Spatial and temporal cross validation strategy for misbehavior detection in C-ITS

A. *Basic safety message*

The basic safety message (BSM) is the core of V2V communications in C-ITS. It contains essential pieces of information required for enhancing the safety of the vehicles by providing sensitive information about the current state of the vehicle to other vehicles in the network. In [1] we can find details about this message that is shared between the vehicles in a broadcast manner. Information such as position, heading, speed, acceleration, brake status is necessary to implement safety in the C-ITS. Unfortunately, these pieces of data might be wrong or maliciously tampered, which can deceive the perception of the vehicle and puts it at risk. We discuss some of these tampering attacks present in the VeReMi dataset in section **??**. The application dataset in this paper is a collection of such BSM messages presented in the next section.

B. *VeReMi Dataset*

C. *Attacks*

I. Evaluation

- **AdaBoost** [2] a boosting method that fits multiple classifiers on the dataset, focussing on difficult cases each time.
- **Decision Tree** [3] the classifier infers a set of rules from the data structure to classify the different samples.
- **Naive Bayes** [4] is a supervised learning method based on Bayes' theorem with the "naive" assumption of conditional independence.
- **Nearest Neighbors** [5] is a supervised learning algorithm that classifies a sample with respect to the dominant class in its neighberhood.
- **Neural Net** [6] is a supervised learning algorithm that models the underlying function of the data. We use a simple one-layer neural network.
- **Random Forest** [7] is an algorithm that combines multiple decision trees to make the classification.

A. *Metrics*

- The precision measure for each attack, how precise the model is, i.e., the proportion of the correctly detected attacks (true positives TP) w.r.t the overall set of predicted attacks (true positives TP + false positives FP).

$$precision = \frac{TP}{TP + FP}; \tag{1}$$

- The recall measure for each attack, how much of the existing attacks are detected by the model, i.e., the proportion of the correctly detected attacks (true positives TP) w.r.t the overall set of attacks (true positives TP + false negatives FP).

$$recall = \frac{TP}{TP + FN}; \tag{2}$$

- The $F_1 - score$ is the harmonic mean of precision and recall.

$$F_1 - score = 2.\, \frac{precision\, .\, recall}{precision + recall} \tag{3}$$

References

[1] J.-W. Kim, J.-W. Kim, and D.-K. Jeon, "A cooperative communication protocol for qos provisioning in ieee 802.11 p/wave vehicular networks," *Sensors*, p. 3622, 2018.

[2] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*. Springer, 1995, pp. 23–37.

[3] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, p. 81–106, 1986.

[4] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, p. 41.

[5] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, pp. 21–27, 1967.

[6] M. H. Beale, M. T. Hagan, and H. B. Demuth, "Neural network toolbox user's guide," *The Mathworks Inc*, 1992.

[7] G. Louppe, "Understanding random forests: From theory to practice," *arXiv preprint arXiv:1407.7502*, 2014.

Type	Description	Size (byte)
DSRCmsgID	Data elements used in each message to define the Message type	1
MsgCount	It can check the flow of consecutive messages having the same DSRCmsgID received from the same message sender.	1
TemporaryID	Represents a 4-byte temporary device identifier. When used in a mobile OBU device, this value is periodically changed to ensure anonymity.	4
Dsecond	Represents two bytes of time information.	2
Latitude	Represents the geographic latitude of an object.	4
Longitude	Represents the geographic longitude of an object.	4
Elevation	Represents an altitude measured by the WGS84 coordinate system.	2
PositionAccuracy	Various quality parameters used to model the positioning accuracy for each given axis.	4
TransmissionAndSpeed	Represents the speed of the vehicle.	2
Heading	The current direction value is expressed in units of 0.0125 degrees.	2
SteeringWheelAngle	Represents the current steering angle of the steering wheel.	1
AccelerationSet4Way	It consists of three orthogonal directions of acceleration and yaw rate.	7
BrakeSystemStatus	Represents a data element that records various control states related to braking of the vehicle.	2
VehicleSize	Represents the length and width of the vehicle.	3

TABLE I: Basic safety message (BSM) information [1]

Features	Description	Symbol
Type	identifier for message type	ID
Reception Time	time BSM was received by the receiver	Rt
Receiver ID	Id of the receiving vehicle	RID
Receiver X position	receiving vehicle x coordinate	RXP
Receiver Y position	receiving vehicle y coordinate	RYP
Receiver Z position	receiving vehicle z coordinate	RZP
Transmission Time	time BSM was emitted by the emitter	Tt
Transmitter ID	Id of the transmitting vehicle	TID
BSM ID	Id of the message	MID
Transmitter X position	transmitting vehicle x coordinate	TXP
Transmitter Y position	transmitting vehicle y coordinate	TZP
Transmitter Z position	transmitting vehicle z coordinate	TZP
Transmitter X velocity	transmitting vehicle x velocity	TXV
Transmitter Y velocity	transmitting vehicle y velocity	TYV
Transmitter Z velocity	transmitting vehicle z velocity	TZV
RSSI	received Signal Strength Indicator	RSSI
Label ID	(0=Normal Behavior)	L

TABLE II: VeReMi dataset for connected and automated vehicles

Label ID	Description	Parameters
1: Constant	Attacker transmits a fixed location	x = 5560, y = 5820
2: Constant Offset	Attacker transmits a fixed, offset added to the real position	Δx = 250, Δy = -150
4: Random	Attacker sends a random position inside the simulation area	uniformly random in playground
8: Random Offset	Attacker sends a random position in a rectangle around the vehicle	Δx , Δy are uniformly random from [-300,300]
16: Eventual Stop	Attacker behaves normally for some time and then attacks by transmitting the same position repeatedly	Stop probability increases by 0.025 each position update

TABLE III: Attack Definition

Attack	split	Ada Boost	Decision Tree	Naive Bayes	Nearest Neighbors	Neural Net	Random Forest
Constant	*Random*	0.999722	0.999705	0.999328	0.998577	0.986341	0.999902
	Temporal	0.999667	0.999648	0.999608	0.998787	0.991516	0.999941
Constant offset	*Random*	0.957469	0.952769	0.526931	0.988706	0.907642	0.986768
	Temporal	0.965516	0.938025	0.539810	0.990687	0.881437	0.976016
Eventual stop	*Random*	0.946040	0.913531	0.882501	0.849532	0.908969	0.943108
	Temporal	0.946748	0.914211	0.879039	0.827747	0.938937	0.939966
Genuine	*Random*	0.990301	0.987103	0.951445	0.980236	0.971481	0.993612
	Temporal	0.991418	0.986339	0.949752	0.981043	0.981876	0.992886
Random	*Random*	0.999132	0.999482	0.999223	0.998043	0.997825	0.999508
	Temporal	0.999047	0.999047	0.999190	0.998888	0.997981	0.999508
Random offset	*Random*	0.969173	0.968529	0.922016	0.879568	0.845031	0.985895
	Temporal	0.972464	0.966641	0.901854	0.885191	0.954407	0.986975

TABLE IV: $F_1 - score$ using different splits and machine learning methods