Cohorts and Groups for Safe and Efficient Autonomous Driving on Highways

Gérard Le Lann

Cohorts and Groups for Safe and Efficient Autonomous Driving on Highways

Gérard Le Lann
IMARA – INRIA Paris-Rocquencourt – France
Gerard.Le_Lann@inria.fr

ISBN 978-1-4673-0048-3, in Proc. of the 3rd IEEE Vehicular Networking Conference, 2011, pp. 1-8. © IEEE

***Abstract*—We introduce constructs aimed at reconciling safety and efficiency for ad hoc highway-centric clusters of autonomous vehicles. The cohort construct is an ad hoc variant of the platoon construct. We show how to enforce safe inter-vehicle spacing in cohorts despite inaccurate vehicle space-time coordinates and failing telemetry capabilities, via neighbor-to-neighbor beaconing based on short range unidirectional communications. Worst-case analytical results are established for safe spacing bounds. A classical spacing algorithm is revisited, and proofs of usability in a discrete time beaconing model are given. Along with the group construct, which is based on prefixing usage of sensing-based solutions with omnidirectional inter-vehicular communications, we present a categorization of safety-critical scenarios. We discuss the benefits resulting from prefixing vehicle maneuvers with vehicle role assignments in safety-critical scenarios.**

Keywords-Safety, Dependability, Vehicular Ad Hoc Networks, Autonomous Driving, V2V Communications.

I. INTRODUCTION

Safety and efficiency are fundamental concerns with intelligent transportation systems (ITS). In this paper, we consider autonomous vehicles forming ad hoc clusters on highways. Efficiency implies compact single-lane clusters and optimized multi-lane maneuvers, resting on the elimination of human reaction latencies. Expected benefits are reduced energy consumption, pollution, accident rates, and travel times. Solutions for maximal compactness have emerged years ago [1], referred to as platoons, and solutions have been proposed for some multi-lane scenarios, considering platoons or vehicular ad hoc networks (VANETs). Conversely, numerous safety-related problems remain unsolved yet, and we are still lacking a general framework for analyzing safety-critical scenarios. Increased efficiency will remain illusory until safety issues are rigorously addressed. This paper is a contribution towards this end.

Lack of structuring constructs is a major impediment to proving certain properties about "vehicle clusters". This is reminiscent of the early days of distributed computing. Emergence of essential results (of theoretical and practical relevance) was made possible with the advent of such fundamental concepts as "atomicity" and "transactions" [2]. We believe there is a fruitful analogy to be drawn in the ITS domain. A cohort is the VANET counterpart of a distributed transaction; members of a cohort shall be immune to state transitions (motions and varying spacing of vehicles that do not change lanes) internal to other cohorts/transactions. Cohort interferences (e.g., lane changes) are the cyber-physical counterparts of reads and writes which concurrent transactions perform on shared variables. Distributed agreement protocols shall serve to enforce desired safety properties despite cohort interferences. Such protocols are akin to concurrency control algorithms used for maintaining data consistency in distributed databases or cloud computing. A group is the VANET counterpart of the temporary binding of variables read and written concurrently, and the "place" where concurrency control/agreement protocols are run. Cross-fertilization of distributed dependable computing and kinematics is essential for addressing combined safety and efficiency issues in ITS.

The system model used throughout this paper is exposed in Section II, with a special focus on safety requirements. The concepts of stationary scenarios and cohorts on the one hand, of transitory scenarios and groups on the other hand, are introduced. Section III is devoted to cohorts and to the companion concept of neighbor-to-neighbor communications based on unidirectional antennas. We focus on the problem of safe inter-vehicle spacing posed by failing telemetry capabilities. A solution is presented, along with analytical results. Section IV is devoted to safety-critical (SC) scenarios and related group constructs, whereby vehicles are assigned roles prior to engaging in risk-prone maneuvers. The on-ramp merging SC scenario serves as an illustration.

II. SYSTEM MODEL

We adopt the usual assumptions. Extensions due to safety considerations are listed separately.

A. On-Board Systems Functions and Terrestrial Referential

In addition to such functions as processing, I/Os, and storage, on-board systems provide autonomous vehicles with three major functions, namely telemetry, geo-localization/time keeping, and omnidirectional communications. In this paper, telemetry refers to all sensing-based solutions and technologies that serve to enforce safe longitudinal inter-vehicle spacing (e.g., radars, lidars, cameras, image processing, sensor fusion). Geo-localization and time-keeping capabilities, typically GPS/GNSS/Galileo devices – referred to as GP – augmented with e.g., dead reckoning and inertial systems (resp., clocks), are in charge of maintaining up-to-date knowledge of 360° space coordinates (resp., time coordinates). The corresponding function is denoted GP+. Inaccuracy of GP+ longitudinal space (resp., time) coordinates is denoted γ (resp., τ). Omnidirectional antennas provide for vehicle-to-vehicle (V2V) and vehicle-to-

infrastructure (V2I) communications over radio broadcast channels within ranges in the order of 250 m, interference ranges in the order of 400 m. Various protocols are being standardized, notably the IEEE 802.11p and IEEE 1609.x standards, which specify 7 channels offering each a throughput of 6 MBits/s, as well as a stochastic multi-access control (MAC) protocol based on collision avoidance (CSMA-CA). Elements of the terrestrial referential, e.g., road-side units, landmarks, as well as their exact space coordinates constitute topological data made available to on-board systems via electronic maps. Such data can be combined with GP+ space coordinates, scene recognition and AI, for achieving lane-level positioning [3]-[4].

B. Implications of Safety

1) Diversified functional redundancy

Safety does not reduce to any of those properties defined in the field of dependable computing [5]. Conversely, a violation of such properties as, e.g., reliability or availability may lead to safety hazards. Every function provided by an on-board system shall be implemented out of diversified redundant hardware, data, and software capabilities, so as to avoid common cause failures. There are numerous solutions permitting to perform detection or/and masking of capability failures, notably through periodic checking and redundancy [5]. However, this does not suffice. Under worst-case conditions, a function may be lost (fail-stopped or detectably erroneous), transiently or permanently. Some other function shall be able to supersede a failed function. This is mandatory in safety-critical domains (e.g., air transportation), where system reliability or availability figures have a lower bound in the order of $1\text{-}10^{-9}$ per hour. Ad hoc vehicle clusters are no exception. A presentation of a complete system solution that would withstand general failure assumptions (e.g., arbitrary failures) or intentional intrusions and attacks is beyond the scope of this paper, where we restrict ourselves to examining telemetry failures.

2) Realistic worst-case assumptions

Meeting safety requirements implies proving specific properties under realistic worst-case assumptions, namely inaccurate space-time coordinates in our case. GP+ inaccuracy γ is in the order of 15 m (GP inaccuracies may exceed 40 m). GP inaccuracy of time coordinates is in the order of 1μs. Safety mandates usage of clocks so as to withstand GP outages. With "good enough" affordable clocks, e.g., intrinsic drift in the order of $0.5\ 10^{-5}$, assuming outages last less than 10 s, time discrepancy 2τ for any two vehicles is in the order of 100 μs.

3) Timeliness and time bounded MAC delays

Since SC scenarios may develop far away from a road-side unit, we do not consider V2I communications. Vehicles involved in some SC scenario are necessarily close to each other. Moreover, since acceptable reaction latencies in SC scenarios are antagonistic with relaying, 1-hop communications are considered. Merits of V2V communications regarding, e.g., early warnings or collision avoidance are discussed in numerous publications. However, a major problem remains open with mobile wireless networks: how to prove that channel access delays are finite and bounded (non stochastic bounds) in the presence of worst-case contention and hidden nodes? To the best of our knowledge, there is no published protocol, be it based on CSMA, CDMA, or TDMA [6], which solves this problem under realistic assumptions (there are no impossibility proofs either). Various MAC protocols such as location-based or space division based protocols rest on assuming that different vehicles in proximate neighborhood necessarily compute different GP+ positioning data, either at the same time or at times approximately equal. This amounts to assuming that γ and τ are negligible. Since safety mandates making the opposite assumption, such protocols cannot be considered for solving the time-bounded MAC delays problem in our system model. In forthcoming papers, we present "deterministic" MAC protocols that guarantee time-bounded access delays (despite γ and τ), for event-driven messages and for periodic beacons exchanged in VANETs. One class of such protocols is based on collision freedom. Another class is based on collision detection and "deterministic" collision-resolution – the protocol in [7] can be extended to work in VANETs, with the cohort or/and the group constructs.

C. Stationary vs. Transitory Scenarios

Although not being part of the system model, the fundamental dichotomy between stationary scenarios and cohorts on the one hand, transitory scenarios and groups on the other hand, is presented in this section since it is a mandatory prerequisite to detailed expositions of these concepts. Hazard analyses and proofs of properties rest on the existence of nominal bounds for deceleration/acceleration rates, velocities, inter-vehicle spacing, as well as failure patterns. Consider a finite bounded set V of vehicles. A scenario is said to be stationary whenever V occupies a single lane and every vehicle behaves within nominal bounds, experiencing no failures or tolerable failures only – failures that shall be handled within V in a non visible manner outside V (an atomicity property). A scenario is said to be transitory whenever (1) V occupies a single lane, and one vehicle at least behaves outside nominal bounds or experiences a fatal failure – other than a tolerable failure, (2) V occupies multiple lanes, one vehicle at least performing lane change maneuvers. In this paper, we consider that telemetry failures shall be tolerable. By definition, vehicles in a stationary scenario cannot enter hazardous states. We will thus keep qualifier SC for transitory scenarios (Section IV). Variables used in this paper can be found in Table I.

TABLE I. VARIABLES AND NOTATIONS

ζ: range of telemetry capabilities
υ^*: radio range of unidirectional antennas
ρ^*: radio range of omnidirectional antennas
γ: inaccuracy of longitudinal space coordinates
τ: inaccuracy of time coordinates
v: velocity
s_{xy}: safe spacing between contiguous cohort members X and Y
c: additional spacing needed for withstanding telemetry failures
σ_{xy}: safe spacing between contiguous cohort members X and Y in the presence of Y's telemetry failure; worst-case $\sigma_{xy} = s_{xy} + c\bullet$
S: safe inter-cohort spacing
π: N2N beaconing period
δ: deceleration rate
η: ratio threshold for non SC deceleration rates, $0 < \eta < 1$

Our notation for bounds is as follows (b is a mute variable): b° for lower bound, b• for upper bound. Since we consider fully autonomous driving, bound values are computed zeroing human driver reaction latencies. Safety regulations stipulate bounds s° and S°. Bounds δ• are smallest upper bound values sustainable by every vehicle. In this paper, due to lack of space, we do not examine SC scenarios related to accelerations. Bounds v• and δ• are monitored, and enforced whenever needed. Values assigned to such bounds are not fixed, since they may depend on highway sections (delimitated by road-side units) and/or temporary local conditions (e.g., weather or highway surface). In the latter case, distributed agreement protocols serve to maintain common knowledge of current bound values (see cohort management).

III. COHORTS

Most often, proofs of safety properties for platoons are intricate and incomplete. Scrutinizing all possible states resulting from failures is hardly feasible [8], a consequence of the complex nature of platoon management, meant to encompass every possible combination of scenarios. The "splitting" of the platoon construct in two distinct constructs (cohort and group) simplifies matters, making it easier to meet proof obligations. As regards VANETs, maximal compactness is not a primary concern, a weakness to be corrected.

A. The Cohort Construct

A cohort is an ad hoc set of no more than r• contiguous vehicles circulating on a single lane endowed with demonstrated compactness *and* safety properties. Contrary to platoons, lane changes are not the province of cohort management. Cohort management is fully distributed (a head plays no particular role), resting on diversified functional redundancy (see Subsection B). On a given lane, a vehicle leaves or joins a cohort simply by decelerating or accelerating. In case some nominal bound is violated, a SC scenario is triggered without prior approval from the cohort head. Cohort membership changes may be triggered concurrently. In a cohort, some contiguous members may form a pre-planned platoon, while others do not. Given that, in addition to compactness, safety is a design driver for cohorts, it is mandatory to prove that no hazards may result from the failure of an on-board system function – telemetry in this paper.

1) Nominal inter-vehicle spacing and inter-cohort spacing

Inter-vehicle spacing has been extensively explored [9], under various car-following models [10] and for mixed vehicle networks [11]. We apply and extend existing work to the cohort construct. Consider two contiguous members X and Y, X preceding Y. Spacing s_{xy} which depends on X's and Y's velocities is such that $s° \leq s_{xy} \leq s•$. Bound s° is derived from safety calculations for smallest velocities (e.g., 2 m for velocities smaller than 20 km/h). Bound s•, which is derived from efficiency calculations (cohort compactness), is reached when $v_x = v_y = v•$ (e.g., 150 km/h). In case Y would detect that its spacing with X is nearing s•, either Y accelerates so as to remain a member of its current cohort, or Y decelerates until s_{xy} reaches value S°, Y becoming head of a cohort. Spacing s_{xy} must meet some optimality tradeoff, reconciling high compactness and high safety, two antagonistic requirements. Inter-cohort spacing is denoted $S_{ct/ch}$, CH standing for the head of a cohort and CT the tail of the preceding cohort – see Fig. 1. $S_{ct/ch}$ has a lower bound S° derived from safety calculations for smallest velocities (e.g., 15 m for velocities smaller than 20 km/h). S• is reached when $v_{ct} = v_{ch} = v•$. Specifying r• and $S_{ct/ch}$ permits to set an upper bound for the number of vehicles that may be involved in a collective rear-end collision would on-board systems of contiguous members experience fatal failures (all functions down) simultaneously. Given that $S_{ct/ch}$ is enforced, a cohort head always stops before hitting the tail of a preceding cohort (the "brick wall" paradigm). Smallest compactness is achieved with cohorts of 1 vehicle each. With autonomous driving, vehicle motions are under the control of on-board systems. Consequently, instantiations of low density patterns can be avoided by enforcing the creation of maximally compact complete cohorts (r• members each) whenever feasible. Most often, s° and s_{xy} are computed ignoring telemetry failures, and there is no distinction made between variables s and S.

2) Rationale for N2N unidirectional communications

In platoons or cohorts, if Y follower of X has its telemetry down, spacing s_{xy} is out of control. V2V communications may be (transiently, permanently) disrupted. Moreover, V2V communications may not be the most efficient solution for superseding failed telemetry, since any such potentially hazardous situation ought to be handled by two contiguous members only. Unidirectional communications are feasible by resorting to small beamwidth radio antennas [12]. With such antennas, restricted to span very short ranges υ* (in the order of 20 m), it is possible to provide any two contiguous cohort members with a semi-private communication channel, a function referred to as neighbor-to-neighbor (N2N) communications, implemented via at least one couple of front-looking and rear-looking unidirectional antennas, operating on channel(s) other than those allocated to V2V communications. Tunable antennas with transmit power proportional to inter-vehicle spacing help in mitigating radio interferences. If needed (unlikely for ranges as small as υ*), highway lane curvatures can be accommodated with steerable antennas. Front-looking antennas and rear-looking antennas can be assigned different channels.

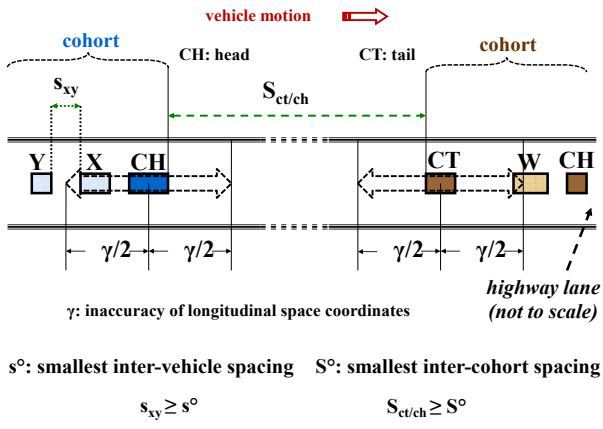

Figure 1. Cohorts, inter-vehicle spacing, and inter-cohort spacing

One may also consider 2 different pairs of radio channels, one (resp., the other) used by vehicles circulating on even-numbered (resp., odd-numbered) lanes. N2N radio interferences can thus be minimized. Nevertheless, some MAC protocol guaranteeing time bounded N2N channel access delays is mandatory (see Conclusions).

3) N2N communications and cohort management

N2N communications consist of messages and beacons exchanged over N2N links. A N2N message is assigned a type, which serves to tell cohort members how to perform relaying inside their cohort. Examples of types are (1) unidirectional upstream (resp., downstream) linear relaying, terminated by the head (resp., the tail), (2) bidirectional linear relaying (simultaneous upstream and downstream relaying), (3) unidirectional and bidirectional circular relaying. By the virtue of N2N messaging, a cohort can be structured as a chain or a virtual ring. It is reasonably easy to devise distributed fault-tolerant agreement algorithms out of such features, which algorithms are essential for cohort management. Cohort members may reach event-driven or repeated agreements [13], [14] on, e.g., new temporary nominal bounds for velocity, mutual spacing, new beaconing periods, such agreements being prompted by, e.g., varied detections of changing environmental conditions. Of great importance are "altruistic" algorithms, whereby members assist each other, notably when involved in a SC scenario. Cohort management is in charge of enforcing r•.

A N2N beacon is shared by two contiguous cohort members only (no relaying). If useful (altruistic algorithms), a N2N beacon may carry data relative to neighbors more than 1 hop away. N2N beaconing is a periodic process. Vehicles need not have access to the GP+ time referential (good timers/clocks suffice). Continuous situational awareness, a rationale for V2V beaconing, rests on assuming that space-time coordinates found in beacons are usable, i.e. inaccuracies γ and τ are not an impediment to safety-centric calculations. This may hold true for cohort heads and tails, due to S°, but not for cohort neighbors. Therefore, a N2N solution shall rest on beacons that do not carry vehicle space-time coordinates (or, if they do, such data shall not be used for enforcing safe inter-vehicle spacing).

B. *Safe Inter-Vehicle Spacing Without Telemetry*

When a platoon or cohort member experiences a telemetry failure, safety mandates immediate stopping on an emergency lane, or steep breaking and reverting to manual driving. This is not necessary with N2N beaconing and spacing algorithms designed to supersede failed telemetry. Due to lack of space, we only provide algorithmic principles and simplified analytical results. Notice that analytical results are mandatory since worst-case bounds cannot be established via simulations.

1) Principles

Consider neighbors X and Y, Y following X. Our discrete time model consists of consecutive time intervals, each of duration π, as shown Fig. 2. Convenient values of π are smaller than 1 s. Let beacon(X, i) stand for the N2N beacon sent (to Y) by X at time t(i). Let β stand for the worst-case delay incurred with processing and transmitting a N2N beacon, possible retransmissions and N2N channel contention delay(s) included. A N2N channel involves a very small number of contenders.

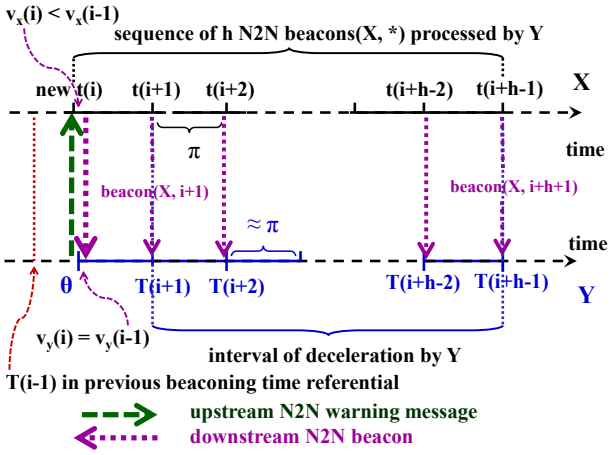

Figure 2. N2N beaconing and worst-case X/Y deceleration scenario

Therefore, MAC level delays are small with N2N beaconing, in contrast with V2V beaconing which involves hundreds of contenders under worst-case traffic density and a high number of lanes (both highway directions). In the order of a few ms, β is negligible compared to π. Thus, for the sake of simplification, let us consider that beacon(X, j) is received by Y at T(j) ≈ t(j). Nevertheless, the processing by Y of incoming beacons is strictly event-driven. In the sequel, T and t are used interchangeably. Let θ stand for the time of occurrence of Y's telemetry failure. By definition, spacing s_{xy} at θ is safe. We want to express σ_{xy}, the counterpart of s_{xy} in the presence of telemetry failures. Therefore, we must find a spacing c• such that $\sigma_{xy}(\theta) = s_{xy}(\theta) + c•$ could be proved to (1) entail a safe X/Y spacing in worst-case telemetry failure conditions, (2) have a lower bound σ° that does not depend on X's or Y's velocities.

Since we study risks X poses to Y past θ, we must consider worst-case scenarios where X decelerates. Let us introduce η such that $\forall t, \delta_x(t) \leq \eta\delta(t), 0 \leq \eta < 1$. The distinction between X decelerating "spontaneously" at a rate at most equal to $\eta\delta(t)$, and X forced to violate this bound matches exactly the distinction between a stationary scenario and a SC scenario. Recall that $\delta(t)$ – nominal bound valid at t and known to X and Y – has δ• as an upper bound. Parameter η permits to establish novel analytical results of practical significance, useable in a straightforward manner by safety authorities.

A round-based algorithm Π supersedes failed telemetry for enforcing safe X/Y spacing. At θ, when Y's telemetry failure is detected by Y's on-board system, Π's initialization thread is run. Y has a choice among 2 options (application-level decisions), to "leave" its cohort, or to remain X's follower. Π is notified which option is chosen, and Π sends two N2N warning messages (upstream to X, downstream to Y's follower, denoted Q, unless Y is cohort tail) carrying "Y's telemetry failed, selected option". If useful (cohort members behind Q may need be warned), the downstream message is typed "unidirectional downstream linear" relaying. Under option "leaving", X is not involved. Q's telemetry being operational, Q decelerates appropriately, i.e. spacing Y/Q is safe. If Q's telemetry is down as well, Q starts a SC scenario. Observe that having 2 telemetry

failures impacting 2 neighbors simultaneously – a fatal failure case – is an event occurring with extremely low probability, likely below the 10^{-9} per hour figure. While Y keeps decelerating, s_{xy} reaches $s\bullet$ (N2N beaconing is effective provided that $\upsilon^* > s\bullet$), at which time Y starts relying on V2V beaconing for enforcing safe inter-cohort spacing $\Sigma_{xy} = S_{xy} + c\bullet$ between itself (new cohort head) and X (new cohort tail). Notice that option "leaving" entails a traffic compactness vs. safety tradeoff biased in favor of safety.

In the sequel, we focus on option "staying". In addition to Q, X is involved, as follows. Upon receiving Y's N2N warning message, X may "freeze" its current velocity $v_x(\theta)$ for some time, or even accelerate, if at all possible. For worst-case analyses, we have to consider that X decelerates at highest rate $\eta\delta\bullet$. Upon receiving Y's N2N warning message, X returns downstream beacon(X, i), i.e. X aligns its time referential for periodic beaconing to θ, yielding new $t(i) = \theta$ (last sequence number used was i-1). Beacon variables are kept in on-board system memory for 1 round at least.

2) Worst-case analysis for option "staying"

Parameter η permits to examine all possible worst-case scenarios, i.e. cases where Y stops while X still is moving, as well as cases where both vehicles either keep moving or stop. Worst-case arises when X and Y circulate at identical velocity v at $\theta-\pi'$, X decelerates at $\theta-\pi'+\mu$, $\mu \approx 0$, i.e. right after sending beacon (X, i-1) quoting $v_x(i-1) = v$, and when $\pi' = \pi$. The (easy) proof is by contradiction. Due to latency π, X/Y spacing is bound to decrease initially. At $T(i) = \theta$, in worst-case conditions, we have $v_y(i) = v$ and $v_x(i) = v-\delta\bullet\eta\pi$. Thus, lost X/Y spacing in interval $[\theta-\pi, \theta]$, denoted c', is equal to $\eta\delta\bullet\pi^2/2$. Past θ, X (resp., Y) keeps decelerating continuously at rate $\eta\delta\bullet$ (resp., $\delta\bullet$). Let c(t), $t > \theta$, stand for the distance travelled by Y minus the distance travelled by X during interval $[\theta, t]$. We have $c(t) = \delta\bullet t\{\eta\pi - t(1-\eta)/2\}$. Highest value of c(t) is reached when its derivative w.r.t. t is 0, i.e. at $t = t^* = \eta\pi/(1-\eta)$. Thus: $c(t^*) = \delta\bullet(\eta\pi)^2/2(1-\eta)$. It follows that the worst-case total lost spacing $c\bullet = c'+c(t^*)$ is:

$$c\bullet = \pi^2\delta\bullet\eta/2(1-\eta).$$

Notice that $c\bullet$ does not depend on velocities. Lower bound of σ_{xy} writes $\sigma^\circ = s^\circ + c\bullet$. Since bound s° holds for velocities below v°, such is the case for σ°. This completes the proof establishing that spacing $\sigma_{xy} = s_{xy} + c\bullet$ is safe for cohort pair {X, Y} in the presence of a Y's telemetry failure.

For any $t > \theta+t^*$, in the worst-case deceleration scenario, $v_y(t) < v_x(t)$. Thus, scenarios where X does not stop prior to Y reaching velocity 0 are such that $v_x(\theta+t^*) > 0$, which condition writes $v - (\pi+t^*)\eta\delta\bullet > 0$ (X starts decelerating at $\theta-\pi$), yielding $\eta < v/(v+\pi\delta\bullet)$. It follows that $v\bullet/(v\bullet+\pi\delta\bullet)$ is the highest possible value of η, denoted $\eta\bullet$. Y may stop prior to having the X/Y spacing reduction reach $c\bullet$, which we call "early stopping". Trivially, the worst-case condition for early stopping is $\pi + v_y/\delta\bullet < \pi/(1-\eta)$, and there is no early stopping whenever $v_y(i-1) > \delta\bullet\eta\pi/(1-\eta)$. Y's early stopping is safe, since every cohort member circulating behind Y either has its telemetry operational, or triggers a SC scenario otherwise. These worst-case results translate as follows in the Π's discrete time model. The number of rounds needed for reaching spacing reduction $c\bullet$ is $h = \lceil 1/(1-\eta) \rceil$, total duration $h\pi$. Illustration with $\pi = 0.5$ s, $\eta = 0.77$, $\delta\bullet = 7$ m/s^2. One finds $c\bullet = 2.93$ m, $t^* = 1.67$ s, $h = 5$, and $h\pi = 2.5$ s. Y early stops if $v_y(\theta) \leq 11.72$ m/s. Assuming $v_y(\theta) = 35$ m/s, one finds $v(t^*) = 23.28$ m/s and velocity 0 may be reached (if ever) in 5 s.

3) Discussion

In the above analysis, we have ignored sensing, computing, and actuating latencies. A detailed worst-case analysis shall also account for the fact that beaconing cannot be strictly periodic (be it N2N or V2V beaconing), due to channel contention. Some quantity derived from time bounds guaranteed by a MAC protocol (negligible compared to π) shall be added to $c\bullet$. Worst-case bounds depend on η. Consequently, meeting highest rate $\eta\bullet$ could be made mandatory by safety authorities, yielding nominal upper bound $\eta\bullet\delta\bullet$ for non safety critical deceleration rates. While X's deceleration rate is not higher than $\eta\bullet\delta\bullet$, X and Y rely on N2N beaconing. In case X must decelerate at a higher rate, X initiates a SC scenario. With a "deterministic" V2V MAC protocol, Y and Y's followers receive X's SC-message within a few milliseconds, saving the (at most) π latency for Y, and the N2N message relaying latency for Y's followers.

Safety is guaranteed with the N2N beaconing solution provided that at most 1 telemetry failure may occur within a cohort, or that coincidental telemetry failures in a cohort occur at times separated by at least t* (assumption H). A vehicle Q that follows Y decelerates at rate $\delta\bullet$ as soon as Y does, thanks to its telemetry function. In case the coverage of assumption H would be estimated smaller than the $1\text{-}10^{-9}$ per hour figure for a cohort of at most $r\bullet$ members (see Subsection B1), it might be impossible to avoid a collective rear-end collision if vehicles simply break, staying on their lane. One shall rely on SC scenarios, notably the "imminent interleaved lane changes" scenario. Vehicles such as X and Y on lane k would issue V2V SC-messages, in addition to sending N2N messages, instructing downstream vehicles on lane k and lanes adjacent to k that they have to coordinate in order to accommodate lane changes to the benefit of vehicles leaving lane k (see Section IV).

4) Spacing Algorithm Π

A full description of algorithm Π is beyond the scope of this paper. Besides the initialization thread run at θ, Π comprises threads run in case Y is made aware of a violation of rate $\eta\delta\bullet$ by X, or when Y is about to stop, among others. Let us focus on Π's thread activated by the arrival of periodic N2N beacons. Since σ_{xy} includes $c\bullet$, any telemetry-based spacing algorithm known to enforce safe and stable cooperative adaptive cruise control can be used within this thread, provided that ratio η or deceleration rates δ can be handled explicitly within such an algorithm. An interesting example is the spacing formula expressed as Eq. (45) in [15] that gives critical warning distances for two contiguous decelerating vehicles. Converted in our notations, original Eq. (45) is as follows:

$$v_y^2/\delta_y < v_x^2/\delta_x - 2\varepsilon v_y + 2s_{xy} \qquad (1)$$

where ε stands for Y's lag time. Given that we consider autonomous vehicles, ε stands for the activation latency of a telemetry-based spacing algorithm (denoted TBS) resting on Eq. (45), rather than for human driver reaction latency as in [15]. The worst-case value of this latency is TBS period of execution. Typical values of ε are much smaller than 100 ms.

When applied to our N2N beaconing model and accounting for telemetry failures, original Eq. (45) becomes:

$$v_y^2/\delta_y < v_x^2/\delta_x - 2\pi v_y + 2\sigma_{xy} \qquad (2)$$

with $\pi > \varepsilon$, σ_{xy} standing for X/Y spacing at t(i), other variables averaged over interval [t(i-1), t(i)], excepted δ_y computed for interval [t(i), t(i+1)]. Proving that algorithm Π resting on TBS matches our worst-case analysis simply consists in establishing those conditions under which (2) implies (1). Let us rewrite (1) and (2) respectively as follows:

$$\{v_x^2/\delta_x - 2\varepsilon v_y + 2s_{xy}\} - v_y^2/\delta_y = V_1,\ V_1 > 0 \quad \text{and}$$

$$\{v_x^2/\delta_x - 2\pi v_y + 2(s_{xy} + c\bullet)\} - v_y^2/\delta_y = V_2,\ V_2 > 0.$$

Trivially, $V_2 > 0$ implies $V_1 > 0$ iff $V_2 \leq V_1$. Simple manipulations lead to the condition sought:

$$(\pi-\varepsilon)\ v_y \geq c\bullet \qquad (3)$$

A number of analytical results can be derived from (3). Firstly, bounds of π are roots of (3), which roots exist under the condition $v_y > 2\eta\delta\bullet\varepsilon/(1-\eta)$. This condition is always satisfied since this lower bound is smaller than v_y's lower bound stated in (3) by an amount equal to $(\pi-2\varepsilon)^2$. Therefore, roots of (3) always exist, and bounds of π are:

$$\pi^\circ = (1-\eta)v_y[1-\lambda]/\delta\bullet\eta, \qquad \pi\bullet = (1-\eta)v_y[1+\lambda]/\delta\bullet\eta,$$

where $\lambda^2 = 1 - 2\eta\delta\bullet\varepsilon/(1-\eta)v_y$, $\lambda > 0$. We have established that (2) implies (1) for any value of π ranging between π° and $\pi\bullet$. Smallest (resp., highest) value of π° (resp., $\pi\bullet$) is obtained with $v_y = v^\circ$ (resp., $v_y = v\bullet$). For a practical utilization, values of π close to π° shall be favored since $c\bullet$ is quadratic in π. Illustration with $\varepsilon = 0.05$, $\delta\bullet = 7.5$, $\eta = 0.77$, and $v_y = 40$. One finds $\pi^\circ = 0.051$, and $\pi\bullet = 3.14$. With $\pi = 1.7$, $c\bullet = 36.28$, whereas $\pi = 0.4$ leads to $c\bullet = 2.01$.

Secondly, knowing π, we know v_y's lower bound $v^\circ(\eta)$, which is: $v^\circ(\eta) = \pi^2\delta\bullet\eta/2(1-\eta)(\pi-\varepsilon)$. For example, with $\pi = 0.5$, $\varepsilon = 0.07$, $\eta = 0.7$ and $\delta\bullet = 7.5$, one finds $v^\circ(0.7) = 5.09$. The physical interpretation is as follows: for velocities below threshold $v^\circ(\eta)$, spacing s_{xy} is kept to lower bound $s^\circ(\eta)$.

In Π, the above results can be exploited as follows. Let $v_x(i)$ and $\varphi_x(i)$ stand, respectively, for X's velocity at t(i) and X's constant deceleration or acceleration rate computed for interval [t(i-1), t(i)]. Rates $\varphi_x(i)$ can be measured or/and trivially derived from distances travelled during interval [t(i-1), t(i)]. Variables read in N2N beacon(X, i) are round number i (modulo some highest value, used for discarding duplicates), $v_x(i)$, and $\varphi_x(i)$. Being current or computed a posteriori, $v_x(i)$ and $\varphi_x(i)$ reflect real X's motions. For our study, $\varphi_x(i) = \delta_x(i)$. Rates $\delta_y(i)$ are computed by Π and communicated to Y. Vehicle Y actuates deceleration command $\delta_y(i)$ during interval [t(i), t(i+1)]. In case the $\delta_x(i)$'s and $\delta_y(i)$'s would not be considered accurate enough, $c\bullet$ should be augmented by some quantity c(δ'), derived from δ', the inaccuracy of rates δ (a reasonably trivial calculation). It would also be necessary to "augment" Π so as to keep σ_{xy} within safe bounds. To this end, a timer shall be added to Π, value Γ(δ'), determined by c(δ'). Past θ+Γ(δ'), Y shall not trust Π any longer for enforcing σ_{xy}. Thus, either Y reaches its destination earlier than θ+Γ(δ'), or at θ+Γ(δ'), Y stops on the emergency lane (a SC scenario), or Y reverts to manual driving, or Y repeats the N2N messaging process run at θ, option field set to "leaving".

5) Conclusions – Summary of worst-case safety conditions

We have established that the ability to withstand telemetry failures entails augmenting inter-vehicle spacings usually considered for platoons by a fixed velocity-independent spacing $c\bullet$. Moreover, pivotal parameters π, $\delta\bullet$ and η can be chosen so as to have $c\bullet$ taking values comparable to spacing values contemplated for platoons, assuming no telemetry failures. Thus, the ability to withstand telemetry failures roughly translates into doubling usual intra-platoon spacing. Inter-vehicle safe spacing lower bound $\sigma^\circ = s^\circ + c\bullet$ is enforced by telemetry, and safe spacing lower bound s° is enforced by N2N beaconing in the event of a telemetry failure. Early stopping, possible if velocity is smaller than $\delta\bullet\eta\bullet\pi/(1-\eta\bullet)$ when telemetry fails, is safe. Upper bound $r\bullet$ is enforced by cohort management. Upper bound $s\bullet$, derived from efficiency calculations and worst-case analyses for accelerating scenarios (rate $\alpha\bullet$ supersedes $\delta\bullet$), is monitored by telemetry and N2N beaconing, which implies $\upsilon^* > s\bullet$. As for inter-cohort safe spacing, lower bound $\Sigma^\circ = S^\circ + c\bullet$ is enforced by telemetry, which implies $\zeta > \Sigma^\circ$. Safe spacing $S_{ct/ch}$ can be enforced by V2V beaconing provided that $\rho^* > S\bullet$ (reached at velocity $v\bullet$), which is feasible. Enforcement of safe spacing $S_{ct/ch}$ by telemetry would imply $\zeta > S\bullet$, which may not be feasible. Finally, since values of π can be chosen (possibly significantly) smaller than 1 s, N2N beaconing achieves cohort compactness higher than feasible with human drivers, for identical or better safety figures, a result which is consistent with our initial goal.

IV. GROUPS

1) Introduction

Our aim here is to introduce new ideas relative to the handling of SC scenarios. Consequently, we simply provide an informal description of the group construct. Detailed presentations of SC scenarios and group forming are given in forthcoming papers. Consider two neighboring vehicles circulating at high velocities that engage in a maneuver involving one lane change at least. With a solution based on sensing-based capabilities only, it may well be that both vehicles get too close to each other before knowing "what to do" for avoiding a collision. Imagine that a coordination algorithm based on V2V communications is run by both vehicle on-board systems, so that each vehicle is assigned a specific role (for example, which goes first and which yields the way) in due time. Individual behaviors can then be deduced from respective roles, thus preventing a risky situation from happening. Thanks to a V2V communication-based solution, respective vehicle maneuvers can be decided sufficiently ahead of time, prior to undertaking such maneuvers, the result being that a sensing-based solution (e.g., scene recognition) is invoked under favorable circumstances. In so doing, feasibility conditions for a correct handling of such scenarios can be improved quite significantly. Moreover, thanks to advance role assignments, maneuvers can be smooth and optimized, which results into savings regarding energy and pollution (no undue accelerations or/and decelerations). We have just described an example of a transitory or SC scenario. Recall that SC scenarios encompass multi-lane scenarios, violations of

nominal bounds assigned to v, δ, η, s, or S in (single-lane) stationary scenarios, as well as occurrences of fatal failures (see Subsection II-C). For example, coincidental failures of the rear-looking N2N communications function of some member and the telemetry function of the following member are handled as SC scenarios. A categorization of SC scenarios is given Table II, based on two criteria (single-lane or multi-lane, originating event is intentional or unintentional). Semantics of SC scenarios is such that a SC scenario always has at least one initiator, denoted Z (Z' added when more than one initiator).

2) Anatomy of a SC Scenario

A type, denoted F, is assigned to every SC scenario. A velocity upper bound is specified for every F, denoted v•(F). A type F SC scenario started by initiator Z is denoted {Z, F}. At some unpredictable time, Z broadcasts a SC-message denoted M(Z, F). All messages exchanged in the course of a SC scenario are SC-messages broadcast over the V2V SC-channel. Three groups are defined with scenario {Z, F}. Group R(Z, F) comprises vehicles which receive M(Z, F). Group E(Z, F) comprises vehicles in R(Z, F) that *may have* to take some active part in {Z, F}; such vehicles are said to be eligible (for becoming actors). Group A(Z, F) comprises vehicles that *do have* to take an active part in {Z, F}, referred to as actors. Every actor plays a specific role, which roles depend on type F. A multi-lane SC scenario comprises 3 algorithmic phases (role assignments), and 2 kinematic phases resting on sensing-based control capabilities, as follows:

- Phase 1, R(Z, F) is created. If not stopped, Z may have to decelerate (velocity not higher than v•(F)). The same constraint applies to members of R(Z, F). Phase 2, E(Z, F) is created. Every vehicle in R(Z, F) runs an F-dependent eligibility test, based on local data and data contained in M(Z, F). An eligible vehicle W responds to Z by broadcasting a SC-message denoted EM(Z, F, W). Phase 3, A(Z, F) is created. Z runs a role assignment algorithm which has SC-messages received from eligible vehicles as inputs. Z broadcasts its decisions (actor/non actor status, roles) relative to each member of E(Z, F).

- In phase 4, actors undertake coarse grain maneuvers inferred from their roles assigned in phase 3. An actor may have to decelerate or to accelerate, or start changing lane. In phase 5, actors perform fine grain maneuvers, under the control of their sensing-based proximity capabilities (notably, side-looking capabilities).

Network protocols and coordination algorithms are run in phases 1, 2 and 3, during which vehicles other than actors cannot change lane(s). (They behave according to cohort management rules.) Durations of these three phases are approximately in the {1-2} seconds range for unintentional scenarios, in the {3-5} seconds range for intentional scenarios. In a single-lane SC scenario, roles can be directly inferred from type F read in SC-message M(Z, F). Therefore, phases 2 and 3 described above are combined into a single phase, ditto for phases 4 and 5.

Via N2N beaconing, every cohort member may be aware of contextual data such as, e.g., ranking, mutual spacing, acceleration/deceleration rate, velocity, length, relative to each of its current neighbors (if any).

TABLE II. EXAMPLES OF HIGHWAY-CENTRIC SC SCENARIOS

Intentional single-lane scenario
Dangerous misbehavior of human driver — Z: vehicle reverted to manual mode, brutal acceleration or deceleration
Unintentional single-lane scenarios
Brutal stopping — Z: vehicle which stops abruptly (the "brick wall" paradigm, e.g. collision with a deer) — Z, Z': vehicles involved in an accident, 1 lane blocked
Sudden acceleration — Z: vehicle which accelerates abruptly (failure of vehicle equipment, of on-board system)
Intentional multi-lane scenarios
On ramp merging — Z: entrant vehicle
Collective lane change — Z, Z': members of a platoon (cohort subset) wanting to move to a different lane in a monolithic fashion
Overtaking — Z: vehicle wanting to overtake (double lane change)
Unintentional multi-lane scenarios
Imminent interleaved lane changes — Z, Z', and so on: same as individual scenario, vehicles wanting to change lane(s) "immediately", interleaved with vehicles on adjacent lane(s)
Emergency stopping (individual multi-lane change) — Z: vehicle wanting to reach the emergency lane as soon as possible

Therefore, one can devise "altruistic" algorithms whereby vehicles assist each other. In a SC scenario, altruism rests on having W's neighbors' contextual data included within message EM(Z, F, W) returned by eligible vehicle W, in addition to W's own data. This provides for natural masking of message losses. For example, assuming uniform distribution of message losses, Z is provided with exact knowledge of group E(Z, F) membership despite message loss ratios as high as 2/3. Acknowledgement based repetitions are not necessary, which eliminates the acknowledgment implosion problem.

On-ramp merging (type ORM) may serve as an illustration (Fig. 3). Z is the entrant vehicle. In phase 2, only those members of R(Z, ORM) which circulate on lane 1 (rightmost or leftmost lane, depending on the country considered) run the eligibility test. E(Z, ORM) includes every vehicle X estimating that it will reach the merging point at a time t(X) comparable to t(Z), estimated time of Z's arrival at merging point quoted in M(Z, ORM). Every eligible vehicle X broadcasts SC-message EM(Z, ORM, X) carrying t(X), its own contextual data, plus neighbors' data (unless X is a cohort of size 1). In phase 3, Z runs some optimization function having SC-messages EM(*) as inputs, and chooses 2 contiguous vehicles, denoted P and Q, P (resp., Q) being assigned the role of Z's predecessor (resp., Z's successor). In case E(Z, ORM) comprises less than 2 vehicles, Z's choice is trivial. As soon as phase 3 is over, Z, P and Q start phase 4 by adjusting their respective velocities so as to make them approximately equal when they reach the merging point. Moreover, P and Q adjust their respective velocities so as to create a "slot" between them, permitting Z to get inserted on the highway. When P, Z and Q are in line-of-sight with each other, phase 5 is started, consisting in fine lane "insertion" tuning. Cohort management is invoked when phase 5 is about to terminate. Z is assigned the rank previously held by Q, and ranks held by Q and its followers are incremented.

N2N beaconing and V2V messaging shall be used jointly, in order to withstand simultaneous occurrences of telemetry and N2N beaconing (possibly transient) failures.

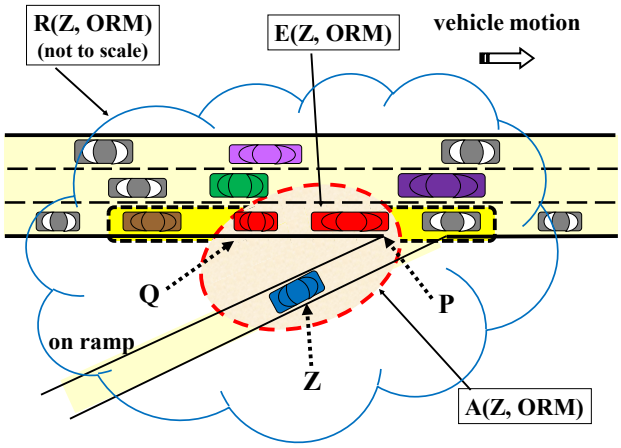

Figure 3. Groups with the On-Ramp-Merging SC scenario

Moreover, discharging V2V channels from handling the traffic load that can be efficiently carried via N2N channels has numerous advantages, notably smaller time bounds for SC-message deliveries, and avoidance of the thrashing phenomenon proper to V2V beaconing in the presence of high density traffic. One outcome of our work is that due to inaccuracies γ and τ, situational awareness based on V2V beaconing is feasible on a per cohort heads and tails basis, for some lower bound r°, not on a per vehicle basis.

V. CONCLUSIONS

We have introduced stationary scenarios and cohorts on the one hand, transitory scenarios and groups on the other hand, novel concepts and constructs aimed at endowing autonomous VANETs and platoons with efficiency and safety properties. Safety implying diversified functional redundancy, we have given an example of a solution to the problem of how to enforce safe inter-vehicle spacing in the presence of failing telemetry functions, and despite inaccurate vehicle space-time coordinates, along with companion analytical results. To this end, we have introduced N2N beaconing as well as a modeling based on a parameter η essential for differentiating SC scenarios from non SC scenarios. It has been established that the additional worst-case spacing c• needed to withstand telemetry failures does not depend on velocities. We have shown how a classical telemetry-based spacing algorithm can be used in a discrete time N2N periodic beaconing framework, and presented proofs of usability with parameter η and N2N beaconing period π. The group forming scheme associated with SC scenarios has been presented, whereby vehicles are assigned specific roles prior to undertaking risk-prone maneuvers, resorting to V2V communications. We have hinted at the communication reliability features intrinsic to groups, and illustrated these concepts with the on-ramp merging SC scenario. Communication-based algorithms and companion property proofs rest on postulating that it is possible to solve the (non stochastic) time-bounded MAC delays problem, a long standing problem with VANETs (with MANETs in general). It turns out that cohorts and groups, not initially devised to that end, are essential cornerstones for solving this problem. Solutions valid in our system model, notably in the presence of inaccurate space-time coordinates and a varying number of lanes, are presented in forthcoming papers. One class of "deterministic" protocols rests on solving the Instantaneous Renaming problem as it arises in mobile wireless networks. Novel solutions are needed, since renaming algorithms designed for static wired distributed systems (e.g., [16]) are inapplicable. Another issue also addressed in forthcoming papers is the dynamic generation of pseudonyms (privacy properties), a problem that bears a strong resemblance with the Instantaneous Renaming problem.

REFERENCES

[1] R.J. Caudill, W.L. Garrard, "Vehicle-follower longitudinal control for automated transit vehicles", Trans. ASME Journal of Dynanic Systems, Measurement, and Control, vol. 99, Dec. 1977, pp. 241-248.

[2] P.A. Bernstein, V. Hadzilacos, N. Goodman, "Concurrency control and recovery in database systems", Addison Wesley Pub., 1987, 370 p.

[3] I. Skog, P. Händel, "In-car positioning and navigation technologies—A survey", IEEE Trans. Intelligent Transport. Systems, vol. 10, 1, March 2009, pp. 4-21.

[4] R. Toledo-Moreo, D. Bétaille, and F. Peyret, "Lane-level integrity provision for navigation and map matching with GNSS, dead reckoning, and enhanced maps", IEEE Trans. Intelligent Transport. Systems, vol. 11, 1, March 2010, pp. 100-112.

[5] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing", IEEE Trans. Dependable and Secure Computing, vol.1, 1, 2004, pp. 11-33.

[6] T.L. Willke, P. Tientrakool, and N.F. Maxemchuk, "A survey of inter-vehicle communication protocols and their applications", IEEE Comm. Surveys and Tutorials, vol. 11, 2, 2nd quarter 2009, pp. 3-20.

[7] J.-F. Hermant, G. Le Lann, "A protocol and correctness proofs for real-time high-performance broadcast networks", 18th IEEE Intl. Conference on Distributed Computing Systems (ICDCS'98), Amsterdam, The Netherlands, 26-29 May 1998, pp. 360-369.

[8] J. Lygeros, D.N. Godbole and M. Broucke, "A fault tolerant control architecture for automated highway systems", IEEE Trans. Control Systems Technology, vol. 8, 2, March 2000, pp. 205-219.

[9] S.E. Shladover, "Longitudinal control of automotive vehicles in close-formation platoons", ASME Journal on Dynamic Systems, Measurement and Control, vol.113, 1991, pp. 231-241.

[10] S. Panwai, H. Dia, "Comparative evaluation of microscopic car-following behavior", IEEE Trans. Intelligent Transport. Systems, vol. 6, 3, Sept. 2005, pp. 314-325.

[11] A. Chakravarthy, K. Song, and E. Freron, "Preventing automotive pileup crashes in mixed-communication environments", IEEE Trans. Intelligent Transport. Systems, vol. 10, 2, June 2009, pp. 211-225.

[12] R. Ramanathan, J. Redi, C. Santivanez, D. Wiggins, and S. Polit, "Ad hoc networking with directional antennas: A complete system solution", IEEE Journal Selected Areas in Communications, vol. 23, 3, March 2005, pp. 496-506.

[13] D. Dolev, N.A. Lynch, S.S. Pinter, E.W. Stark, and W.E. Weihl, "Reaching approximate agreement in the presence of faults," Journal of the ACM, vol. 33, 3, July 1986, pp. 499-516.

[14] D. Dolev, C. Dwork, and L. Stockmeyer: "On the minimal synchrony needed for distributed consensus", Journal of ACM, vol. 34, 1, January 1987, pp. 77–97.

[15] G.M. Fitch et al, "Safety benefit evaluation of a forward collision warning system: final report", NHTSA DOT HS 810 910, 2008, 100 p.

[16] M. Okun, A. Barak, and E. Gafni, "Renaming in synchronous message passing systems with Byzantine failures", Distributed Computing (Springer), vol. 20, 6, 2008, pp. 403-413.