Mitigating Catastrophic Failure at Intersections of Autonomous Vehicles

(Short Paper)

Kurt Dresner and Peter Stone
University of Texas at Austin
Department of Computer Sciences
Austin, TX 78712 USA
{kdresner,pstone}@cs.utexas.edu

ABSTRACT

Fully autonomous vehicles promise enormous gains in safety, efficiency, and economy. Before such gains can be realized, safety and reliability concerns must be addressed. We have previously introduced a system for managing such vehicles at intersections that is capable of handling more vehicles and causing fewer delays than traffic lights and stop signs [2]. While the system is safe under normal operating conditions, we have not discussed the possibility or implications of unforeseen mechanical failures. Because the system orchestrates such precarious "close calls" the tolerance for such errors is small.

In this paper, we introduce safety features of the system designed to deal with these types of failures, and perform a basic failure mode analysis, demonstrating that without these features, the system is unsuitable for deployment due to a propensity for catastrophic failure modes.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Miscellaneous

Keywords

multiagent systems, intelligent transportation systems, intersection control, autonomous vehicles

1. INTRODUCTION

Fully autonomous vehicles promise enormous gains in safety, efficiency, and economy for transportation. By eliminating driver error, some estimates suggest as much as 96% of all automobile accidents can be prevented [4]. Even if each accident were substantially worse, autonomous vehicles would effect an overall improvement in safety.

Traffic intersections are a compelling problem for multiagent systems. Often sources of great frustration, intersections are a sensitive point of failure as well as a major bottleneck in automobile travel. While fully autonomous open-road driving was demonstrated over ten years ago, events such as the DARPA Urban Challenge prove that city driving, including intersections, still pose substantial difficulty to AI and intelligent transportation systems (ITS) researchers.

In previous work, we proposed a reservation-based multiagent framework for managing vehicles at intersections, including both human-driven vehicles and fully autonomous vehicles [2]. Instead of using traffic lights, autonomous vehicles "call ahead" to arbiter agents stationed at intersections and reserve the space-time to pass. When a vehicle obtains a reservation, it can proceed through the intersection without stopping. By coordinating the actions of many such vehicles, the system dramatically decreases time spent stopped or slowing. However, this increased efficiency is precarious. The system orchestrates "close calls", with vehicles missing each other by small (but adjustable) margins[1]. Figure 1 shows a screenshot from our project website depicting a particularly busy intersection.

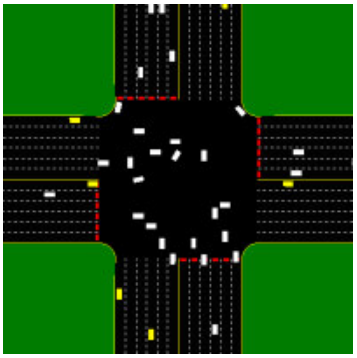


Figure 1: A screenshot from our project website showing a busy intersection with a lot of "close calls."

While the system is safe in the face of communication failures, we have not previously addressed mechanical failures or "freak" accidents. In a world without vehicle malfunctions, this would be little cause for concern. However, one can easily imagine an otherwise ordinary problem, such as a flat tire or a slippery patch of road, quickly becoming a nightmare.

Even though the vast majority of automobile accidents can be blamed on driver error (or in some cases, the limita-

[1]Our project website includes videos of our simulations that demonstrate this phenomenon: `http://www.cs.utexas.edu/~kdresner/aim/`

Cite as: Mitigating Catastrophic Failure at Intersections of Autonomous Vehicles (Short Paper), Kurt Dresner and Peter Stone, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1393-1396.

tions of human drivers), if individual incidents are a hundred times more deadly, no reasonably achievable reduction in incident frequency will effect an overall improvement. However, if in the rare event of an accident, the total damage can be kept under control—perhaps at most a few times as many as normal—then, as a whole, riding in automobiles will be a safer experience than it is today.

2. BACKGROUND INFORMATION

Our multiagent intersection control mechanism involves two classes of agents. *Driver agents* pilot vehicles, while *intersection managers* are arbiter agents stationed at each intersection that control access to that intersection. To cross the intersection, driver agents must first obtain approval from the intersection manager.

2.1 Communication Protocol

In our communication protocol, driver agents "call ahead" to the intersection manager using a REQUEST message [1]. In addition to the physical characteristics and capabilities of the vehicle, REQUEST messages include the driver agent's intended direction of travel and estimates of its time and velocity of arrival. The intersection manager uses this information, along with an *intersection control policy* to decide whether to grant the reservation. To grant the reservation, it responds with a CONFIRM message containing restrictions the vehicle must obey. The intersection manager can also use these restrictions to make a counter-offer. The driver agent's acceptance is implicit; once the intersection transmits the CONFIRM message, the vehicle "has" the reservation. To reject the request, the intersection manager responds with a REJECT message. No vehicle may enter the intersection under any circumstances without a reservation.

A vehicle with a reservation is guaranteed safe passage, provided it crosses in accordance with its reservation. If the driver agent cannot meet the reservation, it sends a CANCEL message. Vehicles can attempt to change reservations using a CHANGE-REQUEST message. This message is the same as REQUEST, but if the intersection manager responds with a REJECT message, the original reservation remains.

2.2 First Come, First Served

Along with our framework, we have introduced several intersection control policies, including some that emulate stop signs and traffic lights. The most efficient policies are based on a "first come, first served" (FCFS) algorithm. FCFS divides the intersection into an $n \times n$ grid of *reservation tiles*, where n is the *granularity*. For each REQUEST, an FCFS policy simulates the trajectory of the vehicle across the intersection using the REQUEST parameters. Throughout the simulation, the policy determines which reservation tiles are occupied by the simulated vehicle, and whether or not any of them are reserved by another vehicle. If no conflicts are detected, the appropriate tiles are reserved for the required times and the intersection manager sends the requesting agent a CONFIRM message with the relevant information. Otherwise, the driver agent receives a REJECT message.

2.3 Safety Guarantees

While this paper focuses on some of the ways our mechanism can react to gross mechanical failures, we first point out the ways in which it compensates for smaller, more common errors. As long as all vehicles follow the protocol and all the technology works as expected, no two vehicles should occupy the same space in the intersection at the same time. At most one vehicle can reserve a particular reservation tile at one time, and vehicles can only cross the intersection in accordance with their reservations. Unfortunately, even under normal operating conditions, this is not quite enough. Communication failures including dropped and corrupted messages, as well as small errors in the vehicle's sensors and actuators could all cause problems. Our mechanism is robust to all of these. The driver agent's implicit acceptance of reservation confirmations limits the worst possible consequence of a dropped or corrupted message to additional delay—not a collision. Buffering in the intersection control policies adds protection against small sensor errors by reserving extra space for vehicles.

3. ADDING A SAFETY NET

A collision in purely autonomous traffic can have any number of causes: software errors in the driver agent, a physical malfunction in the vehicle, or even meteorological phenomena. Currently such factors are largely ignored for two reasons. First, an exclusively human-populated system, with generous margins for error, is not as sensitive to small or moderate aberrations. Second, none of these causes are significant with respect to driver error. According to a study from the 1980's, vehicle and road issues alone were responsible for fewer than 5% of accidents [4]. However, in the future of autonomous vehicles, it is exactly these issues which will be the prevalent causes of collisions. The safety allowances are adjustable—given a maximum allowable error in vehicle positioning, buffers can be extended to handle that error—but no reasonable adjustment can account for gross mechanical malfunction like a blowout or failed brakes. Because these issues are infrequent, we believe the intersection control mechanism will be acceptable even if individual occurrences are slightly worse than accidents today. As we will show, without the safety measures presented in this section, the system is prone to spectacular failure modes, sometimes involving dozens of vehicles.

3.1 Assumptions

We make several important assumptions about the capabilities of intersection managers and driver agents. We assume that intersections can be equipped with a wireless communication device with enough strength and bandwidth to communicate with hundreds of driver agents simultaneously. We also assume that the intersection manager has access to sufficient computational resources to process all the messages from these driver agents and respond to them quickly. Because our simulator can execute all the driver agent and intersection manager algorithms in real time, in one process on a desktop computer, we believe this is a realistic assumption. Finally, we assume that vehicles can be similarly outfitted, both in terms of communication and computation, and that these vehicles have access to GPS navigation equipment, detailed electronic maps, short-wave radar, lidar, and any sensing technology required to determine location and sense surrounding objects and vehicles. These assumptions are reasonable given current technology.

In order to reduce the average number of vehicles involved in a crash from dozens to one or two, we make one additional assumption—that the intersection manager is able to detect when something has gone wrong. While this assumption is

non-trivial, it is reasonable. There are two basic ways by which the intersection manager could detect that a vehicle has encountered a problem: the vehicle can directly inform the intersection manager, or the intersection manager can observe the vehicle's status. In the event of a collision, a device similar to that which triggers an airbag can send a signal to the intersection manager. Such devices already exist in aircraft to emit distress signals and locator beacons in the event of a crash. Using cameras or other sensors, the intersection manager could detect a vehicle that is not where it is supposed to be. However, this method of detection is likely to be much slower to react to a problem. Each approach has advantages and disadvantages, and a combination of the two would most likely be the safest. Ideally, whenever a vehicle violates its reservation in any way, the intersection manager should become aware as soon as possible.

Our protocol also includes a DONE message that vehicles transmit when they complete their reservations. One way to sense when a vehicle is in distress is to notice a missing DONE message. This approach has a major drawback: the intersection manager may not be able to notice the missing message until some time after the incident has occurred. We study the effects of such a delay in our experiments.

3.2 Incident Mitigation

When a vehicle deviates significantly from its planned course through the intersection, resulting in physical harm to the vehicle or its presumed occupants, we refer to the situation as an *incident*. Once an incident has occurred, the first priority is to ensure the safety of all persons and vehicles nearby. Because we expect incidents to be infrequent, re-establishing normal operation of the intersection is a lower priority and the optimization of that process is left to future work.

3.2.1 Intersection Manager Response

Once the intersection manager is notified of an incident, it immediately stops granting reservations. Subsequent REQUESTs are rejected without consideration. Because the protocol requires robustness to dropped messages, reservations cannot reliably be revoked—no self-interested agent would acknowledge receipt of such a message. However, given our assumptions, in such a situation the intersection manager can signal to the vehicles that an incident has occurred. This signal is sent via a new EMERGENCY-STOP message. This message lets vehicles know that an incident has occurred and that no further reservations will be accepted. Furthermore, vehicles able to come to a stop before entering the intersection should do so, and vehicles in the intersection should no longer assume that "close calls" will not result in collisions. Ideally, all vehicles receive the message and take appropriate actions, including those holding approved reservations. However, as we will show, even if some messages are lost or ignored, the intersection will still be safer.

3.2.2 Vehicle Response

The driver agent also has a role to play once an incident has taken place. Normally, when a vehicle approaches the intersection, it ignores any vehicles sensed in the intersection. What would appear to be an imminent collision on the open road is almost certainly an engineered "close call" in the intersection. Once our driver agent receives the EMERGENCY-STOP message, it disables this behavior and may brake to avoid hitting other vehicles. If the vehicle is not in the intersection, it will try not to enter, even if it has a reservation.

Our first inclination was to make all driver agents that receive the signal immediately decelerate to a stop. However, this is actually less safe. If all vehicles come to a stop, those that would otherwise have cleared the intersection without colliding may find themselves stuck in the intersection—another obstacle for other vehicles to hit. This is especially true if the initial incident takes place on the edge of the intersection where other vehicles are unlikely to become involved. Stopping all the other vehicles in the intersection would make the situation much worse. However, if a driver agent does detect an impending collision, it is allowed to take evasive actions or apply the brakes. Our driver agent brakes if it believes a collision is imminent.

4. EXPERIMENTAL RESULTS

In this section, we present an initial evaluation of our claims using a custom simulator described in our earlier work [2, 3]. Due to space limitations, we include only our main result. In future work, we intend to present a more complete empirical evaluation including experiments with other lane configurations, human drivers, alternate metrics for estimating overall damage, and delayed notification of incidents.

4.1 Experimental Setup

With the great efficiency of the reservation-based system comes extreme sensitivity to error. While the buffering can protect against minute discrepancies, it cannot hope to cover gross mechanical malfunctions. To quantify the effects of such a malfunction, we created a simulation in which individual vehicles could be "crashed", causing them to immediately stop and remain stopped. When a vehicle that is not crashed comes into contact with one that is, it becomes crashed as well. While this does not model the physics of individual impacts, it allows us to estimate how a malfunction might lead to collisions.

To include malfunctions in all different parts of the intersection, we trigger incidents by choosing a random (x, y) coordinate pair inside the intersection, and crashing the first vehicle to cross either the x or y coordinate. After initiating an incident, we simulate 60 additional seconds, recording any further collisions. Using this information, we construct a *crash log*. For each step of the remaining simulation, the crash log indicates how many vehicles were crashed on or before that step. By averaging over many such crash logs for each configuration, we construct an "average" crash log, which gives a picture of a typical incident.

These experiments include scenarios with either 3 or 6 lanes in each of the four cardinal directions (results for 4 and 5 lanes were similar). Vehicles are spawned with equal likelihood in all directions, and are generated via a Poisson process which is controlled by the probability that a vehicle will be generated at each step. Vehicles are generated with a set destination—15% of vehicles turn left, 15% turn right, and the remaining 70% go straight. The leftmost lane is always a left turn lane, while the right lane is always a right turn lane. Turning vehicles are always spawned in the correct lane, while non-turning vehicles are spawned only in other lanes. The traffic level averaged 1.667 vehicles per second per lane in each direction. This works out to 5 total vehicles per second for 3 lanes, and 10 total vehicles per

second for 6 lanes. We chose these settings as they are toward the high end of the spectrum of manageable traffic for the intersection manager. While we wanted traffic flowing smoothly, we also wanted the intersection full of vehicles to test situations that lead to the most destructive collisions.

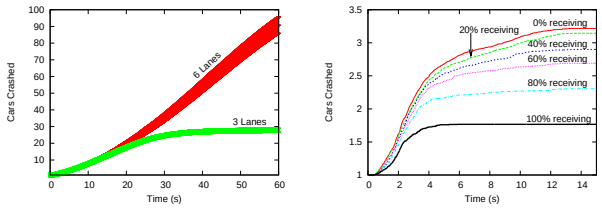
4.2 How Bad Is It?

As we suspected, the average crash log without our safety measures is grisly. Driver agents must ignore their sensors while in the intersection, because many of the "close calls" appear to be impending collisions. Unable to react to the incident, vehicles careen into one another until crashed vehicles protrude from the intersection. Figure 2(a) shows that with 6 lanes, the rate of collisions does not abate until over 70 vehicles have crashed. A minute after the incident begins, vehicles are still colliding. With 3 lanes, the intersection is much smaller, and thus it fills much more rapidly; by 50 seconds, the number of collided vehicles stabilizes.

4.3 Reducing the Number of Collisions

Our safety mechanism has two main components. First, the intersection manager stops accepting reservations. Second, the intersection manager sends a EMERGENCY-STOP message to all vehicles. However, some vehicles might not receive the signal. To explore this, we intentionally disabled some of the vehicles' ability to receive the EMERGENCY-STOP message. A parameter in our simulator controls the fraction of vehicles created with this property, and we investigated the effects of varying this parameter.

With the safety measures in full effect, the number of vehicles involved in the average incident decreases dramatically. Figure 2(b) shows the effects of our safety system on intersections with 6 lanes, with the proportion of receiving vehicles varying from 0% to 100% in increments of 20%. Even with no vehicles responding to the message, the overall number of vehicles involved in the average incident declines by a factor of almost 30. As expected, when more vehicles receive the emergency signal, fewer vehicles crash. Figure 2(b) shows the first 15 seconds of the crash log, because no collisions occurred more than 15 seconds after the incident started.



(a) No safety measures (b) With safety measures

Figure 2: Average crash logs (2(a) includes 95% confidence interval) with and without safety measures. In 2(a), both the 3- and 6-lane scenarios are shown. In 2(b), 0, 20, 40, 60, 80, and 100% of vehicles receive the Emergency-Stop message.

5. CONCLUSION

We believe these experimental results raise a very important issue. Computerized systems are held to an extremely high standard. Such systems cannot just be safer for the average user; they must be the very paragon of safety. In our experiments, we showed that the number of vehicles involved in individual incidents can be drastically reduced by virtue of some of the safety properties built into our intersection control mechanism. When all vehicles received the warning, over 60% of incidents involved only one vehicle: the vehicle we intentionally crashed. In the worst case considered here, no vehicles received the warning. As a result, approximately 3.25 vehicles crashed in the average incident. If we make the extremely conservative assumption that accidents today involve only one vehicle, even this worst-case will be safer overall if we can reduce incident frequency by 70%. A 2002 Federal Highway Administration report attributed over 95% of accidents to driver error [4]. This figure is for all driving—not just intersection driving, in which driver error is a more common cause of accidents. Because autonomous vehicles will all but eliminate driver error, our data indicate that safer and more efficient automobile travel is entirely realizable. While our results point out important vulnerabilities, they also demonstrate that our proposed modification could allow the mechanism to attain the same levels of efficiency without compromising safety.

Autonomous vehicles are a fascinating and exciting development. Before the benefits of this technology can be realized, more must be done to ensure the safety of the passengers that will use them on a daily basis. We believe we have accomplished a portion of this important work. Our failure mode analysis calls attention to the need for keeping an eye toward safety throughout the development of the algorithms and protocols that will control the transportation systems of the future. Further analysis will be necessary, first in simulation, and ultimately with physical vehicles.

Acknowledgments

This research is supported in part by NSF CAREER award IIS-0237699 and by the United States Federal Highway Administration under cooperative agreement DTFH61-07-H-00030. Computational resources were provided in large part by NSF grant EIA-0303609.

6. REFERENCES

[1] K. Dresner and P. Stone. Multiagent traffic management: A protocol for defining intersection control policies. Technical Report UT-AI-TR-04-315, The University of Texas at Austin, Department of Computer Sciences, AI Laboratory, December 2004.

[2] K. Dresner and P. Stone. Multiagent traffic management: An improved intersection control mechanism. In *The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 471–477, Utrecht, The Netherlands, July 2005.

[3] K. Dresner and P. Stone. Sharing the road: Autonomous vehicles meet human drivers. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, pages 1263–68, Hyderabad, India, January 2007.

[4] W. W. Wierwille, R. J. Hanowski, J. M. Hankey, C. A. Kieliszewski, S. E. Lee, A. Medina, A. S. Keisler, and T. A. Dingus. Identification and evaluation of driver errors: Overview and recommendations. Technical Report FHWA-RD-02-003, Virginia Tech Transportation Institute, Blacksburg, Virginia, USA, August 2002. Sponsored by the Federal Highway Administration.