Evaluation of anomaly detection for in-vehicle networks through information-theoretic algorithms

Mirco Marchetti, Dario Stabili, Alessandro Guido and Michele Colajanni
University of Modena and Reggio Emilia, Italy
Email: {mirco.marchetti, dario.stabili, alessandro.guido, michele.colajanni}@unimore.it

***Abstract*—This paper evaluates the effectiveness of information-theoretic anomaly detection algorithms applied to networks included in modern vehicles. In particular, we focus on providing an experimental evaluation of anomaly detectors based on entropy. Attacks to in-vehicle networks were simulated by injecting different classes of forged CAN messages in traces captured from a modern licensed vehicle. Experimental results show that if entropy-based anomaly detection is applied to all CAN messages it is only possible to detect attacks that comprise a high volume of forged CAN messages. On the other hand, attacks characterized by the injection of few forged CAN messages attacks can be detected only by applying several independent instances of the entropy based anomaly detector, one for each class of CAN messages.**

I. Introduction

Almost all activities in modern vehicles are controlled by software, including many safety-relevant functionalities such as steering, braking, traction and cruise controls. Common city cars are equipped with about 80 Electronic Control Units (ECUs) that continuously exchange messages through a Controller Area Network (CAN). This technology has been first deployed on licensed vehicles in 1989, and was designed to satisfy all functional requirements of the automotive industry, without taking information security into account.

Several studies already highlighted severe security issues of the CAN bus, that offers no protection against attackers willing to sniff, spoof and forge arbitrary CAN messages [1], [2]. Moreover, the protection of in-vehicle networks is becoming increasingly important as more and more manufacturers are producing *always connected* cars, equipped with infotainment and telematics systems that leverage SIM cards and mobile technologies to constantly send and receive data through Internet connections. Always connected cars pave the way for remote exploitation [3] as already demonstrated in [4], where attackers have been able to take over safety-relevant functionalities of an unmodified licensed vehicle from the Internet, without requiring physical access to the CAN bus.

Several works already proposed the application of anomaly detection algorithms to analyze CAN messages [5], [6], [7], [8], [9], [10], [11] looking for evidences of attacks and other illicit activities, but often with very limited evaluation over real in-vehicle network traffic. To address this issue, this paper proposes an experimental evaluation of the effectiveness of an anomaly detection algorithm based on the computation of entropy [12] and applied to the CAN messages exchanged through in-vehicle network.

In particular, the proposed anomaly detection algorithm is trained over CAN traffic traces captured from a modern licensed vehicle during several hours of driving over public motorways, subject to real traffic conditions. Attacks are then simulated by injecting different kind of forged CAN messages at variable rates in these traffic traces, thus mimicking attackers that are trying to compromise safety-relevant functions of the vehicle while driving at high speed. Finally, we evaluate the effectiveness of the entropy-based algorithm by verifying its ability to identify anomalies in the CAN traces that include forged messages. To the best of our knowledge, this is the first paper that provides an extensive evaluation of the effectiveness of entropy-based anomaly detection algorithms applied to real CAN traffic gathered from an unmodified licensed vehicle in real driving conditions.

The remainder of the paper is organized as follows. Section II discusses related work and outlines the main novelties of this paper. Section III introduces the concept of entropy in information theory and describes the entropy-based anomaly detection algorithm. Section IV provides a detailed description of the experimental evaluation and discusses all experimental results. Finally, Section V concludes the paper and outlines future works.

II. Related work

This paper relates directly to two research fields: anomaly detection and information security for automotive environment.

A. Anomaly detection

The field of anomaly detection has been explored widely by the related work [13]. Among all algorithms proposed in the literature, this paper assesses the effectiveness of an information theoretic anomaly detector [14], based on the computation of entropy [12].

The application of entropy-based anomaly detectors to identify several kind of attacks and anomalies in computer networks has already been proposed in several papers. In particular [9] shows the effectiveness of entropy-based approaches for identifying network traffic anomalies caused by the propagation of Internet worms. Other works [15] apply the same family of detectors to mobile cellular traffic.

While previous work provided extensive experimental evaluation of entropy-based anomaly detectors to computer networks, we highlight that in-vehicle networks are characterized by completely different workloads, and CAN messages have different features with respect to IP packets (as an example, CAN messages do not include source and destination addresses). Hence this paper provides a novel contribution by evaluating entropy-based anomaly detection over real CAN messages flowing through the CAN bus of a modern unmodified vehicle.

B. *Information security for automotive*

A more recent and less established research area addresses information security issues applied to electronic control units (ECUs), infotainment and telematics systems that are common in modern vehicles.

Several papers presented different kind of attacks that can be performed by sniffing and injecting messages over the CAN bus [1], [16], [2], as well as by mangling messages used by the Tire Pressure Monitoring System [17] and passive keyless entry systems [18]. Of particular interest are papers describing attacks carried out remotely by exploiting the permanent Internet connection of *always connected* cars [3], [4].

These papers motivated several research efforts focusing on improving the security of modern vehicles. Some works in this area propose new cryptographic libraries or hardware modules to support confidentiality and integrity through encryption [19], [20], [21]. While promising, these approaches are unpractical since they would require expensive modifications to all ECUs.

More related approaches propose the realization of anomaly detectors to analyze CAN traffic and identify possible anomalies related to attacker activities. Several different detectors are proposed in [10] and [11], however their proposals are not experimentally evaluated. Other popular approaches for anomaly detection in in-vehicle networks are based on Support Vector Machines (SVM) [5], [6], [7].

The previous work that relates more closely to this paper is [8], that proposed the application of entropy-based anomaly detection algorithms to in-vehicle networks. However, their experimental evaluation is very limited, and spans over just 15 seconds of CAN traffic including only a single class of CAN messages that are not safety-relevant.

To the best of our knowledge, this is the first paper that proposes an entropy-based anomaly detection algorithm for in-vehicle networks and that includes an extensive experimental evaluation to assess its effectiveness. In particular, experiments carried out in this paper include several hours of CAN traffic captured from an unmodified licensed vehicle over several hours of driving in a public motorway with real road and traffic conditions.

III. Entropy-based anomaly detection for CAN traffic

Anomaly detection [13] can be defined as the process of analysing a set of data aiming at identifying patterns that differ significantly from the expected *normal* behavior. These patterns are defined as anomalies, and often translate to relevant and actionable information about security and safety characteristics of a monitored environment.

Entropy-based anomaly detection algorithms characterize the normal behavior of a set of data based on their level of statistical entropy [12]. The entropy $\mathcal{H}$ of a dataset comprising i different symbols is defined according to equation 1:

$$\mathcal{H} = \sum_{i} p\left(i\right) \log_2 \left[\frac{1}{p\left(i\right)}\right] \tag{1}$$

where $p\left(i\right)$ represents the probability of occurrence of the i_{th} symbol. In information theory, entropy represents the amount of information conveyed in the dataset, expressed in bits. As an example, a dataset composed by only one symbol has $\mathcal{H} = 0$ independently of its length, meaning that it conveys 0 bits of information. On the other hand, a dataset containing n independent and identically distributed symbols has $\mathcal{H} = log_2(n)$. $\mathcal{H}$ also represents the expected amount of information conveyed by each message belonging to the dataset. The value of $\mathcal{H}$ is also used to measure the randomness of an information source.

The use of entropy as a mean to describe the normal behavior of an information source relies on the following underlying assumptions:

- the entropy of messages generated by the information source exhibits stable statistical characteristics;
- relevant anomalies (that is: anomalies that should be detected by the algorithm) introduce significant deviations in the statistical characteristics of the entropy.

A. *Statistical characterization of CAN entropy*

As a preliminary analysis to verify the applicability of entropy-based anomaly detection to the CAN bus, we evaluate the level of entropy of messages exchanged over the CAN bus of a licensed vehicle.

In this paper we analyzed data gathered from the main CAN bus of a 2011 Ford Fiesta®. The vehicle has been instrumented with a custom CAN bus logger realized with a Genuino UNO® prototyping board, a CAN bus shield and a data logger shield that writes CAN messages to a SD memory card. The sniffer can be connected directly to a CAN bus or to the OBD-II diagnostic port (mandatory in all european licensed vehicles since 2001). When connected to the OBD-II interface the CAN-bus logger reads all messages of the main CAN bus of the vehicle (usually referred to as the powertrain CAN bus, or high-speed CAN bus) that can be accessed through pins 6 and 14 of the OBD-II connector. A picture of the prototype CAN-bus logger is shown in Figure 1.

Thanks to our custom data logger we acquired a trace containing all messages flowing on the CAN bus during a 4-hour long motorway trip. The whole trace contains about 48 million CAN messages (about 3.3k messages per second) having 45 distinct IDs.

To characterize the entropy of this data set we analyzed the first 30 minutes of the trace. We divided this trace into non overlapping time windows of 1, 0.5 and 0.1 seconds, and

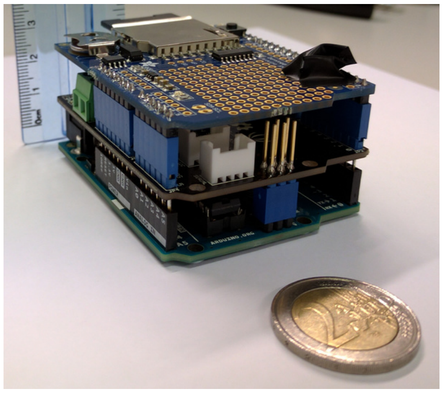

Fig. 1. Prototype CAN-bus logger based on Genuino UNO® board.

used equation 1 to compute the entropy of the set of CAN messages included in each time window. Hence we generated three different time series representing the evolution of entropy (y-axis) over time (x-axis) for the three different time granularities. These three time series are shown in Figure 2.

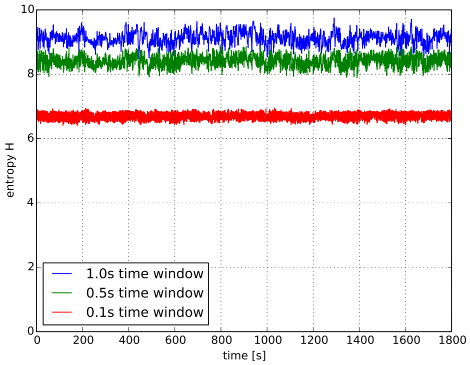

Fig. 2. Evolution of CAN entropy computed over time windows of 1, 0.5 and 0.1 seconds.

We can observe that the value of entropy is quite stable, and independent of specific driving conditions (such as changes in speed, sudden brakes, road turns, activation of turning lights). As expected, entropy computed over larger time frames is higher than entropy computed over shorter time frames, that include fewer messages.

To identify a suitable criteria for anomaly detection we analyzed the distribution of entropy values. The resulting histogram is shown in Figure 3, where the x-axis represent entropy values discretized in bins and the y-axis represents the number of time windows that fall within each bin.

Figure 3 refers to a subset of 100 seconds of CAN messages, and entropy values are computed with a time window of 0.5 seconds. Similar distributions are achieved for all time granularities and are not shown for space reasons. We can observe that the distributions of entropy values are similar to the normal distribution.

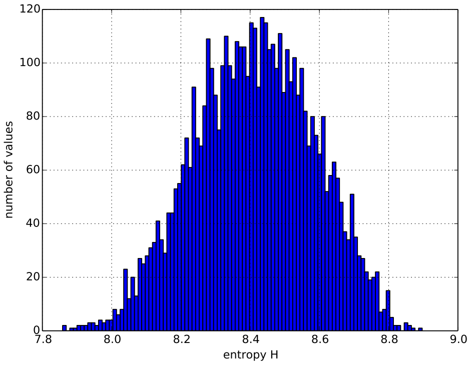

Fig. 3. Distribution of the CAN entropy measured with a time window of 0.5 seconds.

B. *Definition and tuning of the anomaly detection algorithm*

Since entropy values appear to be rather stable over time and distributed according to a normal distribution, we propose an anomaly detection algorithm based on the assumption that entropy values that are too distant from the average entropy are unlikely, and should be considered as anomalies. For each of the three time granularities considered in Section III-A, we compute two descriptive parameters: the average entropy value μ_e and its standard deviation σ_e. For each time window t, the anomaly detection algorithm leverages equation 1 to compute $\mathcal{H}_t$, that is the entropy of all CAN messages included in t. An anomaly is raised if $\mathcal{H}_t$ is not included in the range $[\mu_e - k\sigma_e, \mu_e + k\sigma_e,]$, where k is a model parameter that defines the sensitivity of the algorithm with respect to deviations from μ_e.

To tune the proposed algorithm we applied it to the second 30 minutes of CAN traffic using an initial value of $k = 1$. We then increased k by one until the proposed algorithm raised no anomalies, meaning that we reached 0 false positives with this training traffic trace. These experiments led to the discovery that lowest value of k that generated 0 false positives was $k = 4$ for all three time granularities.

C. *Detecting anomalies partitioning by message ID*

An alternative approach, already proposed in [8], is to focus on messages having the same ID, rather than considering all CAN traffic together. The main idea behind this approach is that entropy characterization for messages with the same ID might be more precise than a global characterization of the CAN bus. Moreover this approach has the added benefit of detecting ID-specific anomalies, rather than generic anomalies. To pursue this approach, we adapt the algorithm proposed in the previous section by considering 45 different windows (one for each message ID). Moreover, since different IDs appear on the bus with different periods, ranging from 0.01 to a few seconds, it is not possible to leverage windows based on time. In this variant the size s of each window is determined by a fixed amount of messages. The proposed algorithm reads all messages from the CAN bus and adds each message to the window identified by its ID. When a window reaches s messages, the entropy $\mathcal{H}^{id}$ is computed. An anomaly is raised if $\mathcal{H}^{id}$ is not included in the range $\left[\mu_e^{id} - k^{id}\sigma_e^{id}, \mu_e^{id} + k^{id}\sigma_e^{id},\right]$, where k^{id} is a model parameter that defines the sensitivity of the algorithm for each id with respect to deviations from the id average entropy μ_e^{id}.

To tune this alternative model it is necessary to compute μ_e^{id} and σ_e^{id} for all 45 message IDs. Then we applied the algorithm to the second 30 minutes of CAN traffic using an initial value of $k^{id} = 1$ for all IDs. We then increased each k^{id} by one until the algorithm raised no anomalies for that id. Figure 4 summarizes the results of the tuning phase for $s = 50$. The x-axis represents possible values of k and the y-axis represents the number of distinct message IDs having $k^{id} = k$. As an example, 11 different IDs have $k^{id} = 1$, 5 have $k^{id} = 2$, 4 have $k^{id} = 3$, and so on. This figure shows that most IDs are characterized by relatively low values of k^{id}. In particular, 24 out of 45 IDs have $k^{id} \leq 4$.

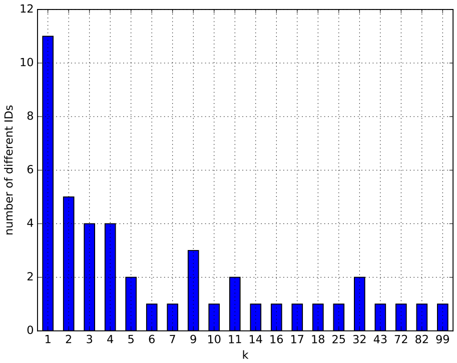

Fig. 4. Number of distinct IDs having $k^{id} = k$.

IV. Experimental evaluation

This section provides an experimental evaluation of the effectiveness of entropy-based anomaly detection for the identification of active attacks realized by injecting forged messages in the CAN bus of a common vehicle. Section IV-A describe the attack scenarios, Sections IV-B shows the effectiveness of entropy-based anomaly detection applied to all messages flowing on the CAN bus and section IV-C evaluates the performance of the same algorithm applied separately to each message ID.

A. Attack scenarios

We consider two different attack scenarios, that mimic activities that an attacker may perform to reverse engineer an unknown CAN bus and to assess vulnerabilities of several ECUs connected to the CAN bus. In both scenarios we assume that the attacker can read and write arbitrary messages to the CAN bus, either through physical access or by exploiting a remote vulnerability [4], [1], [22].

The first scenario represents a replay attack, in which an attacker that is able to sniff legitimate messages flowing on the CAN bus tries to compromise its security by replaying several instances of the same message [16], [2].

The second scenario assumes that the attacker is looking for vulnerabilities by applying common fuzzing techniques [23], [24]. Hence, the attacker injects forged messages having a valid ID (that is, an ID that has been previously observed through passive sniffing) and a random payload.

To simulate the two attack scenarios we injected forged messages in the last three hours of the traffic trace, that were not previously used for building and training the entropy-based anomaly detector. We generated attacks of different intensity by varying the frequency with which the attacker inject forged messages over the CAN bus. In particular, for each scenario we generated 6 different *attack traces* in which attack messages were injected every 1, 0.5, 0.1, 0.05, 0.01 and 0.005 seconds.

B. Anomaly detection applied to all messages

We analyzed all attack traces with three different instances of the anomaly detection algorithm proposed in Section III-B, one for each time granularity presented in Section III-A (1, 0.5 and 0.1 seconds).

For the first attack scenario, experimental results demonstrated that all three models were unable to raise any alerts for the attack traces containing forged messages every 1, 0.5, 0.1 and 0.05 seconds. In all these instances, the insertion of identical messages lowers the entropy, but it remains within the range $[\mu_e - 4\sigma_e, \mu_e + 4\sigma_e,]$. For the attack traces that contain one forged message every 0.01 and 0.005 seconds, all three models detected anomalies, since the entropy fell below the value of $\mu_e - 4\sigma_e$. As an example, Figure 5 shows the entropy computed over 100 seconds of the 0.01 attack trace compared with the anomaly detection model trained with a time granularity of 0.1 seconds. The y-axis represents the value of entropy, while the x-axis represents time. The two horizontal dashed lines show the higher and lower thresholds identified by the anomaly detection model trained with a time granularity of 0.1 seconds. Analogous charts related to models trained with different time granularities are very similar and they are not included for space reasons.

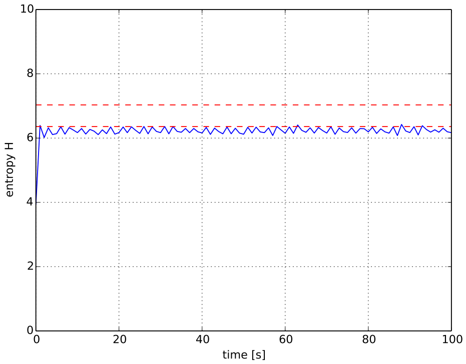

Fig. 5. Entropy of the 0.01 attack trace compared to anomaly thresholds.

Similar experiments were conducted for the second attack scenario. By inserting messages with a random payload, this attack causes the level of entropy to increase. However, all three anomaly detection models failed to detect anomalies for the 1, 0.5, 0.1, 0.05 and 0.01 attack traces. Only the most intense attack, in which the forged message was injected every 0.005 seconds, raised the entropy above the $\mu_e + 4\sigma_e$. Figure 6 shows the entropy of the 0.005 attack trace as measured by the model trained with 0.1 seconds time granularity. As in the previous case, analogous charts for different models are very similar and they are not included for space reasons.

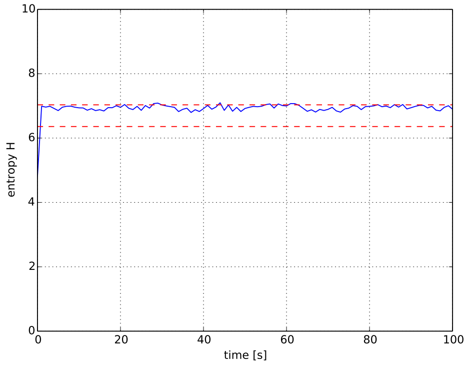

Fig. 6. Entropy of the 0.005 attack trace compared to anomaly thresholds.

C. Anomaly detection applied to each ID

To evaluate the effectiveness of the anomaly detection approach based on message IDs, we repeated both attack scenarios at all 6 injection rates 45 times. Each time we injected attack messages with the same message ID, thus executing all attack scenarios with all IDs. All these attack traces have been analyzed with the anomaly detection algorithm proposed in Section III-C. In these experiments we set a value of $s = 50$ messages. Detection results are summarized in Table I.

TABLE I
NUMBER OF DISTINCT IDS FOR WHICH AN ANOMALY WAS DETECTED FOR THE TWO ATTACK SCENARIOS AND FOR DIFFERENT ATTACK RATES.

	1.0	**0.5**	**0.1**	**0.05**	**0.01**	**0.005**
First scenario	40	40	41	42	42	42
Second scenario	40	40	40	40	40	40

Columns of Table I represent different attack rates, while the two rows refer to the two attack scenarios. Each cell contains the number of distinct IDs for which at least one anomaly was generated. Experimental results show that detection performance is almost independent of the attack rate, and that this approach is able to detect low-rate attacks for the majority of IDs. However, for the few IDs that have the highest values of k^{id} (see Figure 4) this approach is ineffective even for the highest-rate attacks.

V. CONCLUSION

This paper proposes and evaluates an entropy-based algorithm for detecting anomalies in CAN messages generated by an unmodified licensed vehicle. In particular this paper includes extensive experimental evaluation based on several hours of CAN traffic captured during driving sessions on public motorways, reflecting real road and traffic conditions.

Our experimental evaluation shows that anomaly detectors based on entropy represent a viable approach for identifying CAN bus anomalies caused by the activity of attackers that injecting messages over the CAN bus. The main benefit of this approach for anomaly detection is the complete independence with respect to the content of CAN messages, hence it can be applied immediately to the CAN bus of any vehicle without the need of proprietary information that is necessary to interpret the semantic of CAN messages. Moreover, experimental results show that the detection performance of entropy-based are independent of the time granularity used by the detection model.

However, experimental results also shows some limitations of entropy-based approaches. In particular, if the entropy-based anomaly detection model is used to analyze all CAN messages, independently of their ID, reliable detection can be achieved only for high-rate attacks, in which the attacker injects hundreds of forged CAN messages per seconds. Detection of low-volume attacks, in which the attacker injects only 1 packet per second, can be achieved by applying entropy-based anomaly detection only to messages having the same ID. However, this approach requires several anomaly detectors (one for each ID) to be executed in parallel. Moreover, this approach proves to be ineffective for a small subset of IDs whose entropy exhibits large variations even in normal conditions.

REFERENCES

[1] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno, "Comprehensive experimental analyses of automotive attack surfaces," in *Proceedings of the 20th USENIX Conference on Security (SEC'2011)*, Aug 2011.

[2] T. Hoppe, S. Kiltz, and J. Dittmann, "Automotive it-security as a challenge: Basic attacks from the black box perspective on the example of privacy threats," in *Proc. of the 28th International Conference on Computer Safety, Reliability, and Security (SAFECOMP '09)*, Sep 2009.

[3] S. Jafarnejad, L. Codeca, W. Bronzi, R. Frank, and T. Engel, "A car hacking experiment: When connectivity meets vulnerability," in *Proc. of the 2015 IEEE Globecom Workshops (GC Wkshps)*, Dec 2015.

[4] C. Miller and C. Valasek, "Remote exploitation of an unaltered passenger vehicle," in *Proc. of the Black Hat 2015 conference*, Aug 2015.

[5] A. Theissler, "Anomaly detection in recordings from in-vehicle networks," in *Proceedings of the 1st International Workshop on Big Data Applications and Principles (BIGDAP 2014)*, Sep 2014.

[6] Y. Gu, A. McCallum, and D. Towsley, "Detecting anomalies in network traffic using maximum entropy estimation," in *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement (IMC '05)*, 2005.

[7] A. Wagner and B. Plattner, "Entropy based worm and anomaly detection in fast ip networks," in *Proc. of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise (WETICE'05)*, Jun 2005, pp. 172–177.

[8] M. Müter and N. Asaj, "Entropy-based anomaly detection for in-vehicle networks," in *Proc. of the IEEE Intelligent Vehicles Symposium (IV 2011)*, Jun 2011, pp. 1110–1115.

[9] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, "An empirical evaluation of entropy-based traffic anomaly detection," in *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement (IMC '08)*, Oct 2008.

[10] I. Studnia, V. Nicomette, E. Alata, Y. Deswarte, K. Mohamed, and Y. Laarouchi, "Survey on security threats and protection mechanisms in embedded automotive networks," in *Proc. of the 43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop (DSN-W 2013)*, June 2013.

[11] M. Muter, A. Groll, and F. Freiling, "A structured approach to anomaly detection for in-vehicle networks," in *Proc. of the Sixth International Conference on Information Assurance and Security (IAS 2010)*, Aug 2010.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.

[13] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[14] W. Lee and D. Xiang, "Information-theoretic measures for anomaly detection," in *Proceedings of the 2001 IEEE Symposium on Security and Privacy (S&P 2001)*, May 2001, pp. 130–143.

[15] P. Fiadino, A. D'Alconzo, M. Schiavone, and P. Casas, "Challenging entropy-based anomaly detection and diagnosis in cellular networks," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM '15)*, Aug 2015.

[16] T. Hoppe and J. Dittman, "Sniffing/replay attacks on can buses: A simulated attack on the electric window lift classified using an adapted cert taxonomy," in *Proc. of the 2nd Workshop on Embedded Systems Security (WESS)*, Oct 2007.

[17] I. Rouf, R. Miller, H. Mustafa, T. Taylor, S. Oh, W. Xu, M. Gruteser, W. Trappe, and I. Seskar, "Security and privacy vulnerabilities of in-car wireless networks: A tire pressure monitoring system case study," in *Proc. of the 19th USENIX Conference on Security (USENIX Security'10)*, Aug 2010.

[18] A. Francillon, B. Danev, and S. Capkun, "Relay attacks on passive keyless entry and start systems in modern cars," in *Proc. of the 18th Annual Network & Distributed System Security Symposium (NDSS 2011)*, Feb 2011.

[19] M. Wolf and T. Gendrullis, "Design, implementation, and evaluation of a vehicular hardware security module," in *Proc. of the 14th International Conference on Information Security and Cryptology (ICISC'11)*, Nov 2011.

[20] J. Pieprzyk, A.-R. Sadeghi, and M. Manulis, "Libra-can: a lightweight broadcast authentication protocol for controller area networks," in *Proc. of the 11th International Conference on Cryptology and Network Security (CANS 2012)*, Dec 2012.

[21] O. Hartkopp, C. Reuber, and R. Schilling, "Macan - message authenticated can," in *Proc. of the 11th Conference on Embedded Security in Cars (ESCAR 2012)*, Nov 2012.

[22] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile," in *Proc. of the 31st IEEE Symposium on Security and Privacy (S&P 2010)*, May 2010.

[23] P. Oehlert, "Violating assumptions with fuzzing," *IEEE Security Privacy*, vol. 3, no. 2, pp. 58–62, March 2005.

[24] M. Sutton, A. Greene, and P. Amini, *Fuzzing: brute force vulnerability discovery*. Addison-Wesley Professional, 2007.