The Effect of Information Uncertainty in Transportation Systems

Sorina Costache Litescu, Vaisagh Viswanathan, Heiko Aydt, Alois Knoll

TUM-I1633
The Effect of Information Uncertainty in Transportation Systems

Sorina Costache Litescu\textsuperscript{a}, Vaisagh Viswanathan\textsuperscript{a}, Heiko Aydt\textsuperscript{b}, Alois Knoll\textsuperscript{c}

\textsuperscript{a}TUM CREATE
\textsuperscript{b}Singapore-ETH Centre
\textsuperscript{c}Technical University of Munich

Abstract

Developments in Intelligent Transportation Systems (ITS), navigation devices and traffic sensors make it possible for traffic participants to not just access real time information regarding the traffic situation but, at the same time, also provide data back to the transportation system. This creates a feedback loop that can have significant consequences on the system performance in terms of total average travel time. In the current paper, we investigate the effect that information inaccuracy caused by different types of sources, can have on the system performance. We first identify three types of uncertainty that can arise in such a system: inaccuracy due to sparsity of data sources, collection inaccuracy and presentation inaccuracy. Subsequently, we use an agent-based microscopic traffic simulation to explore the effects that each type of inaccuracy can have on the transportation system. Experiments reveal certain surprising observations. Firstly, less than twenty percent of the traffic participants need to be data sources for optimal system performance. We also discover that lower precision of information presented to participants is sufficient and, in certain cases, better for system performance. This can have important implications on how information is displayed in navigation devices.

Keywords: Information Uncertainty, Participatory Sensing, Human Complex Systems, Information Propagation, Dynamical Information, Traffic Dynamics, Transportation Systems, Congestion,
1. Introduction

Novel technologies and applications on smart devices not only enable commuters to access real-time information, forecasts, and navigation guidance but also to contribute with their traffic data. Surveys show that, in most cases, drivers trust traffic information from smart devices and follow navigation recommendations provided to them [1]. Even when this information is highly detailed and accurate, complex and unexpected dynamics can emerge in such transportation systems. This is due to the massive participation of commuters as both sources for collecting data and consumers of the traffic information [2].

However, uncertainty, sometimes called inaccuracy or noise, can arise in the information passing through this system either at the time of collection, processing, or presentation. In the current paper, we explore the different kinds of noise and the effect that they can have on an ITS of the future.

In the kind of ITS discussed above, information is collected from different types of sensors. This information is then aggregated and processed to recreate a model of the traffic state. Eventually, the relevant parts of this traffic state information are transmitted to commuters through their in-vehicle information systems or personal smart-devices. There are several points in this process in which information inaccuracy may occur either because the collected data produces incomplete information or because the information loses some of the precision during processing and display. This is discussed in more detail in Section 3. It is important to understand the effects that inaccuracy can have as it can affect not just the actions of a few individual commuters but also the performance of the transportation system as a whole.

There have been several studies of the impact of noise on different complex systems. It is interesting to note that there have been some counter-intuitive discoveries suggesting that noise can have a potentially beneficial effect in many non-linear systems - both artificial or natural. An example of the former is the constructive effect of inaccuracy shown in technical systems where noise enhances the information transfer efficiency [3]; similar examples in natural
systems include discoveries in brain function, carrier signals, animal avoidance and feeding [4]. Section 2 provides a more detailed discussion of some of these studies. The objective of this paper is to analyse whether such effects of feedback loops can be found in transportation systems.

Traditionally, when building ITS systems and navigation devices, the effort consists in providing faster and more accurate traffic recommendations and real time predictions [5]. In general, improving the accuracy comes at a certain cost. For example, in order to get more accurate information, either more sensors have to be installed or more high-quality sensors have to be used. Both of these come at a financial cost. In another example, consider the information that is displayed on a traditional navigation device. The designer has to take care to present information in a way that can be easily understood [6][7] and within the constraints of the display device. This generally means that trade-offs have to be made in terms of what roads are displayed and what information regarding these roads is displayed (elevation, speed, etc.). To make these decisions in areas ranging from sensor infrastructure development to navigation device design, it is important to understand the acceptable levels of noise in traffic information.

The contributions of this paper are two-fold: firstly, we introduce a general source-based classification of different kinds of inaccuracy that can occur in data processing in an intelligent transportation system; secondly, we do a microscopic simulation based analysis into the effects that these different types of inaccuracy sources can have on the system and identify the acceptable levels for different types of uncertainty.

2. Related Work

Previous research has analysed the effect of traffic information on a transportation system. It has been shown that the information content, for example consisting of certain routes proposed for the traffic participants [8][9] to achieve either individual or global social optimum performance or using local or global details of the traffic network when determining the routes [10][11], has an effect
on the traffic. Providing inappropriate information to the traffic participants sometimes leads to undesirable situations such as one-sided congestion [12]. In [13], the authors analyse how the information quality and its accuracy influences traffic, unlike the other mentioned studies where information is error free. It was shown that drivers using forecast information, even with inaccuracy, produces a better impact on the traffic performance than present information. When providing predicted information with errors to a larger share of users the improvement in performance is bigger. In this paper, we explore this issue further by first categorizing different errors and analysing the effect that each error can have.

There are several studies that challenge the traditional view in information processing that noise degrades efficiency, and show that controllable noise can even be considered an additional engineering tool [14]. The influence of noise from information transmitted in the form of packages shipped between nodes of hierarchical networks is presented in [3]. The experiments were performed on artificial tree networks, scale-free networks and in a real network formed by email addresses of employees. Two types of noise are considered and shown to have a positive influence: one type dealing with a random part of packets paths and one originating from random changes in the initial network topology. In a similar vein, we deconstruct the different kinds of noise that can arise in a transportation system and analyse both their positive and negative implications.

Besides artificial systems, noise affects the natural complex systems as well. An example of noise influencing pedestrian movement simulation is presented in [15]. The authors describe the formation of pedestrians lanes. The number of lanes depends on the width of the street, on the pedestrian density, and also on the noise level. Animal behaviour is also influenced by the existence of noise, as explained in [16], [17] and [18]. Counter-intuitively, locusts increase the noisiness of their movements in response to a loss of alignment by the group.

In [4], the effect of noise is described in the context of Stochastic Resonance, a statistical phenomenon resulting from the effect of information processing and transfer. This phenomenon is compatible with neural models and brain func-
tions. In [19], the potential benefits of noise in nervous systems (human motor behaviour) was examined. Neural networks formed in the presence of noise are more robust and explore more states; this facilitates learning and adaptation. Moreover, noise induces stochastic facilitations in auditory brainsteam neuron models [20]. In [21] the authors discuss how nature has actively exploited the beneficial effect of noise by creating noise-assisted processes for achieving robust and efficient energy transfer.

A review of existing literature shows that errors and noise present in complex systems can have significant effects on its performance. Inspired by such observations, [14] introduced a new paradigm of noise-engineering. In the following sections, we explore how noise can be introduced in the information that passes through a transportation system and the impact that it can have.

3. Information in the context of transportation systems

Figure 1: Overview of a schema of an Information Control System interacting with a Transportation System where information is affected by uncertainty due to sparsity inaccuracy, collection inaccuracy and presentation inaccuracy.

In a transportation system, traffic information is obtained from data collected by sensors. These sensors can be fixed (e.g. inductive loop detectors,
radars, infra-red or acoustic) or mobile devices (e.g. smart phones, navigation devices, etc.) within vehicles. This information is aggregated and processed, some times through several layers, before it is presented back to the commuters through their different information systems (in-vehicle entertainment system or smart-phones). This forms a feedback loop since commuters are both consumers and producers of this information.

Figure 1 illustrates the feedback loop. On the left side of the image, is the front-end of a human complex system (HCS), in this case a transportation system with roads, traffic participants, vehicles sensors and so on; on the right side, is the information control system (ICS) that works at the back-end of the transportation system. HCS provides data to the ICS and also eventually utilizes the information that the ICS provides. The ICS is responsible for cleaning the raw data from the HCS, aggregating it, processing it and present it to the traffic participants. It is important to note that in our categorization, the processing system of the information presentation devices like smart phones or in-vehicle information displays are also part of the ICS as they determine how information is received by the user.

The information in this system can be seen to flow through three stages: input, processing and output. Information inaccuracy can arise in any or all of these stages. We define the uncertainty in each of these stages, based on their most common underlying cause, as: inaccuracy due to sparsity, processing and display.

During the input stage, the real world traffic status is converted into raw data by the different kinds of sensors. It would be practically impossible to observe every single point of the real world system due to the large number of high quality sensors that would be required. We term the inaccuracy that arise due to this lack of coverage of sensor networks as sparsity inaccuracy. Sparsity inaccuracy would be impossible to avoid completely in practice; however, it is useful, even vital, to discover the minimum coverage required for optimal system performance.

During the processing stage, the raw data is converted into information that
can be used to reconstruct the traffic state and, eventually, to a form that is presented back to the HCS. Inaccuracy can occur at different steps of this process. This could start right from low resolution sensors, to improper cleaning and inefficient algorithms for aggregation or traffic state reconstruction. This uncertainty is difficult to avoid but they become smaller over time as technology advances. Since the uncertainty manifests in the information system in the same way as inaccurate traffic state reconstruction, we classify them together as *collection inaccuracy*.

In the final step of the process, this traffic state information is presented back to the commuters through their smart devices. It would be impossible to display the state of the complete traffic system to the user. Thus design decisions have to be taken as to what information is displayed and in what resolution. For example, when displaying a map for navigation with congestion information, the roads with a range of high speeds may be marked in green and others in red; or there could even be a color gradient from red to green for a range of speeds. Lower resolution information may mean that it is easier for the user to process a larger amount of information (several roads at the same time) and it would probably also be technically easier to display this information. We term these types or errors due to trade-offs in how information is presented as *presentation inaccuracy*. It is crucial to understand these in order to create better smart devices for ITS.

In other previous research [13], the authors categorize traffic errors in the case of predicted information as: routes not precisely estimated, simulation model imperfection, current traffic condition not exactly monitored, driver’s route choice behaviour not understood. We believe that the new categorization of inaccuracy based on sources proposed in this paper is essential to study the impact of information uncertainty and noise on modern transportation systems that consist of mobile sensors, ITS and smart navigation devices. Furthermore, this categorization can help in gaining a better understanding of the modern and future transportation systems. Engineering solutions can eventually be developed that leverage on information as a control tool integrated in ITS. In
the following section, we present a methodology for exploring in more detail the different types of impact that each of these systems can have.

4. Computational Model

A real world scenario for studying the impact of noise is difficult to implement as it requires, among others, a massive rate of participation of the drivers both as sources and users of traffic information. It would also be difficult to study each of the different types of errors in isolation. In order to do this, we use a simulation based approach. The computational model that we use for the traffic flow, congestion formation, data collection and information dissemination has been described in detail in our previous study [2]. Here, we first present a brief overview of this model and subsequently introduce the new parts introduced in this paper, i.e. how the different types of information inaccuracy sources or noise are simulated.

The transportation system is simulated using an agent-based microscopic traffic simulation. The agents know the road network, perform route calculations and move forward on their route with a certain speed and acceleration determined by a time-stepped car following model (Intelligent Driver Model IDM [22, 23] in our particular case). A road \( Y \) is characterised by a tuple of minimum speed, maximum speed and road length: \( \text{Road}_Y = (v^\text{min}_Y, v^\text{max}_Y, L_Y) \). Our objective is to analyse the effect of inaccurate information dissemination in the presence of congestion. For this, as in the previous study, we introduce repeated stochastic disturbances in the traffic flow to create a controlled scenario with persistent congestion.

Each agent uses Dijkstra’s algorithm to determine the route from the source to destination. The estimated speed on each road is used as the weight for the Dijkstra’s algorithm. Informed and uninformed agents are contrasted by modifying this estimated speed. **Uninformed agents** use the maximum speed on the road (thus assuming free flowing traffic); while **informed agents** use, for each lane, the current average speed on the road, calculated as the average of speeds...
reported by the agents currently on that road. In this way, congested roads tend to have a lower priority in the informed driver’s choice. The percentage of informed agents in a scenario is denoted by the letter $p$.

Additionally, we simulate the effect of the three types of inaccuracy introduced in Section 3. **Sparsity inaccuracy** are simulated by varying the percentage $s$ of agents that provide information about their current situation. **Collection inaccuracy** and **presentation inaccuracy** generally manifest in the form of lower resolution information. We simulate this by dividing the speed range $[0, v_{max}]$ into $n$ bins and reporting the middle value of the chosen bin rather than the actual value. As the number of bins increases the information resolution and accuracy increases. We call the collection inaccuracy bins $n_c$ and the presentation inaccuracy bins $n_p$.

An example of how the real values are affected by 2 error bins is presented in Figure 2. First, we identify the interval (bin) in which the real value belongs. Instead of using the real value, a value equal to the *bin center* is used. For instance, in a real world example, consider how information about average speed on roads is reflected on a traffic map. For values corresponding to the first *bin*, the roads are coloured in red and for values in the second bin, the roads are coloured green.

![Figure 2: Example of information uncertainty introduced by using 2 error bins. The precise speed value is approximated with a value from the corresponding bin (Bin1 in this case).](image)

5. Experimental Setup

The experimental setup is similar to the one described in our previous study [2]. We consider a simplified scenario using a road network as shown in
Table 1: Main parameters used in the experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Min Value</th>
<th>Max Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>percentage of sources</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>$n_c$</td>
<td>number of bins for collection inaccuracy</td>
<td>1 [bin]</td>
<td>19 [bins]</td>
</tr>
<tr>
<td>$n_p$</td>
<td>number of bins for presentation inaccuracy</td>
<td>1 [bin]</td>
<td>19 [bins]</td>
</tr>
<tr>
<td>$p$</td>
<td>percentage of informed agents</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>$L_B$</td>
<td>length of Road B</td>
<td>625 [m]</td>
<td>1250 [m]</td>
</tr>
</tbody>
</table>

Figure 3: Agents move from origin to destination. They have two route choices: $Road_A = < 11 [m/s], 19 [m/s], 500 [m] >$ and $Road_B = < 11 [m/s], 19 [m/s], L_B >$. Agents are created by a Poisson process with a mean inter arrival time of 1700 [ms]. We simulate 40 minutes (approximately 1000 agents simulated). From this amount, we consider the last 800 trips, giving a warm-up period of 10 minutes.

We simulate a constant level of congestion generated by local disturbances. To create a disturbance, a random vehicle $i$ driving on the disturbance segment of the road is chosen every 2 [s] and forced to brake ($v_i = 0 [m/s]$). The car accelerates gradually and once again reaches full speed, thus causing congestion. In Table 1 we present the main parameters of the experiments. Each experiment is repeated 10 times.

To quantify the effect of information dissemination, we define the network...
performance $T$ as the average travel time of all agents in one experiment.

$$T = \frac{1}{F_t} \sum_{i=0}^{F_t} t_i,$$

where $t_i$ is the trip duration of agent $i$, $F_t$ is the fraction of agents (last 800 agents) that complete their trip.

We define an information impact indicator to quantify the impact that each of the three types of inaccuracy produce on $T$. We consider that information is affected only by one type of inaccuracy at a time.

$$I_{LB} = \max(T_{ref}(LB) - T_{i,p}(LB)),$$

where $i \in (i_{min}, i_{max}]$, $p \in (0, 100]$. In the case of sparsity errors, collection errors and presentation errors $e$ refers to $s$, $n_c$ and $n_p$, respectively. For each $LB$ we calculate the maximum impact across all levels of informed agents and all values of $e$. $T_{ref}(LB)$ is calculated for $i = i_{min}$ and $p = 0\%$. $I_{LB}$ quantifies the maximum change on $T$ when compared to $T_{ref}$. For sparsity errors, $i_{min} = 0\%$ (no sources) and $i_{max} = 100\%$ (every vehicle is a source). For collection and presentation errors, $i_{min} = 19[\text{bins}]$ (information is error free) and $i_{max} = 1[\text{bin}]$.

The case with no noise corresponds to $19[\text{bins}]$ as the maximum speed on roads is $19[m/s]$.

6. Results

In this section, we use the metrics introduced in Section 5 to analyse the impact that different types of inaccuracy have on the traffic performance. First, we show how variation in the network topology (varying the length of the alternative road $L_B$) can impact the different types of inaccuracy. Next we explore how the different kinds of errors influence the traffic performance.

In our previous study [2], we evaluated the impact that route recommendation based on accurate information can have on the traffic. In particular, we evaluated the impact that a range of values of $L_B$ have on performance.
Here, we do a similar analysis to find the impact of introducing inaccuracy to information.

It is interesting to note that, all three types of inaccuracy produce an effect on $T$ (defined in Equation 1) for these particular values of $L_B$, as illustrated in Figure 4. For this we calculate the information impact indicator $I_{L_B}$ (defined in Equation 2). We observe that the information impact is decreasing for bigger lengths of the alternative road for all types of error. Sparsity inaccuracy produce a bigger impact on performance than presentation and collection inaccuracy. It is surprising to note that collection and presentation inaccuracy have a similar impact on the traffic situation. However, this is only natural as both these types of inaccuracy manifest in the same way i.e. the speed based on which decision is made is quantised (just to different degrees).

Next, we choose the case of $L_B = 875[m]$ to further analyse the effect of information uncertainty as this case provides a significant improvement when we vary $s$, $n_c$ and $n_p$ (as shown in Figure 4). These values are plotted in Figure 5. We discover that, in most cases, having more than 20% agents as sources produces marginal to no improvement. The only exception is when $p=100\%$ where we see the surprising effect that decreasing the inaccuracy produces a reduction in traffic performance. We refer to the former as Case A and the latter as Case B.
Figure 5: The average travel $T$ (performance defined in Equation 1) depending on $s$. It reflects the effect of sparsity inaccuracy on the traffic situation. $L_B = 875(m)$. No collection or presentation errors are considered.

(а) $F_A$ and $F_B$ depending on $s$ for $p = 40\%$. $F_A$ and $F_B$ represent the fraction of agents that select Road A or Road B.

(b) STD (standard deviation) of $F_A$ and $F_B$ depending on $s$ for $p = 100\%$. Figure 6a we notice that for $s = 0\%$ (when it is assumed that the speed on the roads is maximum) most of the traffic participants select Road A. As $s$ increases, $F_A$ decreases while $F_B$ increases.

Figure 6: Explanation of the effect of information sparsity inaccuracy on $T$ for $p = 40\%$ and $p = 100\%$.

In the previous study, we observed that the biggest effect on performance was seen for $p = 40\%$ of the drivers using information (error free in that case). Thus, to explain Case A, we choose the same scenario where $p = 40\%$. We define $F_A$ and $F_B$ as the fraction of agents that select either Road A or Road B. In Figure 6a we notice that for $s = 0\%$ (when it is assumed that the speed on the roads is maximum) most of the traffic participants select Road A. As $s$ increases,
the accuracy of the recommendations increase and more drivers are redirected to Road B; this results in improving T. As the percentage of sources increases above 20%, there is only marginal improvement in the additional information gained and as such T does not change much.

In order to explain Case B, we calculate the standard deviation (STD) of $F_A$ and $F_B$. In Figure 6b we notice that STD of $F_A$ and $F_B$ increases with an increasing number of sources. A higher STD for $F_A$ and $F_B$ is reflected in a destabilisation of the transportation system; this is due to an extensive use of information. The higher STD means that the recommendations from Dijkstra’s algorithm change more frequently for a higher level of resolution. Some informed agents are recommended to select Road B, even though this recommendation becomes invalid very soon. Nevertheless, despite receiving newer information, agents that are already on Road B are unable to change to Road A. So, there are too many agents that are stuck on the long route, resulting in a negative impact on T.

To summarise, modifying the amount of sources for data collection affects data precision and this is reflected in traffic recommendations. The recommendations determine the number of agents that select one route or the other thus influencing T. The fact that a massive number of drivers use navigation recommendations produces a destabilization of the system and a decrease in T.

Next we consider the effect of collection inaccuracy and presentation inaccuracy on T. We consider the case of $s = 100\%$ for $L_B = 875[m]$. In Figure 7 we show that, increasing precision or the number of bins $n_c$ and $n_p$ for the collected data and for the displayed information. In most of the cases, produces either a small improvement ($< 2s$) or it has no effect on T.

However for the case of $p = 100\%$ where there is a massive usage of information, we notice some counter-intuitive behaviour: for a better precision in information (less inaccuracy) T decreases. This means that when most participants have access to information, then a better precision (in both $n_c$ and $n_p$) reduces system performance (T increases). It is also interesting to observe that, for collection inaccuracy, increasing the precision beyond a certain value
Figure 7: The average travel $T$ (performance defined in Equation 1) depending on the inaccuracy or noise introduced either at collection or display, $L_B = 875(m)$, $s = 100\%$.

(i.e. $n_c > 4[\text{bins}]$) has almost no effect on the system performance. In the case of presentation inaccuracy the same effect appears only for $n_p > 10[\text{bins}]$.

The higher value of this threshold for presentation errors is because the same level of resolution or precision in information that is used by the participant is obtained for fewer numbers of bins in the case of collection inaccuracy. To understand this, consider the case of there being two bins for collection, i.e. the processing stage gives a value of either $NC_1$ or $NC_2$. The speed that is reported to and used by an informed participant is the average of this value across all participants with sensors. Thus, if there are two drivers, the value of $n_p$ for the informed driver would be three, as there are three values $NC_1$, $NC_2$ or $\frac{NC_1+NC_2}{2}$ that may be reported. Thus, a collection inaccuracy of $n_c$ translates to a much smaller presentation inaccuracy.

To explain the counter-intuitive effect of noise for the case of $p = 100\%$, we define $F_A$ and $F_B$ as the fraction of agents that select either Road A or Road B in the entire simulation. In Figure 8 we present that standard deviation ($STD$) of $F_A$ and $F_B$. We notice that $STD$ increases with increasing the number of bins. This means that, the right level of noise produces stabilization in the overall traffic situation, this having a positive effect on the overall performance.
7. Conclusions and Future Work

New advancements in Intelligent Transportation Systems and navigation devices enable commuters to access real-time traffic recommendations and at the same time provide data about their trips. Such systems are expected to play a key role in solving major traffic problems in cities [24, 25]. ITS systems process collected traffic data and provide information to drivers as navigation recommendations. Information can be affected by different levels of inaccuracy or uncertainty, this having an impact on the overall traffic performance.

In this study, we first classify the information inaccuracy present in modern transportation systems as sparsity, collection and presentation inaccuracy. We analyse how each type of inaccuracy source affects the overall performance of a transportation system. Also, we investigate how the amount of traffic participants that use inaccurate information can influence the overall performance. This reveals an interesting insight into how information dissemination strategies and smart devices should be developed.

Interestingly, our results show that in most of the cases, only a small fraction (<20%) of the traffic participants is necessary to provide data for collection in order to have the best traffic performance. For the case when there is a massive participation both as sources and consumers of information, the traffic performance actually decreases. In general, noise in the form of collection or
presentation inaccuracy decreases the traffic performance. However, when the traffic participants massively use the navigation recommendations, we show that noise can actually produce an improvement in the traffic situation. Beyond a certain limit, increased precision of information does not have a corresponding increase in traffic performance.

Our findings are relevant in the context of Intelligent Transportation Systems, where a major effort is invested in providing information with higher precision. Our study helps improving such systems by offering relevant insights on how different levels of information inaccuracy can impact the overall traffic performance. Our experiments reveal the amount of sensors or probe vehicles necessary to collect data that provides the best traffic performance. We determined the acceptable level of inaccuracy during information processing. The study on presentation inaccuracy gives a target for improving the design of information dissemination devices. In future studies, more advanced experiments dealing with information and uncertainty can be performed using realistic traffic networks and travel patterns. Also, more detailed human behaviour models may reveal the exact way in which people choose to use traffic recommendations.

8. Acknowledgements

This work was financially supported by the Singapore National Research Foundation under its Campus for Research Excellence And Technological Enterprise (CREATE) programme.

References


