# VeReMi扩展：一个用于比较评估VANET中不良行为检测的数据集

## 论文题目

VeReMi Extension: A Dataset for Comparable Evaluation of Misbehavior Detection in VANETs

VeReMi扩展：一个用于比较评估VANET中不良行为检测的数据集

发表时间：2020.06

会议：IEEE International Conference on Communications，CCF-C

## 研究背景和目的

|  |  |
| --- | --- |
| 研究背景  （原因） | VeReMi是车联网安全领域的唯一公开数据集（截止2022.7） |
| 研究目的 | VeReMi Extension基于VeReMi做了进一步改进 |
| 研究方法/创新点 | 1. 在车辆的物理层上设计并实现了一个真实的传感器误差模型 2. 实现了更大更复杂的攻击集。 3. 新的攻击可以操纵消息频率和数字证书以及消息内容。 4. 描述并实现了一组局部似然性检测器和一个简单的融合检测机制。然后根据这种错误行为融合机制对数据集进行测试。 |

## 数据集生成

### 19种不当行为模型

在本研究中，我们还旨在通过从文献[17]中使用的模型聚合的一组新攻击来扩展VeReMi攻击库。不当行为分为故障和攻击。前者是由OBU或车辆传感器故障引起的非恶意行为，而后者是车辆故意发送错误信息的恶意行为。

1） 故障：位置故障、速度故障、消息延迟

位置故障，通常是定位系统故障（例如GPS）造成的。可能表现为以下四种用例之一：

1. 恒定的错误位置
2. 各时间步均为随机位置
3. 在真实位置上的恒定偏移
4. 在真实位置上的随机偏移

速度故障，可能是OBU错误或物理传感器故障造成的。速度故障的产生与前面描述的位置故障类似：

1. 恒定的错误速度
2. 各时间步均为随机速度
3. 在真实速度上的恒定偏移
4. 在真实速度上的随机偏移
5. 消息延迟，可能是由于网络开销大或车载处理单元成本低或速度慢造成的。这些消息包含所有正确的数据和所需的信息，但从现实中以延迟Δt发送。

2） 攻击：

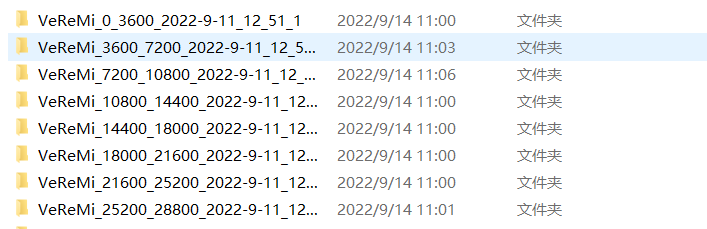
1. DoS攻击，包括车辆发送频率高于相应IEEE或ETSI标准设置的限制的消息。
2. DoS Random，是所有消息字段都设置为随机值的DoS攻击。这可能是一种充斥网络并阻止真实消息被广播的策略。这种攻击也可以在sybil 模式中执行，攻击者更改其在每个发送消息上的身份以避免检测。
3. 数据重放，包括发送以前从特定目标邻居那里收到的信息。重放的信息是用攻击者的证书签名的。它可以在Sybil模式下执行，攻击者在每个新选择的目标上改变其身份以避免被发现。
4. 破坏性攻击，是对以前收到的随机邻居的数据进行信息重放。它也可以是一种淹没网络并阻止真正的信息被广播的策略。这种攻击也可以在Sybil和DoS模式下执行（DoSDisruptive、DoSDisruptiveSybil）。
5. 最终停止，是一种车辆通过冻结位置值和设置速度值为空来模拟突然停止的攻击。
6. 交通拥堵Sybil，是一种旨在创造一个虚假的交通拥堵。攻击者通过为每辆假车保持一个新的身份和正确的信息频率，在选定的位置生成一个假车网格。

该表的Num编号符合VeReMi\_Extension的文件命名

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Num | AttackType | Description | 释义 |
| 故障 |  | ConstPos | 攻击者广播错误的固定位置 | 固定位置 |
|  | ConstPosOffset | 攻击者广播一个与实际位置有固定偏移的位置 | 固定偏移 |
|  | RandomPos | 攻击者在模拟区域内发送一个随机位置 | 随机位置 |
|  | RandomPosOffset | 攻击者在车辆周围的矩形中发送一个随机位置 | 随机偏移 |
|  | ConstSpeed | 攻击者广播错误的固定速度 | 固定速度 |
|  | ConstSpeedOffset | 攻击者广播一个与实际速度有固定偏移的速度 | 固定速度偏移 |
|  | RandomSpeed | 攻击者在一定范围内发送一个随机速度 | 随机速度 |
|  | RandomSpeedOffset | 攻击者在真实速度的上随机加入偏移 | 随机速度偏移 |
| 12 | DelayedMessages | 消息正确，但从现实中以延迟Δt发送 | 消息延迟 |
| 攻击 | 9 | EventualStop | 车辆通过冻结位置值和设置速度值为空来模拟突然停止 | 停止攻击 |
| 10 | Disruptive | 对以前收到的随机邻居的数据进行信息重放 | 破坏性攻击 |
|  | DataReply | 重复发送以前从特定目标邻居那里收到的信息 | 数据重放 |
|  | Dos | 发送频率高于相应IEEE或ETSI标准设置的限制的消息  发送间隔0.5s、0.25s等，正常是1s | Dos攻击 |
|  | DosRandom | 所有消息字段都设置为随机值的DoS攻击 | 随机Dos |
|  | DoSDisruptive | Dos攻击+破坏性攻击 |  |
|  | GridSybil | 一辆车有多个假名，发送消息的频率也更高 | 假名攻击 |
|  | DataReplySybil | DataReply+ Sybil |  |
|  | DoSRandomSybil | DosRandom+ GridSybil |  |
|  | DoSDisruptiveSybil | Dos攻击+破坏性攻击+Sybil |  |
|  |  | MixALL | 混合了上述所有攻击 | 混合攻击 |

注：每种攻击都有两次模拟，一个模拟处于高峰时间（7h-9h），另一个模拟位于低峰时间（14h-16h），MixALL只有一个00h-24h的模拟，因此总共有39个数据集。

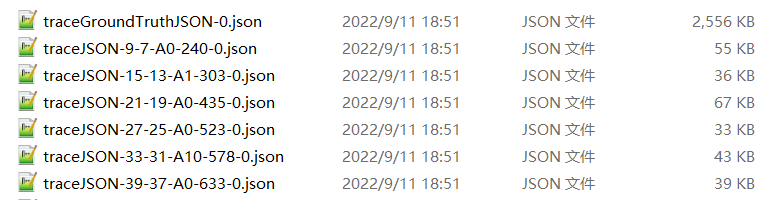
### 数据集命名规则



以VeReMi\_0\_3600\_2022-9-11\_12\_51\_1为例：

发送时间在0s\_3600s (0~1h)之间，数据集生成时间在2022-9-11 12:51:1

子文件：



a) traceGroundTruthJSON-7.json：地面实况文件，记录所有车辆发送的消息（正常消息和异常消息）。

b) traceJSON-9-7-A0-240-7.json：车辆日志文件，记录着当前车辆接收到的附近车辆发送的 BSM 数据，包括type2和type3类型的消息。9-7：车辆ID, ID-2；A0：行为正常的车辆，A1-A19表示各类异常行为的车辆；240：车辆生成时间 / 接收和发送的开始时间； 0h：车流开始时间。

可通过traceJSON文件名来找到所有恶意车辆的ID。

### 字段格式

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名 | 字段解释 | 数据格式 | 数据解释 |
| 1. type | 日志类型 | 2,3,4 | 2：接收  3：发送并接收  4：发送 |
| 1. rcvTime | 接收时间戳 | double |  |
| 1. sendTime | 发送时间戳 | double |  |
| 1. sender | 发送者ID | Int |  |
| 1. senderPseudo | 发送者假名ID | Int | Sybil攻击中，一个sender可能有多个senderPseudo |
| 1. messageID | 唯一的消息ID | Int | 日志的主键 |
| 1. pos | 位置矢量 | [x. , y. , z.] | 空间坐标 |
| 1. pos\_noise | 位置噪声矢量 | [x. , y. , z.] | 一般为[0,0,0] |
| 1. spd | 速度矢量 | [x. , y. , z.] | z一般为0 |
| 1. spd\_noise | 速度噪声矢量 | [x. , y. , z.] | 一般为[0,0,0] |
| 1. acl | 加速度矢量 | [x. , y. , z.] | z一般为0 |
| 1. acl \_noise | 加速度噪声矢量 | [x. , y. , z.] | 一般为[0,0,0] |
| 1. hed | 航向矢量 | [x. , y. , z.] | z一般为0 |
| 1. hed\_noise | 航向噪声矢量 | [x. , y. , z.] | 一般为[0,0,0] |
|  |  |  |  |

注：四种误差计算公式见小节：传感器误差模型

### 数据集解释

以下特征均通过程序验证，不用质疑其准确性，除非数据集发生变化。

1. 可通过traceJSON文件名来找到所有异常车辆ID。比如traceJSON-15-13-A1-25207-7.json，表示sender为15的车辆为异常车辆（异常类型为1）
2. 异常sender发送的一定是异常消息，满足3个条件才算是异常消息：发送类型为3，sender为异常车辆，messageID没有出现过（防止重复）。
3. 条件4：异常sender在GT文件中的消息和在车辆日志文件中的消息不一致，即异常sender发送消息时会修改GT文件的消息内容。
4. 消息是否异常取决于sender是否异常，而非判断其是否出现在异常文件中。
5. 同一messageID会同时出现在多个车辆中，因为消息是广播的，异常车辆和正常车辆都可以接收，但是同一文件只会出现一次**。**
6. 此外，traceJSON中有的消息traceGroundTruthJSON却不一定有：车辆空消息，可能是地面真值文件信息丢失。
7. 反之traceGroundTruthJSON有的traceJSONs也不一定有：地面真值空消息，发送的消息可能无车辆接收。
8. 因此，traceGroundTruthJSON中没有的消息甚至可能在不同的traceJSON出现多次，需要避免重复统计，格外注意。
9. 将traceGroundTruthJSON的消息数视为总消息数量，只考虑traceGroundTruthJSON中的异常消息和真实消息，有如下数量关系：
10. 异常消息、真实消息的范围扩展到全部traceJSONs中，有如下数量关系：

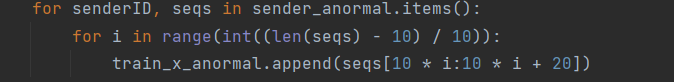
### 生成消息标签和时间序列

打标签思路：先通过traceJSON文件名来找到所有异常车辆ID， 需要注意的是，异常车辆发送的都为异常消息，且不同车辆可能会接收到同一异常消息，因此需要甄别，避免重复，然后遍历车辆日志文件的所有消息，满足3个条件才算是异常消息：发送类型为3，sender为异常车辆，messageID没有出现过

时间序列生成：

遍历GT文件生成正常序列，遍历所有车辆日志文件生成异常序列：

以sender为单位生成字典：{sender:消息序列}



然后以20为时间序列长度，10为步长生成时间序列，时间序列数：

对于一个有88条消息的车辆，能够生成 = 条时间序列。

各数据集结果：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Num | AttackType(0708) | 真实消息：异常消息 | 真实序列：异常序列 |
| 故障 | 1 | ConstPos | 189402:52380 | 16287:4454 |
| 2 | ConstPosOffset | 189402:52380 | 16287:4454 |
| 3 | RandomPos | 189402:52380 | 16287:4454 |
| 4 | RandomPosOffset | 189402:52380 | 16287:4454 |
| 5 | ConstSpeed | 189402:52380 | 16287:4454 |
| 6 | ConstSpeedOffset | 189402:52380 | 16287:4454 |
| 7 | RandomSpeed | 189402:52380 | 16287:4454 |
| 8 | RandomSpeedOffset | 189402:52380 | 16287:4454 |
| 12 | DelayedMessages | 189402:52380 | 16287:4454 |
| 攻击 | 9 | EventualStop | 189402:52380 | 16287:4454 |
| 10 | Disruptive | 189402:52380 | 16287:4454 |
| 11 | DataReply | 189402:52380 | 16287:4454 |
| 13 | Dos | 298051:155862 | 27141:14798 |
| 14 | DosRandom | 298609:156714 | 27191:14870 |
| 15 | DoSDisruptive | 298149:155868 | 27151:14796 |
| 16 | GridSybil | 352887:207824 | 32635:19987 |
| 17 | DataReplySybil | 189402:52380 | 16287:4454 |
| 18 | DoSRandomSybil | 244099:104474 | 21745:9647 |
| 19 | DoSDisruptiveSybil | 244099:104474 | 21745:9647 |
|  | 正常：异常 | Sum of above | 4198120:1566156 | **369339:141647** |

MixAll\_00\_24数据集结果：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Num | AttackType | 消息数 | 序列数 |
| 故障 | 0 | Genuine | 3050501 | 269578 |
|  | ConstPos | 38707 | 3309 |
|  | ConstPosOffset | 38221 | 3270 |
|  | RandomPos | 37754 | 3224 |
|  | RandomPosOffset | 36847 | 3127 |
|  | ConstSpeed | 36864 | 3144 |
|  | ConstSpeedOffset | 37234 | 3181 |
|  | RandomSpeed | 36671 | 3113 |
|  | RandomSpeedOffset | 36665 | 3118 |
| 12 | DelayedMessages | 36614 | 3103 |
| 攻击 | 9 | EventualStop | 37382 | 3173 |
| 10 | Disruptive | 37885 | 3234 |
|  | DataReply | 38331 | 3283 |
|  | Dos | 116120 | 11052 |
|  | DosRandom | 112726 | 10710 |
|  | DoSDisruptive | 114689 | 10908 |
|  | GridSybil | 150180 | 14445 |
|  | DataReplySybil | 38654 | 3304 |
|  | DoSRandomSybil | 72708 | 6710 |
|  | DoSDisruptiveSybil | 73244 | 6764 |
| 正常:异常 | | Sum of above | 3050501：1127496 | 269578：102172 |