ChatGPT

Analysis of Misbehavior Report JSON

The **Metadata** section provides context for the report. It includes fields like `senderId`, `reportedId`, `generationTime`, `senderRealId`, `reportedRealId`, and `attackType` [1]. Here, `senderId` and `reportedId` are the (pseudo) identifiers of the reporter and the suspected node, respectively; `senderRealId` and `reportedRealId` are their true identities. `generationTime` is the timestamp when the report was created. The `attackType` string indicates the scenario (e.g. `"ConstPos"`), aligning with the DARE attacker model [2] [3]. For instance, `attackType = "ConstPos"` denotes a **constant-position** misbehavior (a position-malfunction case) by the reported vehicle [3]. This field and the IDs allow the Misbehavior Authority to link multiple reports to the same attacker and identify its nature. In DARE, all misbehavior reports follow this schema [1], ensuring the local reporter, target, time, and attack category are clearly recorded.

BsmCheck

The **BsmCheck** section summarizes the local detection outcome. In DARE's framework, each vehicle computes simple plausibility and consistency checks on received BSMs and derives a trust score [4]. Algorithmically, the minimum check value (`cmin`) is converted into a trust metric; if the global trust falls below a threshold, the vehicle is deemed malicious [4]. Thus, `BsmCheck` likely contains the result of this process (for example, a boolean "misbehaving" flag or a trust score). A low trust (near 0) or an explicit misbehavior flag would indicate that the local checks found anomalies. These checks yield scores in [0,1], where 1 means fully plausible and 0 means certainly malicious [5]. In practice, a `BsmCheck` entry might record this aggregated result of all per-message checks. It reflects the local detection logic: e.g. if the reporter's combined checks triggered a report (as per Algorithm 1 in DARE [4]), then `BsmCheck` would show that outcome. In summary, this section encapsulates the local misbehavior decision (trusted vs. flagged) based on the collected BSM evidence and plausibility tests.

BSMs

The **BSMs** section lists the Basic Safety Messages collected by the reporter from the suspected vehicle. Each BSM entry includes fields such as `"CreationTime"`, `"Pos"`, `"Speed"`, and `"Accel"` [6]. For example, `"Pos":[x,y,z]` gives the 3D position coordinates, while `"Speed":[vx,vy,vz]` and `"Accel":[ax,ay,az]` give the velocity and acceleration vectors at that time. These raw kinematic values form the evidence of behavior. Under normal operation, these values should change consistently: position should move over time if speed is non-zero. In a **ConstPos** scenario, we observe that the reported GPS coordinates remain (nearly) constant across successive BSMs even though the `Speed` field may be nonzero. This trend – unchanged `Pos` with varying `Speed` – is a clear red flag. It implies the vehicle claims to be stationary (constant position) while its reported speed suggests motion. The presence of this pattern in the BSM list is exactly the kind of anomaly that plausibility checks are designed to catch. For instance, if all `Pos` vectors are identical but the `Speed` vectors change, it indicates the attacker is falsifying data by transmitting a fixed location. This supports the classification of `attackType =`

`"ConstPos"` (constant position attack) and suggests the reporter recognized the inconsistency as misbehavior.

BsmChecks

The **BsmChecks** section contains the computed plausibility/consistency scores for each BSM listed above. Each element has fields like `"rangePlausibility"`, `"posPlausibility"`, `"posConsistency"`, and `"speedPlausibility"` [7]. In DARE (via F2MD), these checks evaluate whether the reported values are physically reasonable. For example, **range plausibility** checks if the claimed distance between vehicles is feasible given transmission constraints; **pos plausibility** checks if the position is in a physically possible location; **speed plausibility** checks if the speed lies within limits (e.g. below maximum road speed) [8]; and **pos consistency** checks if changes in reported position match the reported speed/acceleration. The scores are on a 0–1 scale (with values near 1 meaning "plausible" and values near 0 meaning highly suspect) [5] [8]. In our report, we expect to see very low (or even negative) values in these fields for the ConstPos attack. Specifically, the constant position with nonzero speed will cause the **posConsistency** check to fail severely. That is, if a vehicle claims to be at the same coordinates (`Pos`) but reports moving (`Speed` $\neq 0$), the computed distance vs. expected displacement will not match, yielding a very low consistency score. The **speedPlausibility** might also drop if the implausible motion is flagged. In contrast, `rangePlausibility` might remain close to 1 if the vehicle was actually within radio range throughout. These low plausibility/consistency values trigger the trust-based decision noted above [4]. Thus, the `BsmChecks` section quantitatively reflects why the vehicle was labeled malicious. In DARE's methodology, values below a threshold (e.g. near 0) indicate detected misbehavior [5] [8].

Attack Classification and Consistency with DARE Model

The report's contents match the DARE attacker model and detection pipeline. In DARE, the attacker is assumed **insider** (a legitimate vehicle with valid credentials) [2]. The ConstPos scenario corresponds to a *position malfunction* in their taxonomy: "the broadcasted position could be fixed" [3]. This is treated as a local attacker case (not a Sybil or collaborative attack), since it involves one vehicle constantly lying about its position. The report's `attackType = "ConstPos"` reflects this categorization. The observed BSM trend (unchanged GPS vs. changing speed) fits exactly this attack model and violates the consistency checks described in DARE [8]. The data in `BsmChecks` (very low consistency scores) backs up the identification. Finally, the report format (Metadata + BSMs + BsmChecks) is precisely what DARE prescribes for an initial misbehavior report [1] [7]. It shows how local detection (simple plausibility checks and a trust threshold [4]) generates an evidence report for the global authority. In summary, every part of the report – from metadata identifiers to the BSM entries and their check scores – aligns with the DARE description of a ConstPos attack by a local insider and the associated detection logic [4] [8].

Sources: The field definitions and check logic are based on the DARE dataset specification [1] [7] and its described misbehavior-detection methodology [8] [4]. The interpretation of plausibility scores follows F2MD's approach [5].

---

[1] [2] [3] [4] [6] [7] [8] dare dataset.pdf
file://file-Y4Z6ncY1uosiD7gcruSDic

5 AI algorithms for detecting faulty data in V2X communication
https://www.msg.group/en/publications/ai-algorithms-for-detecting-faulty-data-in-v2x-communication