Chapter 2
Data Sources and Management

2.1 Introduction

For the last twenty or so years transportation system managers have received data detailing system performance from a range of infrastructure-based sensors such as inductive loops or CCTV cameras. While these sensors have enabled performance increases to be garnered from transportation networks, the data they provide is limited in scope, timeliness and availability. However, the explosive growth in mobile devices and communications network capability has created a fertile ground for new developments in transportation-oriented sensing. It is the purpose of this chapter to first take a look back at extant infrastructure-based sensors and then move on to positioning, communications, participatory sensing and finally data management issues. These, as discussed in Chap. 1, form the basis for the first tier of the DMII which underpins the new transportation paradigm outlined in this book.

2.2 Traffic Detection and Surveillance Systems

The ability to have a complete picture of the environment is an essential ingredient for making intelligent operational decisions. In this section, the most commonly deployed infrastructure-based traffic sensors are discussed before moving on to in-vehicle and mobile sensors and concluding with a discussion of some newer traffic sensing technologies.

2.2.1 Sensors for Traffic Monitoring

Traffic detection uses a variety of technologies including acoustic, video, ultrasonic, microwave radar, infrared and other modalities. The earliest instances of automated

P. (Vonu) Thakuriah and D. G. Geers, *Transportation and Information*, SpringerBriefs in Computer Science, DOI: 10.1007/978-1-4614-7129-5_2,

detection and surveillance in transportation were for the measurement of road traffic, prior to which measurements were made manually. The authors of [22] note that the monitoring of roads via Closed Circuit Television (CCTV) was first implemented in the US in Detroit, Michigan, in 1961. At around the same time the Chicago Area Expressway Surveillance Project pioneered the use of inductive loop detectors [23], and these remain the primary method of vehicle detection on Chicago's highways to this day.

It is convenient to define two classes of sensors: those that are essentially static in location, such as inductive loops, and those that are highly mobile such as probe vehicles or wirelessly connected sensors. The former are commonly associated with traditional ITS applications and will be discussed first. The latter are a fundamental enabling technology for LBS and Cooperative ITS (C-ITS). Of course, mobile sensors may be used in static scenarios and some traditionally fixed sensors (at least in the road transportation sector) such as cameras and radars may also be used as mobile sensors. Further, static sensing systems may be further classified as either in-road or over-road.

Currently the vast majority of traffic data is captured using static and point-detection sensors such as inductive loops. The adherence to point-detection sensors is largely due to the longevity of traffic control systems and infrastructure with the two predominant systems—SCATS [24] and SCOOT [25]—having been is use and development for more than 25 years. Commercial efforts to produce viable loop replacements have met with only limited success. The reliance on such signal control systems has perhaps hampered both the development and deployment of enhanced sensors and the acceptance of newer traffic control algorithms that rely on a more detailed approximation of the traffic state than can be achieved using inductive loop data alone.

2.2.1.1 In-Road Sensors

For many years inductive loops have been taken as the gold standard for traffic detection despite being disliked by road maintainers who see the required saw cuts as areas of potential sheet failure. From the installation and maintenance point of view, any in-road detection system requires traffic lanes be closed, resulting in traffic disruption and consequently higher costs. The major infrastructure-based sensors are:

Inductive Loop Detectors: The most widely deployed vehicle sensor is the inductive loop [26] which consists of a conductive wire loop (rectangular or circular geometries of greater than 1 m perimeter are typical) buried in the road. The buried coil forms part of the inductive component of a tuned electrical circuit. Vehicles passing over the loop cause the inductance to change and this is detected by associated road-side mounted electronic circuitry. Inductive loops are extremely accurate but can suffer from excessive failure (due to loop breakage) in environments where freeze-thaw cycles are common.

Loops are commonly used to count vehicles, detect presence and to infer headway and occupancy. A pair of loops (sometimes referred to as a split loop) may be used to determine vehicle speed by noting the time at which the vehicle passes over each loop. Speed resolution is a function of the data sampling rate of the detection system and falls as vehicle speed increases. Once the speed is known, vehicle length can be estimated, allowing crude vehicle classification.

Magnetometer: During the last decade or so, in-road battery powered two-axis flux-gate magnetometers [26] have become available. These devices are typically placed under the road surface in the middle of a traffic lane. Battery life under normal usage has been estimated at 10 years [27]. These sensors communicate with the roadside equipment using low power wireless, typically in the 900 MHz or 2.4 Ghz range, and so provide for simpler installation than required for inductive loops. Physically, flux-gate magnetometers sense the change (direction and magnitude) in the Earths magnetic field produced by the presence of a nearby ferrous metal object such as a car.

Like loops, magentometers can detect vehicle presence and so can be used for vehicle counting. Similarly, a pair of closely spaced magnetometers may be used for speed measurement and crude vehicle classification.

2.2.1.2 Over-Road Sensors

All over-road systems suffer from the effects of weather, lighting and occlusion to greater or lesser extents. No single system will ever function "perfectly" under all conditions and much research in the data fusion regime has been entered into. The major technologies used for over-road sensing are:

Passive Infrared: Passive infrared (PIR) detectors contain non-imaging sensing elements that respond to the self-generated and reflected component of infrared radiation from an object. All objects with finite temperatures emit radiation in the infrared part of the electromagnetic spectrum and these sensors tend to operate in the 10 μm wavelength range. Electronics in the passive infrared sensing system establish a nominal background temperature and any object that differs in temperature form the background will be detected. In the case of vehicle sensors, the nominal background is the road surface. Multiple sensing elements are used to create multiple detection zones on the road and passive infrared sensors are able to measure vehicle speed and size in the same way as an inductive loop pair. The PIR is typically gantry mounted over the lane that is being observed, and receding or approaching traffic may be monitored [28].

Active Infrared: There are two types of active infrared vehicle detection systems [26]. One uses LEDs, operating at around 850 nm (in the near-infrared) to illuminate the scene and operates in a non-imaging mode with a receiver similar to the PIR described above. The other uses an infrared laser which is scanned across the detection area. The IR receiver builds up a range image of any objects in the detection area. Active IR systems may also be used in a side-fire configuration and can cover many traffic lanes. Some active IR systems have also been certified for speed enforcement in certain jurisdictions.

Acoustic: Passive acoustic vehicle detection systems are effectively microphone arrays that monitor traffic and vehicle sounds [29]. Such sensors can not only detect vehicle presence but can also, perhaps surprisingly, be used for vehicle classification [30].

Ultrasonic: In these systems, ultrasonic sound waves (with frequencies greater than 20 kHz) are projected into the detection area and the returned signal reflections detected. Such systems may be mounted overhead or side-mounted [31, 32]. Similar to radar in many respects, ultrasonic detection systems report vehicle presence from the reflected signal and vehicle speed from the Doppler shift. Doppler shift is the small change in frequency caused when a wave (electromagnetic or sound) is reflected (or emitted) from a moving object. Doppler shift is explicitly linked with vehicle motion and stationary vehicles cannot be detected [33].

Video Systems: Many public transportation operators and roads authorities have installed video surveillance systems over the last thirty or so years. The earliest systems used predominantly commercial-grade monochrome closed circuit television (CCTV) cameras and in some, slow-scan (several seconds per frame) equipment was used (the analog images were transmitted over leased copper lines in the audio range). The cameras would typically be installed at known trouble spots and the feeds displayed on a video wall located in a Transportation Management Center (TMC) and monitored by human operators. With the correct software, video traffic monitoring systems can provide traffic information from presence through to vehicle identification, with the caveat that the data provided may have a measurable error rate. While many of the monochrome CCTV cameras have been replaced by color CCTV cameras and lately by network (or IP) cameras, they are still predominantly viewed by human operators despite significant recent advances in computer vision technology (discussed below).

Clearly, camera and camera-like sensor systems provide a much richer data stream than that available from the simple in-road point sensors described above. However, the automated generation of useful traffic information from video data, such as vehicle count, queue length or speed, is a non-trivial problem that has been actively studied by the computer vision and ITS research communities since at least the late 1970s (see [34–36] and their references). Such systems vary in complexity from inductive loop "replacement" systems to sophisticated vehicle trackers and classifiers. The latter class of systems also lend themselves to the development of automated incident detection systems [37–40].

There are many challenges to address in these systems. Deployed systems must work robustly in real-time under all weather and lighting conditions. When coupled to an actuated traffic light control system, such devices must know when they are not working so the signals can fall back to fixed-time operation in order to prevent potential crashes caused by frustrated (and delayed) drivers of vehicles that have not been detected.

The techniques used to locate and track vehicles in video frames cover the full gamut of those available to the computer vision community. Traditional motion-detection based techniques using interframe differencing [41] or background subtraction [42, 43] have either been used directly to infer vehicle presence or as preprocessing steps prior to applying more advanced methods [44, 45]. Of course, motion-based methods are indiscriminate and essentially respond to pixel intensity change (albeit with various levels of sophistication) and so cannot readily distinguish a shadow on the road from a vehicle. Indeed much effort has gone into solving this problem [46–48].

Using motion-detection to locate regions or "blobs" of interest, however, is a practical means of vastly reducing the in-frame image regions over which more sophisticated and computationally more expensive vehicle detection algorithms must be run. The more sophisticated and robust methods usually rely on having an internal vehicle model, and use pattern matching techniques to distinguish vehicles from non-vehicles and image background.

The pattern matching formalism generally relies on identifying a set of "invariant features" which are common to all vehicles, of which headlights are a good example [49]. The typical process consists of extracting many image patches containing headlights and many not containing headlights. These "positive" and "negative" examples are then input into the training phase of a supervised learning system of which there are numerous examples [50]. The more fashionable currently are Support Vector Machines [51] and the numerous variations on the Boosting algorithm [52]. The output of the learning system is a function such that when an image patch is input, the output will be "true" if it is a headlight and "false" otherwise. Of course, this is a gross oversimplification as no classifier is perfect. There will be "false negatives" and "false positives" and there is no guarantee that the system will "generalize" to images captured using different cameras or under different illumination conditions. Automatically extracting the image patches to input into the classifier is also non-trivial if not performed naively using exhaustive search over the image. Of course, similar techniques may be used to locate pedestrians and other objects of interest in video feeds.

Continuing on with the headlight example, the vehicle model consists of the vehicle headlights and their spacing in world coordinates, which despite its simplicity allows for some distinction between vehicle classes. The term "world coordinates" requires explanation. Video images taken by a single camera are a projection (usually through some optical system comprising lenses and aperture stops) from the three-dimensional world observed by the camera onto the two dimensional image plane (CCD sensor, film, etc.). If the camera parameters (roughly height, pitch, roll and focal length) are known, then the parameters of the projective transform from world coordinates to image coordinates can be computed [53]. The inverse transform will map image coordinates to world coordinates and thus pixel separation in the image can be converted into real world distances. In the single camera case there is some ambiguity in the transform and care must be taken to ensure the correct mapping is used. The use of two cameras in a stereo pair removes the ambiguity in the transform but requires the additional complexity of matching the slightly disparate images from two cameras before being able to compute the parameters of the required projective transform. There is at least one commercial product that uses a stereo camera for road side monitoring of pedestrian crossings [54].

Computer vision also forms the basis for Automatic Number Plate Recognition (ANPR) systems. If used between two points, such systems can be used to measure travel time and conduct trip origin-destination (O-D) studies [55].

Radar: There are two types of radar systems that have been widely deployed for measuring traffic parameters from the curbside: Doppler and Frequency Modulated Continuous Wave (FMCW). Both provide all basic traffic data and they are discussed in turn.

In Doppler radar, super-high frequency radio signals (at frequencies near 10.5 and 24.0 GHz as permitted for traffic detection by the Federal Communications Commission (FCC)) are transmitted into the traffic stream and the reflected signals detected and processed to determine their Doppler shift [56] and hence vehicle speed. Stationary vehicles cannot be detected.

In FMCW radar, the extremely high radio frequency signals are modulated by a triangular waveform so that the frequency of the transmitted signal varies slightly with time (this is the definition of frequency modulation). The signal reflected from the target is mixed with the transmitted signal to produce a beat of much lower frequency. By processing the beat signal it is possible to extract the range and speed of the vehicles illuminated by the radar [33].

Laser-Pulsed: Pulsed laser systems are typically referred to as LIDAR (an acronym for Light Detection and Ranging) systems and they provide the full gamut of basic traffic data. In these systems, near-infrared laser pulses, typically at the kilohertz rate, are used to illuminate vehicles and the reflections detected. Time of flight of the pulses is used to determine range and the time between successive pulse returns is used to determine vehicle speed. With these systems it is possible to build up a profile of the vehicles passing through the detection zone.

Other Technologies: Various other technologies have been researched [57, 58] to measure the same physical properties of vehicles (electric/magnetic properties, reflectance, sound, etc) as the sensors described above. Weigh-in-Motion systems deserve special mention because unlike the sensors described above they are primarily used for road access control and pricing rather than real-time traffic management. Most systems are based on conventional load cell technology but recently optical fiber-based systems have been proposed [59, 60]. An interesting possibility would be the dual use of a fiber sensor as an inductive loop replacement (magnetic field sensor) and simultaneously for weigh-in-motion.

2.2.2 In-Vehicle Sensors

The number of sensors in a car has also rapidly increased, with sensors for the powertrain (which monitors vehicle energy use, drivability and performance, and involves the engine, transmission and onboard diagnostics), chassis (the main control functions of which are monitoring vehicle handling and safety, and involves the steering system, suspension, vehicle braking and stability) and, finally, for the body (which monitors occupant needs including occupant safety, security, comfort and information) [61].

One direction in automotive systems is in the area of Intelligent Vehicle Technology, a concept typically associated with the development of autonomous vehicle functionality for Unmanned Ground Vehicles (UGV). The key attributes of intelligent vehicles include the following: (1) the ability to sense the vehicles own status as well as its environment; (2) the ability to communicate with the environment; and (3) the ability to plan and execute the most appropriate manoeuvers [62]. Applications of intelligent vehicle technologies to the automotive sector are often seen as the next generation of vehicle safety systems. Specifically, for applications within the automotive industry, "Intelligent Vehicle" systems are defined as systems that sense the driving environment and provide information or vehicle control to assist the driver in optimum vehicle operation [63]. Overall, different data about the driving environment can be obtained through any combination of sources such as on-board video cameras, radars, lidars, digital maps navigated by GPS, communication from other vehicles or the highway infrastructure.

Vehicle-mounted radars, typically operating at 77 GHz, are used in adaptive cruise control systems, overtaking assistance systems and rear-end collision warning systems (which may include brake assistance). Many vehicles are now available with parking sensors that detect nearby objects when the vehicle is travelling at low speed. Some vehicle manufacturers extend these systems to semi-automatic (the driver controlling the speed of the vehicle) reverse parallel parking and recently to 90° reverse parking [64].

2.2.3 Some New Sensing Modalities

Additional sensing systems are now available which are either being used currently or hold potential for future use.

Toll Tags: Many vehicles are now equipped with active Radio Frequency Identification (RFID) tags (known as Electronic Toll Collection or ETC tags in the industry), which are primarily used for free-flow tolling applications. Travel time and O-D information between receiver gantries may be easily obtained by matching tag IDs [65]. However, many jurisdictions will only allow partial tag IDs to be released, which introduces a source of noise in the data and reduces the overall accuracy. The use of ETC information opens up a whole range of privacy issues, many of which are discussed in Sect. 2.6 below.

Bluetooth: Bluetooth is a short range (less than 100 m) wireless technology that is most commonly used to connect fixed and mobile devices together. A typical use is to allow hands-free operation of a mobile phone while driving using factory installed or third party devices. Each Bluetooth radio broadcasts a unique identifier known as the media access control (MAC) address.

By using road-side mounted equipment it is possible to detect the MAC addresses of Bluetooth radios in passing vehicles. Matching the received MAC addresses between spatially separated road-side receivers enables travel time and O-D surveys in an identical manner to using ETC tags [66, 67]. It is to be noted that Bluetooth MACs contain no information that can be used to identify the owner or user of the device, and so privacy is less of a concern. Such systems are now available commercially.

Mobile Phones and Mobile Devices: The rapid growth in the adoption of personal connected devices such as mobile phones, Personal Digital Assistants (PDAs) with wireless connection, handheld devices, tablet PCs (such as the iPad) and smartphones has brought about a sea-change in ubiquitous information generation, sharing and delivery. Mobile phones have sensed traffic since the 1990s, with drivers calling by phone to report traffic incidents to emergency management services. In the last decade, the convergence of several technologies has enabled connected mobile devices to become critical sensors for mobility services, particularly on an ongoing, organized basis.

Many current smartphones have sensors for position, sound, video, acceleration, ambient lighting orientation and proximity. Such devices have large memories and significant processing power. Additional functionality is made possible through third party "apps" which are able to access many of the internal systems through well-defined Application Programming Interface (APIs) [68, 69] which is discussed further in Sect. 2.3.3.

Underlying technologies that have positioned connected mobile devices to become an important sensing approach include being location-enabled (either through built-in GPS or through cell tower multilateration), WiFi connectivity, and integrated functionalities such as the ability to take high-resolution photographs of incidents and special events that have the ability to disrupt traffic. Two important benefits of connected mobile devices are their high penetration rate which enables sensing at significant granularity and density, so as to become "pervasive" sensors of the transportation environment on an as-needed basis; and their ability to be connected, by special instrumentation, to secondary sensors, for example, for air pollution or noise-level monitoring [70], and integration with wirelessly-connected wearable body sensors for health monitoring. Some of these applications are discussed in Chap. 3.

In essence, mobile network-connected devices can act as real-time probes in the transportation network and for location-monitoring for social networks and many other applications. For the developers of in-car navigators and on-line mapping systems, this new source of data has enabled a range of applications from traffic congestion warnings to travel-time estimation.

Biometric Sensors: Biometric sensors are body-wearable sensors that collect, store, and share relevant data. Developments in wearable sensors have led to diagnostic as well as monitoring applications, which allow physiological and biochemical sensing, as well as motion sensing [71].

A new generation of wearable biometric sensors measure Electrodermal Activity (also known as skin conductance or galvanic skin response) for monitoring arousal associated with emotion, cognition, and attention. Another type of biometric sensor is mobile Photoplethysmography (PPG) for screening of cardio-vascular pathologies that may find increasing safety, security and travel quality applications in transportation. Applications are in the areas of mobile health and remote health telemonitoring, for example, while driving.

People as Sensors: One of the most active research areas covered by this book is on the use of humans as sensors. Information generated by users is generally called User-Generated Content (UGC) and can primarily occur *proactively* when users generate primary data on events, concepts or activities of interest; or *retroactively*, by analysts who process secondary user-submitted data that is published using social media tools such as Web 2.0 tools, blogs, microblogs and so on. Several different sub-modes can be differentiated within these modes: for example, proactive UGC can result from idea or design competitions, participatory sensing (where users voluntarily report events witnessed or share mobility experiences) or by opportunistic sensing (where users agree to be "tracked" and allow movement, trajectory or speed data to be transmitted to peers or to a central server). We review these modes of people-centric sensors in Sect. 3.3.

2.3 Transportation-Oriented Communications Systems

During the period of explosive growth of in-vehicle sensor technology a similar exponential growth in wireless communications technology has also taken place. The ability to share raw sensor data or derived information with other vehicles and travelers in (or near) real time enables a range of previously unforeseen transportation applications.

It is beyond the scope of this work to go into the details of mobile data communication standards and techniques [72]; however, it is worth noting that travelers can be permanently connected to an Internet Protocol (IP) network with a bandwidth of several megabits per second for less than $10 per month; and this cost will only fall as time goes on. The development of these systems is led by the large telecommunication companies and follows consumer demand.

2.3.1 Communications Access for Land Mobiles

A systems architecture based on IP Version 6 [73]—Communications Access for Land Mobiles (CALM) [74]—is being developed for continuous communications between vehicles (Vehicle-to-Vehicle or V2V) and between vehicles and infrastructure (V2I) which spans a range of data rates and latencies by seamlessly switching wireless mode (physical layer) based on need. For example, messages pertaining to safety should have priority over messages reporting headlight status. Of course, pedestrians can be easily accommodated through connections to nomadic devices such as cell phones. A highly simplified functional diagram of the CALM architecture is shown in Fig. 2.1.

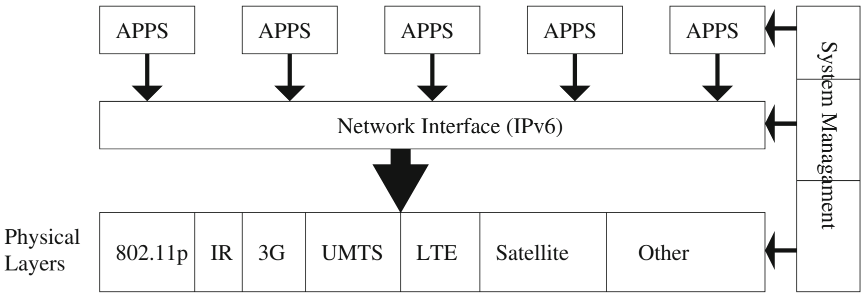

Fig. 2.1 Simplified CALM architecture

As can be seen from Fig. 2.1 CALM can easily accommodate wireless mobile data services such as the General Packet Radio Service (GPRS), Universal Mobile Telecommunications Systems (UMTS) and Long Term Evolution (LTE) over GSM; and the various data transmission additions to CDMA. Satellite communications may also be readily included, as can any IPv6 capable bearer.

The now standardised low latency component of CALM is designated CALM M5 (ISO 21215) which extends IEEE 802.11p (known as WAVE—Wireless Access in Vehicular Environments—in the US) and is colloquially referred to as Dedicated Short Range Communications (DSRC). WAVE is part of the IEEE 802.11 family of wireless standards [75] of which the familiar consumer WiFi (802.11a/b/g/n) is a member. Operating at 5.9 GHz, DSRC offers a communications range of up to 1 km. A complying DSRC implementation based on CALM M5 or 802.11p broadcasts vehicle position and other parameters at 10 Hz in the so-called Cooperative Awareness Message (CAM—colloquially referred to as DSRC Beacon Messages). It will not be long before mobile phones are equipped with DSRC radios and so link pedestrians into the mix.

It is envisaged that vehicles and other travelers will form part of a wireless Mobile Ad Hoc Network (MANET) based on DSRC. The adjective ad hoc is included to indicate that the clearly decentralized network does not rely on any existing infrastructure and members (nodes) may come and go over time. Further, each node forwards messages it receives to their ultimate destination based on a routing algorithm. Mobile routing algorithms are the subject of much research.

When specialised to vehicles, MANETs are given the moniker VANET [76] for Vehicular Ad Hoc Network. Interest in VANETs has been largely motivated by safety and traffic efficiency considerations and the possibility of opening up a whole new area of innovative applications. Research in VANETs has focused on the inherently unfavorable characteristics of the wireless communications environment in which VANETs have to perform, the fair and efficient use of available bandwidth in a totally decentralized and self-organized network, the high mobility and scalability requirements, the wide variety of environmental conditions in which these networks have to perform, and the potential security and privacy considerations of receiver and sender. Research into these areas has significantly increased in the last two decades [77–79].

Several prominent, primarily government initiatives have provided opportunities to researchers to study the technical and socio-economic aspects of VANETs, such as the Connected Vehicles (formerly IntelliDrive^{SM}) program in the US, and four integrated projects within the 6th Programme Framework of the EU in areas that touch the field of VANET: COOPERS [80], CVIS [81], PReVENT [82], and SAFESPOT [83].

2.3.2 Machine-to-Machine (M2M)

With the ever increasing number of network-centric sensors and systems with large computing power, the requirement for human intervention in the sense-communicate-control chain falls, ultimately to zero and the need for direct Machine-to-Machine (M2M) interaction becomes self-evident. According to [84], M2M represents a future where "billions to trillions of everyday objects and the surrounding environment are connected and managed through a range of devices, communication networks, and cloud-based servers".

The initial concept of the "Internet of Things" (IoT) attributed to Ashton (and promulgated by The Auto-ID Labs [85]) which was based on RFID tags barely scratched the surface of what is now possible with the advent of 802.11 and the low power IEEE 802.15.4 ("Zigbee") [86] wireless protocols.

Mobile devices with cameras capable of reading QR codes (somewhat coincidently invented by Denso Wave, a subsidiary of Toyota, for tracking vehicles during manufacture [87]) also provide novel ways of connecting unintelligent objects into the mix. Of course standardization has played, and continues to play, an enormous role in the adoption and deployment of such technologies. In the context of this work it is clear that v2x systems; and traffic and transportation sensor systems fit neatly into the M2M (or extended IoT) paradigm.

Connecting large numbers of sensors together enables much greater knowledge of the system state but brings with it significant challenges especially in device management and control. The limited address space provided by IPv4 is clearly a problem of the past but must be handled when dealing with legacy systems. Sensor nodes which "sleep" for significant periods do not fit into the conventional IP network modalities. Nor does the small packet size prescribed for Zigbee. However, neither of these challenges is insurmountable [88, 89].

The proprietary Urban Operating System (UOSTM) [90] and other middleware systems such as Ice from ZeroC [91] provide M2M integration of services by means of a programming platform and operating infrastructure for the development of sensor networks of intelligent devices embedded in urban systems and distributed sensing applications that run on them.

The Web of Things (WoT) redefines the IoT to use the standard web protocol HTTP and the RESTful architecture [92] for communicating between devices. Not only does this make it easier for programmers to connect to a wide range of devices through well-known means, it also makes it easier for human interaction with potentially unfamiliar devices through familiar technologies such as the Web Browser.

Even then human interaction with the WoT can be challenging. How should the data be presented? What visualization techniques are helpful when the dimensionality runs into the billions? Where, when and how should humans intervene?

2.3.3 Application Programming Interface

The data collected by sensors of all modalities must be made available to programmers and users if it is to be useful in developing new and novel systems. Programmatic interfaces have tended to be formalized through an Application Programming Interface (API). Formally, an API is a source code level interface that enables programmers to connect software components together. APIs may be programming language and operating system dependent or independent even though they usually provide an interface to an object-code library targeted for a specific architecture. In the network-centric world of today the concept of Service Oriented Architectures (SOAs) has arisen. An SOA is a programming paradigm in which a network-based service (usually a web—either html or xml—service) provides specific functionality. Software as a Service (SaaS) is an extreme example of a SOA but the input or output of a (web) service do not have to be human readable. The power of SOA lies in the ability to connect multiple services together in order to obtain functionality beyond that which the component services provide separately. Clearly the component services of a SOA need a well-defined and documented API in order to be useable.

In fact, designing usable APIs is critical for any data organization exposing a programmatic user interface. One of the most important reasons is that highly usable APIs can drive adoption and sustained use of a particular technology and initial encounter with a poorly designed API can discourage users from ever adopting a particular technology. However, many data sources which are useful for mobility services are not yet associated with APIs. Sometimes such data is available through Rich Site Summary [93] (RSS) or its (almost) successor, Atom Syndication [94, 95], feeds.

A number of transportation agencies have released public APIs to access web feeds (see [96] and [97] for example). Judicious use of the reader's favorite search engine will reveal many more.

2.3.4 Positioning Systems

An essential ingredient for many of the above applications is the need for location-awareness through positioning systems. It was noted in 1993 that the first automotive in-vehicle navigation systems introduced in 1984 utilized a combination of dead reckoning and map matching to track a vehicle's movement over the road network [98]. Dead reckoning was accomplished with differential wheel sensors and a magnetic flux-gate compass. Map matching required highly accurate digital road maps and sophisticated software to correlate vehicle motion with the road network and to deal with the many subtle complexities involved. New sensors and technologies including inclinometers, gyroscopes, satellite-based navigation, inverse Loran, roadway electronic benchmarks, electronic odometers and ABS wheel sensors have been considered to supplement or replace the special wheel sensors and magnetic compass, and to simplify or eliminate map matching.

Modern outdoor positioning systems are either satellite or mobile-phone based. The satellite based systems (such as the currently available US Global Positioning System (GPS) and Russian GLONASS systems, and the European Galileo or Chinese BeiDou-2 systems that are under development) are generally more accurate and available anywhere on the surface of the Earth. These satellite systems require line-of-sight visibility from the satellite to the receiver, which may be unattainable in urban areas (the so called "urban canyon" effect) or in dense forests. Under ideal conditions, satellite systems exhibit position errors of the order of ten meters or less.

Satellite-based Positioning: The US GPS may be treated as the archetypal satellite based positioning system. The other systems named above use the same fundamental positioning technology based on intersecting range-spheres from several satellites. In general, three satellite signals are required for an accurate position fix. In its raw form, GPS has an accuracy of about 10 m horizontally and 20 m vertically, which is sufficient for navigation but not for vehicle collision avoidance.

GPS accuracy may be augmented by broadcasting corrections to GPS positions using terrestrial or satellite based transmitters [99]. The terrestrial system is termed differential GPS (DGPS) and the satellite based system, which is available throughout the continental US and Alaska, is known as the Wide Area Augmentation System (WAAS) [100]. Both systems rely on computing GPS corrections from accurately surveyed reference stations and differ only in the means of disseminating the corrections. WAAS typically enhances GPS positional accuracy to 1.0 m horizontally and 1.5 m vertically over the reception area. Terrestrial DGPS has notionally zero error at the reference site with the error growing linearly with distance from the reference.

A novel technique for GPS correction, called the Quasi-Zenith Satellite System (QZSS) [101], is being developed in Japan and will ultimately use a three satellite constellation to provide sub-meter class positioning. As well as transmitting DGPS data, QZSS will also transmit supplemental GPS signals. The satellite orbits are designed so that at least one satellite will be visible at an elevation of 70° or more above the horizon over the Japanese islands, thereby reducing the effect of urban canyons on positioning accuracy.

Mobile Phone based Positioning: Mobile phone based positioning systems rely on signal multilateration between several mobile phone towers and the handset [102, 103]. Typical accuracy is of the order of fifty meters. Positioning at such poor resolution is unsuitable for navigation but is suitable for other less stringent location-based service requirements. It should also be noted that the majority of smart phones are equipped with in-built GPS receivers and use a combination of multilateration and GPS to provide accurate positions. Additionally, most GPS equipped handsets utilize Assisted GPS (A-GPS) [104] in which the mobile phone service provider transmits various GPS related data to the handset thereby speeding up the time the user takes to obtain a GPS position.

Other Issues: For some cooperative safety applications, however, such as side-impact collision warnings, the accuracy achieved by the above technologies is not sufficient, and research into real-time kinematic (RTK) positioning systems is being conducted for improved relative positioning at centimeter-level accuracy [105]. Similar accuracy may be possible by broadcasting differential corrections from DSRC road-side units or on sub-carriers of commercial radio or television signals. Positioning is also possible through technologies such as WiFi and Bluetooth, although they have been used to a lesser degree in transportation.

Indoor Positioning: One area of recent research interest is the area of indoor positioning, a problem that is motivated by GPS signal attenuation inside buildings. These types of applications are relevant to transportation particularly for mobility planning through mobile devices inside rail and transit terminals; and airports, for pedestrians looking to walk through public buildings to shorten their travel time, and for E911 emergency planning. Augmented Reality (AR) technologies have been used to supplement location-awareness in indoor positioning. AR is a powerful user interface technology that augments the users environment with computer generated entities and is defined by three important aspects [106]. It blends the real and virtual within a real environment, is real-time interactive and registered in 3D. In contrast to virtual reality, which completely replaces the real world, augmented reality displays virtual objects and information registered to real world locations. These technologies have significant potential for persons with visual impairments and other types of disabilities.

2.4 Methods to Add Intelligence to Sensor Data

Virtually all sensing technologies require processes to extract intelligence from the raw data, which may not be useful for transportation applications by themselves.

Information Extraction Methods: Methods that translate raw output from sensors are as numerous as the sensors themselves, and many of the mature technologies have been extensively studied in other disciplines. Nevertheless, what works and what does not work with many of these mature detection technologies have undergone significant study and testing by the transportation community, given considerations of purchase, maintenance and repair costs, the types of traffic parameters that can be monitored (for example, some types of sensors such as inductive loop detectors can measure traffic volumes, speed with the use of two loops, vehicle type, occupancy, and headway between vehicles, but not the density of vehicles over a stretch of highway, whereas others, like video cameras, can measure all these parameters), the need to be obtrusive in some cases and unobtrusive in others and operational performance under different weather conditions (for example, see Weil et al. [22], for an evaluation of sensors for the purposes of automatically detecting traffic incidents).

Information extraction methods can result in a range of intermediate measures, such as simple measures of counts (of persons, cars, pedestrians during some time interval) to more complex frequencies (number of occurrences of specific words, number of interactions among persons during a day, or trips to a certain destination between a certain period of time), continuous measures such as speeds (including time mean speed or averages of speeds of cars or other moving objects crossing a point, or space-mean speed, which the average of the velocities by which moving objects crossed a segment of a road), trajectory, or weight (for weigh-in-motion technologies primarily used for trucks), as well as classified measures into nominal categories (automatic detection of a car versus a truck; or identification of a GPS sensor in a slow-moving bus versus a pedestrian) or ordinal categories (such as different intensities of traffic congestion such as heavy congestion or stopped traffic or precipitation levels such as thunderstorms versus light rain).

An example of an information extraction processes that has received considerable attention in the transportation literature is Video Image Processing (VIP) which converts raw video streams into traffic data. Automatic recognition of speech has been used for many years in phone-based traveler information systems, for controlling in-vehicle functions, for sending text messages or name dialing through mobile phones, and by operators to respond to calls about public emergencies.

More recently, web-based and social media information sources have led to the generation of massive amounts of structured and unstructured (text, video, still images) "Big Data" that are currently active areas of research. Transportation, as does any other sector, generates a large amount of material in unstructured free-form text, such as plans, regulations, policies, technical documentation, and media and user-generated web content. Content analysis has been used on text material for qualitative assessment text-based content, but the use of text mining, using multi-disciplinary aspects of information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning, and data mining, has been used to a lesser degree, although there are some examples of using text mining to develop ontologies of urban planning documents [107] and of microblogs for real-time event monitoring [108].

One development stimulated by the generation of mobility data such as GPS trajectories and advances in data mining technologies is mobility mining whereby digital trajectories are processed to understand people's mobility and activity patterns, for purposes such as mobile commerce, location and context-based search and advertising, early warning systems, traffic planning and management and route prediction.

Analysis: In most cases, raw information from sensors has little meaningful end-use value and needs to be processed in various ways for different purposes. Mobility analytics uses sensor data using a range of methods to utilize data generated by the underlying sensor and communication technologies. Virtually every aspect covered in this book is associated with information processing and analysis which draw from methods in transportation engineering and planning, operations research, management, computer science, geographic information science, and several social sciences and related disciplines, to process the raw data and detect and understand patterns in mobility data and to build services, plans and management strategies, using the intelligence derived.

Information Fusion: Information fusion, of which the perhaps more familiar Sensor or Data Fusion are subsets, is a well-researched topic concerning the exploitation of data from multiple sources including sensors, databases, user generated content, etc. in order to obtain a better view of the world than would be possible from any single source alone. In a work such as this, it is barely possible to scratch the surface. It should be noted data fusion is termed data integration in the GIS domain.

In essence sensor or data fusion is a principled way of combining data from multiple sensors to yield better results than any single sensor could produce on its own [109]. A simple example is afforded by considering an in-road inductive loop and a calibrated road-side video based vehicle tracking system whose field of view includes the loop. Both sensors are able to provide estimates of vehicle speed within some error bounds. Combining the speeds using classical Bayesian data fusion techniques will improve both the accuracy of the measured speed and its variance [110, 111]. In general, a sensor model will include an error distribution which is essentially a formalization of how good the sensor is. When formulated as a probability density function, this becomes the prior probability in the Bayesian framework. More advanced fusion techniques such as Kalman or particle Filters can also be used to fuse multiple sensor inputs [112, 113].

Event Stream Processing and Complex Event Processing: Event stream processing and complex event processing [114, 115] are two relatively new research areas that are potentially very useful in the transportation regime. An Event Stream is defined to be a linearly (usually by time) ordered sequence of events, say bus arrival times at a particular stop. It would be possible, for example, to predict the arrival time of buses at the stop from the past arrival times or to raise a red flag when two buses arrived less than five minutes apart. An Event Cloud is a partially ordered set of events (which may be unbounded), where the partial orderings are imposed by the causal, timing and other relations between the events. In other words, an Event Cloud can consist of many Event Streams. A simple example of an Event Cloud would be the train arrival times at a transportation interchange and the bus departure times form the same interchange. A more complex example would be the Event Cloud consisting of all available data from the transportation network of a city.

Complex Event Processing (CEP) sits inside the discipline of data fusion but includes methods such as rule-based inference and adaptive neural networks which are usually seen as part of Artificial Intelligence (AI) or machine learning. The essential aim of CEP is to look for patterns that correspond to events that are significant. An example would be the inference of a burst water main, with high probability, if there is a continuously “on” loop detector present at an intersection *and* a bus fails to arrive at a stop nearby *and* there is a drop in water pressure in a main pipe. Determining that a collection of simple events forms a reportable super event is a significant research challenge. Information fused from multiple sources may form one or more of the event streams in a CEP system. Ligozat, Vetulani and Osinski [116] propose a methodology for dealing with CEP in a spatio-temporal setting. Despite being concerned with monitoring for security purposes the same processes may be extended to the transportation domain with appropriate modifications.

2.5 Initiatives and Programs Through Technology Integration

There are several initiatives and programs through which the use of information technology in transportation can be organized. As discussed in Chap. 1, there is no stark line that divides the technology initiatives discussed in this section into mutually exclusive categories. They all utilize information technology, sensor data and communications technology but differ in the way they are organized and managed (public versus private) and the extent to which they emphasize aspects of government-owned and managed infrastructure versus privately initiated applications utilizing advancements in technology

Cross-cutting these technology initiatives are different theoretical and research initiatives that examine various aspects in the transformation of transportation systems, human behaviors and society, as information itself, or information-based solutions become available.

2.5.1 What Happens to All that Data?

The sensors and sensor systems described above will clearly generate huge amounts of data that have both spatial and temporal content as emphasized in the concept of Tier 1 of the DMII. When used in transportation, such databases must function on two different time-scales that embody the movement of travelers throughout a geographic region using the various transportation modes and the slow changes to the transportation network itself. These two time-scales would correspond, for example, to the GPS position of a traveler and the map updates provided by a GPS vendor respectively.

Regular databases cannot handle such data efficiently and a whole class of specialised spatio-temporal databases that utilise special-purpose spatial and temporal data models have arisen [117, 118]. Most commercial [119–121] and many Open-Source databases [122, 123] support spatiotemporal data either natively or through add-ons. Such systems are standardized through the Open Geospatial Consortium [124]. These systems all support spatial queries and predicates (Is there a restaurant within 1mile of (x, y)?), distance functions, etc.

From the design point of view, spatio-temporal databases generally consist of a spatial database combined with a temporal model that allows temporal integrity of the data to be maintained and enables temporal queries. The most commonly used spatial data structures are based on the R-Tree [125], which represents spatial data as a tree-structured hierarchy of minimum bounding rectangles (the idea generalizes to higher dimensions by using minimum bounding boxes). There are some temporal features in the latest Structured Query Language (SQL) standard (ISO/IEC 9075:2011) but they are far from being routinely available.

It should be noted that most transportation data for transportation operations, traffic measurements and vehicular applications, are useful for only a finite period and that internal spatiotemporal database operations would expunge (or at least move to a backup device) expired data for reasons of storage and search efficiency. The data that has expired for real-time applications, could of course be useful for looking for transportation trends, patterns and associations, and as inputs to the planning process, together with information from Tiers II and III of the DMII, for example. These possibilities are examined in Chap. 4 (Sect. 4.4.2).

2.6 Privacy, Trust and Security

Privacy, trust and security are concepts that are essential to human societal interactions. In the case of mobility services, the private information possessed by a traveler potentially comprises their identity, current location, origin and destination of travel, journey time, locational preferences and so on. LBS for transportation usually require a traveler to surrender their (potentially extremely) accurate spatial and temporal location to a transportation service provider in order to obtain a useful service. However, the potential for misuse of this data exists. In the now classic paper, *Geoslavery* [126], Dalton and Fisher discuss numerous situations where knowledge of position is real power and it is imperative for designers and providers of LBS, with the help of government legislation, to curtail such misuse. Such data should only be sent to a trusted organization via a secure communications channel where it will be stored securely.

Privacy is a fundamental human right. However, privacy is also not a static, immutable constant. People are likely to trade off some privacy protection in return for utility gained, security benefits or risks minimized. Such trade-off notions have been explored in various contexts (for example, [127] in the case of data publishing and [128] in the case of rail travel security). Threats to locational privacy include the risks of unauthorized access to raw location data, location information about a person that is secondarily derived or computed using the raw data, hijacking the location transmission channel, and identification of the person who is using generating location information while using a web service or mobile device.

There are three fundamental approaches to addressing locational privacy: legal, consumer awareness and technology-based. Legal and policy-based strategies and privacy principles (ability to opt out, user consent, data protection) as well as issues of consumer education and awareness are described in detail in Sect. 4.2.2 Technology-based approaches are numerous; one particular methodology that has gained considerable support to address location privacy threats is the idea of privacy-by-design to ensure that collected data is only accessed for the purposes which the user agreed. The idea is simple: privacy concepts must be built into physical and software systems, and business processes from the ground up using Privacy Enhancing Technologies

(PETs). Application of this simple idea would help to reduce the number of privacy breaches that are occurring more-and-more frequently.

While a number of PETs for locational privacy have been proposed ([129, 130] are examples from a voluminous literature), commonly used approaches utilize pseudonyms for anonymity (anonymizing proxy for all communication between users and applications) and location protection by degrading the accuracy of the user's location using geographical and temporal masking or cloaking, and encryption. These approaches, while guaranteeing to a certain degree that precise location information transmitted by a user cannot be easily used to re-identify the subject, also have numerous issues; for example, de-identification in a system using pseudonyms is possible by correlating where any given pseudonym spends most of its time and who spends more time than anyone else at any given location [129], which has implications for frequency of pseudonym update. PETs for mobility application are likely to integrate such privacy-enhancing techniques. For example, in the mix-zone approach, geographical regions are considered where no provider can trace user movements. Users, upon entering a mix-zone, receive a pseudonym that changes when they exit the zone. Users also exit the mix-zone in an order different from their order of arrival. In this way, the identities of users entering the mix-zone at the same time are mixed in a way such that the mapping between their old and new pseudonyms is not revealed.

Other approaches utilize communication aspects to address locational privacy. For example, the Mist routing protocol [131] for mobile users address the problem of routing messages to a user's location while keeping the location private from the routers and the sender by utilizing a set of "mist routers" organized in a hierarchical structure that effectively creates a "mist" to conceal from the system and other users by using pseudonym-based routing. The Onion Routing protocol [132] encrypts messages with multiple keys to form an "onion" around the message. The result is that Onion routers remove a layer of encryption to uncover routing instructions to send to the next router where this is repeated, thereby sending the message through several network nodes. This approach can deter the discovery of both the source and the destination information of the packet.

It is also possible to address locational privacy directly in position sensing systems. For example, the Cricket system [133] is an RFID and ultrasound-based indoor positioning system where location sensors are placed in the user's mobile device so that users listen for their position thereby avoiding disclosure of their location during the location-determination process and without having the need for a centralized location-determining system.

Vehicular networks are designed to be open to all participants. In many cases the users of such systems will be unaware (and may not care) of what data are being transmitted and to where. The 802.11p standard specifies that broadcast messages must not be traceable to a specific on-board unit, that in-vehicle units change IP address when they move to another road-side unit, and that MAC addresses be randomly generated from a local address space. Messages are also encrypted using a public key infrastructure with the certificates being distributed by a trusted authority (more than likely a Department of Transport (DoT) or some other government agency).

The purpose of all these measures is to create positional anonymity and prevent tracking. But is this really achievable? A DoT recording DSRC CAM transmissions would be able to track a vehicle based on some simple traffic flow principles even through MAC and IP address changes, and an anonymised vehicle whose origin and destination are repeatedly (perhaps daily) the same geographic coordinates is really not anonymous.

Being wireless and IP based, VANETS are subject to all the usual network security threats: address spoofing, man-in-the-middle attacks, and so on. In the case of safety related issues, signal jamming (even accidental due to arc-welding equipment for example) must be seriously considered if drivers begin to rely heavily on the system for their personal safety. Sending fake messages from a vehicle could create traffic chaos. Authentication cannot prevent a stolen vehicle being used to generate fake messages. Certainly, information fusion and complex event processing will help to rule out some fake messages and to add weight to genuine messages, but these techniques will not work in all cases. Recent work on privacy in vehicular networks are given in [134, 135].

Ultimately, such systems are built on trust. If the user trusts the organization that is receiving her personal data and is not using it to violate privacy and other concerns, then she will continue to use the system. Issues of trust management are explored further in Sect. 4.2.5.

http://www.springer.com/978-1-4614-7128-8

Transportation and Information
Trends in Technology and Policy
Thakuriah, P.V.; Geers, D.G.
2013, X, 127 p. 3 illus., Softcover
ISBN: 978-1-4614-7128-8