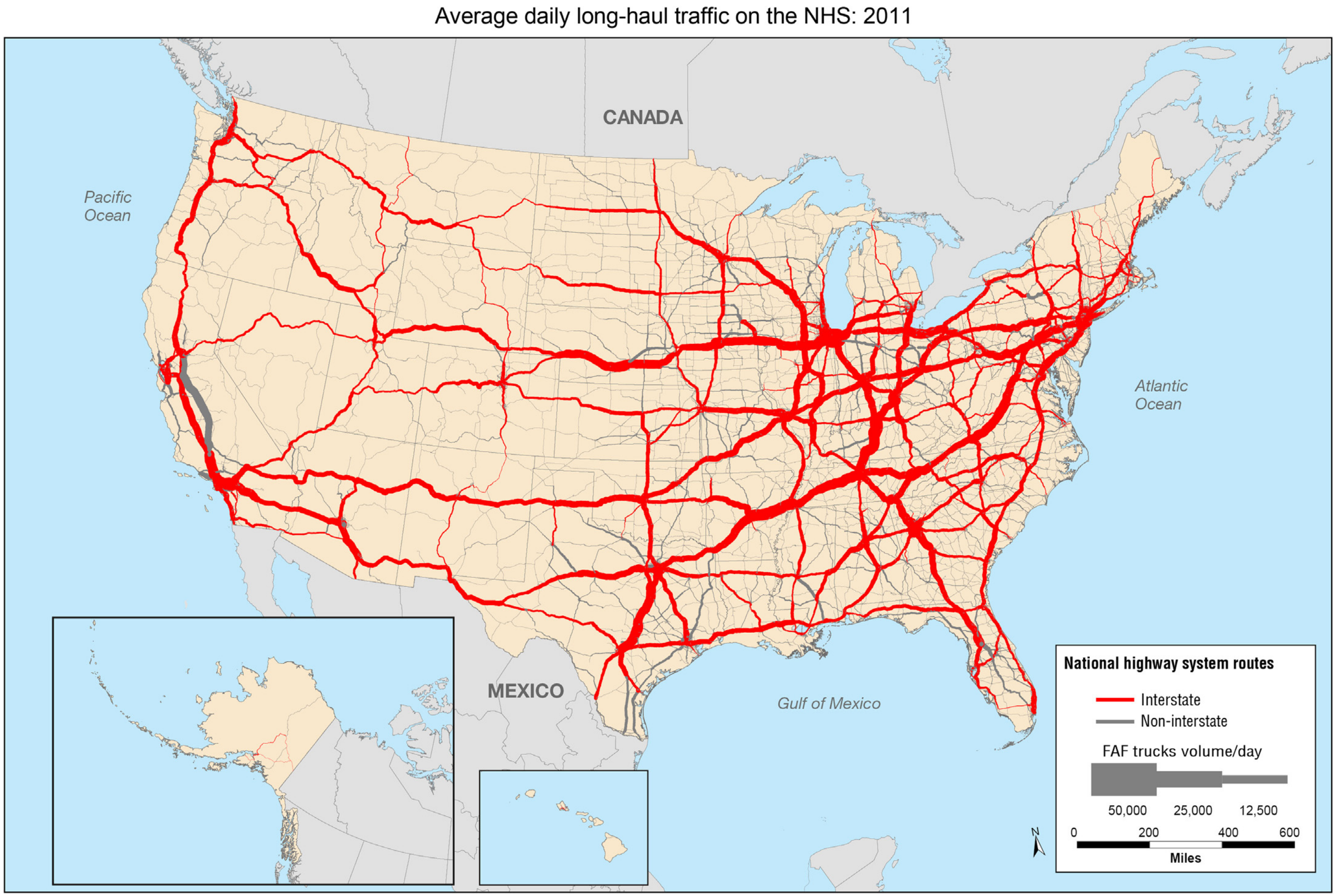

FIGURE 1.1

Average daily long-haul traffic in 2011 on the national highway system.

Source*: U.S. Department of Transportation, Federal Highway Administration, Office of Freight Management and Operations, Freight Analysis Framework, version 3.4.* ⟨*http://www.ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/nhsavglhft2011.htm*⟩*, 2013 (accessed 02.10.16)*

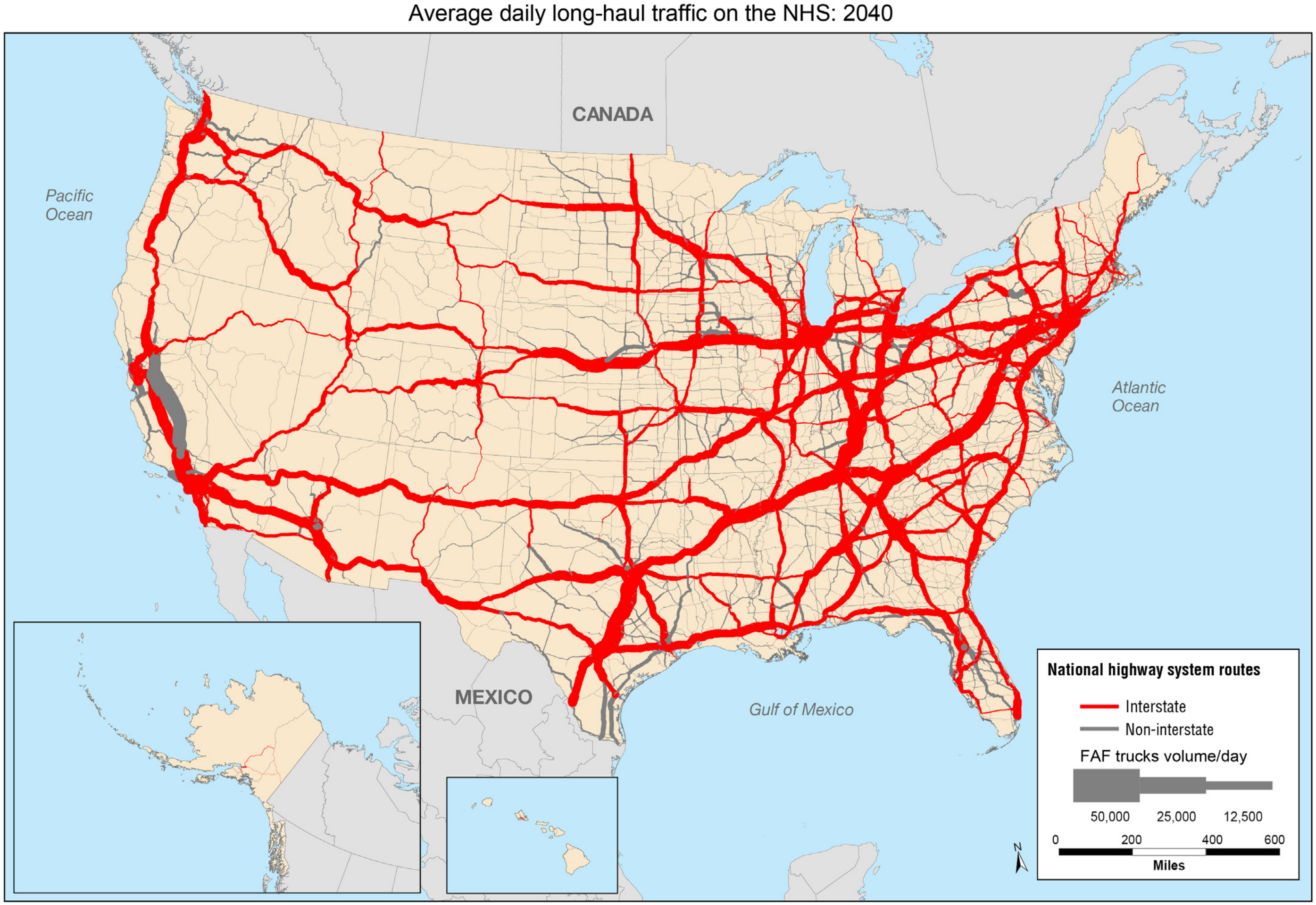

FIGURE 1.2

Average daily long-haul traffic in 2040 on the national highway system.

Source*: U.S. Department of Transportation, Federal Highway Administration, Office of Freight Management and Operations, Freight Analysis Framework, version 3.4.* ⟨*http://www.ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/nhsavglhft2040.htm*⟩*, 2013 (accessed 02.10.16).*

infrastructure, and the human operators. The rapid growth in the scale and complexity of ITS data requires creating data infrastructure and analytics to support the effective and efficient usage of the enormous amount of data that are collected, processed, and distributed for different ITS applications. Batch and stream processing are just two different processing models available. For example, batch processing of very large datasets can be used to create a descriptive illustration of the freight transportation in a given region in a given week by calculating the metrics of interest and producing the results for display in a chart. However, if the application is to provide an up-to-the-minute prediction of traffic flows and incidents, then the data stream must be processed in real-time. Hadoop [20] is a scalable platform for compute and storage that has emerged as a *de facto* standard for Big Data processing at Internet companies and in the scientific community. Many tools have been developed with Hadoop, including tools for parallel, in-memory and stream processing, traditional database approaches using SQL, and unstructured data engines using NoSQL. The Hadoop environment also includes libraries and tools for machine learning, all of which are described in Chapter 5, Data Infrastructure for Intelligent Transportation Systems.

Important problems faced by ITS involve addressing issues of security and privacy. The various layers of the ITS architecture; physical, network, and the application layers, can be configured to provide security, the detailed descriptions of which are provided in Chapter 6, Security and Data Privacy of Modern Automobiles. Privacy is of particular importance in ITS because of the nature of data collection. As more data, particularly timestamped location data, are aggregated to meet complex needs of ITS, there is the likelihood that these aggregated data will reveal information about an individual's daily behaviors, relationships, and work or recreational behaviors unavailable through single set of data alone. The individual must understand the implications of allowing access to certain data, and the organization must aggregate the data to ensure the integrity of individual privacy when the behavior of a community or region is the subject of study.

1.3 ITS ARCHITECTURE: THE FRAMEWORK OF ITS APPLICATIONS

Understanding the framework of ITS applications is a prerequisite to perceiving the different data system components of ITS. An ITS architecture offers a common framework to (1) plan, (2) define, and (3) implement different ITS applications. An ITS architecture also defines the information and data flow through the system and associated standards to provide particular ITS services. For example, the United States National ITS Architecture offers general guidance to ensure interoperability of systems, products, and services. A key goal is to ensure interoperability through standardization while ensuring that the architecture will lead to the deployment of ITS projects even as information and telecommunications technology advances. USDOT initiated the task of defining and developing the US national ITS architecture in 1993, and this scheme must now be used in all ITS projects in order to receive any federal funding [21]. An integrated ITS architecture developed for a region that follows the national ITS architecture can leverage national standards and shared data sources. By doing so, costs are reduced for collecting, processing, and disseminating of data, and duplication of effort is reduced when implementing multiple ITS applications. The national ITS architecture offers systematic guidelines to plan, design and implement ITS applications to ensure the compatibility and interoperability of different ITS components.

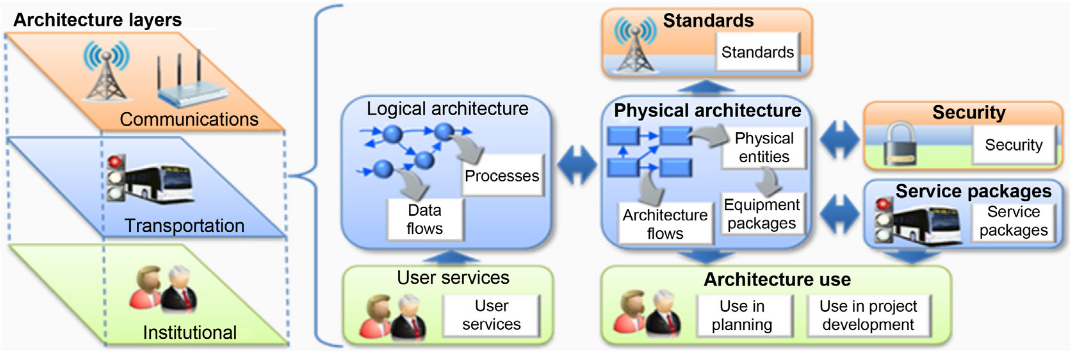

FIGURE 1.3

Different components of US National ITS architecture.

Source: *The Architectural View.* ⟨*http://www.iteris.com/itsarch/html/menu/hypertext.htm*⟩, *2016 (accessed 01.07.16) [24].*

Other developed countries have undertaken similar efforts to develop a national ITS architecture. In Europe, efforts toward a European ITS Architecture began in the 1990s, and a launch of the completed scheme occurred in October 2000 [22]. In Japan, an ITS architecture was developed in 1999 [23]. It was initiated by the multiple government agencies and Vehicle, Road, and Traffic Intelligence Society (currently known as ITS Japan). Prior to the development of each of these architectures, the following criteria were first determined: key stakeholders, application functions, the physical entities where the functions reside, and the information flow between the physical entities.

The US National ITS Architecture consists of three layers: (1) institutional, (2) transportation, and (3) communication layer (as shown in Fig. 1.3). The institutional layer defines policies, funding incentives, and processes to provide institutional support and to make effective decisions. The transportation layer, which is the core component of the ITS architecture, defines the transportation services (e.g., transit signal priority, vehicle safety monitoring), and it includes subsystems, interfaces, functions and data definitions for each transportation service. The communication layer defines communication services and technologies for supporting ITS applications. The US national architecture has the following primary components:

- User services and user service requirements
- Logical architecture
- Physical architecture
- Service packages
- Security
- Standards

Each component of US national ITS architecture is described in the following sections.

1.3.1 USER SERVICES AND USER SERVICE REQUIREMENTS

For the ITS architecture, user services can be considered as the first building block, and these define what the system is required to do. User services are described from the perspective of the users or stakeholders. Initially, the user services were defined by the joint effort of USDOT and ITS America, with the input of diverse stakeholder groups. User services support the establishment of high level transportation services that address identified transportation problems. At first, 29 user services were defined based upon the consensus of industry. To date, the total number of user services is 33, and they are grouped into the following user service areas: (1) travel and traffic management, (2) public transportation management, (3) electronic payment, (4) commercial vehicle operations, (5) emergency management, (6) advanced vehicle safety systems, (7) information management, and (8) maintenance and construction operations. It is necessary to define a set of functions to accomplish these user services. For example, to define the speed of a roadway based on the traffic condition, the traffic needs to be monitored and then data collected by monitoring the traffic flow will be used to predict the speed for the roadway segment. A set of functional statements, which is used to define these different functions of each of the user services, is called user service requirements. Each user service requirement contains a "shall" statement. A new user service requirement is required to be defined, if an agency needs to perform a function and it is not mapped to the existing user service requirements. These user service requirements provide a direction to develop functional processes and information flows[1] of the ITS services instead of acting as mandates to the system/architecture implementers.

1.3.2 LOGICAL ARCHITECTURE

The logical architecture is outlined by a set of activities, functions, processes, information, and data flows as a response to the user service requirements in the US national ITS architecture. The objective of the logical ITS architecture is to define the functional processes and information or data flows of the ITS, and provide guidance to generate the functional requirements for the new ITS applications. A logical architecture does not depend on any technology and implementation. It does not determine where the functions are performed, by whom the functions are performed, or identify how the functions are to be implemented. Using the data flow diagrams, ITS functions are described. Fig. 1.4 shows a simplified data flow diagram for ITS [25]. The rectangles represent the terminators,[2] the circles representing the functions, and the lines connecting the circles and rectangles representing the data flows. Circles representing the functions in the data flow diagram can be decomposed further at lower levels. Process Specification is the lowest level of decomposition.

1.3.3 PHYSICAL ARCHITECTURE

Based on the logical architecture, the physical architecture is developed, and it is composed of the physical subsystems and architecture flow. The physical architecture describes in which way the system should provide the necessary functionality, assigns the processes to the subsystems and

[1]In a physical architecture, any information exchanged between subsystems, and between subsystems and terminators is known as information flow.

[2]Terminators are the boundaries of the architecture. Terminators are people, systems, and general environment which interface to ITS.

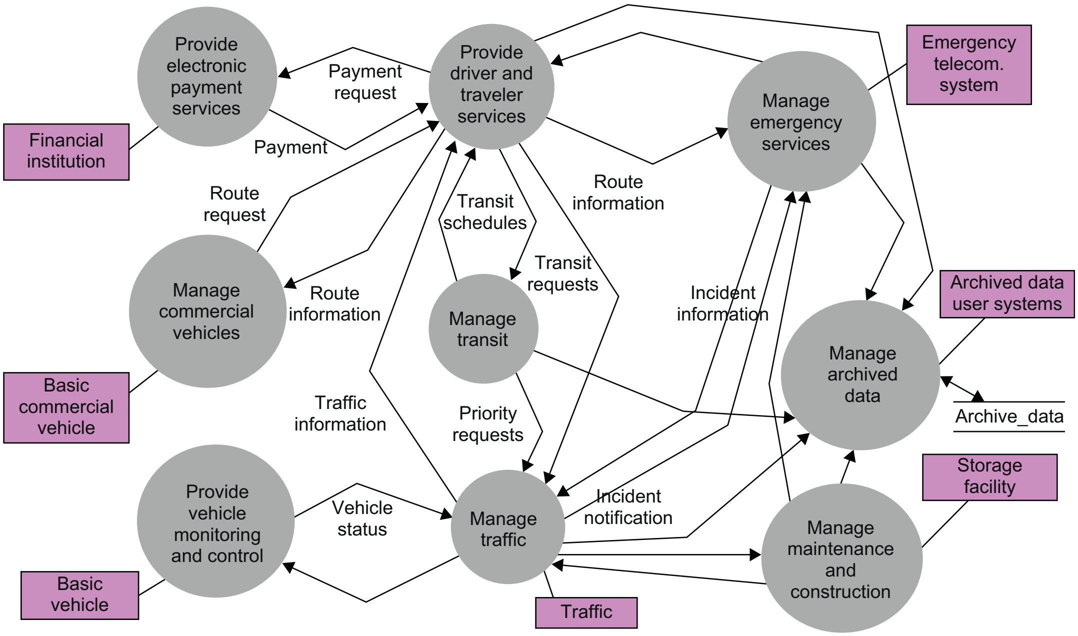

FIGURE 1.4

Data flow diagram.

Source: *Iteris, National ITS Architecture Glossary.* ⟨*http://www.iteris.com/itsarch/html/glossary/glossary-l.htm*⟩*, 2016 (accessed 01.07.16)* [25].

terminators in the ITS architecture. The subsystems (as shown in Fig. 1.5), which are the physical entities of the architecture, are grouped into four classes:

1. Centers, which provide specific functions for the transportation system including management, administrative and support functions;
2. Roadside subsystems, which are spread along the road network and used for surveillance, information provision, and control functions;
3. Vehicles, including driver information and safety systems; and
4. Travelers, who use mobile and other devices to access ITS services before and during trips.

The primary component of the subsystems are equipment packages (as shown in Fig. 1.5), which collect the same type of processes of an individual subsystem to make them an implementable package. The data flows from the logical architecture flow from one subsystem to the other. Data flows are grouped together into architecture flows (as shown in Fig. 1.5). The interfaces/data communication required between subsystems and terminators are defined by the architecture flows and their communication requirements are outlined in different ITS standards.

1.3.4 SERVICE PACKAGES

Service packages offer a service-oriented view of the National ITS Architecture. These service packages are designed to accommodate real world transportation problems. Within the physical

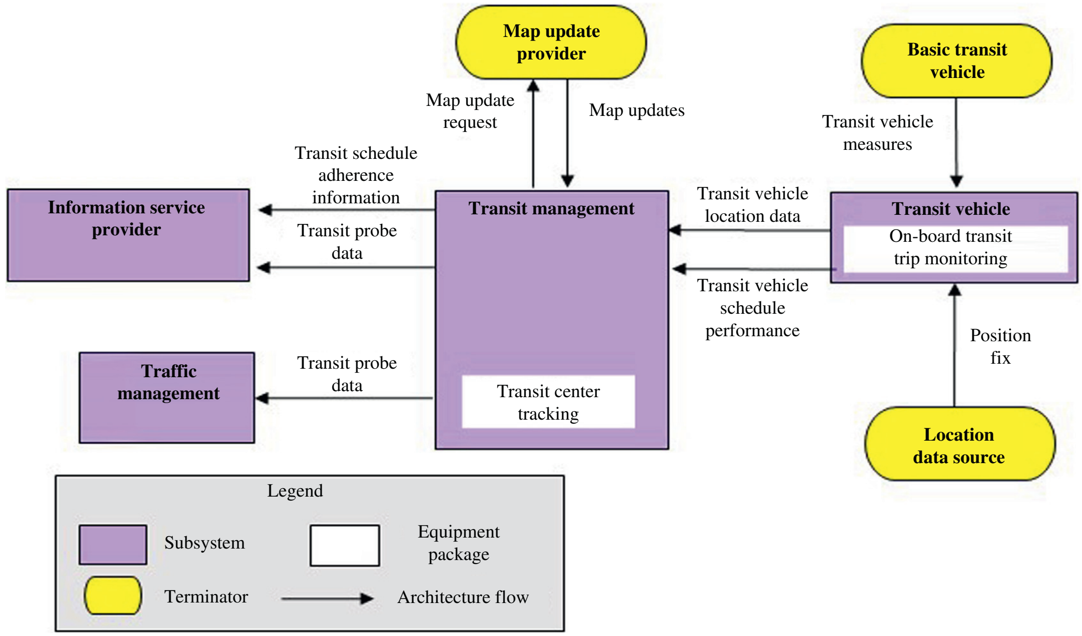

FIGURE 1.5

Transit vehicle tracking service package.

Source: *APTS01-Transit Vehicle Tracking.* ⟨*http://www.iteris.com/itsarch/html/mp/mpapts01.htm*⟩, *2016 (accessed 24.10.16) [26].*

architecture, service packages address specific services. For example, transit vehicle tracking service is provided by the transit vehicle tracking service package. In order to provide a desired service, a service package combines multiple subsystems, equipment packages, terminators, and architecture flows. As an example, Fig. 1.5 shows the transit vehicle tracking service package. Using an automated vehicle location system, this service package monitors transit vehicle location. In this service package, there are four subsystems which include (1) the information service provider, (2) traffic management, (3) transit management, and (4) transit vehicle. Also, this service package has three terminators that include (1) basic transit vehicle, (2) map update provider, and (3) location data source. The Transit Management Subsystem has three tasks, which are (1) processing the information of transit vehicle position, (2) updating the transit schedule, and (3) making real-time information available to the other subsystem, information service provider.

1.3.5 STANDARDS

USDOT envisioned an open ITS environment and ITS Standards are fundamental for this goal. Standards help to integrate independently operated components to provide an interoperable system. Standards ensure the system's interoperability at various levels (e.g., local, regional, and national levels) without impeding technological advancement. The standards development organizations are supported by the USDOT's ITS Joint Program Office (JPO). Both the logical and physical architecture provide the foundation to develop standards. The identified architecture flows (from physical architecture) and data flows (from logical architecture), and the way in which the information is exchanged across different interfaces need to be standardized. Multiple organizations have participated in ITS standards activities, such as the American Association of State Highway and Transportation Officials, the American National Standards Institute, the American Public Transportation Association, the American Society for Testing and Materials, the Institute of Electrical and Electronics Engineers, the Institute of Transportation Engineers, the National Electrical Manufacturers Association, and the Society of Automotive Engineers.

1.3.6 SECURITY

Security defines the protection of the surface transportation infrastructure and information, and it provides the security services and mechanisms to achieve this high-level goal. To collect and distribute information, today's surface transportation system highly depends on information technologies for advancing the mobility and safety of the overall system. In the National ITS Architecture, security is represented in two ways: (1) Securing ITS and (2) ITS Security Areas. The foundation of the ITS security systems is "Securing ITS." Different components (e.g., subsystems, architecture flows) must be protected to provide reliable application services. There are four different areas for Securing ITS, which include: (1) information security, (2) ITS personnel security, (3) operation security, and (4) security management. On the other hand, multiple security areas exist that define how ITS can be used in detecting, and responding to security threats and events on the transportation systems. These security areas include: (1) disaster response and evacuation, (2) freight and commercial vehicle security, (3) HAZMAT security, (4) ITS wide area alert, (5) rail security, (6) transit security, (7) transportation and infrastructure security, and (8) traveler security. These eight ITS security areas are supported by the "Securing ITS" security services. For example, a transit surveillance system can be considered to explain these two security aspects, which includes a control center and CCTV cameras. Control center can only control the cameras. Any sensitive camera images cannot be disclosed to any unauthorized person from Securing ITS perspective, and must be protected. These considerations are addressed as part of "Securing ITS." From another perspective (i.e., ITS Security Area perspective), the transit surveillance system provides a deterrent and a response tool to advance the transportation system security, which is defined in "Transit Security."

1.4 OVERVIEW OF ITS APPLICATIONS

To "enhance American productivity through the integration of advanced communications technologies into the transportation infrastructure and within vehicles," ITS uses advanced computing and communications technologies to address transportation problems, and advance the safety, mobility and environmental aspects of surface transportation systems [19]. For example, Georgia Navigator, operated by the Georgia DOT has implemented an Advanced Traffic Management System (as shown in Fig. 1.6). The Navigator has managed traffic in Metro Atlanta with traffic cameras, ramp meters, changeable message

FIGURE 1.6

Georgia Navigator, a Traffic Management Center (TMC) in Atlanta, GA, United States.

signs, and a traffic speed sensor system since 1996. The data (e.g., traffic condition, lane closure, trip time) collected by the Navigator system can be used to enable different ITS applications.

ITS application deployments have a higher return on investment when compared to costly traditional infrastructure-based road development [27]. The underlying goals for these ITS applications are to reduce congestion, improve safety, mitigate adverse environmental impacts, optimize energy performance, and improve the productivity of surface transportation. An overview of different ITS applications is provided in this section.

1.4.1 TYPES OF ITS APPLICATIONS

ITS applications are broadly classified into three categories: mobility, safety and environmental. ITS mobility applications are intended to provide mobility services such as shortest route between origin-destination pair considering different factors (e.g., distance, time, energy consumption) in a data-rich travel environment based on information collected by the ITS data collection technologies. By adjusting traffic signals, dynamically managing transit operations, or dispatching emergency maintenance services, these applications can help transportation management centers

Table 1.2 Example ITS Applications

Application Type	Application Name	Goal	Data Source	Data Users
Mobility	Transit signal priority	To advance real-time transit system performance	Transit vehicle traffic signal	Transit management center Traffic management center
Safety	Vehicle safety monitoring	• To detect critical elements of the vehicle • To alert the driver about any potential dangers	Vehicle on-board system	Vehicle safety monitoring system
Environmental	Environmental probe surveillance	To collect data from vehicles to infer real-time environmental conditions	Vehicle on-board systems	Weather service Maintenance and construction management center

monitor and manage transportation system performance. The ITS safety applications, such as providing a speed warning at a sharp curve or slippery roadway, will reduce crashes by providing advisories and warnings. These applications include vehicle safety application (e.g., vehicle safety monitoring, driver safety monitoring), emergency management (e.g., emergency routing). The instant traffic congestion information can help a traveler make informed decisions that in-turn decrease the environmental impact of day-to-day trips. Travelers can avoid congestion by taking alternate routes or by rescheduling their trips, which in turn can make the trips more eco-friendly.

The three ITS applications (mobility, safety, and environmental) are shown in Table 1.2. Each example is listed with its goal, data sources, and data users. Note that the data sources include both vehicle and infrastructure sources, and that “users” include human users, centers, and vehicle system.

Each ITS application has a set of stakeholders, which may vary depending upon the ITS application. For example, the variable speed limits application, described below, has stakeholders that include public or private transportation agencies (or both), law enforcement authorities, emergency management services, and vehicle drivers. Cooperation by these stakeholders is critical in the successful design, deployment and management of any ITS application.

A brief case study of an example ITS application, a variable speed limits system, which is one widely implemented ITS application, is presented here. A variable speed limits system uses traffic devices and sensors such as loop detectors, video cameras, and probe vehicles to monitor the prevailing traffic and weather conditions. The application determines the appropriate speed limits to be posted on variable message signs with goals that include safety improvement, congestion reduction, vehicle energy usage minimization, and air pollution reduction. This application is particularly critical for ensuring traffic safety since the posted speed limits are only applicable under noncongested traffic and good weather conditions. When the conditions are less than ideal, for example, during peak rush hour or inclement weather, then the safe operating speed is below the posted speed. Variable speed limits systems use real-time data about the traffic speed, volume, weather information, road surface conditions to determine safe speed.

The variable speed limits application illustrates how different ITS components (sensors, motorists, and ITS centers) interact with each other to achieve a specific purpose. An ITS center, such as a Traffic Management Center (TMC), receives and stores data as input and provides intelligent decision support. In a TMC, the variable speed limits application receives data from ITS devices and sensors, calculates the variable speed limits for a given corridor, and communicates the speed limits to road users via variable speed limits signs. The application is typically monitored and managed centrally at a TMC. The collected data characteristics can vary based on the data collection devices. From a Big Data analytics perspective, data arrive in a stream from sensor data sources on the roadway or in the vehicles. An appropriate infrastructure at the TMC is used to aggregate the data, statistical methods are used to measure the anomalies, and trend analysis is used to measure the traffic flow. Machine learning methods are used to predict future trends and the application sets suitable speed limits after processing the raw data in real-time.

An ITS application can offer multiple services. The US national ITS architecture presents the concept of a service package, where several subsystems, equipment packages, terminators and architecture flows are combined to provide a desired service for stakeholders [24]. For example, the US national ITS architecture has identified the variable speed limits as a service package, which consists of two subsystems as shown in Fig. 1.7. The traffic management subsystem (a center subsystem), included in a transportation facility management center, supports monitoring and controlling of roadway traffic. This subsystem exchanges data with the other subsystem in the variable speed limits service package, which is the roadway subsystem. The roadway subsystem includes the roadway equipment (e.g., traffic detectors, environmental sensors, traffic signals) distributed throughout a corridor for traffic monitoring and roadway management. Here the variable speed

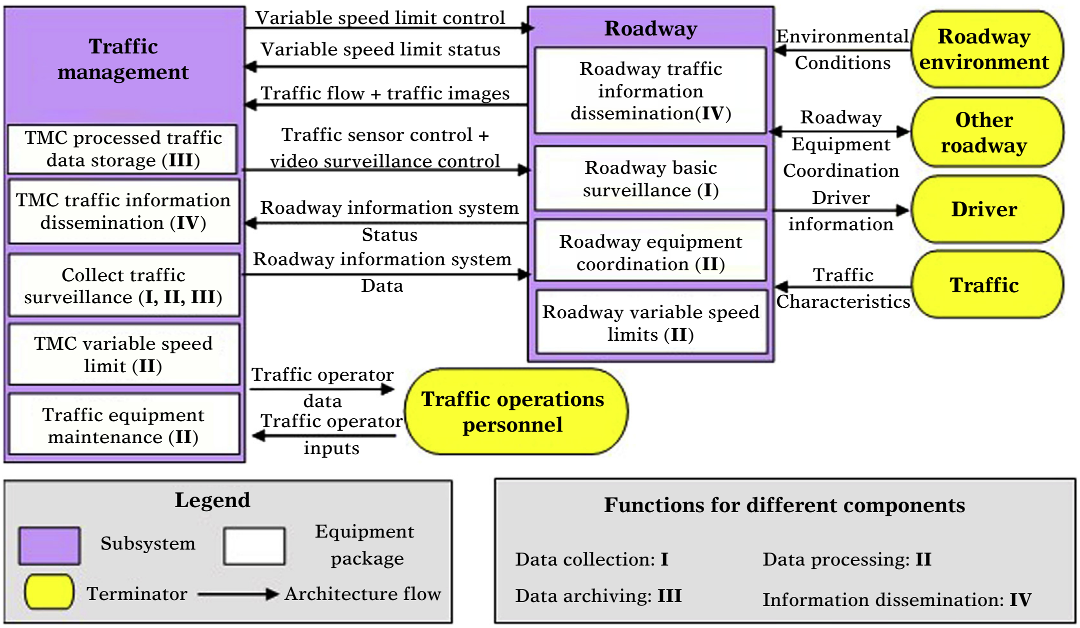

FIGURE 1.7

Variable speed limits service package.

Source: *Adapted from ATMS22-Variable Speed Limits. ⟨http://www.iteris.com/itsarch/html/mp/mpatms22.htm⟩, 2016 (accessed 24.10.16) [28].*

limit service package is supporting the application of setting variable speed limits to promote safety, and improve operational and environmental conditions.

Four functions are performed by the variable speed limits service package: data collection, data processing, data archiving, and information dissemination. Data collected from the roadway, the roadway environment, and traffic are forwarded to the traffic management subsystem. The roadway environment produces data about the physical condition and geometry of the road surface. Data produced also include roadway conditions such as ice, fog, rain, snow, or wind. Data from traffic include real-time vehicle population that provide the traffic flow, and traffic images required for surveillance.

The variable speed limits application has been under continuous evolution since its introduction in 1960 in the United States [29]. Indeed, one of the most recent iterations in use on a section of I-5 in the state of Washington has reduced crashes by 13% [30]. A similar version in New Jersey has significantly decreased the average traffic speeds in adverse weather and traffic conditions and associated weather-related accidents [30].

1.4.2 ITS APPLICATION AND ITS RELATIONSHIP TO DATA ANALYTICS

The USDOT has developed CVRIA to ensure the uniformity in the Connected Vehicle (CV) early deployments. Following this architecture, multiple CV pilot deployments (e.g., Wyoming, Florida, New York [31]) of different CV applications are under development and several CV field demonstrations were developed. For example, a CV field demonstration was performed by Clemson University researchers, where they demonstrated three CV applications: (1) collision warning, (2) queue warning, and (3) traffic mobility data collection in the ITS Carolinas Annual Meeting 2015. Multiple wireless communication technologies—DSRC, cellular/LTE and Wi-Fi were used to demonstrate that different communication technologies can be seamlessly integrated to support the diverse CV application requirements.

In CVRIA, USDOT has identified 97 CV applications, which are categorized into four groups: (1) environmental applications, (2) mobility applications, (3) safety applications, and (4) support applications [12]. Cooperative Adaptive Cruise Control (CACC) is one of the CV mobility applications. Following the conventional cruise control (CCC) systems and adaptive cruise control (ACC) systems, the CACC application represents an evolutionary advancement that utilizes vehicle-to-vehicle (V2V) communication to synchronize CV movement in a vehicle platoon. The physical architecture of this application is shown in Fig. 1.8. The physical architecture includes physical objects, application objects and information flows between application objects, which are required to support the application's functional requirements.

There are four different physical objects in this application: (1) traffic management center, (2) ITS roadway equipment, (3) roadside equipment (RSE), and (4) vehicle on-board equipment (OBE). Each physical object has some specific functions. Functions are classified into four different types, from the perspective of data analytics: (1) data collection, (2) data processing, (3) data archiving, and (4) data dissemination. For example, "RSE-traffic-monitoring" function of the RSE monitors the basic safety messages (BSMs) that are transferred between CVs. This function performs the data processing task, and calculates traffic flow measures based on the collected BSMs.

The information flows between application objects have two contexts: spatial context and time context. The spatial contexts are classified into five categories and the time context is classified into four groups (as shown in Fig. 1.8 and Table 1.3). For example, "traffic-flow" information flow from the CACC application represents the flow of raw/processed data from the ITS roadway equipment to the TMC. Therefore, the information flow characteristics are "local" in spatial context as

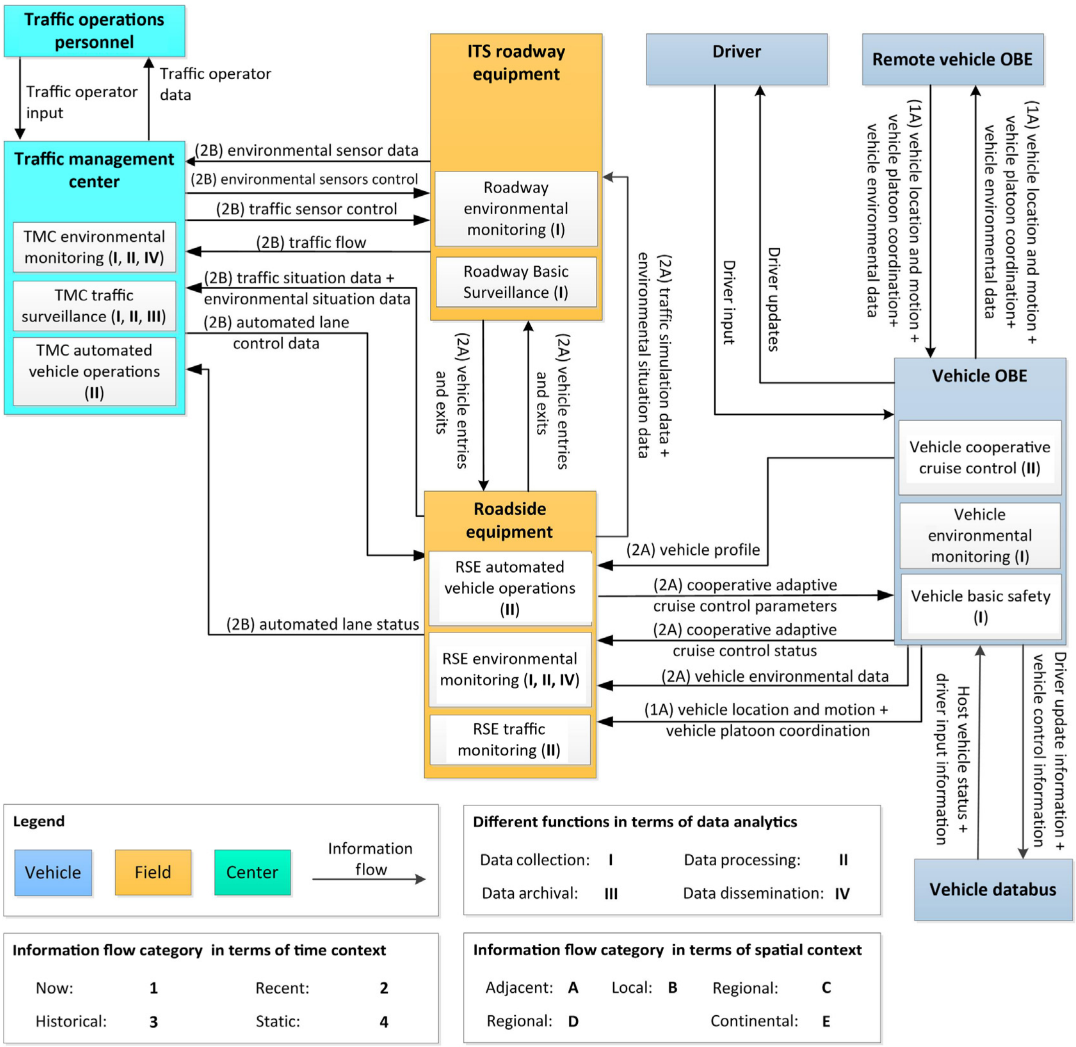

FIGURE 1.8

Physical architecture of Cooperative Adaptive Cruise Control (CACC) application including different functions and information flow characteristics in terms of data analytics.

Source: *Adapted from CVRIA, Cooperative Adaptive Cruise Control.* ⟨*https://www.iteris.com/cvria/html/applications/app8.html#tab-3*⟩, *2014 (accessed 09.10.16) [32].*

the distance between the TMC and ITS roadway equipment is expected to be within 3 km [32,33], and "recent" in time context as the information needs to be transmitted between 1 s and 30 min, which can vary depending on the application requirement. Consequently, it is challenging to deliver data at the same time satisfying different CV application requirements, which necessitates the design of Big Data analytics for a connected transportation system.

Table 1.3 CVRIA Information Flow Characteristics

Information Flow Characteristics	CVRIA Data Flow Category	Characteristic Value Description [33]
Spatial context	Adjacent (A)	0–300 m
	Local (B)	300 m–3 km
	Regional (C)	3–30 km
	National (D)	30 km to National
	Continental (E)	Continental U.S.
Time context	Now (1)	<1 s
	Recent (2)	1 s–30 min
	Historical (3)	30 min–1 month
	Static (4)	>1 month

As discussed in Section 1.1, within the scope of CV and intelligent infrastructure, Big Data is defined by: (1) Volume—Massive volume of data collected from millions of CVs on our roadways, RSE, and internet data (example problem 1 will help the readers to understand how large volume of data is generated from GPS-enabled ITS devices); (2) velocity—The high arrival rate of GPS sampling data, social media messages, and data generated from vehicle on-board devices; (3) variety—The differences within the industry standards, sampling rates, and data types; and (4) veracity—The potential for missing or erroneous data due to environmental conditions, equipment failures, or malicious intent.

Example Problem 1

In order to get localization information, real-time mobility data can be collected from GPS-enabled ITS devices. For example, consider 20,000 people use GPS-enabled devices in a city. Assume that the minimum size of a GPS record is 20 bytes (2 8-byte values of type double for latitude and longitude and 1 4-byte value for time stamp), and data are collected at most once every 10 seconds (i.e., 8640 samples per device per day). What is the amount of stored data in gigabytes (GB) that would be collected in a day? (For storage, 1 GB = 2^{30} bytes)

Solution:

$$\begin{aligned}\text{Daily stored GPS data collected from the city (GB)} &= \text{Size of a GPS record (GB)} \cdot \text{number of samples per device per day} \cdot \text{number of GPS-enabled devices}\\ &= \frac{20}{2^{30}}\ \text{GB} \cdot 8640 \text{ samples per device per day} \cdot 20{,}000 \text{ device}\\ &= 3.219 \text{ GB per day}\end{aligned}$$

It shows, for a city with 20,000 GPS-enabled ITS devices, daily 3.219 GB of data can be generated which we need to further process to extract relevant localization information.

1.5 INTELLIGENT TRANSPORTATION SYSTEMS PAST, PRESENT, AND FUTURE

Here, a brief history of ITS is detailed from its nascent development in the 1960s, to its current iteration as an integral part of all modern surface transportation systems, to the envisioning of how the next generation of ITS will evolve during the 21st century.

1.5.1 1960'S AND 1970'S

The ITS era began in the United States following the development of the Electronic Route Guidance System (ERGS) [27]. The purpose of this program was to provide the motorist with route guidance information through the electronic navigation equipment installed in vehicle and at the respective intersections. First, the motorist entered a trip destination code into the in-vehicle equipment which was then transmitted to the equipment installed at the instrumented intersections. The trip destination code was then decoded and a routing instruction was transmitted back to the vehicle. Following the translated symbol or word messages, the driver then performed the required maneuver at the upcoming intersection.

The automatic route control system or ARCS, developed in the 1970s, was substantially more complicated and automated compared to the ERGS [31]. Unlike the ERGS, the ARCS continuously measured and compared the coordinates of the vehicle's location with the coordinates of the predetermined route, and later provided guidance (audio, visual, and/or printed instructions). The digital processing and logic unit was considered the heart of this system, which was programed to analyze the output of the speed and direction sensors in real-time, and compute and compare the route of the vehicle with the route signature recorded on the tape cartridge, and issue instructions accordingly.

Developed in Japan, the comprehensive automobile traffic control system was mainly a communication system that links (1) in-motion vehicles, (2) RSE and (3) a central data processing center [27]. Information is transferred from in-vehicle transmitters to the central computer control via the RSE. Based on the collected information, the central computer continuously monitors the traffic on arterials and major intersections, and at each intersection drivers were instructed about the optimal route and emergency, and driving advisories were forwarded directly to each vehicle. The Comprehensive Automobile Control System (CACS) was tested in a pilot program successfully in 1977 [34].

The Autofahrer Leit and Information System (ALI), developed in Germany in the mid-1970s, was similar to the CACS in that it was a dynamic route guidance system based on loop detector-collected real traffic condition data. The information was made available to the vehicle drivers though an on-board display.

1.5.2 1980'S AND 1990'S

The pace of ITS application deployment accelerated in the 1980s, with technological advancements (e.g., the introduction of efficient memory storage and computer processing power).

At the beginning of the decade two European projects were begun: (1) the Program for a European Traffic System with Higher Efficiency and Unprecedented Safety (PROMETHEUS), funded by the consortium of European automotive manufactures; and (2) the Dedicated Road

Infrastructure for Vehicle Safety in Europe (DRIVE) which was sponsored by the European Union. The Road/Automobile Communication System (RACS) initiated in Japan in 1984 which shaped the basis for the car navigation system that exists today [27]. The RACS focused on connecting vehicles and RSUs with radio communication, whereas the roadway facilities and TMCs were connected with wire network.

In 1987 the USDOT began the "Mobility 2000" program which evolved into the Intelligent Vehicle-Highway Systems (IVHS) program under the Intermodal Surface Transportation Efficiency Act (ISTEA) enacted in 1991 [31]. The purpose of the ISTEA was to promote the safety, capacity and efficacy of the US transportation system, while minimizing the adverse environmental impacts. Also, in that year, ITS America was initiated to improve the utilization of advanced ITS technologies in US surface transportation systems. As a nonprofit organization, ITS America acts a policy making and advocacy platform for public and private sector stakeholders, and collaborates with similar organizations in other countries.

To facilitate ITS deployments, the European Road Transport Telematics Implementation Coordination Organization (ERTICO) was initiated in 1991 in the form of a private–public-partnership-based organization with ITS stakeholders to improve the secure, safe, clean, orderly, and comfortable movement of both traveler and goods in Europe with the widespread ITS deployment.

Three years later, in April 1994, the Fourth Framework Program of the European Union was created to develop the transportation telematics and ITS applications. In that year on the other side of the world, ITS Japan was established in cooperation with five Japanese government ministries to work with national and international transportation organizations. In 1996, the Vehicle Information and Communication Systems (VICS) began operation in Tokyo and Osaka to provide traffic information to motorists that was retrieved from the national Highway Traffic Information Centre and disseminated through road-side beacons and FM broadcasts

In the mid-1990s, ISTEA mandated the development of an automated highway system with the mission of developing a system in which automated vehicles will operate without direct human involvement in steering, acceleration, and braking. These automated vehicles can be autonomous in that they use only vehicle sensors, and connected, using connectivity between vehicles and roadside infrastructure wirelessly. The National Automated Highway System Consortium (NAHSC), composed of nine core public and private agencies, was an evolution of the earlier scheme. Developed by the USDOT, the NAHSC's work concluded with the use of 20 fully automated vehicles in operation on 1–15 in San Diego, California in the Demo 1997 [31].

1.5.3 2000'S

In 2001, the Vehicle Infrastructure Integration (VII) research program was initiated by the USDOT to identify the potential use of the DSRC-for both V2V and vehicle-to-infrastructure (V2I) communication. In 2004, the Federal Communications Commission published an order which established standard licensing and service rules for DSRC in the 5.9-GHz band [31]. DSRC provides a wireless communication link to share information between both vehicles and roadside infrastructure, which can be used for protecting the traveling public's safety. As a continuation of the Advanced Vehicle Control Systems envisioned by Mobility 2000, USDOT initiated a 5.9 GHz-based VII proof-of-concept. In 2008, the USDOT conducted a test to investigate the technical feasibility of V2V and V2I applications in Michigan and California test beds.

In this period, significant progress was made for vehicle automation. In Europe, Volkswagen developed the Temporary Auto Pilot (TAP) system. Based on the driving situation, surrounding situation analysis, the condition of the driver and the status of the system, TAP provides the optimal automation degree for the driver on roads at speeds between 0−130 km/h. The optimal automation degree will prevent crashes due to human errors by a distracted driver. During this time DARPA Challenge series, a first-of-its-kind race to stimulate the development of self-driving vehicles, took place in the United States. DARPA challenge was funded by the Defense Advanced Research Projects Agency. This agency is a research institute of United States Department of Defense. Regarding the autonomous passenger cars, three DARPA events were held in 2004, 2005, and 2007. Later in 2009, Google officially started the Self-Driving Car project.

1.5.4 2010'S AND BEYOND

In the United States, recent ITS research and deployments have focused on connected and automated vehicles. In a CV environment, vehicles use a number of different communication technologies (such as DSRC) to communicate with the other surrounding vehicles (V2V) and roadside infrastructure (V2I) [35]. In 2012−13, the CV Safety Pilot Model Deployment occurred in Ann Arbor, Michigan. As participants in the CV Pilot Deployment program, USDOT announced three CV deployment sites in September 2015. These sites include corridors from Wyoming, New York, and Florida [31].

To promote automated vehicle research, the USDOT's ITS Joint Program Office has developed a 2015−19 Multimodal Program Plan, in which a regulatory framework for autonomous vehicle operation on public roads was established in California. Same legislations are being considered in Florida, District of Columbia, Nevada, and Michigan. In 2011, Japan initiated ITS Spot program using ITS spots (DSRC radio) to support (1) Dynamic Route Guidance, (2) Driving Safety, and (3) Electronic Toll Collection. Following that 2011 program, Japan announced an Automated Driving System Research (ADSR) Program in May 2014. The purpose of this program is to develop and verify the automated driving system (ADS) for safe operations on public roads. The government's goal through the Japan Revitalization Strategy is to make a test installation of ADS by 2030. This system includes the development of technologies to generate a dynamic map and prediction data, and to enhance the sensing capability [36].

In Europe, the "Horizon 2020" program outlined in the ERTICO Automated Driving Roadmap includes the framework for safe automated road transportation. Indeed, many European countries, particularly the United Kingdom, Germany, and France, are already active in autonomous vehicle systems research within their own jurisdictions [37]. The United Kingdom recently completed a regulatory review to remove any possible barriers for testing autonomous vehicles on UK roadways. Round Table Automated Driving (RTAD) is formed to support automated driving on German roads. However, the testing of automation technology has already started by vehicle manufacturers in Germany. Other developed countries like South Korea, Canada, Australia, and Singapore are also conducting autonomous vehicle research and development. A brief overview of the ongoing ITS development initiatives in the United States, Japan and Europe is summarized in Table 1.4.

By 2050, urban population will be approximately 66% of the total world's population, an increase of 16% from the 2008 census data. With more people living in urban areas, cities will face extreme transportation challenges characterized by managing safety and air pollution under conditions of excessive traffic congestion and inadequate infrastructures. ITS applications will become even more critical in these challenging future scenarios. Connected vehicles will alleviate traffic congestion and increase traveler safety and environmental benefits. Autonomous vehicles will become available to

Table 1.4 ITS Development in the United States, Japan, and Europe

	1980 and Earlier	1981–90	1991–2000	2001–16	Beyond 2016
United States	ERGS, ARCS	Mobility 2000	IVHS, ITS America	VII	Multimodal Program Plan for Vehicle Automation
Japan	CACS	RACS	VICS, ITS JAPAN	ITS Spot, ADSR	ADS
Europe	ALI	PROMETHEUS	DRIVE, ERTICO	RTAD	Horizon 2020

the mass population to further improve mobility efficiency and safety. Technology applications such as traveler information and demand-specific ride sharing services like Uber and Lyft, and the growth of shared-use mobility applications will help to alleviate transportation issues. Smart and connected cities will emerge as a system of interconnected systems, including transportation, residencies, employment, entertainment, public services, and energy distribution. Developing such an all-encompassing future system, however means emphasizing:

- The importance of utilize emerging capabilities that demonstrate the potential to transform transportation at the same time that user and citizen privacy is protected;
- The evolution of standards and architectures to ensure that technological advancements are reflected, and the maintenance of backward compatibility and interoperability of different ITS components;
- The development of a workforce of transportation professionals trained to capture, manage and archive data collected from the smart city system;
- Public agency acceptance of the integration of data analytics via public–private partnerships to provide public agency transportation professionals with the required skillsets to manage these systems.

1.6 OVERVIEW OF BOOK: DATA ANALYTICS FOR ITS APPLICATIONS

The purpose of this book is to prepare an educated ITS workforce and tool builders for the data analytics-enabled ITS. To do so, the overview of data analytics for ITS detailed in this chapter involves a discussion of ITS as a data-intensive application, the sources of ITS data, an overview of Big Data analytics and computational infrastructure needed to support data analytics in ITS. The chapter also describes ITS applications and ITS history. ITS architecture has also been discussed as the framework of ITS applications. Many countries, including the United States, Japan, and European countries, are actively performing research and innovations regarding ITS advancements.

To support Big Data for ITS applications, high performance computing facilities are required as more and more data sources are emerging. Many high performance computing facilitates are available to support for Big Data research. For example, Titan is the fastest supercomputer in 2016 within the United States. Titan is built by Cray and is located Oak Ridge National Laboratory. It uses both conventional central processing units and graphics processing units. Such data analytics research facilities will help to manage large volume of data collected from multiple ITS devices.

(*Continued*)

(CONTINUED)

Titan, Oak Ridge National Laboratory—The fastest US supercomputer.

Source: *Titan (supercomputer).* ⟨*https://en.wikipedia.org/wiki/Titan_(supercomputer)#/media/File:Titan_supercomputer_at_the_Oak_Ridge_National_Laboratory.jpg*⟩, *(accessed 09.10.16).*

Information technology companies that are leaders in Big Data analytics are also some of the largest companies in the world, including Google, Facebook, Twitter, Amazon, Apple, and others. These companies build massive data centers to collect, analyze and store the enormous amount of data. The figure below represents the servers of Facebook data center. This data center is located in Oregon, United States.

Facebook Data Center, Oregon, United States.

Source: *Intel Team Inside Facebook Data Center, by Intel Free Press (CC BY 2.0 (http://creativecommons.org/licenses/by/2.0)), via Wikimedia Commons ⟨https://commons.wikimedia.org/wiki/File:Intel_Team_Inside_Facebook_Data_Center.jpg⟩, (accessed 11.11.16).*

The remaining chapters of this book provide a comprehensive study of data analytics for ITSs. The book is divided into two parts. The first part of this book, Chapters 2–7, covers fundamental topics in data analytics and is the starting knowledge needed for someone new to the area of data analytics. The description of the fundamental of data analytics in Chapter 2. Data Analytics: Fundamentals, provides an introduction to functional facets of data analytics, evolution of data analytics and data science fundamentals. In Chapter 3, Data Science Tools and Techniques to Support Data Analytics in Transportation Applications, the tools for data analytics are discussed and several tutorial presentations are provided. In Chapter 4, The Centrality of Data: Data Lifecycle and Data Pipelines, the data lifecycle and data pipeline detail an understanding of the variety of data that is available for ITS and how different data must be managed and maintained differently. A comprehensive overview of the current data infrastructure, and the tools and systems required for the ingestion, storing, and transformation of today's large scale data is provided in Chapter 5, Data Infrastructure for Intelligent Transportation Systems, followed by the discussion about the often overlooked problems of the security of ITS data, and privacy concerns for ITS data in Chapter 6, Security and Data Privacy of Modern Automobiles. A discussion of data visualization tools walks the reader through both the principles of data visualization and example use of tools and interactive data visualization exercises in Chapter 7, Interactive Data Visualization. Those interested in understanding the landscape of data analytics in ITS are encouraged to study all of these chapters.

The second part of this book, Chapters 8–12, cover additional topics in data analytics for a professional workforce in ITS. A beginning reader may read these chapters selectively, and a thorough study of all of these chapters will be solid preparation for the ITS data analytics professional. Chapter 8, Data Analytics in Systems Engineering for Intelligent Transportation Systems, covers systems engineering of ITS and gives an introduction of the major tools and languages used in this field. The development of a new ITS application is a complex systems engineering task. Also included are the systems engineering task description and the systems engineering process, and a detailed tutorial and case study using the Architecture Analysis and Design Language (AADL).

Chapters 9–11 provide case studies and examples of data analytics in several important areas: safety applications, intermodal freight transportation applications, and social media applications. Together these chapters prepare the reader with tools for solving data analytics problems in a variety of ITS settings. Finally, Chapter 12, Machine Learning in Transportation Data Analytics covers the major machine learning methods of relevance to ITS.

EXERCISE PROBLEMS

1. Identify possible user service requirement for implementing the Transit Signal Priority application in your area. Develop a data flow diagram and map the data flow diagram to a physical architecture. Show the traceability between user service requirement, logical and physical architecture.
2. Provide a detail description of the Traffic Signal Control application in terms of four functions (i.e., data collection, data processing, data archiving, and information dissemination), which are described in this book.
3. Describe the 5 V's of Big Data and give examples from ITS.

4. Explain how the sources of ITS data available today create Big Data from the perspective of the 5 V's as compared to historical sources of ITS data.
5. Identify and describe different emerging data collection technologies for the automated vehicle systems. How these data collection technologies differ from the traditional ITS data collection technologies such as loop detectors and CCTV camera?
6. Describe the complexities of modern ITS in terms of data analytics. How does the data analytics of automated vehicle system differ from the current data analytics?
7. What types of data collection technology are mostly used by your local transportation agencies? Do the local transportation agencies require any Big Data analytics infrastructure to process the collected data?
8. You need to quantify the data generated from GPS-enabled ITS devices from two cities: Dhaka and Istanbul. The numbers of GPS-enabled devices in Dhaka and Istanbul are 6,000,000 and 12,754,334, respectively. Assume that the minimum size of a GPS record is 20 bytes. In a typical GPS map-matching process, the GPS data collection rate for one device can be as high as once every 10 s (i.e., 8640 samples per device per day) or as low as once every 2 min (i.e., 720 samples per device per day). Calculate (1) the amount of daily GPS data in GB collected with a high data collection rate and (2) the amount of daily GPS data in GB collected with a low data collection rate. (For storage, 1 GB = 2^{30} bytes). Use the following equation to calculate the daily GPS data:

$$\text{Daily stored GPS data(GB)} = \text{Size of a GPS record(GB)} \\ \cdot \text{number of samples per device per day} \cdot \text{number of GPS-enabled devices}$$

REFERENCES

[1] D. Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical report, META Group, 2001.

[2] J. Dorsey, Big Data in the Drivers Seat of Connected Car Technological Advances. ⟨http://press.ihs.com/press-release/country-industry-forecasting/big-data-drivers-seat-connected-car-technological-advance⟩, 2013 (accessed 18.07.16).

[3] K. Lantz, S.M. Khan, L.B. Ngo, M. Chowdhury, S. Donaher, A. Apon, Potentials of online media and location-based Big Data for urban transit networks in developing countries, Transportation Research Record, J. Transport. Res. Board 2537 (2015) 52−61.

[4] B. Vorhies, The Big Deal About Big Data: What's Inside—Structured, Unstructured, and Semi-Structured Data. ⟨http://data-magnum.com/the-big-deal-about-big-data-whats-inside-structured-unstructured-and-semi-structured-data/⟩, 2013 (accessed 17.07.16).

[5] Caltrans, Commercial Wholesale Web Portal. ⟨http://www.dot.ca.gov/cwwp/InformationPageForward.do⟩, 2016 (accessed 17.07.16).

[6] A. Luckow, K. Kennedy, F. Manhardt, E. Djerekarov, B. Vorster, A. Apon, Automotive big data: applications, workloads and infrastructures, in: Proceedings of the IEEE International Conference on Big Data, Santa Clara, CA, USA, IEEE, 2015.

[7] The OSI Model's Seven Layers Defined and Functions Explained. ⟨https://support.microsoft.com/en-us/kb/103884⟩, 2014 (accessed 02.10.16).

[8] G. Rucks, A. Kuzma, How Big Data Drives Intelligent Transportation? ⟨https://www.greenbiz.com/blog/2012/08/15/how-big-data-drives-intelligent-transportation⟩, 2016 (accessed 15.06.16).

[9] BITRE, New Traffic Data Sources—An Overview. ⟨https://bitre.gov.au/events/2014/files/NewDataSources-BackgroundPaper-April%202014.pdf⟩, 2014 (accessed 17.07.16).

[10] W. Guo, Z. Wang, W. Wang, H. Bubb, Traffic incident automatic detection algorithms by using loop detector in urban roads, Recent Patents Comput. Sci. 8 (1) (2015) 41–48.

[11] B.A. Coifman, Vehicle Reidentification and Travel Time Measurement Using Loop Detector Speed Traps, Institute of Transportation Studies, Berkeley, CA, 1998.

[12] CVRIA, Connected Vehicle Reference Implementation Architecture. ⟨http://www.iteris.com/cvria/html/applications/applications.html⟩, 2016 (accessed 15.07.16).

[13] K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, E. Shook, Mapping the global Twitter heartbeat: the geography of Twitter, First Monday 18 (5) (2013).

[14] S. Bregman, Uses of social media in public transportation, Trans. Res. Board 99 (2012) 18–28.

[15] CDOT, Survey Manual, Chapter 4, Aerial Surveys, Colorado Department of Transportation. ⟨https://www.codot.gov/business/manuals/survey/chapter-4/chapter4.pdf⟩, 2015 (accessed 17.07.16).

[16] S.M. Khan, Real-Time Traffic Condition Assessment with Connected Vehicles, M.S. Thesis, Clemson University, Clemson, SC, 2015.

[17] U.S. Department of Transportation, Federal Highway Administration, Office of Freight Management and Operations, Freight Analysis Framework, version 3.4. ⟨http://www.ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/nhsavglhft2011.htm⟩, 2013 (accessed 02.10.16).

[18] U.S. Department of Transportation, Federal Highway Administration, Office of Freight Management and Operations, Freight Analysis Framework, version 3.4. ⟨http://www.ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/nhsavglhft2040.htm⟩, 2013 (accessed 02.10.16).

[19] Intelligent Transportation Systems (ITS), Joint Program Office (JPO), Connected Vehicle Reference Implementation Architecture. ⟨http://www.iteris.com/cvria/docs/CVRIA_Feb_Workshop_Presentation_Slides.pdf⟩, 2016 (accessed 07.10.16).

[20] Hadoop: Open Source Implementation of MapReduce. ⟨http://hadoop.apache.org/⟩, 2016 (accessed 20.07.16).

[21] FHWA, National ITS Architecture and Standards Final Rule. ⟨http://ops.fhwa.dot.gov/its_arch_imp/asflyer.htm⟩, 2016 (accessed 07.10.16).

[22] Frame, Frame Architecture—Background. ⟨http://frame-online.eu/first-view/background⟩, 2016 (accessed 17.06.16).

[23] T. Yokota, R.J. Weiland, ITS System Architectures for Developing Countries, Transport and Urban Development Department, World Bank, http://siteresources.worldbank.org/EXTROADSHIGHWAYS/Resources/ITSNote5.pdf, 2004.

[24] The Architectural View. ⟨http://www.iteris.com/itsarch/html/menu/hypertext.htm⟩, 2016 (accessed 01.07.16).

[25] Iteris, National ITS Architecture Glossary. ⟨http://www.iteris.com/itsarch/html/glossary/glossary-l.htm⟩, 2016 (accessed 01.07.16).

[26] APTS01-Transit Vehicle Tracking. ⟨http://www.iteris.com/itsarch/html/mp/mpapts01.htm⟩, 2016 (accessed 24.10.16).

[27] L. Vanajakshi, G. Ramadurai, A. Anand, Intelligent Transportation Systems Synthesis Report on ITS Including Issues and Challenges in India, Centre of Excellence in Urban Transport IIT Madras, Chennai, TN, 2010.

[28] ATMS22-Variable Speed Limits. ⟨http://www.iteris.com/itsarch/html/mp/mpatms22.htm⟩, 2016 (accessed 24.10.16).

[29] D. Warren, Text from “Variable Speed Limits” PowerPoint Presentation. ⟨http://www.ops.fhwa.dot.gov/wz/workshops/accessible/Warren.htm⟩, Washington, DC, 2015 (accessed 08.10.16).

[30] FHWA, Weather-Related Variable Speed Limit Case Studies. ⟨http://safety.fhwa.dot.gov/speedmgt/ref_mats/fhwasa12022/chap_6.cfm⟩, 2015 (accessed 09.10.16).

[31] A. Auer, S. Feese, S. Lockwood, History of Intelligent Transportation Systems. ⟨http://www.its.dot.gov/history/index.html⟩, 2016 (accessed 15.06.16).

[32] CVRIA, Cooperative Adaptive Cruise Control. ⟨https://www.iteris.com/cvria/html/applications/app8.html#tab-3⟩, 2014 (accessed 09.10.16).

[33] Southeast Michigan Test Bed 2014 Concept of Operations. ⟨http://www.iteris.com/cvria/docs/SE_Michigan_Test_Bed_2014_ConOps_-_v1_-_2014-Dec_29.pdf⟩, 2014 (accessed 09.10.16).

[34] T. Mikami, CACS-Urban Traffic Control System Featuring Computer Control, Inafips, 1265, IEEE, 1899.

[35] CAAT, Automated and Connected Vehicles. ⟨http://autocaat.org/Technologies/Automated_and_Connected_Vehicles/⟩, 2016 (accessed 17.07.16).

[36] T. Yamamoto, Automated driving activities in Japan, in: Road Vehicle Automation 2, Springer International Publishing, 2015, pp. 17–28.

[37] ERTRAC, Automated Driving Roadmap. ⟨http://www.ertrac.org/uploads/documentsearch/id38/ERTRAC_Automated-Driving-2015.pdf⟩, 2016 (accessed 17.07.16).