Transportation Planning and Technology, December 2007
Vol. 30, No. 6, pp. 627–642

ARTICLE

Detection Algorithms of Intentional Car Following on Smart Networks: A Primary Methodology

BANIHAN GUNAY

School of the Built Environment, University of Ulster, UK

(Received 3 April 2006; Revised 9 July 2007; In final form 21 September 2007)

ABSTRACT This paper explores the possibility of detecting certain movements of vehicles that might provide useful information for crime investigations. It is known that existing car following models are interested in microscopic interactions between vehicles in randomly formed pairs. The present work, however, introduces the concept of macroscopic analysis of vehicle positions on a network and the idea of seeking if these movements exhibit any meaningful relationships. First of all detection algorithms are produced for two possible types of detection: (a) was a particular vehicle followed by any vehicle? and (b) did a particular vehicle follow any vehicle? These algorithms assume that every link in the network is equipped with some sort of vehicle identification or tracking device and the identities of all vehicles, such as their number plates, are fed into the program. Then a simulation program is developed to implement the first algorithm (Type (a)), as an example, to visualise the concept. Since the present paper is a preliminary and basic approach to the problem, a number of issues and details requiring further research, together with the directions which could be taken, are also identified and discussed.

KEY WORDS: Automatic number plate recognition; detection algorithms; IntCarFol; crime investigation

Correpondence Address: School of the Built Environment, University of Ulster, Jordanstown Campus, Newtownabbey Co. Antrim BT37 0QB, UK. Email: b.gunay@ulster.ac.uk

ISSN 0308-1060 print: ISSN 1029-0354 online © 2007 Taylor & Francis
DOI: 10.1080/03081060701698268

Introduction

The concept of car following in traffic engineering deals with the analysis of how one vehicle follows the one in front. Basically these theories are based on microscopic modelling of two consecutive vehicles moving in the same lane without overtaking. The speed of the following vehicle and hence its longitudinal location is determined by the stimulus caused by its leader travelling in front. When the leader accelerates or decelerates the follower, too, changes their speed to adjust the safe following distance. This adjustment takes place after a time delay which is a function of the reaction time of the following driver and the vehicle's accelerating/decelerating capability. The most well-known car following models are the GM, Forbes, Pipes and Gipps models that are reviewed in Holland (1998), Brackstone and Mc Donald (1999) and Gunay (2007). This type of following is random in a sense that the follower–leader pairs are formed randomly without any particular intention. Any vehicle that happens to be travelling in front of the follower becomes its leader. Also, the existing car following models are interested in microscopic interactions between vehicles travelling along a link. Therefore in simulation studies, after every overtaking or lane changing and after every junction, new leader–follower pairs are introduced and all the parameters are updated accordingly irrespective of vehicle identities. The present paper, however, introduces a different type of following, which needs to be distinguished from the former. We use the term 'intentional car following' for the latter, which deals with the order that the individual vehicles travel rather than the longitudinal safe stopping distance between them. Therefore, intentional car following will be treated in a more macroscopic fashion. Whereas to know the order in which vehicles travel may have no importance from the vehicle dynamics point of view of traffic engineering, it may be a vital piece of information for crime investigations, and existing (and developing) vehicle identification systems have the potential to offer this piece of information.

Vehicle tracking technologies, such as automatic number plate recognition (ANPR), have been advancing rapidly. In the near future it will be possible to track nearly all vehicles on a network by means of various methods available. Because of the small percentage of vehicles which are fitted with on-board vehicle identifications devices, ANPR currently continues to dominate the available vehicle tracking technologies. ANPR is a mass surveillance method that uses optical character recognition on images to read the licence plates on vehicles. Systems can now scan number plates at around one per second on cars travelling up to 160 km/h (Wikipedia, 2007). Over the past few years, vendors,

technologies and even number plate designs have evolved considerably. Various companies and governmental organisations have been using ANPR in various areas, such as speed cameras, car parks, gated communities, airport security, traffic monitoring, and CCTV systems. Research has mainly concentrated on the accuracy of the recognition and the technological improvements of these systems (c.f. Nelson, 2005). Hanif *et al.* (2002) looked at developments of algorithms for optimally locating surveillance technologies with an emphasis on ANPR readers by maximising the benefit that would accrue from measuring travel times on a transportation network. Dixon and Rilett (2005) used information gathered from automatic vehicle identification systems to help estimate short-term trip origin–destination (OD) matrices in an urban environment. They proposed a method for using sample link choice proportions and sample OD matrix information derived from AVI data sampled from a portion of vehicles to estimate population OD matrices with these data collection points acting as the origins and destinations. Based on a comparison of OD estimation accuracies, with and without these data, they found that using vehicle identification data in OD estimation improves significantly the accuracy of the OD matrix estimates. When these technologies become widely available in practice, crime investigation services could benefit more from these systems provided they have certain tools. This paper introduces a new application area of these systems by developing detection algorithms to identify those vehicles that may be in a suspicious following act on a road network. Consider an investigation where the police want to find out whether a particular vehicle was intentionally followed by a vehicle, or whether a vehicle intentionally followed any other vehicle. Even approximate answers to these questions, leading to resolving a crime, could be an invaluable clue.

Before considering the detailed discussions, characteristics of intentional car following will be defined as follows: In order for a vehicle to follow another one intentionally over a period of time and over a number of links on a network, this following act should possess certain characteristics:

1. The travelled path from an origin to a destination will be the same.
2. The follower will not pass the leading vehicle, even if there are opportunities.
3. If the leader stops for a reason, including traffic lights, the follower will stop too.
4. If the leader follows a non-shortest route, the follower too will take that route.
5. If the gap between the leader and the follower widens, the follower will speed up to close it.

6. For most of the time, both vehicles should be on the same link or the follower should be on the previous link (depending on link length) so that the leader can be seen. If, however, there is a middle link between the leader's and the follower's current links it will be impossible to see and follow the leader.
7. When the leader leaves a link by making a turn at a junction, where both left and right turns are possible, in order for the follower to see the leader leave, the follower must be on the link which leads to that particular junction.
8. The distance between the leader and the follower will be sufficiently short, i.e. to be on the same link is not sufficient for long links.

Stochastic Nature of the Phenomenon

Consider a small network shown in Figure 1, where the letters A to L refer to the origins and destinations. Assuming that all the links have some sort of vehicle recognition devices which are all integrated with each other, let us consider two vehicles leaving Origin A, travelling to Destination G, and call these vehicles *n*, the front vehicle, and *m*, the following vehicle. We also make the following assumptions: (a) all destinations have the same attraction values; (b) overtaking is not possible on any link; (c) all links in the network have the same flow levels, meaning same link travel times, and this piece of information is known by drivers; and (d) both vehicles always receive the same green signal at traffic lights.

Let the path shown in bold be the travelled route of both vehicles, and we know that the vehicles arrive at all the links along this route in

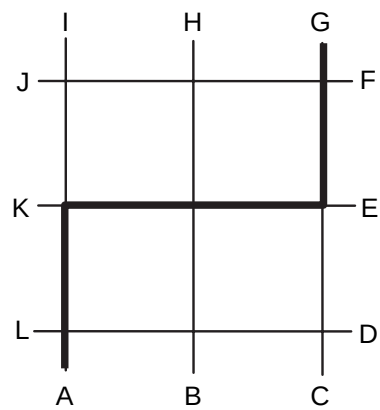

Figure 1. A small network and the travel path of the subject vehicle and its follower

the same order, i.e. vehicle *n* first then vehicle *m*. Then, let the question be to investigate whether: (i) both vehicles took the same route to reach G by pure coincidence; or (ii) the destination of vehicle *n* was G but vehicle *m* followed the front vehicle intentionally to the same destination.

The probability that the front vehicle, *n*, was not intentionally followed by vehicle, *m*, is the same as the probability that both vehicles happen to travel to the same destination by following exactly the same route by coincidence. Mathematically speaking, in general terms, this probability is equal to the products of the chance of two independent events taking place: (a) travelling to the same destination; and (b) following the same path. In more specific terms, for our example, the total number of destinations is 11, i.e. 12–1, since A is the origin for both vehicles and cannot be a destination (remember that all the destinations have the same attraction values). And the number of all possible shortest routes from A to G giving the same journey time is 6 (remember that all the links are the same length). Hence, the chance of this coincidence in question is

$$\frac{1}{11} \times \frac{1}{6} = \frac{1}{66}$$

There are certain (major) factors affecting the magnitude of this probability:

i. Attraction values of destinations: If other destinations have lower attractions than G, the probability that the first and the second vehicles will both go to G by coincidence will increase. If *vice versa*, the chance will be smaller.
ii. Heterogeneous link travel times: Variations in link travel times result in variations in the number of alternative shortest routes between an origin and a destination. Hence, if there are too many (same length) alternatives, but the vehicles *n* and *m* still take the same route, then the probability of being in an unintentional following is small. If *vice versa*, i.e. few alternatives and both vehicles take the same route, the chance in question will be high.
iii. Overtaking opportunities: Single lane roads and heavy flows reduce overtaking opportunities, hence increasing the chance in question of unintentional following. However, if there are multi-lane roads with light flow rates, but vehicle *m* remains behind vehicle *n* all the time, then a higher possibility of intentional following is implied.
iv. Variation in the desired speeds of drivers: In connection with the above factor, if the desired speed difference between the leading and the following drivers is big and overtaking is allowed, but vehicle *m* remains behind vehicle *n* at all times, then the possibility of being in

an unintentional following act will be low. If *vice versa*, the chance in question will be high.

v. Driver information about the network and traffic conditions: When current network and traffic conditions are known by drivers there will be less random route choices, meaning that if two vehicles travel together by making wise decisions in terms of choosing the shortest route, the possibility of being in an intentional following act will be small. However, on the other hand, if the two drivers have no or little information about the current network and traffic conditions, but still travel together throughout their entire journey, then the chance of being in an intentional following will be high.

vi. Network size: The bigger the network, the higher the number of alternatives/variations, and hence, the more random vehicle movement expectations. In other words, in a large network, if the two vehicles (n and m) travel together following the same path, the likelihood of a coincidence will be small. In small networks, however, it will be quite likely that two vehicles travelling together in the origin link, they will still be in a similar following situation a few links later (m may not be immediately behind n, but still a follower). In short, generally speaking, for small networks it is more difficult to distinguish random following and intentional following, whereas the phenomenon becomes stochastic more easily in large networks.

vii. Junction operations: Each junction is a breaking point for an intentional following journey, except grade separated intersections. At priority and roundabout junctions, the follower, m, may not be able to use the same gap that n has accepted. Similarly at signalised junctions both vehicles may not receive the same green. Depending on how far the following vehicle is forced to lag behind the leader by these junctions, generally speaking, the more the number of junctions in a network, the lower the possibility of joint movements of n and m. Therefore if both vehicles pass through many junctions but still arrive at each link more or less together, then the probability of an unintentional following will be small.

The above discussions were based on the assumption that every link is a smart link. Networks can be classified into two types: those that are assumed to be fully covered with vehicle recognition systems and those that are partially covered with these systems. In other words, the former assumes that every link in the network has some sort of vehicle (or number plate) recognition device, meaning that the path of every individual vehicle is known without any missing link. In the latter, the links equipped with these systems will be called 'a smart link' and the ones without any recognition camera, etc. will be called 'a blind link' throughout the paper. Since these technologies

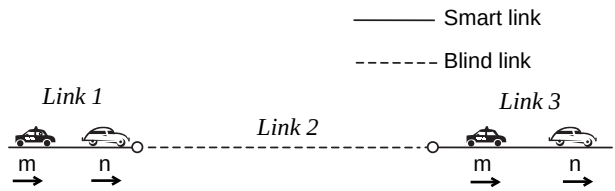

Figure 2. Serial links

are still in the developing stage, a more realistic approach should be the analysis of networks consisting of both smart as well as blind links. But since this will be left for future work, only a brief description of partial coverage will be given here. To start from the simplest situation, consider the network shown in Figure 2, where *Links 1* and *3* are smart links (i.e. they have vehicle recognition devices), and *Link 2* is a blind link (i.e. it does not have such a device). If the blind link is a single lane road with no overtaking allowed, the (*m, n*) vehicle pair in *Link 1* will appear definitely in the same order in *Link 3*. Therefore it will be impossible to suggest that *n* was intentionally followed by *m*. However, if *Link 2* is a two lane road, overtaking and lane changing will be possible especially at light and moderate flows. The possibility of vehicle *n* being overtaken by vehicle *m* in *Link 2* normally depends on the desired speeds of the drivers of these vehicles and the link lengths. Hence the order in which the vehicles will appear in *Link 3* now possesses a stochastic nature, depending on the difference between the desired speeds of the drivers and the length of *Link 2*.

Let us consider Figure 3, where the stochastic nature of the problem discussed above gains a new dimension. At *Node 1* the drivers have a choice of taking *Link 2* or *3* to reach *Link 4*. This choice normally depends on the link travel times of these two alternatives, i.e. the shorter the more attractive. This is, of course, based on the assumption that the drivers have reasonable information about

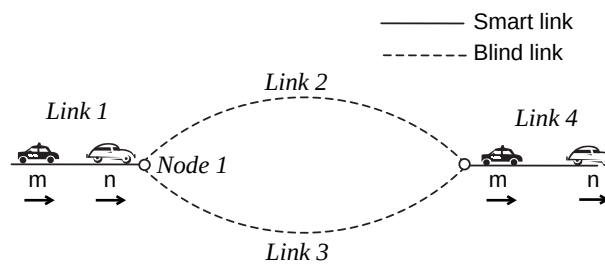

Figure 3. Parallel links

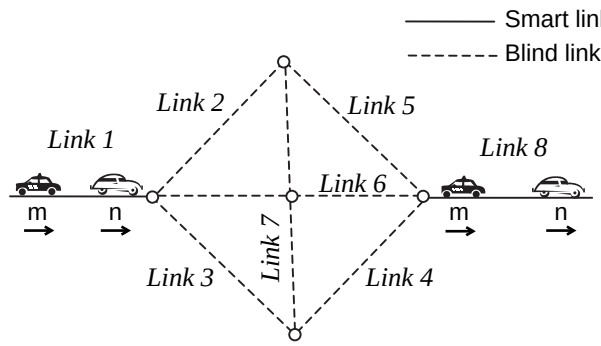

Figure 4. Multiple blind link alternatives

the flow (congestion) levels on *Links 2* and *3*, before turning left or right at Junction (node) 1. Compared to the previous case (Figure 2), given everything else remains the same, the probability that vehicle *m* will be behind vehicle *n* in *Link 4* will be less. It should be remembered that in order to be in an intentional following act, vehicle *m* does not have to be immediately behind vehicle *n*.

With increasing number of decision nodes (junctions) in the network, the arrival order of these vehicles at the next smart link will be affected (Figure 4). Thus, it can be stated that the larger the network (in terms of both the number and the length of links), the more random the phenomena (alternative paths and OD patterns) would be. This means that analytic approaches can no longer be used to deal with the stochastic nature of the problem.

Careful and detailed work is needed for the analysis of the above factors. This requires huge amounts of empirical data to be collected and modelled to understand better the roles that each of these factors play, and this will be the subject of ongoing research. The next section, however, provides algorithms for two basic types of following scenarios and introduces a piece of software.

Detection Algorithms

There are three possible types of intentional following. These are (a) was a particular vehicle followed by any vehicle? (b) did a particular vehicle follow any vehicle? and (c) was there any pair in an intentional following act? Figure 5 shows the flow chart of Type (a), where the following abbreviations are used. This algorithm assumes that the intentional follower, if any, starts this following act on link 'a'.

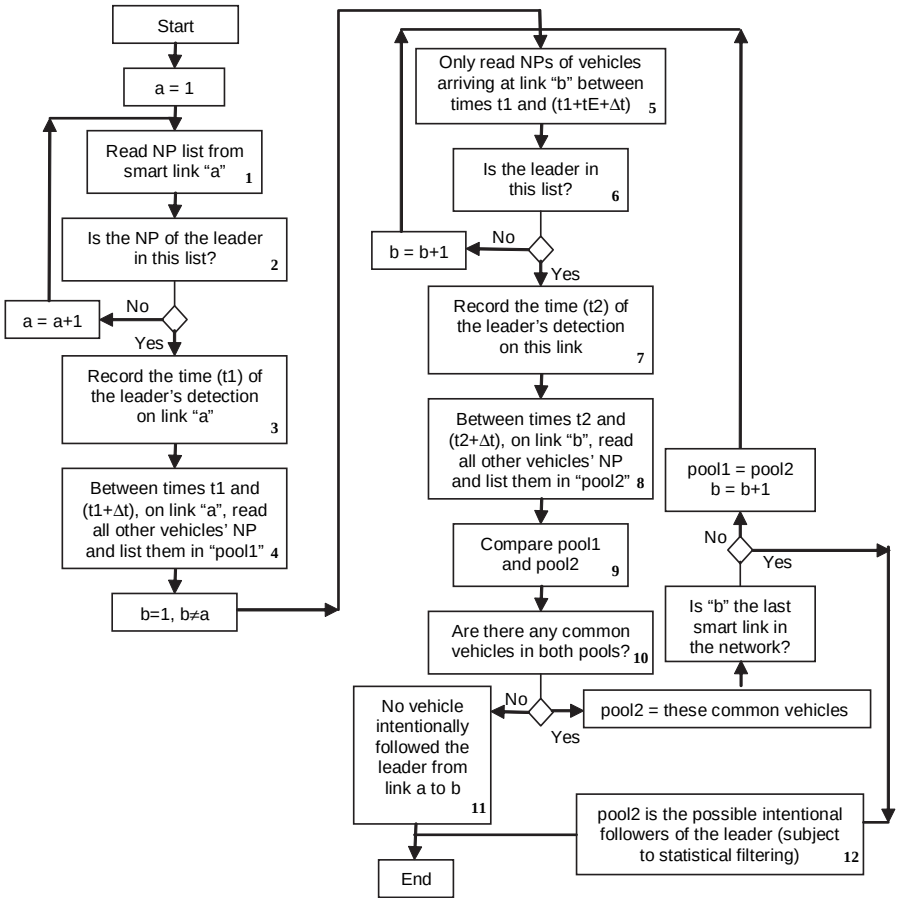

Figure 5. Flow chart of the detection of Type 1 following

NP = number plate;
Δt = max time span for the intentional follower to lag behind the leader but still able to see the leader;
tE = expected travel time from link 'a' to link 'b' or from link 'b_n' to 'b_{n+1}';
SV = subject vehicle.

In Type (a), the identity of one of the two vehicles (i.e. the leader) is known. Whenever this subject vehicle is seen by the system, the detection process is triggered (Box 3 of the flow chart). Then a number of vehicles behind the subject vehicle are monitored (Box 4). These vehicles arrive at a link after the arrival of the subject vehicle at t1. The

time ($t1+\Delta t$) is the cut-off time for recording the number plate of these potential intentional followers. In this paper, we are not interested in the determination of Δt but its magnitude can be up to 10–20 seconds. Boxes 5 and 6 are for finding the next smart link visited by the subject vehicle (i.e. leader). This is done by checking every link, other than 'a', which is within the proximity of link 'a'. In order to reduce the computing time for this checking process, the scan time is limited by ($t1+tE+\Delta t$), where tE is the expected travel time between link 'a' and this link, 'b', or between the two links that are visited consecutively. In order to ensure this, tE must be long enough to cover the arrival of very slow moving vehicles at link 'b'. In Box 8, similar to Box 4, a second group of potential followers are recorded for link 'b'. These two groups (pools) are compared to eliminate those vehicles that are not detected on link 'a' during ($t1+\Delta t$). If this comparison gives no common vehicle, the algorithm is ended, stating that this subject vehicle was not intentionally followed by a vehicle from 'a' to 'b'. If, however, there are common vehicles, the algorithm moves to the next smart link to carry out the same procedures until all links are visited. Box 12 lists these potential intentional followers. It will then be up to the user of the tool to filter this list to distinguish the less and more likely followers (refer to section 'Stochastic Nature of the Phenomenon').

In Type (b), the identity of the potential follower (i.e. the subject vehicle) is known. Whenever this subject vehicle is seen by the system, the detection process is triggered (Box 3 in the flow chart in Figure 6). Then a number of vehicles in front of the subject vehicle are monitored (Box 4). These vehicles arrive at a link before the arrival of the subject vehicle at t1. The time ($t1-\Delta t$) is the cut-off time for recording the number plate of these potential leaders. Similar to Type (a), above, we are not interested in the determination of Δt, but its magnitude can be up to 10–20 seconds. Boxes 5 and 6 are for finding the next smart link visited by the subject vehicle. This is done by checking every link, other than 'a', which is within the proximity of link 'a'. In order to reduce the computing time for this checking process, the scan time is again limited by ($t1+tE-\Delta t$), where tE is the expected travel time between the two links consecutively visited. In Box 8, similar to Box 4, a second group of potential leaders are recorded for link 'b'. These two groups (pools) are compared to eliminate those vehicles that are not detected on link 'a' during ($t1-\Delta t$). If this comparison gives no common vehicle, the algorithm is ended, stating that this subject vehicle did not intentionally follow any vehicle from 'a' to 'b'. If, however, there are common vehicles, the algorithm moves to the next smart link to carry out the same procedures until all smart links are visited. Box 12 lists these potential leaders. It will then be up to the user of the tool to filter this list to distinguish the less and more likely followers.

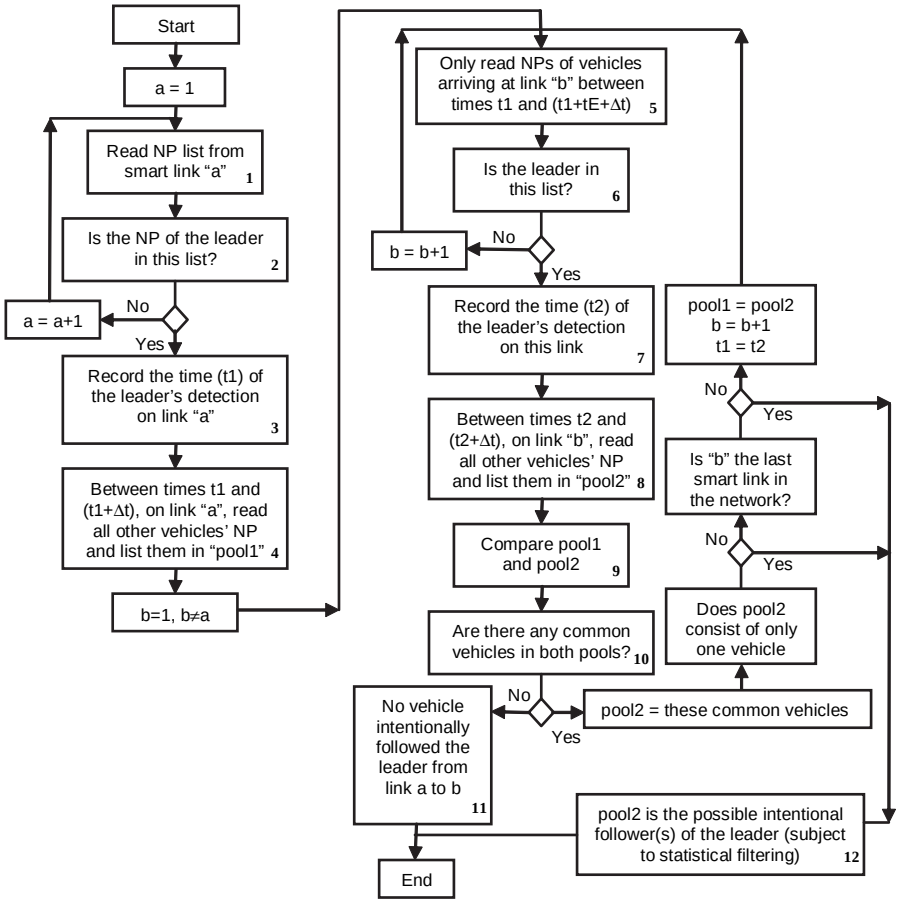

Figure 6. Flow chart of the detection of Type 2 following

In Type (c), both vehicles are unknown, hence all the vehicles in the entire population on the network have to be monitored. This is the most difficult detection type as it requires huge amount of resources of time and computing capacity, hence it will not be treated in the present paper.

In small networks, these algorithms can be implemented manually. For example, the user can look at the lists of number plates obtained from a few points on the network and check to see if there is any suspicious following activity. However, in large networks, with many ANPR points, a piece of software can quickly identify these potential followers. Therefore we developed a prototype package called Int CarFol, first introduced in the present paper. The program is based on a deterministic approach and does not simulate the stochastic nature of the phenomenon discussed in section 'Stochastic Nature of the

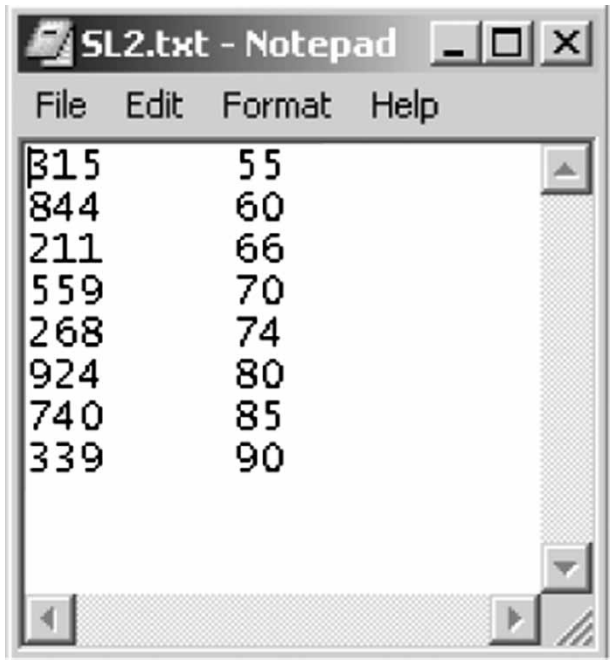

Figure 7. A text file representing vehicle identities and their arrival times on a smart link

Phenomenon'. But further improvements will include this dimension as well. The software is able to read up to 10 smart link files. These files are called SL1.txt to SL10.txt, and stored in the same folder as the IntCarFol.exe file. Each file consists of two sets of artificial data: Column 1 is an array of integers representing number plate readings of vehicles and Column 2 is the time of their arrival (in seconds) at the point where the ANPR camera is assumed to be present. Figure 7 shows an example where identities and arrival times of eight vehicles are obtained from a link.

To run the software, the user first chooses the number of smart links where the ANPR data are obtained. Then the number of vehicles in the longest SL file is entered. This is the number of lines in the longest of these .txt files. Then the uniform link lengths in metres are chosen and the user enters the number plate of the subject vehicle that might be intentionally followed by another vehicle. When all the Smart Link buttons are pressed the table displays the number plate of each car in the order of their arrival in the first column of each Smart Link and the time of arrival in the second column. Finally, an average speed for the traffic on the network is chosen. When the Run button is

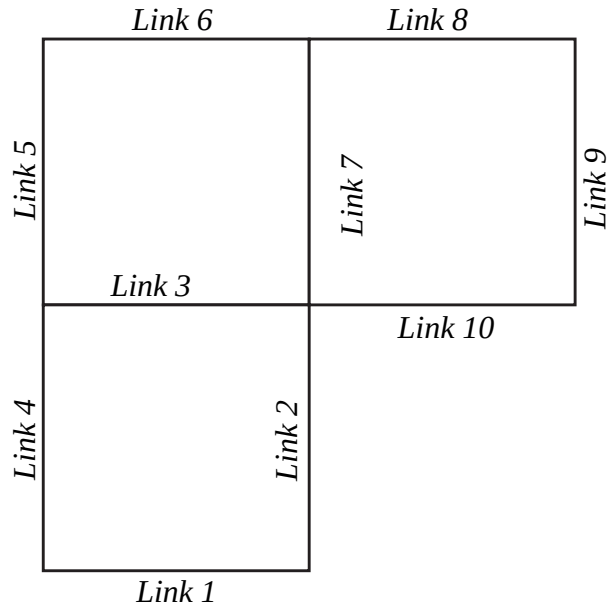

Figure 8. A small network

clicked, the program does the required calculations outlined in the flow chart given in Figure 5. As an output, the program lists the links that are visited by the subject vehicle and the IDs of all the other vehicles that were a reasonable distance behind the subject vehicle along this path. The current version of the program makes the following assumptions: (i) the subject vehicle is first seen in Smart Link 1; (ii) all the links are in the same length; (iii) there are maximum of 10 smart links in the network; (iv) all links in the network are smart (full coverage); and (v) the software works off-line (handling real time will be subject to future research).

The use of the software can be illustrated through a simple example. Consider the network shown in Figure 8 and let us assume that there are 16 vehicles captured by the ANPR on various links at various times. In order to allow manual checking of the results, a small vehicle population was chosen. The program, however, can handle many more vehicles with ease. The number plates of these vehicles and the time of their arrival at these links are shown in Table 1.

Let 'Vehicle 315' be the subject vehicle. It is clear from the table that this vehicle has travelled along Links 1, 2, 7, 8 and 9. We want to find out if this vehicle was followed by any other vehicle along this path. First of all, for each link the number plates of each vehicle and their arrival times

Table 1. Arrival times of artificially generated vehicles at certain links, in seconds

Number plates of vehicles	Link 1	Link 2	Link 3	Link 4	Link 5	Link 6	Link 7	Link 8	Link 9	Link 10
315	5	55					100	160	198	
844	9	60					105	164	205	
211	12	66					108	170	191	
559	20	70					111	178	209	244
268	25	74					125	189		
924	33	80	111	170						
740	35	85					120			
339	41	90			199	166	131			
222			252		231	160				269
333			119	94						140
444			130	100			187	220	257	
555			135	103				230	215	159
666			138		170	212		233	254	
777			100		133	163				75
888				119	141	179				
999			144						195	171

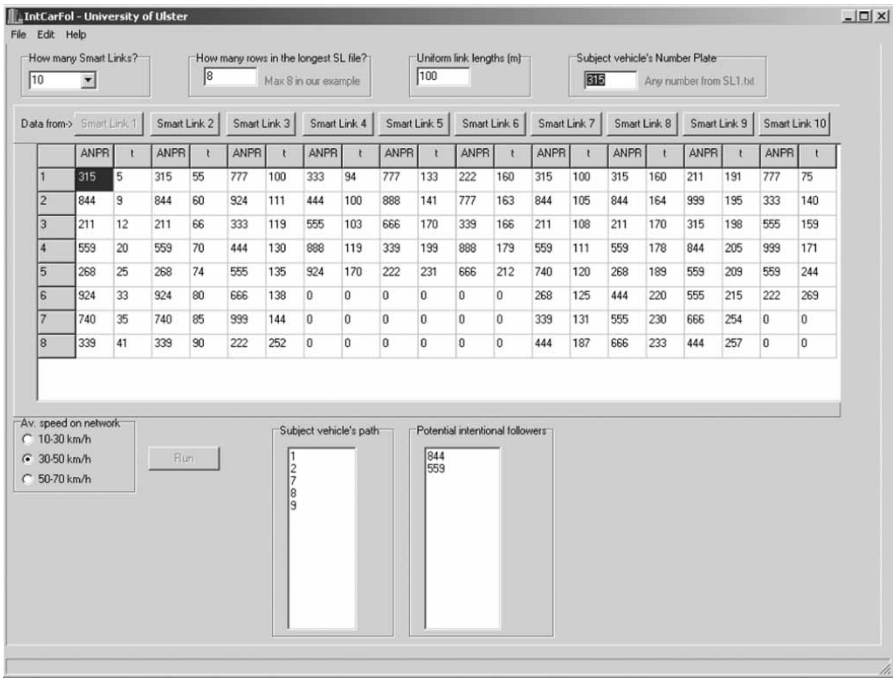

Figure 9. A snapshot of the IntCarFol software

are saved in text files. One of these files is shown in Figure 7, as an example. According to the description given above, when the program is run, the outcome is shown in the two bottom boxes (Figure 9). Thus, according to the program, the subject vehicle took path 1, 2, 7, 8 and 9, and was followed by vehicles 844 and 559.

Concluding Remarks and Research Directions

This paper has introduced the issue of intentional vehicle following. The stochastic nature of the factors affecting the variability of the phenomenon have been discussed. Especially when large amounts of data need to be handled, filtering the data before storage becomes vital to be able to eliminate unintentional following, so that only the identities of likely intentional followers are stored. Three types of detections were presented and two of them were detailed by providing algorithms. Furthermore, for illustration, trial software was developed to mimic Type 1 detection.

The eight defining characteristics of intentional following, summarised in the first section, are critically important in the design of such algorithms. Although it is possible to refine these items, the present

paper has established the fundamentals of the idea. If, however, the existence of such a system in a particular area is known to the public, the system can be tricked. For example, the intentional follower may use different number plates each time or overtake the leader temporarily. If it is not made known to the public, or at least the presence of the cameras is disguised, the efficiencies will improve. But of course, to have such a system in place is better than not having it. Another limitation, regarding the software developed, is that at the moment it can handle only a small number of files (assumed to represent smart links) in an off-line fashion. Besides, for simplicity, the program did not take into account the stochastic effects of some of the factors mentioned in the second section. Further improvements will be able to deal with more links for various detection types, ideally on-line in real-time, with more factors involved. Valuable databases to serve interested authorities will then be achieved especially if stored continually in real-time.

A further contribution of the paper is related to the existing registration plate reading technologies, which are reasonably good, but due to problems like poor visibility, dirt on the number plate and similar looking letters (such as *D* and *O*, *S* and *5*), image processing algorithms sometimes give errors. However, if the readings are handled in pairs of vehicles, as introduced in the paper, the next camera downstream can be used to verify the reading of the previous camera, leading to improved accuracies in recognition.

References

Brackstone, M. & McDonald, M. (1999) Car-following: a historical review, *Transportation Research Part F*, 2(4), pp. 181–196.

Dixon, M. P. & Rilett, L. R. (2005) Population origin–destination estimation using automatic vehicle identification and volume data, *Journal of Transportation Engineering*, 131(2), pp. 75–82.

Gunay, B. (2007) Car following theory with lateral discomfort, *Transportation Research Part B*, 41(7), pp. 722–735.

Hanif, S. D., Desai, J., Rakha, H. & El-Shawarby, I. (2002) *A Discrete Optimization Approach for Locating Automatic Vehicle Identification Readers for the Provision of Roadway Travel Times* Technical Report, November, Virginia Tech.

Holland, E. N. (1998) A generalised stability criterion for motorway traffic, *Transportation Research-B*, 32(2), pp. 141–154.

Nelson, L. J. (2005) *Licence and Registration ANPR Focus Part one*. Traffic Technology International, Oct/Nov.

Wikipedia (2007) Automatic number plate recognition. Available at: http://en.wikipedia.org/wiki/ANPR (accessed 20 September 2007).