Abstract: The current and expected future proliferation of mobile and embedded technology provides unique opportunities for crowdsourcing platforms to gather more user data for making data-driven decisions at the system level. Intelligent Transportation Systems (ITS) and Vehicular Social Networks (VSN) can be leveraged by mobile, spatial, and passive sensing crowdsourcing techniques due to improved connectivity, higher throughput, smart vehicles containing many embedded systems and sensors, and novel distributed processing techniques. These crowdsourcing systems have the capability of profoundly transforming transportation systems for the better by providing more data regarding (but not limited to) infrastructure health, navigation pathways, and congestion management. In this paper, we review and discuss the architecture and types of ITS crowdsourcing. Then, we delve into the techniques and technologies that serve as the foundation for these systems to function while providing some simulation results to show benefits from the implementation of these techniques and technologies on specific crowdsourcing-based ITS systems. Afterward, we provide an overview of cutting edge work associated with ITS crowdsourcing challenges. Finally, we propose various use-cases and applications for ITS crowdsourcing, and suggest some open research directions.

Keywords: automatic sensing; mobile crowdsourcing; ITS; smart cities; smart vehicles; spatial crowdsourcing

1. Introduction

Crowdsourcing is a platform that manages dissemination of tasks by a recruiter—the entity operating the crowdsourcing platform for compensated workers to perform. With the rise of the Internet and personal mobile devices (PMDs), crowdsourcing can help solve complex tasks that machines cannot, at scale [1]. The scaling ability depends on the emergence of the Internet of things (IoT) and distributed data storage architectures. The increase in mobile devices, higher data bandwidth, and ultra reliable and low-latency communications in current and next generation cellular networks provide more opportunities to develop powerful mobile crowdsourcing frameworks. On the other hand, researchers have recently designed data gathering and dissemination infrastructure for intelligent transportation systems (ITS), such as roadside units (RSU), vehicular social networks (VSN) [2], and even unmanned aerial vehicles (UAV) [3] that strive to apply current and future widespread broadband connectivity to the Internet for data-driven improvement of transportation systems.

ITS-based crowdsourcing that takes advantage of the aforementioned infrastructure and its dense connectivity has the potential to provide a massive influx of new data for ITS operators to work with. The data influx will aid in a variety of verticals such as infrastructure design and improvement decisions, real-time navigation, safety system operations (e.g., traffic and collision monitoring), and will improve the interconnectivity of vehicles. Consider Waze—a traffic crowdsourcing system that allows users to provide information about traffic conditions, providing other users the ability to make more...
informed data-driven decisions regarding how to navigate along their trajectory safely and efficiently. While Waze is effective, it is limited by human data inputs, which may be inaccurate, suffer from a very limited useful time-frame, and require passengers to actually input the data to avoid unsafe distracted driving. By leveraging ITS and sensors on smart vehicles, data collection can be automated, increasing collection efficiency, volume, and data quality. Vehicular social networks can improve connectivity amongst ITS infrastructure, providing new channels for intelligent data dissemination [2,4]. The increased connectivity provides more conduits for data to be collected, improving performance of crowdsourcing systems. The improvement in ITS communications technology has lead to increased potential for the amount of useful information to be disseminated. This evolution in technological progress, visualized in Figure 1, started with an era of no communication between vehicles, to radio-based communications devices such as cellular phones, to integrating sensors within vehicles, then adding the potential for connecting the vehicle sensors via broadband connection to infrastructure, and formation of ad-hoc networks, with the final stage being autonomous vehicles that are capable of passively gathering data.

![Progressive ITS Technology Advances](image)

**Figure 1.** A diagram highlighting the evolution of Intelligent Transportation System (ITS) infrastructure over time, from an initial basis of minimal/no communications technology for the average driver to fully automated ITS data collection systems. The advances are positioned alongside a red exponential growth curve in order to convey the explosive growth of data generated by these new advances in ITS crowdsourcing technologies, highlighting the need for ITS crowdsourcing systems to be engineered with scalability in mind.

The current work in engineering these data systems is motivated by the need to develop robust, resilient, secure, future-proof, and scalable infrastructure. Unstructured data is generated on the scale of exabytes on a daily basis [5]. This will be further compounded with the expected rapid growth in IoT device usage in many places, including the transportation sector. The combination of the rapid growth in these intertwined domains along with the impending ocean of data from ITS crowdsourcing systems potentiates the need for new avenues of recruitment of workers for mobile, spatial, and passive collection crowdsourcing tasks in harmony with already existing technology and methods.

In this paper, we first present the ITS crowdsourcing architecture, and then discuss them in further detail by highlighting the different crowdsourcing types: mobile (i.e., using mobile devices...
such as smartphones), spatial (i.e., tasks are tied to a specific location or region), and passive sensing (i.e., the data is collected automatically by smart devices) crowdsourcing, along with promising applications of each. Next, we discuss the technology and techniques that will lead to enabling efficient ITS crowdsourcing applications, such as advances in ITS communications technologies and cloud computing. We accompany the discussion with selected simulation results showing how these techniques bring about benefits for users of some crowdsourcing-based ITS systems. Afterwards, we provide a high-level overview of the state of the art in ITS crowdsourcing, by reviewing the different recent work led by researchers to address shortcomings in this area. From there, we consider the logical future directions needed for improvement, development, and deployment of automated real-time crowdsourcing systems. These systems will provide the basis for many useful applications, such as advanced traffic monitoring, highly granular roadway mapping, and other systems that will facilitate the emergence of autonomous vehicle technology on the roadway.

The rest of this paper is organized as following. Section 2 outlines the main crowdsourcing architectures that are utilized in ITS. Section 3 discusses the technology that enables functionality of these systems supported by selected simulation results. Section 4 overviews the state of the art. Section 5 discusses potential future research directions and challenges. Finally, Section 6 concludes the paper.

2. ITS Crowdsourcing: Architecture and Variants

This section takes a high-level look at some of the different kinds of ITS-based crowdsourcing applications in mobile, spatial, and passive sensing crowdsourcing. In ITS crowdsourcing, the task requesters are oftentimes drivers or people attempting to make transportation decisions such as what route to take, where to park, etc. In the future, they may also be smart/autonomous vehicles. The information they request serves to aid in their decision-making. Workers are usually other people through mobile phones or automated vehicles/UAVs/IoT devices that perform tasks in diverse locations across the transportation system at the request of the aforementioned task requesters. Figure 2 illustrates a basic example of a crowdsourcing system architecture. We use the example of a mobile crowdsourcing system architecture due to its prevalence today. Figure 3 shows an example of how all three crowdsourcing variants, discussed in this section, could work in parallel in an urban setting.

![Figure 2](image-url)  
**Figure 2.** A diagram showing the basic architecture of a mobile crowdsourcing platform. Requesters submit tasks to the platform. Tasks are assigned to workers on the online platform, who utilize their Personal Mobile Devices (PMDs) to collect data, and transmit it to the cloud, where platform operators may extract it and utilize it for models, visualizations, and/or decision-making.
Figure 3. A diagram illustrating all three crowdsourcing frameworks in action. The red star is the location of a spatial crowdsourcing task. (A) through (C): Examples of spatial crowdsourcing. (D): Example of passive sensing. (E): An Unmanned Aerial Vehicle (UAV) performing an active sensing task. (F): An example of mobile crowdsourcing.

2.1. Mobile Crowdsourcing

According to Statista (https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/), just under 2.5 billion people globally use smartphones. The explosive growth of smartphone use over the last decade has put mobile computing into the hands of nearly half of the world’s population in just over a decade. The large user-base of smartphone technology leads to an ever-growing set of potential crowdsourcing opportunities in many domains, including transportation.

Mobile crowdsourcing is when an assigned task is performed by workers on mobile platforms (e.g., a smartphone). Waze and Uber are both contemporary examples of mobile crowdsourcing applied for transportation; Waze allows for users to input road conditions for the benefit of others, and Uber is a ride-share service operated through mobile technology. Other novel systems aim to leverage social media data to predict incidents [6], which may have promise in dissemination of information related to car accidents and traffic. Many people driving in unfamiliar places rely on applications such as Google Maps or Apple Maps to navigate. The authors of [7] developed a mobile crowdsourcing platform called CrowdNavi that addresses a weakness in these maps applications—they often give incorrect navigation instructions near the destination. CrowdNavi works by tracking how drivers performing as the workers of the crowdsourcing framework exit off of a road toward their destination. By utilizing the behavior of multiple drivers entering from different approaches, the system improves upon the directions provided by the traditional Maps applications, improving the experience for other drivers.

Whenever drivers encounter congestion, they typically attempt to change to an alternate route to avoid the congestion, hence making them poorer potential workers for crowdsourcing frameworks to determine congestion in real-time. In urban settings, where some of the worst congestion can occur, there is often a large fixed-route bus system as part of the public transportation system. Since these buses operate on fixed routes, their drivers cannot re-route the bus’ path to avoid congestion. The authors of [4] proposed a crowdsourcing platform that takes advantage of the
fixed paths of the bus system and delegates the task of measuring the bus’ trajectory. If the bus has an unusual trajectory (i.e., is too slow), then it provides the system operators notice that there may be congestion. This information can also be disseminated quicker to other drivers, providing them more time to re-route and achieve a better user-equilibrium. Other methodologies have been developed to measure and predict traffic congestion based on mobile crowdsourcing techniques. The authors of [8,9] developed systems that utilize data mining techniques and social media to predict traffic. As mentioned before, the authors of [10] developed a relationship-based data dissemination architecture. In conjunction with automated data collection, this type of system could be used to pass the congestion information to drivers that frequently traverse along the congested route to make a detour. One commonality amongst the systems described above is that they are designed to assign one type of task to the workers, possibly through heuristics such as the ones developed by the authors of [11].

2.2. Spatial Crowdsourcing

Spatial crowdsourcing is when a crowdsourcing task requires workers to be at a specified location while completing the task [12,13]. Mobile and spatial crowdsourcing have a significant amount of overlap. PMDs provide the power of connectivity and computation anywhere the user brings the device, assuming a strong enough network connection. Indeed, the mobility of smartphones along with ubiquitous connectivity lends itself as an effective platform for spatial crowdsourcing tasks. While mobile devices are the primary setting for spatial crowdsourcing now, the rise of embedded IoT technology, especially in vehicles allows for the growth of new spatial crowdsourcing systems that work more passively than mobile crowdsourcing tasks [14].

In [11,15], parking location monitoring systems are an example of spatial crowdsourcing. In the envisioned platform, workers report open parking spaces in certain locations. This information can assist drivers in finding street parking, and provide other information regarding their destination, reducing wasted time searching for a parking space. Information regarding the location of parking in proximity to an end-user’s destination will reduce energy and time wasted looking for parking, and will aid drivers in making efficient navigation decisions.

Spatial crowdsourcing may also be utilized for panoramic tour construction [16,17]. By stitching together multiple images from an allocation and creating a panoramic tour, platforms such as Google Maps may be able to update their street-level views much more rapidly, removing the need for dedicated vehicles and workers to perform the task of taking pictures along fixed, assigned routes. Beyond stitching crowdsourced images for panoramic tours, spatial crowdsourcing may be utilized for generating 3D maps for autonomous vehicles [16]. With the utilization of crowdsourced data for this application, costs associated with 3D map generation can be drastically reduced, enabling more detailed, useful maps for future autonomous vehicle navigation.

2.3. Automatic Sensing Crowdsourcing

Automatic sensing crowdsourcing takes advantage of sensors embedded into smart vehicles to passively collect data while operating that is sent to a central system for processing and use. One of the main objectives of passive collection tasks is to schedule the collection based on the vehicle trajectories [11]. Other passive collection tasks may be utilized to track vehicle fleets in real-time [18].

Although this example utilizes a mobile phone for proof-of-concept, the system developed in [19] is an example of automated sensing crowdsourcing. The system leverages smartphones (playing the role of future embedded sensors) mounted on the dashboard of a vehicle to monitor the quality of pavement. In areas with high volumes of traffic, a system like this can collect a very large amount of data (including redundant data that can smooth out individual observations). This can be leveraged by ITS operators to make more informed decisions on how to improve existing roadways. In addition, trip data can be incorporated with this to determine the best ways to re-route traffic when the roadway does require physical repairs. Another example is that of automated lane imaging [20], which can be
utilized to gain information on the width and condition of roads automatically, and can be leveraged into machine learning algorithms that aid in navigation for autonomous vehicles [21].

By automating data collection, the system becomes more reliable and safer, as less bias would exist in the data collected, and humans would not be distracted from vehicle operation by the data collection processes. The abundance of passively-collected data may even bring the future of fully automated transportation and logistics systems to fruition by providing researchers and practitioners more robust datasets to train machine learning models that aid in future traffic control systems. These developments would make roadways safer and more efficient for end-users, and would promote increased economic productivity in the regions they serve.

3. What Enables Its Crowdsourcing?

Advances in communications and cloud computing are resulting in those technologies becoming the main technological underpinning of ITS crowdsourcing systems. Increased connectivity increases the size and number of gateways to the main cloud systems, which act as scalable storage and computing power centers. The adaptability of these technologies in concert lead the technologies to act almost as a living system, that can respond in real time to shifts in use and demand. In this section, we examine technologies that enable automated ITS crowdsourcing and investigate some practical use-cases.

3.1. ITS Technologies Enabling Crowdsourcing

ITS technology powered by IoT devices such as embedded sensors in smart vehicles and new communications infrastructure stand to provide an untapped basis for mobile crowdsourcing applications from the exponential growth of inter-connectivity between all nodes of the overlaid communications network. The rise of 5G networks can lead to all sorts of high-bandwidth network topologies connecting smart vehicles in a VSN [2] (V2X), such as vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-pedestrian (V2P), and vehicle-to-drone (V2D). See Figure 4 for an illustration of all of these communication paradigms working in concert together. A common challenge associated to the effective exploitation of ITS technologies to enable automated crowdsourcing tasks is its efficient infrastructure planning. In [22], we developed a smart placement of RSUs assisting crowdsourcing ITS infrastructure in a way to maximize the effective coverage of transmission architecture, while considering solar-powered RSUs and an amortized budget. We also developed heuristics designed to find a suboptimal configuration faster than solving a mixed integer programming problem [23] and to adjust RSU coverage to unexpected events [24]. Figure 5 shows RSU placement through our optimization framework. The aim of the framework is to determine an effective placement of RSUs in order to maximize the sum of the “demand” of all the covered points, where the “demand” could refer to any metric of interest for ITS planners, given an amortized budget consisting of periodic operational costs and amortized capital expenditures. In other words, this framework was purposely designed to be a general problem, so the “demand” metric can be freely changed based on the planner’s choice. In our example, we defined the “demand” as a normalized linear combination of randomly sampled data meant to represent traffic fluctuations, along with car accident data sourced from NYC OpenData (https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95). In our case, overall coverage efficiency is 92.5% given the provided by budget. The framework places RSUs to cover highly demand points of interest. The coverage of low demand points is a consequence of them existing in the same general region as high demand points. The placement of RSUs aids in intelligent infrastructure planning, providing increased connectivity. The improvement in the number and quality of connection gateways, along with smart vehicles armed with a dizzying array of sensors and increased autonomous driving capabilities, provide new means of collecting transportation-related data for ITS mobile crowdsourcing applications.
Figure 4. This figure provides a high-level look at cloud technology in ITS crowdsourcing systems. The blue arrows represent wired or mobile broadband connections, the red arrows represent Vehicle to Drone (V2D) and Vehicle to Infrastructure (V2I) connections, the white arrows represent Vehicle to Vehicle (V2V) connections, and the green arrows represent VSN relationships.

Figure 5. Roadside Units (RSUs) form a key component of V2I Networks. This figure demonstrates optimal placement of RSUs in Lower Manhattan, where the color of the points corresponds to connectivity demand, the red triangles correspond to RSUs, and the gray ellipses represent their coverage area. The coverage efficiency of this particular configuration is 92.5%.
Automated sensors on vehicles, UAVs, and RSUs can collect data that could not be manually collected by humans in parallel with human-produced data, and provide wireless transmission access for the data to the cloud. When the data is aggregated, it could be applied to aid ITS planners for decision-making, or can be used by end-users on the roadways to make more informed decisions while in transit.

### 3.2. Vehicular Social Networks

A VSN is a type of ad-hoc, dynamic connection between users of smart vehicles in an ITS based on common factors such as destination, route, and location [2,4]. In the context of ITS crowdsourcing, VSNs may be leveraged for operational applications. VSNs can be used to gather information on traffic congestion, how vehicles interact with each other on the roadways, and may be used to relay other information from one user to another. VSNs offer the advantage of connectivity based on shared interests, much like ordinary social networks (OSN). The social aspect allows for more granular sharing aspects between users, providing a mechanism for more accurate and useful information dissemination.

We illustrate a benefit of mobile and automated sensing crowdsourcing for ride-share recommendation systems in Figure 6. The proposed recommendation system utilizes automated sensing and mobile-crowdsourcing techniques to improve traditional ride-hailing and regular curbside ride-sharing services by introducing collaboration among drivers, improving quality of service by reducing customer wait time and wasted searching for new fares [25,26]. In this comparison, we evaluate the performance of the proposed recommendation system for both regular and ride-hailing taxi services. There are four cases: Two traditional taxi services without recommendations (Regular Trad. and Ride-hailing Trad.) and two taxi services based on our proposed recommendation system (Regular Recom. (VSN), and Ride-hailing Recom.). With VSN, the taxis can instantaneously share their data through the vehicular social networks, which allow drivers to have more knowledge about the customers’ demand and the traffic situation. We demonstrate around 10–15% reductions in customer waiting times for traditional curbside services, 20–25% reductions in waiting time for ride-hailing services, 60% reductions in extra distance driven by workers while searching for new fares in curbside services, and 25% reductions in extra driven distance by ride-share workers. This is a significant savings for both the customers and the taxi drivers that are exploiting crowdsourcing services through VSN.

![Figure 6](image-url)

**Figure 6.** Visualizing how mobile crowdsourcing can improve ride-sharing services. The red and black bars correspond to traditional taxi pickup strategies (street- and ride-hailing, respectively), where the other colored bars correspond to crowdsourced information for driver cooperation. We see improvements in idle time for customers and drivers between traditional (red and black), vs. crowdsourced (green and pink) approaches.

### 3.3. Cloud Techniques Enabling Crowdsourcing

The cloud is a network of computers that work in coordinated fashion to process data or store it in parallel. Cloud computing infrastructure stands to serve as the basis for many ITS crowdsourcing
platforms; IoT and communications technology discussed in the previous section go hand-in-hand with the scalability offered by advances in cloud computing technology and techniques. Cloud computing technologies and techniques such as edge computing [27], distributed file processing, storage, task management [28–30], worker reliability metrics [31], and trust systems [32] can streamline the system. Edge computing distributes work away from the centralized control system by leasing computations out to devices on the edge of the network (e.g., workers’ PMDs or other embedded computers in the network, such as smart vehicles, RSUs, or UAVs) to optimize the transmission of data across the network for prevention of bandwidth issues. Distributed file-systems allow for cloud storage and resources to scale up or down in response to demand, and worker reliability metrics will ensure that the system is not polluted with useless data. In this section, we touch on some key characteristics of the cloud that enable ITS crowdsourcing. Figure 7 conveys a basic image of how these cloud systems mat work in practice.

**Figure 7.** A diagram representing the relationships between devices in edge computing. The data warehousing and bulk processing make up the core, while other collection and small processing tasks take place in the IoT and mobile devices on the “edge” of the network. This serves to reduce overhead in the central cloud infrastructure.

### 3.3.1. Edge Computing

Computers “on the edge” of a cloud computing network are far away from the centralized data centers that make up the heart of the system’s computing and storage capabilities [33]. While edge computing devices are less powerful than the powerful servers in the datacenter, they exist to offset the main bottleneck now present in cloud computing: data transmission across the network. Figure 8 illustrates the trade-off in action. In Figure 8, “slow edge GPU” means the edge computers in the simulated network have a benchmark speed of processing 250 MB/s of image data, which is analogous to LiDAR point set data—this benchmark speed is based on the Nvidia Jetson System on Chip (SoC). The “med.” and “fast” speeds correspond to 500 MB/s and 1000 MB/s, respectively, which would be present in more powerful edge computing devices that may consist of crowdsourced computing power similar to the Folding@Home project (https://foldingathome.org). The figure shows that in the “fast”
case, 60% of the tasks are assigned to the edge computers, as their speed offsets the high latency associated with sending all tasks to the cloud, leading to overall lower system latency even though the cloud computing center is much more powerful than the edge computers [34]. Computing tasks can be performed on the edge to minimize the amount of network bandwidth taken up by transmission of data to the center of the network, eschewing a traditional server-client framework for a much more scalable, flexible network topology [27].

ITS crowdsourcing frameworks likely will involve many spatial and mobile tasks being performed in parallel amongst the many users of the future transportation systems where transmission of information across the many connected entities will be massive. Edge computing will allow for distribution of computing and storage tasks to be outsourced away from the centralized servers to devices much closer to data sources, reducing bandwidth consumed by transmissions, reducing loads on the central servers, and speeding up the latency of a task’s completion.

![Diagram](image.png)

**Figure 8.** The graph illustrates at how the speed of the edge servers affects the average processing latency of Elevated LiDAR (ELiD) data, and how much data is sent to the cloud. When the edge servers can handle the load in a reasonable amount of time, not all of the data is sent to the cloud for processing.

3.3.2. Distributed Storage

Technologies such as Apache Hadoop allow for a distributed filesystem to run on a connected network of computers in the cloud, essentially acting as one massive data storage system. This distributed storage allows for “easy” large-order scalability simply by adding (or subtracting) computational resources, acts as a redundancy system in the case of the failure of a node in the network, and distributes data across the system in a way to reduce latency.

A cloud-based data pipeline that shares data amongst members of a VSN based on factors such as similarity of driving patterns was proposed in [10]. In conjunction with edge computing and scaling systems such as Hadoop, this type of architecture can further reduce bandwidth consumption by data in the network, along with developing more personalized and spatially segmented services.

An example of an ITS crowdsourcing technology that requires high computational power and storage is the use of shared LiDAR for autonomous vehicles. Elevated LiDAR (ELiD) [35] is touted as an alternative or at least complementary system to LiDAR units placed on-board, in order to reduce the amount of computational and cost overhead required to safely operate an autonomous vehicle. One challenge for ELiD systems is to allocate their real-time generated data to edge and cloud servers for processing of the raw point cloud data in order to minimize the latency and speed up their operation to provide vehicles instantaneous information, e.g., 3D mapping. In Figure 8, we provide the performance of a backhaul latency-minimizing framework that considers low communication latency to edge computers with less-powerful resources (leading to larger processing latency), and longer communication latency but shorter processing times to the cloud. Figure 8 visualizes this trade-off—as the processing speed of the edge computers increases, average processing latency for each LiDAR decreases. Optimizing the task-server allocation for ITS crowdsourcing applications is important to enable smooth operation of the system.
3.4. Automated Crowdsourcing Management

The complexity of ITS crowdsourcing frameworks, large set of tasks, and the numerous potential workers requires systems in place to allocate tasks across the platform. Spatial tasks involve their own set of unique challenges regarding task allocation [36]. The authors of [37] developed an automated assignment system for mobile crowdsourcing frameworks based on relationship metrics, and network service parameters.

Automated management systems would go hand-in-hand with the cloud systems mentioned earlier. By automatically optimizing the task allocation, the tasks can be divided spatially and results can be passed through the network at optimized access points and data can be pipelined, collected, and stored in locations that are geographically optimize to be best-suited for use based on the areas it was collected at.

Recent developments of cloud infrastructure is laying the groundwork for ITS’ to act as a massive crowdsourcing platform. Incorporating traditional transportation infrastructure into the IoT will make transportation systems into cloud computers, providing ITS operators a much greater amount of information for decisions that would improve the quality of the transportation experience for anyone who utilizes the roadways in the future.

3.5. Reliability Scoring

When humans are in the loop of crowdsourcing systems, oftentimes there is uncertainty on how reliable human-crowdsourced responses are. Until automated systems are robust enough, humans stand to play roles in ITS crowdsourcing. With ITS data, reliability is a key metric of focus, as faulty information can lead to potentially fatal mistakes.

Researchers developed a probabilistic model for predicting the reliability of human crowdsourcing workers [31]. Development of reliability systems such as this stand to not only assist in the improvement of the quality of human-produced data, but also automatically collected data—probabilistic systems can be used as a type of anomaly detection to filter out any corrupted input data from transmission errors. While not specifically a cloud technology, this type of reliability metric can be utilized in conjunction with other cloud infrastructure to minimize bandwidth consumption by faulty data or by malicious actors performing denial of service attacks.

The work in [38] addresses the reliability scoring and automated crowdsourcing management problems in crowdsourcing-based real-time vehicle navigation where routes are continuously updates according to other workers’ feedback. Oftentimes, there is uncertainty surrounding the accuracy of data reported by workers in crowdsourcing systems. Figure 9 visualizes how accounting for uncertainty may influence navigation decisions. The red path corresponds to a traditional navigation following the shortest path taken by a driver, which has an obstruction that forces the driver to reroute. In [38], an algorithm designed to account for these types of blockages, automatically using other road users’ feedback, finds a new optimal path for the driver (blue path), and the green path is the optimal path when also attempting to minimize the risk associated with traversing along paths where the data collection is not as reliable. Hence, the crowdsourcing platform proposes to the driver to follow a longer route to avoid the risk to be trapped in a congested route due to its limited knowledge.
3.6. Security and Privacy

Whenever data is processed in the cloud/edge and transmitted over the internet, measures must be taken in order to protect the identities and data of crowdsourcing workers. Part of the aim of privacy and security researchers in this area is to ensure “privacy-by-design” [39]. End-to-end (E2E) encryption [40,41] is vital to harden communications across the network, so malicious actors cannot passively collect information in transit. In addition to E2E encryption, obfuscation techniques related to the location metadata of the data collection agent [42–45] are necessary when the data is being processed at the end of its journey through the communications network. These obfuscation techniques have also been explored in the edge as well [46].

4. Overview of the State of the Art

In this section, we briefly review the current state of ITS crowdsourcing. While doing so, we categorize work based on the concept explored or problem addressed, and summarize the general contribution made. In addition, we mark which crowdsourcing variant that each topic is most closely associated with. The summary of our findings are shown in Table 1.
Table 1. Literature review summary.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Reference(s)</th>
<th>Contribution</th>
<th>Crowdsourcing Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mobile</td>
</tr>
<tr>
<td>Architecture/Planning</td>
<td>[1] Survey paper of mobile crowdsourcing: examples of systems and definition of what mobile crowdsourcing is</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[12,13] Surveys of spatial crowdsourcing: definitions and examples</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[32] Provides a trust management mechanism as a framework to secure vehicular social data, a cloud-based architecture for VSNs</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[2,4] Survey of VSN</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[28,29] Distributed architecture for mobile crowdsourcing management</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[3] UAVs in the ITS: their role in increasing network connectivity</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[22–24] RSU placement: optimal, heuristic, and adjustment to unexpected events</td>
<td>X X X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[35] Proposal of automated elevated LiDAR to reduce autonomous vehicle computational overhead</td>
<td>X X X</td>
<td>X</td>
</tr>
<tr>
<td>Applications</td>
<td>[15] Crowdsourced monitoring and reporting of available street parking</td>
<td>X X X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[20] Crowdsourced data collection of road lane data</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[8,9] Social media-based traffic prediction with NLP</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[6] Social media data mining applied to detect roadway incidents</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[7] Development and testing of a last-mile navigation system to add to existing mapping programs</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[19] Infrastructure health monitoring</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[18] Advanced tracking of public transportation vehicles</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[16,17] Generating a 3D map of surroundings based on crowdsourced data</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[4] Utilization of mobile devices on fixed bus routes to measure and visualize traffic congestion</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Concept</td>
<td>Reference(s)</td>
<td>Contribution</td>
<td>Crowdsourcing Variant</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mobile</td>
</tr>
<tr>
<td>Task Recruitment</td>
<td>[38,47]</td>
<td>Advanced routing strategies beyond minimizing time traveled; powered</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>through crowd-sourced data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[31]</td>
<td>Ratings system for crowdsourced workers</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[36]</td>
<td>Spatial Crowdsourcing Task assignment survey</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[37]</td>
<td>Crowdsourcing assignment for mobile crowdsensing</td>
<td>X</td>
</tr>
<tr>
<td>Edge Computing</td>
<td>[33,48]</td>
<td>Edge computing architectures for mobile crowdsensing</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[10,49]</td>
<td>Efficient data processing and handling on the edge</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[27]</td>
<td>Survey of IoT; discussion of edge computing</td>
<td>X</td>
</tr>
<tr>
<td>Security/Privacy</td>
<td>[40,41]</td>
<td>End-to-end encrypted mobile crowdsourcing</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[42–45]</td>
<td>Obfuscation of identifying information of crowdsourcing by reducing</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>granularity of data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[46]</td>
<td>Obfuscation of data in edge computing servers</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>[39]</td>
<td>Perspectives and suggestions of engineering “privacy by design” in smart</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>vehicles</td>
<td>X</td>
</tr>
</tbody>
</table>
The literature we reviewed can be broadly broken down into five main categories. The Architecture/Planning category is mainly survey papers that define and explore one of the crowdsourcing models in depth, or are papers that focus on the planning/infrastructure that underpins crowdsourcing systems. We consider which class of crowdsourcing each paper contributes most to. The ITS crowdsourcing applications category contains papers that are systems that apply one or more of the crowdsourcing architectures to solve an ITS problem. The task recruitment category contains work that mainly focuses on efficiently selecting workers, assigning tasks relevant to the system, and rating their performance. The edge computing category focuses on the edge computing networks along with distributed storage that plays a vital role in the operation of a crowdsourcing system in real time. We conclude with the security/privacy category, which focuses on work that is designed to keep these crowdsourcing systems safe for people to participate, and to ensure that data is collected ethically.

The architecture/planning set of papers mainly focuses on the definitions of the ITS crowdsourcing systems as well as planning systems for the implementation of the crowdsourcing architecture and systems. In order to understand the core aspects of crowdsourcing systems, we considered surveys of mobile [1] and spatial crowdsourcing [12,13], which gave us broad insight into how these systems operated, and how they may work in ITS. After, we reviewed papers pertaining to the current technologies that will underpin these crowdsourcing systems in ITS. These include VSN [2,4], which will provide the main communications backbone for ITS crowdsourcing systems. In addition, we considered VSN components. We looked at the planning of RSU installation [22–24], along with UAVs as flying RSUs [3], trusted cloud computing systems [32], automated sensing systems [14,35], and decentralized control/access systems [28,29].

After reviewing the systems that serve as the basis for ITS crowdsourcing, we review several promising applications. These include street parking monitoring and reporting [15], lane data collection [20] and detection [21] social-media based traffic [8,9] and incident detection [6], advanced route planning based on on-board sensors [47] and considering uncertainty [38] and crowdsourced last-mile navigation data [7], and 3D mapping [16,17]. We chose these application examples due to their potential capability of significantly improving the performance of existing transportation systems. Crowdsourcing applications cannot work well without efficient task recruitment and verification processes. We reviewed techniques developed by authors to address this. Work has been done to develop multi-objective recruitment systems [11], task allocation solvers [30,36,37], ratings systems to reward good workers and incentivize poor performers to improve [31], and a system to take into account uncertainty associated with data collection [38].

We then moved on to looking at more cutting-edge advances related to ITS crowdsourcing. Edge computing is a recent development that aims to utilize computers closer to the data collection sources to minimize bandwidth utilization in the network [27]. We reviewed papers that related on applying edge computing in crowdsourcing applications [33,48], and how data processing on the edge is more efficient [10,49].

Finally, we considered privacy and security measures that are being undertaken to protect crowdsourced worker’s identities. Security and privacy must be built into the system while it is being developed [39]. End-to-end encryption is utilized to safeguard communications in transit [40,41], and obfuscation methods are used to anonymize data during mission-critical processing tasks [42–44,46].

Our review encompasses all of these areas related to ITS crowdsourcing to get an idea of how it is a system of systems. The architecture survey papers provide the system-level view. Application papers show the broad set of subsystems that will be powered by ITS crowdsourcing systems. The recruitment papers consider the human aspect of the system. The edge computing papers highlight technological advancements that continue to improve the viability of ITS communications systems. Finally, we consider security and privacy papers, as these are a critical subsystem that ensures the systems are safe in practice.
5. Future Research Directions and Perspectives

ITS infrastructure, cloud computing, and crowdsourcing frameworks provide for countless opportunities to apply data for solving or managing current transportation system inefficiencies. With this, we discuss systems and applications that take could securely and ethically leverage the plentiful bounty of vehicle and mobile-generated data to make transportation systems safer and more reliable. We see that in general, the outcomes of addressing these research directions will manifest as robust automated ITS crowdsourcing systems that will be a core technology in the operation of autonomous vehicles.

5.1. Future Research Directions

5.1.1. Worker Recruitment

The size and scope of crowdsourcing platforms have the potential to lead to massive-scale operations. This requires an efficient worker recruitment system in order to prioritize the most productive workers, as well as provide proper incentive. Without such a system, the platform would not have the ability to efficiently scale. Since scheduling is a hard problem (especially joint scheduling of multi-task optimization), and the base of workers is only expected to grow very fast, researchers should focus on developing efficient algorithms that can work in real-time to schedule workers, as many mobile and spatial crowdsourcing tasks also are time-sensitive, as the data collected will feed into control systems for autonomous vehicles.

5.1.2. Privacy and Security

The user’s geo-spatial data is considered as sensitive as people are willing to share their daily movement and routines. To gain their trust, securely engineered ITS crowdsourcing systems must be built, in order to prevent malicious actors from exploiting the systems to steal sensitive data data. A highly effective way to reduce the attack surface of the system should incorporate end-to-end encryption to protect data en route from man-in-the-middle attacks. In addition, to foster trust, the users should have the right to decide what kind data they are willing to share. As mentioned earlier, there is a great deal of work in implementing these measures into ITS crowdsourcing systems, but much more effort is needed to counter the constant innovation of attacking soft points in the system.

5.1.3. Data Fusion

The crowd-sourced data may be a potential stream of training data for autonomous vehicles. In conjunction with the data relevant to car-following models [50], computer vision systems can be used to detect other entities on the road. The combination of the data streams would allow vehicles to synthesize different types of data to consider the interactions and therefore a more complex feature space before making decisions, potentially leading to better performance. The main risk of both the data fusion system and the advanced car-following system is over-fitting to the training data, due to the highly complex feature space. Over-fitting would lead to poor generalized behavior, and would arrest the benefit of developing these models.

5.1.4. Data Filtration on the Edge

In proposed ITS networks, oftentimes RSUs and UAVs are considered to be network gateways. Due to the expected spike in data throughput of the network, it would be prudent to consider methods of filtering data input to the network prior to the data actually passing through the gateway. By distributing data filtration on the “edge” of the cloud, it would free up the central computer nodes for storing data and funneling it into useful metrics for ITS planners, as well as reducing the likelihood of corrupted or useless data from entering the system in the first place. In addition, it would reduce the bandwidth overhead of the communications network, as less data would be transmitted though
the system. Edge computing also has the ability to reduce overall data processing latency, which will be a boon for autonomous vehicle performance.

A couple of challenges are present when considering filtration of data on the edge. Anomaly detection systems would be a viable candidate for the edge filtration system; however, these require a certain amount of training data to be able to build a proper distribution of the data input. The type of data inputted into the system will also determine which kind of anomaly-detection system to use (e.g., for images, an auto-encoder network may be the best type of system to use [49]). Additionally, the devices on the edge are likely to be embedded and therefore resource-scarce devices. The anomaly detection system would have to be painstakingly optimized to ensure that the embedded devices are not overworked. If the devices on the edge are scheduled to power on/off periodically to conserve energy, a system must be put in place to filter the data upstream in the cloud network.

5.2. Its Crowdsourcing Use Cases

5.2.1. Infrastructure Monitoring for Improvement

The authors of [19] developed a mobile phone-based sensing system to detect bumps and other anomalous behavior while driving a car to measure the quality of the road surface that vehicles drive on. A logical step forward from this idea would be to consider how smart vehicles equipped with an array of specialized sensors may perform more granular monitoring of transportation infrastructure. Combined with an automated spatial crowdsourcing framework, drivers can be tasked with driving along roadways, bridges, tunnels, etc. to provide data on the state of key infrastructure. This type of system could assist ITS operators in deciding which infrastructure improvement projects to target first, and to plan for diverting traffic away from that area based on the volume of traffic recorded through the area of interest.

The main challenges to such a system would be how to utilize sensors or on-board mobile phones for these tasks, and as an extension, how to measure the reliability of the sensors and overarching system. Additionally, there must exist some form of incentive for drivers to drive along roadways that may not have as much traffic to gather more data, especially if those pathways are an inconvenience to drive along. These concerns may be addressed as autonomous vehicle share on roadways increases; they may be automatically routed along desired parts of the network, and will come with integrated sensor technology that will have the capabilities of recording what the task requester needs.

5.2.2. Real-Time Decision-Making

The data influx from the roadways can be used for real-time monitoring of traffic systems. The combination of vehicles armed with an array of sensors, robust high-bandwidth communications technology, and mobile phones can be used to upload volumes of information regarding traffic flow, average speed, etc. For example, this information can be leveraged for a real-time navigation system designed to detect unusual disruptions in traffic patterns could be used to re-route vehicles away from that road segment, minimizing increases in travel time across the system.

A challenge in implementing this is how to detect service disruptions in the first place. While traffic congestion is likely periodic (e.g., rush hour periods) and can be modeled from historical data, accidents are mostly unpredictable events, and diverting traffic around them are reactionary processes. Additionally, automatic accident detection must lead to updates as quick as possible to ensure that the people planning on taking a trip through the blocked path may avoid it and find a more suitable route in a reasonable amount of time. Assuming some form of automated control system, this data could be used to solve a user equilibrium problem to guide how traffic is routed. Distributed computing systems may be the key to reducing the decision making latency within a real-time scope.
5.2.3. Improved Car Following Models

In agent-based traffic simulation, vehicle following models play a vital role in simulations, as these models form part of the basis of the rules governing the behavior of each individual agent (vehicle). With vehicle-based crowdsourcing systems, the movement of vehicles in an area can be tracked simultaneously with the movement of its surrounding vehicles. This can be used to develop stronger car-following models that consider more than the traditional models that are based on just the vehicle directly in front.

The main challenge in leveraging crowd sourced data for car following models would be how to tie vehicle movement data together. Consideration of dozens of vehicle trajectories per a small road segment over a very short period of time would be very expensive computationally for even small-scope problems. Additionally, the complexity of training a predictive model would be quite large, especially if deep learning is utilized to mitigate the challenge of engineering such a complex feature space.

6. Conclusions

In this paper, we discussed the efforts of researchers to develop the foundation infrastructure for data-driven ITS. Along with the discussion, we provided some simulation results to demonstrate how technologies powering ITS mobile crowdsourcing systems can be used to improve the quality of experience and safety for ITS users. Afterward, we reviewed state-of-the-art advances in this area. Finally, we discussed potential research directions to consider for further improving the performance and viability of these ITS crowdsourcing systems. Leveraging ITS for crowdsourced data has the potential for revolutionizing transportation systems. Increased connectivity and granular, data-driven decision making will make roads safer and more efficient, enriching the lives of end-users.

Author Contributions: Contributions to this manuscript can be summarized as follows: conceptualization, M.C.L., and H.G.; software, M.C.L. and X.W.; writing—original draft preparation, M.C.L.; writing—review and editing, H.G.; visualization, X.W. and M.C.L.; supervision, H.G. and Y.M.; project administration, Y.M.; funding acquisition, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

9. Yan, H.; Yu, D. Short-Term traffic condition prediction of urban road network based on improved SVM. In Proceedings of the 2017 International Smart Cities Conference (ISC2), Wuxi, China, 14–17 September 2017; pp. 1–2. [CrossRef]


