AN ANALYSIS OF NAVIGATION AND COMMUNICATION ASPECTS FOR GNSS-BASED ELECTRONIC FEE COLLECTION

Rafael Toledo-Moreo*, José Santa†, Miguel A. Zamora-Izquierdo†, Benito Úbeda† and Antonio F. Gómez-Skarmeta†

*Technical University of Cartagena
30202 Cartagena, Spain
rafael.toledo@upct.es

†University of Murcia
30100 Murcia, Spain
(toledo | josesanta | mzamora | bubeda | skarmeta)@um.es

ABSTRACT

Global Navigation Satellite Systems (GNSS) appear to be today the most flexible and cost-efficient technology for deploying electronic fee collection (EFC) systems. However, the deployment of a successful GNSS-based EFC system implies a number of GNSS issues such as the availability, accuracy and reliability of the measurements, as long as some others related to possible aiding sensors, reference maps and communication aspects. This paper focuses on these aspects, providing analyses and recommendations, which can be useful for service providers and project evaluators.

INTRODUCTION

The payment methods for road usage have received a great attention during the past two decades. More recently, new advances in ICT (information and communication technologies) have encouraged researchers all around the world to develop automatic charging systems aiming at avoiding manual payments at toll plazas while enabling administrations to deploy charging schemes capable to reduce congestion and pollution. The recent application of Global Navigation Satellite Systems (GNSS) on these charging platforms can present important advances, and the research community in ITS (Intelligent Transportation Systems) is aware of this.

Although charging systems for road use have been called in many different names, the two most extended have been toll collection and road user charging (RUC), which were established considering the prime two reasons for deploying these systems (13). Firstly, toll collection was initially employed for charging the users of certain road infrastructures, with the aim of recovering the costs in construction, operation and maintenance. Many studies defend the application of this economic model to finance future road networks (20), instead of using public taxes or charging vehicle owners with a periodic fee (this is the case of Spain, for instance). On the other hand, road user charging has been the term used when the final aim of the system is not only to obtain revenue for road deployment expenses, but also to modify certain traffic behaviors in order to reduce pollution or congestion (among others) (8). The application of ICT to automate the charging process has introduced new terms, such as electronic tolling or electronic toll collection. In practice, many authors in the literature use all
these terms indistinctively.

During the past years, dedicated short-range communications (DSRC) have been a key technology to automate the charging process on roads. By means of an on-board transceiver, the vehicle is detected when passing toll points. In real deployments there are usually speed limitations, since the communication channel between the on-board unit (OBU) and the roadside unit must be maintained for a while to allow the exchange of charging data. However, DSRC-based solutions present important problems, such as the cost of deploying roadside equipments when new roads want to be included in the system (a scalability problem) and a lack of flexibility for varying the set of road objects subject to charge. In this context, GNSS is lately considered as a good alternative. Essentially, GNSS-based RUC use geographic positions to locate vehicles in charging areas or roads, and this information is sent to the operator’s back office to finally create the bill. The European Union is promoting the European Electronic Tolling Service (EETS) (7) as an interoperable system throughout Europe. This is based on a number of technologies, as it is shown in Figure 1, although three of them are essential:

- Satellite positioning, GNSS;
- Mobile communications using cellular networks (CN);
- DSRC technology, using the microwave 5.8 GHz band.

Several standardizations actions concerning electronic fee collection have been already considered by the European Commission, such as the security framework needed for an interoperable EETS, to enable trust between all stakeholders, and the definition of an examination framework for charging performing metrics.

![Figure 1. Several elements comprise a GNSS-based electronic fee collection system.](image)

Currently, some of the most important deployments of electronic RUC already use GNSS. In Switzerland, the LSVA system (also known as HVF for the English acronym of Heavy Vehicle Fee) complements a distance-based model that uses odometry and DSRC to check vehicle routes with GPS measurements. The role of GNSS in the German Toll Collect system is more remarkable, since GPS positions are used to identify road segments. Nevertheless,
other extra mechanisms are used to assure vehicle charging in places where the GPS accuracy cannot guarantee the road identification. This problem has been analyzed for a potential deployment of a GNSS-based RUC in Denmark (22), comparing the GPS performances obtained in 2003 and 2008. Although availability and accuracy problems had limited the usage of GNSS for RUC in the city of Copenhagen in 2003, more recent results showed that advances in receiver technology and updates in the GPS system made possible this application in 2008. This study supports this thesis primarily on the rise of the number of satellites in sight. In our opinion, these results must be taken with caution, since the experiments does not analyze how many satellites in view are affected by non-line-of-sight (NLOS) multipath. In The Netherlands, the plans for creating a distance-based charging system for all vehicles on all roads also consider GNSS as the potential base technology (6).

As it has been aforementioned, the accuracy of the position estimates is one of the main concerns towards the application of GNSS for RUC. It is necessary to provide a confidence level that assures that the estimate of the vehicle location is close enough to the real one with a certain high probability. This is the reason why the integrity concept is receiving a great attention in GNSS-based RUC these days (12). Additionally, the communication subsystem, crucial for EFC, has not been properly attended in the literature so far. It is important to keep in mind that payment transactions cannot be completed if charging information does not arrive to credit and control centers. This paper focuses on the aspects of these two subsystems, navigation and communication, for EFC purposes.

**GNSS-BASED ELECTRONIC FEE COLLECTION**

In GNSS-based RUC, information from the GNSS sensor is used to locate vehicles at charging places. The use of GNSS as the main positioning technology to charge users for the road usage, has several benefits related to flexibility and deployment costs:

- A minimum set of roadside units would be needed for enforcement purposes, at least.
- OBU capabilities can be as simple as reporting GPS positions to a processing center, or as complex as calculating the charge and reporting payment transactions.
- A software-based OBU allows for software updates, reducing maintenance and system upgrade costs.
- GNSS sensors are cheaper and cheaper, and its performance is increasing.
- Cellular networks, which are the main communication technology considered, have a wide coverage, more than enough data rates for RUC, and decreasing costs that are also subject to agreements with operators.

Due to the flexibility of GNSS-based RUC, multitude of approaches can be designed to charge users. As main distinction factor, GNSS-based RUC solutions can be classified according to the tariff scheme used in the system. According to the literature (4,5), three tariff schemes can be distinguished:

1) *Discrete charging:* In this case toll events are associated to the identification of road
objects subject to be charged. This group includes single object charging (bridges, tunnels, etc.), closed road charging on certain motorway segments, discrete road links charging, cordon charging, or zone presence charging.

2) Continuous charging: The tariff is calculated based on a cumulative value of time or distance. Distance-based charging and time-in-use charging are included in this group.

3) Mixed charging: A combination of aforementioned approaches is used. An example of this tariff scheme is charging for cumulative distance or time considering a different price for each road segment.

MEASURING THE PERFORMANCE OF GNSS-BASED EFC

A clear definition of the performance requirements for a road user charging system is needed for two main reasons. First of all, the industrial consortiums that apply for a deployment must be equally evaluated and the final choice must be based on the goodness of each solution according to some previously established performance needs. Secondly, the interests of users and authorities must be guaranteed.

Performance requirements must be described in such a way that any possible implementation that fulfills the needs may be under consideration and verifiable by means of field tests. Thus, requirements must be independent of the technology and internal calculations for charging. As the authors of (4) claim, the issue of the positioning errors must be addressed by the proposed system, but not directly evaluated by the third part examiner that will evaluate all the proposals.

The description of the performance requirements depends on the final charging scheme. Since it is likely that any final charging scheme is based on a combination of continuous and discrete ones, let us analyze briefly both cases here.

For a discrete charging scheme, there are only four possible cases: a correct detection (CD), a correct rejection (CR), a missed detection (MD) and a false detection (FD). Last two cases cause undercharging and overcharging respectively. Because the consequences of a MD and a FD are not the same, it is necessary to analyze these effects separately, and not by a single index of overall correct detection rate. Therefore, there must be two different performance requirements to avoid overcharges (for users) and to ensure revenues by avoiding undercharges (for authorities). Furthermore, it must be decided whether the requirements must be satisfied any time, for any trip in any scenario and under any circumstance, or it is enough if the average and some statistical parameters show that the overall errors or overcharge and undercharge are within desirable thresholds. The latter may lead to persistent errors in the bills of some users who repeatedly drive trajectories not well covered by the RUC system, due for example to bad satellite visibility conditions in the area. These special cases should be handled as exceptions, because it cannot be accepted that a system does not treat fairly every user.

Analogously, for continuous schemes two parameters are also needed to protect the interests
of both users and service providers. Inspired by the notation of the navigation community (14), some authors introduce the concepts of charging availability and charging integrity (3). Charging availability can be defined as the probability that the charging error is within a desirable error interval. This parameter protects the interest of both the user and the toll charger, since it covers positive and negative errors (overcharges and undercharges respectively). Its main mission is to provide the toll charger with a level of warranty that the user will pay for the road infrastructure usage. On the contrary, charging integrity can be defined as the probability that the error is not over an upper limit; this is, that the user is not overcharged, and its value must be more restrictive than the charging availability (this is why we claim that the main objective of charging availability is to protect the interests of the authorities).

Since the charging integrity cannot be compromised, the developers must find a way to be aware of the reliability of every charge. In case of reasonable doubt, it is preferable not to charge, rather than to charge wrongly. For this reason, some integrity indexes must be calculated to verify the certainty of the charges. If integrity indexes inform of a possibly unsafe charge and the user is finally not charged, the probability associated to charging availability becomes smaller, but not the one linked to charging integrity. On the contrary, if the user is charged wrongly, both values of probabilities become smaller and the charging availability and integrity are compromised. The tuning of the integrity indexes must be done in such a way that it satisfies the needs regarding availability and integrity. If this tuning cannot be found, the system is incapable of providing the aimed level of reliability and it must be disregarded. Although a good estimation of the integrity parameters is crucial for the developers, this aspect must neither appear in the definition of performance requirements, nor being tracked during the evaluations. It must be understood only as an internal parameter that eventually affects the charging availability and integrity.

Finally, one must bear in mind that the performance indexes coming from both discrete and continuous schemes must be transformed into a unique performance parameter, based for example on the impact of each error (discrete or continuous) on the eventual charge. This is necessary since despite the fact that the proposals coming from the industry could be based on different charging schemes, there must be a possible direct comparison for all of them and the final system must be seen as a sole charging system independent of the scheme particularities. Furthermore, the integration of continuous and discrete performance indexes turns into essential for mixed charging schemes.

### GNSS PERFORMANCE ISSUES

The main technological drawback of GNSS-based RUC is the performance of the GNSS sensor. The lack of availability of the GPS signals at places where there is no line of sight with satellites is a remarkable problem in urban canyons, tunnels or mountain roads, for instance. A research assignment demanded by the Dutch Ministry of Transport, Public Works
and Water Management (23) focuses on the accuracy and reliability of distance and position measurements by GNSS systems. The trials involved 19 vehicles during one month, and concluded that during the 13% of the traveling time there was no valid GPS position, although the overwhelming part of the unavailability was due to time to first fix (TTFF).

Highly related to this, the continuity of the GPS services is also dependent on military decisions of the US government, since GPS is not a pure-civil navigation system. Moreover, the accuracy of the position estimates, although it has been improved thanks to enhancements in the space segment and in the receiver technology, is still not fully reliable to decide whether or not a user must be charged for supposedly using a road. Although some performance problems can be compensated (satellite clock bias, signal propagation delay, etc.), others such as multipath effects in the user plane are not yet modeled and degrade the accuracy in urban canyons above all. All these problems can reduce the performance of a liability critical service such as RUC. The analysis made in (22) for GNSS positioning accuracy shows that its 95% level is 37 m. Nevertheless, this number must be taken with caution when considering RUC applications, because many other factors apart from the GPS inaccuracies themselves can affect this result, such as inaccuracies in digital maps or errors in the map-matching process.

The consequences of the positioning errors in the system performance would not be so severe if current GNSS devices would provide a fully meaningful value of the reliability of the positioning: its integrity. In this case, although the performance of the system may diminish, its integrity remains and users would be protected against overcharge. It is then up to the authorities to decide whether or not the expected performance is good enough to deploy the system, in other words, to ensure the revenue of the investment. However, current integrity values provided for GNSS devices are inappropriate.

An approximation to provide integrity in GNSS-based positioning is given by the Receiver Autonomous Integrity Monitoring (RAIM) algorithm. This technique, initially created for aerial navigation, is based on an over-determined solution to evaluate its consistency, and therefore it requires a minimum of five satellites to detect a satellite anomaly, and six or more to be able to reject it (9). Unfortunately, this cannot be assumed in usual road traffic situations, especially in cities (19). In addition, the RAIM method assumes that only one failure appears at once, something feasible in the aerial field, but not in road scenarios: it is usual that several satellite signals are affected by simultaneous multi-path propagations in an urban area.

Satellite Based Augmentation Systems (SBAS), such as EGNOS (European Geostationary Navigation Overlay Service) or WAAS (Wide Area Augmentation System), also offer integrity calculation. By means of the information about the GNSS operational state, broadcasted by GEO satellites, it is possible to compute a parameter of system integrity (17,2). However, this approach does not consider local errors such as multipath, which are of key importance in terrestrial navigation.

Due to these problems, in the last years some authors have suggested new paradigms to...
estimate the system integrity (12,21). In concrete, (12) shows an interesting approach for integrity provision based solely on GNSS that obtains interesting results. Figure 2 illustrates the solutions provided by two different approaches (14,38) (among a number of the literature) for position integrity. The red line represents the HPL (Horizontal Protection Level) estimated by using the information provided by EGNOS. HPL does not include local errors at the user plane (such as multipath) or the contribution of the aiding sensors. The green line shows the HIT (Horizontal Integrity Threshold) values along the trajectory. HIT represents the confidence on the horizontal position estimated by the filter that fuses the sensor data (this could be a particle filter or a Kalman filter, for instance). In Figure 2, HIT does not show the peaks that appear in HPL caused by bad GNSS coverage, since HIT follows errors models that consider the vehicle and the aiding sensors. In this way, although HIT does not consider EGNOS integrity information for each satellite, it usually offers a better estimation of the real performance of the navigation system, since a multi-sensor approach (which supports periods of GNSS absence) is considered.

Figure 2. Two different solutions for navigation integrity over the vehicle trajectory (in blue).

PERFORMANCE OF THE COMMUNICATION LINK

There is another part of the OBU of key importance to perform payment transactions in GNSS-based EFC: the communication subsystem. Independently of the OBU capabilities to save the vehicle route or calculate the toll fee, a communication channel is necessary to send this information on time to a processing center. The performance of this link is also relevant, since payment transactions must be committed on the back office. Among the different communication technologies used in vehicle telematics, the most
considered for GNSS-based EFC are cellular networks (CN). These are highly useful for EFC, since they offer vehicle to infrastructure communications by means of a wide deployment on road networks. Other communication technologies, such as DSRC (Dedicated Short Range Communications) or WiFi, are less suitable, since they are against the flexibility principle of GNSS road-pricing schemes, but they can be considered as complementary technologies.

A set of communication performance parameters must be assessed in a GNSS-based EFC system, in order to include the necessary OBU capabilities to cache/calculate charging information or choose a telecom operator. Main performance issues of cellular networks include (16) network coverage, network capacity, mobility conditions or vehicle speed.

Network availability is one of the main drawbacks of CN systems. Telecom operators do not offer the same service over the terrestrial surface. In urban environments, the CN coverage is excellent, due to the number of base stations where the mobile terminal can connect. In rural areas, however, the CN deployment is poor, or even null in some places. In CN connections, it is also important to differentiate between two important concepts regarding network availability: coverage and capacity. Coverage can be understood as the possibility of the mobile terminal to use the network, because in this exact location, an operator has deployed (or not) the necessary infrastructure. However, even under good coverage conditions, the user can be rejected to establish a call or a data connection if the capacity of the network has been exceeded. Depending on several technological issues, such as modulation, frequency allocation and time slot scheduling, this effect has a different behavior. This way, the number of users who are concurrently using the network also restricts the CN availability. Since capacity or coverage problems can appear suddenly, the OBU must be ready to cache all charging information, which will be sent as soon as a data link is available again.

Apart from potential access to the network, some problems arise due to mobility effects. Handoffs between base stations are also important, due to potential decrease of performance in the process. If the mobile terminal is moving at locations far away from the base station without performing a handoff, poor latency and throughput results can be obtained (1). However, this effect is not remarkable in EFC environments these days, since most of the charged road networks are highways, where handoffs are less frequent. Nevertheless, the distance between two physical edges in the communication is not the only noticeable effect of mobility. Interference with other radio equipments and especially bad orography conditions could also cause communication problems in CN systems.

Finally, the speed is also a noticeable issue in cellular networks (11). At the physical level, effects such as Doppler shift, Rayleigh fading and multipath propagation limit the bit rate allowable in CN at high speed. At link level, handoff issues must also be considered. Because the handoff process takes time (depending on the type), vehicles at high speed could have problems in places where the base station density is too high. For this reason, a high rate of base stations at highways is favorable, apart from the obvious reduction in deployment costs.

The tests shown in Figure 3 illustrate some of the previous communication issues using a
UMTS (Universal Mobile Telecommunications System) link. The on-board system considered in the field trials (15) uses the cellular network to send periodical messages to a server connected to Internet at the infrastructure side. The first graph shows the delay values for each message sent from the OBU. At a first glance, it is noticeable how mobility conditions cause continuous delay peaks.

Figure 3. An evaluation of a vehicle-to-infrastructure communication test using the UMTS cellular network.

There are three main problematic areas in the results shown in Figure 3, which are noticeable in the groups of delay peaks observed in the first graph. The first one, between times 150 s and 200 s, comprises a road stretch where the car drives near a parking area covered with a metallic roof. Since radio signals are partially reflected, the network performance decreases. The second problematic area is more evident, between times 540 s and 700 s. The vehicle reaches the farthest position from the CN base station used for the connection and several buildings also block the line of sight with it, provoking a severe coverage problem. After leaving this area, the vehicle comes back towards the base station, where the network performance is good again. However, just before reaching the base station, the vehicle turns and goes across a third problematic zone between times 900 s and 950 s. At this location, a small hill between the vehicle and the base station decreases the channel quality. The graph
that illustrates the cumulative distribution function (CDF) of the delay results shows that values between 180 and 240 ms comprise more than 90% of the messages. The rest of latency values are distributed in a quasi-logarithmic trend, since high latencies are less and less common. The last graph in Figure 3 clearly illustrates this distribution of values, showing a histogram plot of the latency results with containers that are situated in a logarithmic scale on the axis of abscissas.

As it has been checked, mobile conditions in real scenarios cause performance variations when a cellular link is used to communicate with a node located in the wired Internet. Due to this, a good EFC system must be capable of caching a good amount of charging information to commit payment transactions when the network is not available. Nonetheless, it is important to realize the good performance of new UMTS networks (a mean delay of 240 ms in the previous test) in places where telecom operators have deployed the necessary infrastructure. This can encourage researchers and engineers to consider this communication technology for telematic services such as road user charging in the next years.

**CONCLUSIONS**

The deployment of GNSS-based road pricing schemes needs to consider the impact of the different technologies involved in such a great system. The performances of the navigation and communication subsystems are found essential for a correct operation of a GNSS-based EFC. This analysis is considered as a first step prior to the assessment of high-level performance measurements, as charging reliability.

For today, it seems really difficult that any solution exclusively based on GNSS for positioning can ensure the high standards for charging reliability for road user charging. For this reason, it is the authors’ opinion that GNSS technology must be supported by aiding information coming from onboard sensors and even digital maps. Moreover, the calculation of integrity factors that indicate the position goodness at any time is considered a key need to measure and track the performance of the navigation system in liability critical services such as road charging, independently of the navigation system used.

Cellular networks have been identified as a good communication technology, due to their wide deployment and recent advances. However, several performance issues in real environments indicate that on-board units must be prepared to overcome eventual communication problems with backend systems. Among all these issues, the coverage and the capacity of the network in road segments are identified as the most important ones. In fact, these are being considered by telecom operators nowadays, installing more base stations or making the most of the available frequency spectrum. Only a reliable communication link can assure that payment transactions are committed at the back office of an electronic RUC system.
ACKNOWLEDGEMENTS

This work has been sponsored by the Spanish Ministry of Transportation through the SATELITES Project (FOM/2454/2007), the Ministry of Education and Science through the SEISCIENTOS Project (TIN2008-06441-C02), the Ministry of Industry through the TIMI Project (2007/CENIT/6731) and the Seneca Foundation, by means of the Excellence Researching Group Program (04552/GERM/06). In addition, the authors would like to thank the colleagues of the GNSS Metering Association for Road User Charging (GMAR) and the rest of advisors for the fruitful discussions.

REFERENCES


12