An Ontology Based Anomaly Detection System for Vehicular Communications

Quentin Ricard, Philippe Owezarski
qricard@laas.fr owe@laas.fr

ERTS - 2020
10th European Congress on Embedded Real Time Software and Systems

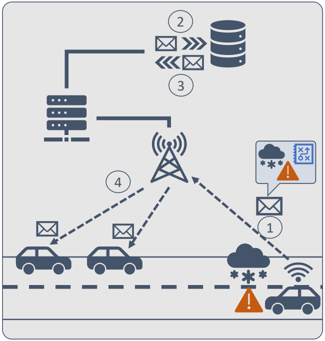
January 29, 2020

Outline

Context

E-horizon project: Continental

- Improving 3 types of services:
  1. Safety:
     - Ex: Weather condition, maximum advised speed;
  2. Fleet monitoring:
     - Ex: fuel consumption, premature wear detection;
  3. User experience:
     - Ex: ETA, points of interest;



- Implies creating a new communication channel between vehicles and the rest of the world

Problem Statement

New attack vector

- Jeep Cherokee, Miller and Valasek [2015]
- Nissan Leaf, Troy Hunt [2016]
- Volkswagen/Audi, Keuper and Alkemad Computest [2018]

How can we prevent these new attacks and protect vehicles?

Anomaly Detection

- Apply existing methods to the automotive field:
  - Adapt algorithms to mobile network traffic;
  - Use relevant communication datasets;

Problem Statement

New attack vector

- Jeep Cherokee, Miller and Valasek [2015]
- Nissan Leaf, Troy Hunt [2016]
- Volkswagen/Audi, Keuper and Alkemad Computest [2018]

How can we prevent these new attacks and protect vehicles?

Anomaly Detection

- Apply existing methods to the automotive field:
  - Adapt algorithms to mobile network traffic;
  - Use relevant communication datasets;

Challenges

Nature of the communications

- Vehicle-related traffic:
  - Built from sensors and actuators of the vehicle.
- User-related traffic:
  - Use of infotainment applications e.g. e-mails, music streaming.

Requirements for the detector

- Online
- Small footprint
- Broad spectrum of detection

Ontological representation of the communications

Flow Class

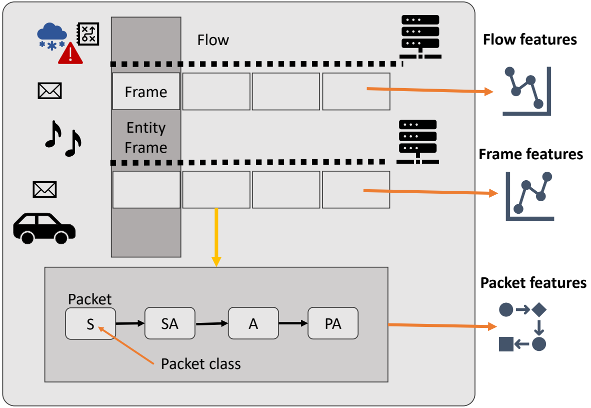
- Differentiated by IPs and ports

Frame Class

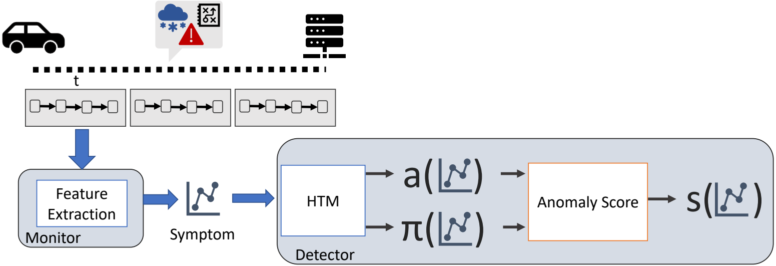
- Collection of packets received during δt window

Packet Class

- Semantic description of the packet



Anomaly Detection Process



HTM algorithm

- Hierarchical temporal memory algorithm Hawkins and Blakeslee [2007]
- Online and unsupervised Ahmad *et al.* [2017]

Anomaly representation and Inference Rules

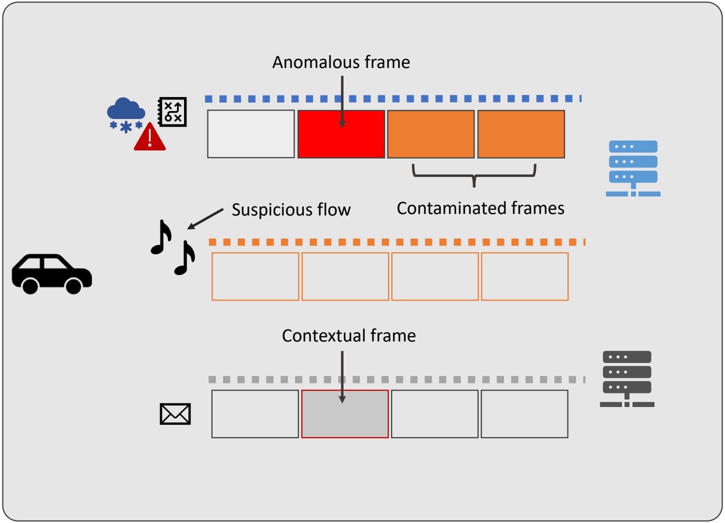
What is considered anomalous ?

- Every frame whose anomaly score > threshold
- Related frames and flows over the same period

Inference Rules

- Contaminated Frames
- Suspicious Flows
- Contextual Frames

Inference Rules



Evaluation

Autobot emulation environnement

- Ricard and Owezarski [2019]
- Communicating applications
- Telemetry and infotainment traffic

Dynamic generation of anomalies

- Port Scan
- DNS Tunneling
- Telemetry anomaly

Autobot

Docker

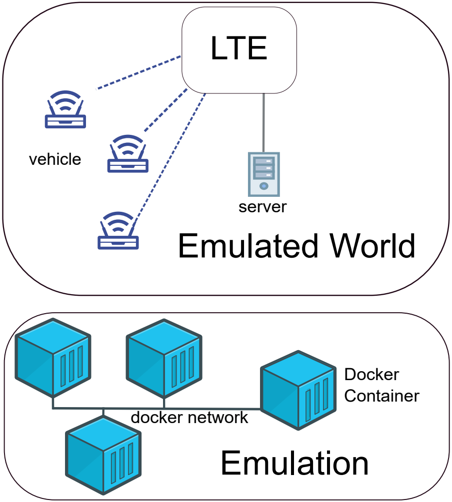
- OS-level virtualization tool
- Containers embed realistic applications

Docker Network

- Connects containers to one another & Internet
- Route packets

Traffic Control

- Emulates cellular network connectivity
- Shapes the traffic



Embedded applications

Telemetry

- Aggregated CAN bus data
- Sensoris message format
- Send messages over MQTT session

Infotainment

- Spotifyd
- Waze

Detection Evaluation

Label	No Ontology	Ontology	
		No Inference	Inference
FPR	2.3% (97/3291)	3.4% (298/8736)	14% (1228/8736)
Scan	0% (0/2)	0.6% (7/1024)	0.6% (7/1024)
DNS	1.4% (3/211)	9.4% (25/264)	39.0% (103/264)
Tele	3.7% (1/27)	27.2% (3/11)	90.9% (10/11)

Frame based detection

- Detection of frames containing anomalies

Comparison with other algorithms

Label	HTM		OCSVM		DBSCAN	
	S1	S2	S1	S2	S1	S2
FPR	6.6%	3.4%	0.16%	37.1%	68.3%	0%
Scan	0.1%	0.6%	97.7%	97.7%	0%	0%
DNS	6.1%	9.4%	0%	96.2%	100%	0%
Tele	9%	27.2%	0%	90.1%	100%	0%

Feature Set

- S1 : 44 features based on Lashkari *et al.* [2017].
- S2 : packets/s, mean packet length in forward direction and average packet size.

Conclusion

Detection Results and ontology

- HTM obtains good results with relatively few features
- Broad spectrum of detection
- Communication model has great impact on the detection

Current and future work

- Feature selection instead of feature weighting
- Reduce false positives using score based on history (Anomaly likelihood)

Thank you for your attention.

References I

Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262:134–147, 2017.

Computest. The connected car-ways to get unauthorized access and potential implications. *Online: https://www.computest.nl/wp-content/uploads/2018/04/connected-car-rapport.pdf*, 2018.

Jeff Hawkins and Sandra Blakeslee. *On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines*. Macmillan, 2007.

References II

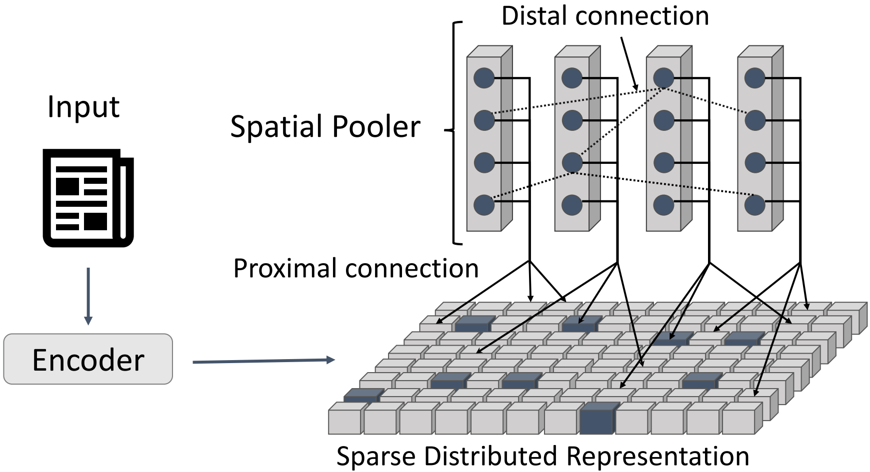
Troy Hunt. Controlling vehicle features of nissan leafs across the globe via vulnerable apis. 2016.

Arash Habibi Lashkari, Gerard Draper-Gil, Mohammad Saiful Islam Mamun, and Ali A Ghorbani. Characterization of tor traffic using time based features. In *ICISSP*, pages 253–262, 2017.

Charlie Miller and Chris Valasek. Remote exploitation of an unaltered passenger vehicle. 2015.

Quentin Ricard and Philippe Owezarski. Autobot: An emulation environment for cellular vehicular communications. In *Proceedings of the 2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications*. IEEE Computer Society, 2019.

HTM


Input
Encoder
Spatial Pooler
Distal connection
Proximal connection
Sparse Distributed Representation

Features

totalfpackets	totalbpackets	totalfpktl	totalbpktl
fpktspersecond	bpktspersecond	flowpktspersecond	flowbytespersecond
minfpktl	minbpktl	maxfpktl	maxbpktl
meanfpktl	meanbpktl	stdfpktl	stdbpktl
varfpktl	varbpktl	totalfiat	totalbiat
minfiat	minbiat	maxfiat	maxbiat
meanfiat	meanbiat	stdfiat	stdbiat
varfiat	varbiat	varflowpktl	varflowiat
minflowpktl	maxflowpktl	meanflowpktl	stdflowpktl
minflowiat	maxflowiat	meanflowiat	stdflowiat

Features stored inside the ontology

Tools

- Docker
  - https://docs.docker.com/v17.09/
- Traffic-control :
  - http://man7.org/linux/man-pages/man8/tc-netem.8.html
- Spotifyd
  - https://github.com/Spotifyd/spotifyd