Apprentissage d'une stratégie d'attaque contre un véhicule autonome

Responsable : Guillaume Hutzler
Contact : guillaume.hutzler@ibisc.univ-evry.fr
Collaboration : Hanna Klaudel (IBISC), Artur Rataj (SystemX), Witold Klaudel (SafeTech Cybernetics)
Mots-clés : véhicules autonomes, analyse de risque, apprentissage automatique, Q-learning

Thématique : Implantation d'un algorithme de Q-learning pour l'apprentissage d'une stratégie d'attaque d'un véhicule autonome afin d'en détecter les vulnérabilités

Le contexte : Dans le cadre des véhicules autonomes, ou plus généralement des *Systèmes de Transport Intelligents*, un enjeu majeur est la protection contre les cyber-attaques. L'ensemble du fonctionnement d'un véhicule autonome est en effet sous le contrôle d'algorithmes qui analysent les données des capteurs (GPS, caméras, ultrasons, LIDARs, etc.) pour reconstituer la « scène » dans laquelle se trouve le véhicule, décident et planifient les actions futures à entreprendre en fonction de la situation, et commandent les actionneurs (moteur, volant, pédales, etc.) permettant de contrôler la trajectoire du véhicule. Tout ceci s'effectue par ailleurs en communication permanente avec les autres véhicules et l'infrastructure routière. De même que n'importe quel objet connecté, un véhicule autonome est donc exposé au risque de cyber-attaque, avec des conséquences qui pourraient être catastrophiques. Il s'agit donc de s'assurer que des attaques simples ne peuvent engendrer que des conséquences bénignes, et qu'à l'inverse le risque d'attaque entraînant des conséquences graves est négligeable. Pour ce faire, l'objectif est d'entraîner un système d'apprentissage automatique, afin qu'il apprenne les stratégies d'attaque les plus efficaces contre un véhicule, et identifier le cas échéant des vulnérabilités du système qu'il faudra corriger.

L'existant : Un véhicule autonome est constitué d'une multitude d'éléments ou modules interconnectés, aussi bien matériels (capteurs, actionneurs, processeurs, matériel réseau, matériel télécom, etc.) que logiciels (systèmes d'exploitation, applications d'analyse, de décision, de contrôle, de communication, etc.). Dans le cadre d'une collaboration avec SystemX, un formalisme a été proposé pour modéliser l'ensemble de ces différents modules et de leurs interactions [2]. Le formalisme permet également de modéliser, quand un module se retrouve dans un état non fonctionnel (*malware*, *bad data*, ou *not available*), le risque que cet état se propage à d'autres modules qui pourraient à leur tour devenir non fonctionnels.

Objectif : Différentes approches, issues des méthodes formelles, sont actuellement explorées [1]. Dans l'hypothèse où ces approches seraient mises en échec, du fait d'un trop grand nombre de modules à modéliser résultant en un espace d'états à explorer trop grand, nous souhaitons explorer en parallèle des approches exploitant des techniques d'apprentissage automatique pour essayer de détecter automatiquement des stratégies d'attaque permettant d'atteindre tel ou tel module vital du véhicule, révélant ainsi des vulnérabilités à corriger. L'approche envisagée se fonde sur les algorithmes de Q-learning [3].

Le travail consistera en :

- Une rapide revue des principales variantes d'algorithmes de Q-learning
- La sélection d'une variante adaptée au problème exposé
- La recherche d'une implantation open source de l'algorithme sélectionné et son adaptation au problème, ou son implantation s'il n'est pas possible d'en trouver une existante
- L'application à un modèle test

[1] Johan Arcile, Raymond R. Devillers, Hanna Klaudel, Witold Klaudel, Bozena Wozna-Szczesniak, "Modeling and checking robustness of communicating autonomous vehicles". DCAI 2017: 173-180.

[2] Witold Klaudel, Artur Rataj, "Towards a Formalisation of Expert's Knowledge for an Automatic Construction of a Vulnerability Model of a Cyberphysical System". ICISSP 2021: 391-398.

[3] Beakcheol Jang, Myeonghwi Kim, Gaspard Harerimana, Jong Wook Kim, "Q-Learning Algorithms: A Comprehensive Classification and Applications". *IEEE Access*, pp. 1-1. 10.1109/ACCESS.2019.2941229, 2019.