In *The Fifth Workshop on Agents in Traffic and Transportation Multiagent Systems* (ATT 08), pp. 78-85, Estoril, Portugal, May 2008.

Mitigating Catastrophic Failure at Intersections of Autonomous Vehicles

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 for transportation. However, before such gains can be realized, a plethora of safety and reliability concerns must be addressed. In previous work, we have introduced a system for managing autonomous vehicles at intersections that is capable of handling more vehicles and causing fewer delays than modern-day mechanisms such as traffic lights and stop signs [3]. 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 very low. In this paper, we make four main contributions. First, we introduce safety features of the system designed to deal with these types of failures. Second, we 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. Third, we give extensive empirical evidence suggesting that not only is this method effective, but that it is so even when normal communications are disrupted. Finally, we provide an analysis of the data indicating that despite the apparent potential for disastrous accidents, autonomous intersection management is likely to improve driver safety considerably.

Keywords

Autonomous Vehicles, Multiagent Systems, Intelligent Transportation Systems

1. INTRODUCTION

Fully autonomous vehicles promise enormous gains in safety, efficiency, and economy for transportation. By taking the responsibility of driving away from humans, autonomous vehicles will completely eliminate driver error from the complicated equation of automobile traffic. By some estimates, driver error can be blamed for as much as 96% of all automobile accidents [11]. Thus, even if each accident were substantially worse, overall autonomous vehicles would represent an improvement in safety over the current situation. With automobile collisions costing the U.S. economy over $230 billion annually, any significant decrease would be a major triumph for artificial intelligence [6].

Traffic intersections are a compelling problem for multiagent systems. Often a source of great frustration for drivers, intersections represent both 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.

We have proposed a reservation-based multiagent framework for managing vehicles at intersections, including both human-driven vehicles and fully autonomous vehicles [3]. Instead of using traditional traffic lights, the mechanism allows autonomous vehicles to "call ahead" to arbiter agents stationed at intersections and reserve space-time in the intersection. When a vehicle obtains a reservation, it can proceed through the intersection without stopping. By coordinating the actions of many autonomous vehicles, the system dramatically decreases time spent stopped or slowing down due to intersections. Because the system heavily exploits the precision sensory and control capabilities of computerized drivers, it offers dramatic improvements in efficiency. However, this increased efficiency is quite precarious. The system orchestrates what can only be described as "extremely close calls", with vehicles missing each other by the smallest (albeit adjustable) margins[1]. Figure 1 contains a screenshot depicting a particularly busy intersection.

While the system is safe in the face of communication failures, we have not addressed the possibility or effects of mechanical failures or unlikely "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 limitations of human drivers), if individual incidents are a hundred times more deadly, no reasonably achievable reduction in incident frequency will effect an overall improvement. How-

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

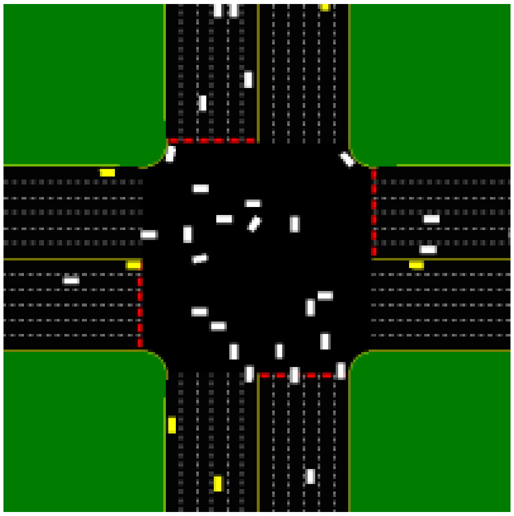


Figure 1: A screenshot showing a busy intersection with a lot of "close calls."

ever, 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.

In this paper, we make four main contributions. First, we introduce safety features of the system designed to deal with these types of failures. Second, we 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. Third, we give extensive empirical evidence suggesting that not only is this method effective, but that it is robust in the face of poor communications. Finally, we provide an analysis of the data indicating that despite the apparent potential for disastrous accidents, autonomous intersection management is likely to improve driver safety considerably.

The remainder of this paper is organized as follows. Section 2 briefly summarizes our reservation system and earlier results. In Section 3, we elucidate the safety mechanisms in the system to deal with potentially catastrophic mechanical failures and argue for their necessity. Section 4 presents empirical evidence evaluating our addition to the system. We discuss these results and their implications in Section 5 and conclude in Section 6.

2. BACKGROUND INFORMATION

Our multiagent intersection control mechanism comprises the interactions of two classes of agents: *intersection managers* and *driver agents* [3]. Driver agents are computer programs that pilot vehicles, while intersection managers are specialized arbiter agents stationed at each intersection that control access to that intersection. In order to cross the intersection, driver agents must first obtain approval from the intersection manager.

2.1 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 of their environs, short-wave radar and lidar systems, and any other sensing technology required for them to accurately and reliably determine their location and sense the objects and vehicles around them. These assumptions are all reasonable given current technology.

2.2 Communication Protocol

A major part of the reservation mechanism is the communication protocol that governs all transmissions between agents [2]. In this protocol, when a vehicle approaches an intersection, the driver agent controlling that vehicle "calls ahead" to the intersection manager, requesting permission to cross. This request comes in the form of a REQUEST message. In addition to parameters describing the physical characteristics of the vehicle, such as its size and performance capabilities, a REQUEST message includes the direction the driver agent would like to leave the intersection, as well as estimates of its arrival time and arrival velocity. The intersection manager can then use this information, along with an *intersection control policy* to decide whether or not to grant the reservation. If it chooses to grant the reservation, it responds with a CONFIRM message containing some restrictions the vehicle must obey in order to cross safely. According to the protocol, the intersection manager can also use the restrictions in the CONFIRM message to make a counter-offer. The driver agents acceptance of the confirmation is implicit; as soon as the intersection transmits the CONFIRM message, the vehicle "has" the reservation described therein. If the intersection manager decides not to grant the reservation, it responds with a REJECT message, which can optionally include a reason for the rejection. According to the rules of the protocol, no vehicle may enter the intersection under any circumstances without a reservation.

Once a vehicle has a reservation, its safety is guaranteed in the intersection, provided it crosses the intersection in accordance with that reservation. If the driver agent concludes at any time that it cannot meet the reservation, it can send a CANCEL message to the intersection manager, at which point the vehicle is no longer considered to have a reservation. Additionally, vehicles can attempt to change their reservations, with a CHANGE-REQUEST message. This message is the same as the REQUEST message, except that if the intersection manager responds with a REJECT message, the vehicle maintains its original reservation.

2.3 First Come, First Served

Our framework includes several intersection control policies, including some that emulate current-day mechanisms like stop signs and traffic lights. However, most of the extremely efficient policies are based around the "first come, first served" (FCFS) algorithm. This algorithm divides the intersection into an $n \times n$ grid of *reservation tiles*, where the parameter n is called the *granularity* of the policy. When it receives a REQUEST, an FCFS policy simulates the trajectory of the vehicle across the intersection using the parameters in the message. Throughout the simulation, the policy determines which of the reservation tiles are occupied by the simulated vehicle, and whether or not any of them are already reserved by another vehicle while the simulated vehicle would occupy them. If no such conflicts are detected throughout the simulation, the appropriate tiles are reserved for the required times, the policy creates the reservation, and the intersection manager sends the relevant information to the requesting agent in a CONFIRM message. Otherwise, the driver agent receives a REJECT message.

While FCFS policies are limited to use by autonomous vehicles only, we have also created a policy called FCFS-LIGHT, which can accommodate human drivers [4]. Briefly, an FCFS-LIGHT policy is similar to standard FCFS, except that it incorporates a *light model*, which both controls a set of physical lights at the intersection, and provides information to the policy about the state of those lights. Areas of the intersection that correspond to conventional paths through the intersection are blocked off from use by autonomous vehicles whenever the light controlling access to that path is green, yellow, or recently red. This creates a *de facto* reservation for any human that might be crossing the intersection based on that green light. While this does allow a sort of "backwards compatibility", it must be noted that by nature, FCFS-LIGHT policies tend to be much less efficient than standard FCFS policies.

2.4 Safety Guarantees

While this paper focuses on some of the ways our mechanism can react to gross mechanical failures, we must 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 be able to occupy the same space in the intersection at the same time. Only one vehicle can reserve any 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. The mechanism is designed to be robust against all of these. The driver agent's implicit acceptance of reservation confirmations means that the worst possible consequence of a dropped or corrupted message is additional delay, and not a collision. Buffering in the intersection control policies adds protection against small sensor errors by reserving space for vehicles as if they were larger than they actually are.

3. ADDING A SAFETY NET

A collision in purely autonomous traffic can have any number of causes, including software errors in the driver agent, a physical malfunction in the vehicle, or even meteorological phenomena. In modern-day traffic, such factors are largely ignored for two reasons. First, the exclusively human-populated system, with its 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 as causes of accidents. In fact, according to a study from the 1980's, vehicle and road issues alone were responsible for fewer than 5% of accidents [11]. However, in the future of infallible autonomous driver agents, it is exactly these issues which will be the prevalent causes of automobile collisions. The safety buffers in our mechanism are adjustable—given some maximum allowable error in vehicle positioning, the 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 types of issues are infrequent, we believe the safety of the intersection control mechanism will be acceptable even if individual occurrences are slightly worse than accidents today. As we will show in Section 4, without the safety measures presented in this section, the system is prone to spectacular failure modes, sometimes involving dozens of vehicles.

3.1 Assumptions

In Section 4, we will show how our modification can reduce the average number of vehicles involved in a crash from dozens to one or two. In order to make this improvement, we must make a few assumptions above and beyond those originally made by Dresner and Stone [3].

3.1.1 Detecting the Problem

First, we assume that the intersection manager is able to detect when something has gone wrong. While this is certainly a non-trivial assumption, it is necessary for any reasonable solution. Simply put, the intersection manager cannot react to something it cannot detect. There are two basic ways by which the intersection manager could detect that a vehicle has encountered some sort of problem: the vehicle can inform the intersection manager, or the intersection manager can detect the vehicle directly. For instance, in the event of a collision, a device similar to that which triggers an airbag can send a signal to the intersection manager. Devices such as this already exist in aircraft to emit distress signals and locator beacons in the event of a crash. The intersection manager itself might notice a less severe problem, such as a vehicle that is not where it is supposed to be, using cameras or sensors at the intersection. However, this method of detection is likely to be much slower to react to a problem. Each has advantages and disadvantages, and a combination of the two would most likely be the safest. The specifics of the implementation are beyond the scope of this paper. What is important is that whenever a vehicle violates its reservation in any way, the intersection manager should become aware as soon as possible. Because our simulations only deal with collisions, we assume that the colliding vehicle sends a signal and the intersection manager becomes aware of the situation immediately.

The communication protocol also includes a DONE message that vehicles transmit when they complete their reservations. One way to reliably sense when a vehicle is in distress would be to notice a missing DONE message. This does have two drawbacks however. First, the DONE message is optional, mainly because there is no incentive for the driver agent to transmit it. Second, the intersection manager may not be able to notice the missing message until some time after the incident has occurred. We hope to investigate this alternative in future work.

3.1.2 Informing the Other Vehicles

The second assumption we make is that there exists a way for the intersection manager to broadcast the fact that something is wrong to the vehicles. We already assume that the intersection manager can communicate with the vehicles, but this new assumption is a bit different. Under normal operating conditions, individual messages each containing multiple pieces of information are transmitted between agents [2]. In case of an emergency, however, the intersection manager needs only to communicate one bit of information: whether or not something is wrong. This can come in the form of a coded signal (to prevent fakes) repeated continuously on a specific frequency. As with the previous assumption, the specifics of the implementation are not relevant to this work. We assume that the intersection manager can transmit such a signal, and that the vehicles receive it. As we will show in Section 4, even without assuming the vehicles receive the signal, it is still possible to drastically improve the safety properties of the mechanism in the face of mechanical failures.

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 these incidents to be very infrequent occurrences, 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

As soon as our intersection manager is notified of an incident, it ceases granting reservations. All subsequent received requests are rejected without consideration. Due to the nature of the protocol, reservations cannot be revoked by the intersection manager. However, given our assumptions, in such a dire situation the intersection manager can signal to the vehicles that an incident has occurred. This signal is repeatedly broadcast, and is not part of the reservation protocol. Ideally, all vehicles would receive the signal and stop immediately, including those holding approved reservations.

This concept extends naturally to policies that can accommodate humans, such as FCFS-Light. Analogous to refusing further reservation requests, upon detecting an incident, the intersection manager immediately turns all lights red. In a real-world implementation, a more conspicuous visual cue could be provided, but semantically it is only important that the intersection inform the human drivers that they may not enter.

3.2.2 Vehicle Response

The driver agent also has a role to play once an incident has taken place. Normally, when a vehicle is approaching the intersection, it ignores any vehicles sensed in the intersection—what might otherwise appear to be an imminent collision on the open road is almost certainly a precisely coordinated "near-miss" in the intersection. Once the driver agent has received the emergency signal from the intersection manager, it disables this behavior. Thus, if something is wrong, and the vehicle is in the intersection, the driver agent will not blindly drive into another vehicle, if it can help it. If the vehicle is not in the intersection, it will not enter, even if it has a reservation.

A first approach might be to make all driver agents that receive the signal immediately decelerate to a stop. However, this is actually less safe. If all vehicles that receive the signal come to a stop, vehicles that would otherwise have cleared the intersection without colliding may find themselves stuck in the intersection—another object for other vehicles to run into. This is especially true if the crashed vehicle is off on the very edge of the intersection where it is unlikely to be hit. Trying to stop all the other vehicles in the intersection would make the situation much worse.

If a driver agent does detect an impending collision, however, it is allowed to take evasive actions or apply the brakes. Since this is a true multiagent system with self-interested agents, we cannot prevent the driver agents from doing so. Thus, our driver agent only brakes if it believes a collision is imminent.

4. EXPERIMENTAL RESULTS

In this section, we present empirical evaluations of our claims using a custom simulator described in our earlier work [3, 4].

4.1 Experimental Setup

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

In order to ensure that we included malfunctions in all different parts of the intersection, we triggered each incident by choosing a random (x, y) coordinate pair inside the intersection, and crashing the first vehicle to cross either the x or y coordinate. This is akin to creating two infinitesimally thin walls, one horizontal and the other vertical, that intersect at (x, y). Figure 2 provides a visual depiction of this process.

After initiating an incident, we ran the simulator for an additional 60 seconds, recording any additional collisions and when they occurred. Using this information, we constructed a *crash log*, which is essentially a histogram of crashed vehicles. For each step of the remaining simulation, the crash log indicates how many vehicles were crashed by that step. By averaging over many such crash logs for each configuration, we were able to construct an "average" crash log, which gives a picture of what a typical incident would produce.

Because our system is compatible with humans, we felt it necessary to also include experiments with the human-compatible intersection control policies [4]. When a significant number of human drivers are present, the FCFS-Light policies do not offer much of a performance benefit over traditional traffic light systems. As such, we limited our experimentation to scenarios in which 5% of the vehicles are controlled by simulated human drivers. These human drivers have a very simple behavior that attempts to maintain a following distance that is proportional to the vehicle's velocity.

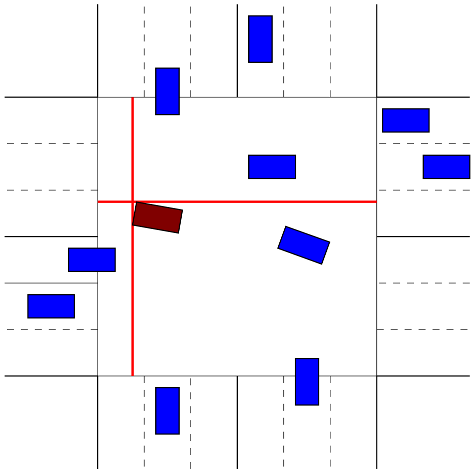


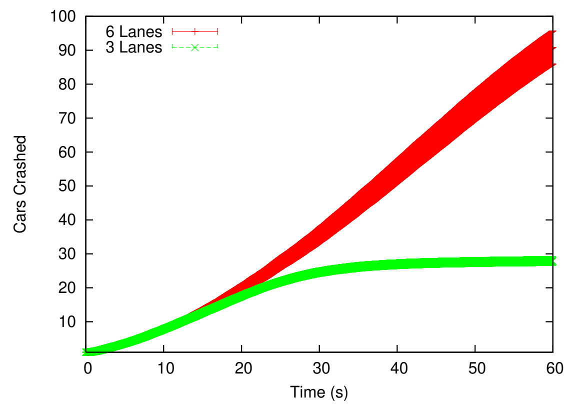
Figure 2: Triggering an incident in the intersection simulator. The dark vehicle turning left is crashed because it has crossed the randomly chosen x coordinate. If a different vehicle had crossed that x coordinate or the randomly chosen y coordinate earlier, it would be crashed instead.

With only 5% human drivers, an FCFS-Light policy can still create a lot of the precarious situations that are the focus of this investigation.

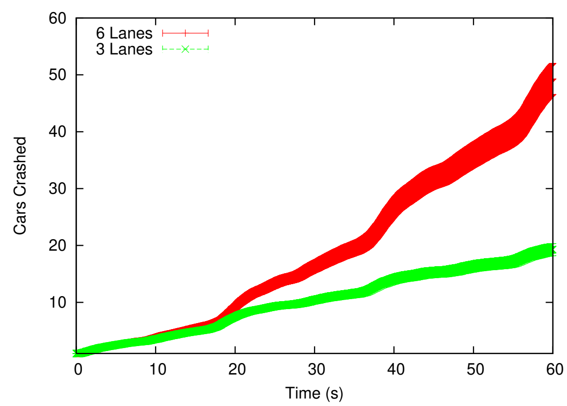
For these experiments, we ran our simulator with scenarios of 3, 4, 5, and 6 lanes in each of the four cardinal directions, although we will discuss results only for the 3- and 6-lane cases (other results were similar, but space is limited). Vehicles are spawned equally likely in all directions, and 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, and non-turning vehicles are not spawned in the turn lanes. In scenarios involving only autonomous vehicles, we set the traffic level at an average of 1.667 vehicles per second per lane in each direction. This equates to 5 total vehicles per second for 3 lanes, and 10 total vehicles per second for 6 lanes. Scenarios with human-driven vehicles had one third the traffic of the fully autonomous scenarios—the intersection cannot be nearly as efficient with human drivers present. We chose these amounts of traffic as they are toward the high end of the spectrum of manageable traffic for the respective variants of the intersection manager. While we wanted traffic to be flowing smoothly, we also wanted the intersection to be full of vehicles to test situations that likely lead to the most destructive possible collisions.

4.2 How Bad Is It?

As we suspected, the average crash log without the safety measures is quite grisly. As explained in Section 3.2.2, driver agents must ignore their sensors while in the intersection, because many of the "close calls" would appear to be impending collisions. Without any way to react the situation going awry, vehicles careen into the intersection, piling up until the entire intersection is filled and crashed vehicles protrude from the intersection. Figure 3 shows that for both 6-lane cases—fully autonomous and 5% human drivers—the rate of collisions does not abate until over 70 vehicles have crashed. Even a full 60 seconds after the incident begins, vehicles are still colliding. In the 3-lane case, the intersection is much smaller, and thus it fills much more rapidly; by 50 seconds, the number of collided vehicles levels off.



(a) All autonomous



(b) 5% humans

Figure 3: Average crash logs (with 95% confidence interval) for 3- and 6-lane intersections, for the system with the safety measures from Section 3 disabled. In 3(a), the intersection manages only autonomous vehicles, while 3(b) includes 5% human drivers.

In both of the scenarios with human drivers, shown in Figure 3(b), the number of vehicles involved in the average incident is noticeably smaller. This outcome is likely the result of two factors. First and foremost, the FCFS-Light policy must make broad allowances to accommodate the human drivers, and thus overall is inherently less dangerous. The characteristic "close calls" from the standard FCFS policy are less common. Second, the simulated human driver agents do not drive "blindly" into the intersection—trusting to the intersection manager—the way the autonomous vehicles do. Also of note in Figure 3(b) is the visible periodicity of the light model portion of the policy. As paths open up for autonomous vehicles due to changes in the lights, they drive unwittingly into the growing mass of crashed cars.

4.3 Number of Collisions

	Fully Autonomous		5% Human	
	3 Lanes	6 Lanes	3 Lanes	6 Lanes
Safety Off	**27.9** ± 1.3	**90.9** ± 4.9	**19.3** ± 1.1	**49.3** ± 2.7
Recv.				
0%	**2.63** ± .13	**3.23** ± .16	**2.23** ± .10	**2.35** ± .13
20%	2.44 ± .13	3.15 ± .17	2.07 ± .10	2.29 ± .13
40%	2.28 ± .12	2.90 ± .16	1.91 ± .10	2.07 ± .12
60%	1.89 ± .10	2.69 ± .15	1.72 ± .09	1.98 ± .11
80%	1.71 ± .08	2.30 ± .13	1.46 ± .07	1.65 ± .09
100%	**1.36** ± .06	**1.77** ± .10	**1.22** ± .05	**1.50** ± .09

Table 1: Average number of vehicles involved in incidents for 3- and 6-lane intersections with Section 3's safety features disabled, and the system intact with various percentages of the vehicles receiving the emergency signal. Even without any vehicles receiving the emergency signal, our modification dramatically decreases the number of crashed vehicles. As more vehicles receive the emergency signal, the amount decreases further.

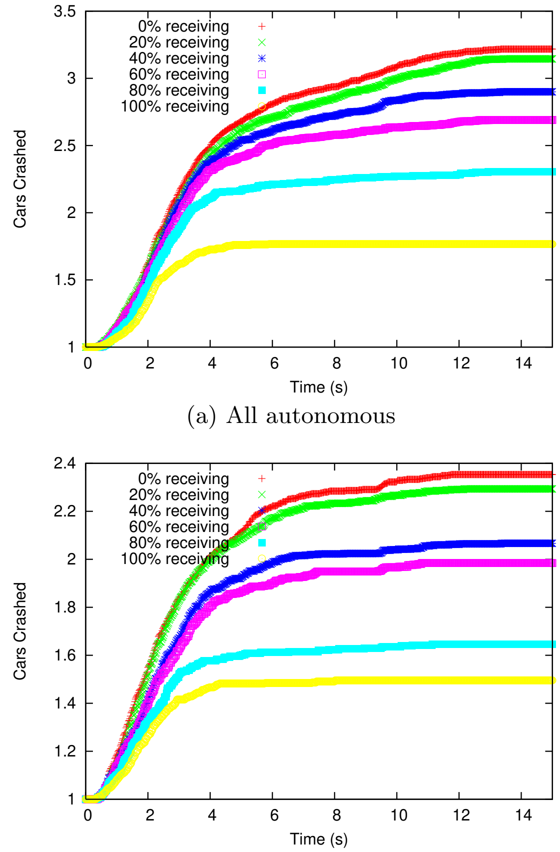
There are two main components to the safety mechanism we described in Section 3. First, the intersection manager stops accepting reservations. Second, the intersection manager emits a signal informing the vehicles that an incident has taken place. There is a possibility that this second part might not always work perfectly; some vehicles might not receive the signal. As part of our experiment, we intentionally disabled some of the vehicles' ability to receive the emergency signal. A parameter in our simulator controls the fraction of vehicles created with this property, and we investigated the effect of varying this parameter on the number of vehicles incidents.

With the safety measures in full effect, the number of vehicles involved in the average incident decreases dramatically. Table 1 shows the numerical results for both the 3- and 6-lane intersections, along with a 95% confidence interval. While we would have liked to include the average crash logs for these runs in Figure 3, they would have been impossible to distinguish from one another. For that reason, we present them in Figure 4.

Figure 4 shows the effect of the safety measures 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 warning signal, the overall number of vehicles involved in the average incident declines by a factor of almost 30 in the fully autonomous scenario, and a factor of over 20 in the scenario with 5% human drivers. As expected, when more vehicles receive the emergency signal, fewer vehicles wind up crashing. The graphs in Figure 4 only show the first 15 seconds of the incident, because in no case did a collision occur more than 15 seconds after the incident started.

4.4 Severity of Collisions

While it is reassuring to know that the number of vehicles involved in the average incident can be kept fairly low, these data do not give the entire picture. For example, compare an incident in which 30 vehicles each lose a hubcap to one in which two vehicles are completely destroyed and all occupants killed. While we do not currently have any plans to model the intricate physics of each individual collision with

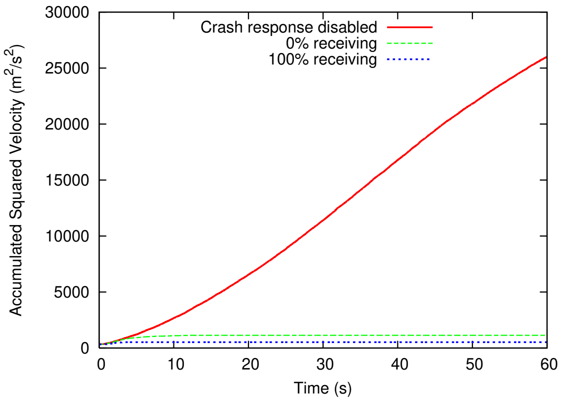


(a) All autonomous

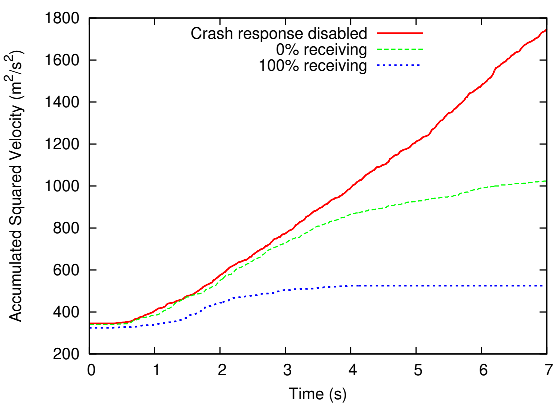
(b) 5% humans

Figure 4: The first 15 seconds of average crash logs for 6-lane intersections with all safety measures in place. As more vehicles react to the signal, safety improves further.

high fidelity, our simulations do give us access to the velocity at which the collisions occur. In the previous example, we might notice that the 30 vehicles all bumped into one another at low velocities, while the two vehicles were traveling at full speed. To quantify this information, we record not only when a collision happens, but the velocity at which it happens. In a collision, the amount of damage done is usually proportional to the amount of kinetic energy that is lost. Because kinetic energy is proportional to the square of velocity, we can use a running total of the squares of these crash velocities to create a rough estimate of the amount of damage caused by the incident. Figure 5 shows an average "damage log" of a 6-lane intersection of autonomous vehicles. Qualitatively similar results were found for the other intersection types, but are not displayed here due to space concerns.



(a) the average incident



(b) zoomed in

Figure 5: Average total squared velocity of crashed vehicles for a 6-lane intersection with only autonomous vehicles. Sending the emergency signal to vehicles not only causes fewer collisions, but also makes the remaining collisions less dangerous.

As Figure 5(a) shows, the improvement by this metric is quite dramatic as well. When no vehicles receive the emergency signal the total accumulated squared velocity decreases by a factor of over 25. When all vehicles receive the signal, it decreases by another factor of 2. Of particular note is the zoomed-in graph in Figure 5(b). Without the emergency signal, the total squared velocity accumulates as if no modification had been made, until the first vehicles stop short of the intersection at around 3 seconds; without a reservation, they may not enter. When the emergency signal is broadcast and all the vehicles receive it, the improvement is almost immediate.

5. DISCUSSION AND RELATED WORK

We believe that these experimental results raise a very important issue. People are often hesitant to put their well-being (physical or otherwise) in the hands of a computer unless they can be convinced that will receive a significant safety benefit in exchange for surrendering precious control. Humans often suffer from the *overconfidence effect*, erroneously believing they are more skillful than others. In a 1981 survey of Swedish drivers, respondents were asked to rate their driving ability in relation to others. A full 80% of those asked placed themselves in the top 30% of drivers [10]. It is this effect that creates the high standard to which computerized systems are held. It is insufficient for such systems to be marginally safer, or safer for the *average* user; they must be the very paragon of safety.

5.1 A Safer System Overall

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. In fact, when all vehicles received the warning, a large portion of the incidents involved only one vehicle: the one we intentionally crashed. Even in the worst case—6 lanes of traffic and no vehicles receiving the warning signals—an average of only 3.23 vehicles were involved. But how does this compare with current systems? If we conservatively assume that accidents in traffic today involve only one vehicle, this represents a 223% increase per occurrence. Thus, all other things being equal, if the frequency of accidents can be reduced by 70%, the the autonomous intersection management system will be safer overall. A 2002 report for the Federal Highway Administration blamed over 95% of all accidents on driver error [11]. The report blamed 2% of accidents on vehicle failures and another 2% on problems with roads. It is important to note that these numbers are for all driving, not just intersection driving. Accidents in intersections are even more likely to be caused by driver error, sometimes by drivers willfully disobeying the law: running red lights and stop signs or making illegal "U"-turns.

Even if we make overly conservative assumptions—that all driving is as dangerous as intersection driving, and that driver error is no more accountable for intersection crashes than it is in other types of driving—our data suggest that automobile traffic with autonomous driver agents and an intersection control mechanism like ours will reduce collisions in intersections by over 80%. We believe that in reality, the improvement will be staggeringly greater.

The technique presented in this paper is just one method for improving the safety of this system's failure modes. More sophisticated methods involving explicit cooperation amongst vehicles may create an even safer system. The main thrust of our discussion is not that this particular safety mechanism is by any means the best possible. Rather, it is that even with this fairly simple response to accidents, the overall safety of the system can be strengthened well beyond that of current automobile traffic—all without sacrificing the benefit of vastly improved efficiency.

5.2 Related Work

To the best of our knowledge, this paper represents the first study of the impact of an efficient, multiagent intersection control protocol for fully autonomous vehicles on driver safety. However, there is an enormous body of work regarding safety properties of traditional intersections. This includes the general—correlating traffic level and accident frequency [9] and analyses of particular types of intersections [1, 5, 7]—as well as plenty of the esoteric, such as characterizing the role of Alzheimer's Disease in intersection collisions [8]. However, because it concerns only human-operated vehicles, none of this work is particularly applicable to the setting we are concerned with in this paper.

6. CONCLUSION

In this paper, we discussed our previously proposed multiagent intersection control mechanism for autonomous vehicles. We believe the mechanism is promising, but we are not willing to sacrifice too much in the way of safety in the pursuit of efficiency. Our empirical results support our hypothesis that the mechanism can attain its high levels of efficiency without compromising on safety.

This work still leaves some unanswered questions. For example, we have examined only one method of disabling vehicles. In the future, we would like to explore other possibilities such as locking a vehicle's steering, simulating a blowout, sticking the accelerator, or disabling the brakes. For this paper, our aim was to initiate incidents that would test the limits of the intersection control mechanism by disrupting as much of the traffic flow as possible. A truly comprehensive failure mode analysis must include a much wider array of potential hazards. While our very conservative estimates indicate that this intersection control mechanism will be vastly safer than current systems with human drivers, we would like to conduct a more detailed study comparing the two, to quantify the improvement more precisely.

Autonomous vehicles, and the promise of easier and more efficient travel that they offer are a fascinating and exciting development. Before the benefits of this technology can be realized, much more work must be done to ensure that they are as safe as possible for the hundreds of millions of passengers that will use it on a daily basis. This failure mode analysis of our autonomous intersection management mechanism 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. In this way, we believe we have accomplished a portion of this important work. Further analysis will of course be necessary, first in simulation, and ultimately with real 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.

7. REFERENCES

[1] J. A. Bonneson and P. T. McCoy. Estimation of safety at two-way stop–controlled intersections on rural highways. *Transportation Research Record*, 1401:83–89, 1993.

[2] 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.

[3] 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.

[4] 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.

[5] D. W. Harwood, K. M. Bauer, I. B. Potts, D. J. Torbic, K. R. Richard, E. R. K. Rabbani, E. Hauer, L. Elefteriadou, and M. S. Griffith. Safety effectiveness of intersection left- and right-turn lanes. *Transportation Research Record*, 1840:131–139, 2003.

[6] National Highway Traffic Safety Administration. Economic impact of U.S. motor vehicle crashes reaches $230.6 billion, new NHTSA study shows. NHTSA Press Release 38-02, May 2002. `http://www.nhtsa.dot.gov`.

[7] B. N. Persaud, R. A. Retting, P. E. Gardner, and D. Lord. Safety effect of roundabout conversions in the united states: Empirical bayes observational before-after study. *Transportation Research Record*, 1751:1–8, 2001.

[8] M. Rizzo, D. V. McGehee, J. D. Dawson, and S. N. Anderson. Simulated car crashes at intersections in drivers with Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 15(1):10–20, 2001.

[9] T. Sayed and S. Zein. Traffic conflict standards for intersections. *Transportation Planning and Technology*, 22(4):309–323, August 1999.

[10] O. Svenson. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2):143–148, February 1981.

[11] 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.