Machine Learning Based Sender Identification for Controller Area Network

Oleg Schell
Karlsruhe Institute of Technology

Marcel Kneib
Bosch Engineering GmbH

29th Crypto Day, 6/7 September 2018

Since its deployment, the Controller Area Network (CAN) bus system emerged to one of the de facto standards for intra-vehicle communication. However, security mechanisms were not taken into account during design phase, which nowadays offer a broad surface for attackers to compromise control units (ECU) and to inject malicious messages. For closing this security gap and to ensure authenticity, the use of traditional security mechanisms such as Message Authentication Codes (MAC) has been researched and implemented. This, however, would lead to an increased communication bandwidth and the need to change existing controllers, which makes this approach inconvenient in practice. Exploiting the unique physical characteristic of an ECU during transmission, Murvay & Groza (2014) were able to identify the source of a message in a CAN-based network.

We use these distinct voltage characteristics to develop a machine learning based sender identification on a resource limited microcontroller to be deployed in a dedicated Intrusion Detection System (IDS). After both time and frequency domain features have been extracted, a machine learning model is trained for the distinction of the connected ECUs. To make a decision on the model to be chosen, we evaluate the suitability of different approaches for the application on an embedded system. Since the voltage characteristics on the network might change because of temperature and aging, the possibility of dynamically adapting the learned models during operation is also investigated. Due to these properties and one's own preferences, an approach can finally be selected considering memory consumption, detection accuracy and computational complexity.

In general, we show that the sender identification on CAN-based networks can be realized on embedded and resource limited hardware using lightweight machine learning approaches. By evaluating data from a vehicle, we achieve identification rates of over 98% with the time required to build a model being in the range of seconds and minutes. The investigation of the adaptability to a simulated voltage drop showed that the approaches can adapt in a changing setting, allowing their use in a dedicated IDS within a vehicular environment.

References

Pal-Stefan Murvay & Bogdan Groza (2014). Source identification using signal characteristics in controller area networks. *IEEE Signal Processing Letters* **21**(4), 395–399.

crypto day matters 29 (2018)
Beiträge vom 29. Krypto-Tag, 06.-07.09.2018, Bosch, Renningen.
doi:10.18420/cdm-2018-29-05