Threat Evaluation Based on Automatic Sensor Signal Characterisation and Anomaly Detection

Anatolij Bezemskij, Richard John Anthony, Diane Gan, George Loukas
Department of Computing & Information Systems
University Of Greenwich
Email: {a.bezemskij, r.j.anthony, d.gan, g.loukas}@gre.ac.uk

***Abstract*—Autonomous cyber physical systems are increasingly common in a wide variety of application domains, with a correspondingly wide range of functionalities and types of sensing and actuation. At the same time, the variety and frequency of cyber attacks is increasing in correspondence with the increasing popularity and functionality of these systems, from in-vehicle driver assistance to smart city infrastructure and robotics. These technologies rely on a variety of sensors, actuating nodes and control communications. Each sensor adds context by which the autonomous system can better understand its environment, but each sensor also provides opportunities for attack, as has been observed in a variety of attacks on different systems. In this paper, we introduce a model to observe signal characteristics, including noise level patterns, on sensor data streams and incorporate this information to differentiate between normal or abnormal behaviour of a robotic vehicle. This model forms the basis of an automated threat detection scheme, which we test using a purpose-built testbed. Experiments are conducted in a controlled environment using stochastic elements to introduce certain levels of randomness during the experiment. The results indicate that the system is able to distinguish the behaviour of a robotic vehicle under different levels of environmental volatility and is able to identify a sensory channel attack against it.**

Keywords–Anomaly detection; Autonomous behaviour; Threat; Cyber-Security; Signature.

I. INTRODUCTION

Detection of cyber threats is an expanding area of study in the embedded systems domain. The need for cyber security has increased significantly and there are many researchers currently working towards cyber-physical security of such systems, such as the decision tree-based approach in [1] using decision trees for anomaly detection, and the behaviour rule specification in [2]. In this paper, we evaluate our robotic testbed system behaviour by monitoring components with instrumentation installed on the system.

Several different attack vectors can apply to cyber-physical systems. We divide these into cyber-physical and physical-cyber. Cyber-physical attacks are attacks in cyberspace that adversely affect the physical space. For instance, an attacker can target the communication between the system and the operator to disrupt normal system operation. In an autonomous system, a system's own autonomy can be used against it to take over control over the autonomous system. Conversely, physical-cyber attacks are the ones performed in physical space to adversely affect cyberspace [3]. A trivial example would be physical damage that would make the network unavailable. A non-trivial example would be an attack consisting of custom laser beams targeting an autonomous vehicle's LiDAR [4], or externally generated noise targeting ultrasonic sensors so as to confuse the vehicle's spatial awareness. Such attacks that manipulate the input to sensor systems with the purpose to affect the operation of a system that depends on them are often referred to as sensory channel attacks.

Previously, there had been little or no consideration for cyber security during the design of safety-critical systems, but this is changing since the practical cyber-physical attacks against vehicles were showcased for the first time a few years ago [5][6]. Ten years ago, the threat level was significantly lower, but now with the availability of electronic devices such as Arduino kits, a variety of sensors that can be used for educational purposes, consumer products and industrial applications are wide spread. With increasing knowledge in this area the threat to such systems increases. An attacker may not necessarily have the intention of disrupting the system; motives can vary and the outcomes can range from small value fluctuations to possible lethal injuries [7]. This shows that there is a need to secure cyber-physical systems.

In this paper, we focus on robotic vehicles, but we believe that our model can be extended for use in unmanned aerial vehicles, other cyber-physical systems where erratic sensing can be the target or an indicator of a sensory channel attack. In Section I-A, a reader will find the discussion on the current state of a research in the cyber-security domain for robotic vehicles. Later in Section II robotic vehicle testbed design is discussed in detail covering its functionality, design specifics and the experiment environment discussed in Section III. The behaviour profile that we use in our methodology is discussed in Section IV, explaining how sensor unique characteristics are formed for the behavioural profile and its format. The methodology itself is explained in Section V using readings from a single data source during an attack. Overall (using all data sources) the robotic testbed methodology performance is discussed in Section VI, followed by the methodology evaluation and conclusion in Section VII.

A. Related Work

Previous research in cyber attack resilience for such systems, has focussed on detection using a variety of techniques such as anomaly detection based on rule specification[7] where state is being defined using pre-defined system functionality. The approach by Vuong et al. [1] shows that it is highly beneficial to monitor not only cyber but also physical metrics to identify cyber attacks [8], for instance to reduce the false positive rate of detection [9]. Various voting algorithms [2] where system nodes are interacting with each other to identify an attack based on behaviour rule specifications have also been proposed. A similar approach has been used by [10] where robotic multi-agents have a reputation based on their observations and try to reach consensus regarding misbehaving robotic agents. Most researchers agree that a cyber-physical system's security has to be improved at the design stage, and for this reason propose the use of more secure communication [11] or the integration of gateway firewalls [6].

Another point of view is to evaluate mission success threats based on the risk of a failure. For instance, Orojloo et al.

have developed a method for evaluation of the security of cyber-physical systems [12] by evaluating the mean time to system security failure with regards to system components and different types of cyber-attacks. Majed et al. [13] have proposed a framework for evaluating cyber threat exposure for energy smart-grids by using attack trees and attack-graphs. A variety of reliability [14] and survivability [15] models have also been proposed for cyber threat evaluation.

Yampolskiy et al. proposed a language describing attacks on cyber-physical systems [16]. This language would enable the impact of certain attacks applicable to specific systems to be described. When it comes to threat analysis, there is little research done on quantification of threats. One example from Sandia National Laboratories [17] uses a threat driven approach for cyber security evaluation of organisations. Some aspects of their findings can be taken into account when a cyber-physical system is evaluated. The majority of these approaches and frameworks take into account an attack based on methods, conditions and impacts.

In most research presented above, researchers have taken into account attack characteristics as input to identify anomalous behaviour. In other words, the type of attack is predefined. Our view is that this limits the practicality and likely effectiveness of a protection mechanism to attacks that have already occurred and are known to the system at hand. Here, we attempt to detect attacks on which we have not already trained our system. Our proposal is to monitor sensor noise data accompanied with a system's knowledge about itself, as input for anomaly detection and to evaluate a possible threat to the system. Such approach will treat attacks as generic entities, therefore introducing dynamic anomaly detection approach, which will continue being applicable as new attack vectors are introduced and the sophistication (and/or number) of attacks increases over time.

II. Robotic Testbed System Design

To facilitate detailed investigation of autonomous techniques to detect cyber-physical attacks, we have built a richly-instrumented robotic vehicle testbed which is shown in Figure 3, with a variety of different sensor types. The control system of the testbed comprises an integrated set of modular embedded systems. It uses a variety of communication protocols that are used in the industry, such as CAN, RS-485, WiFi and ZigBee. This system was intentionally built integrating technologies that are used by the industry so as to be representative of a large subset of deployed systems. We conduct a variety of real-world relevant experiments and evaluate the system within the cyber security domain.

TABLE I. ROBOTIC TESTBED INSTALLED EQUIPMENT

Feature	**Purpose**
CAN bus	Internal communication
ZigBee	External communication
WiFi	Media streaming
Compass Bearing	Navigation correction
DC Motors	Movement
Ultrasonic Rangers	Collision avoidance

System components produce signals and feedback that is used by other system components to change overall system behaviour. Several components that are mentioned in Table I produce instrumentation data which is used as cyber or physical domain indicators. The combination of such indicators can produce additional meta-data that can be used to identify a particular behaviour of a system, as we describe later. All sensor data is generalised and is treated as a data source. Processing is distributed across the various embedded processors on the testbed platform.

One type of processing node is an AVR-CAN development board, several of which are used to host specific sensors. If such a node is responsible for navigational tasks, the node will listen for data sources with related data and act appropriately. A variety of sensing or actuating components share their processing node. For instance, a single node is responsible for processing bearing, pitch and roll sensors. Overall the system contains six processing nodes, five of which are AVR-CAN development boards clocked at 16 MHz, and one STK300 Kanda board powered by Atmel ATMega1281 chip clocked at 8 MHz.

System components allow the robotic vehicle testbed to undertake a variety of autonomic tasks, such as navigation based on the logical mission layer that represents a sequence of steps given to the testbed. Sensors allow the vehicle to navigate autonomously in an environment using the compass bearing to keep track of the direction, ultrasonic rangers for collision detection and avoidance, and pitch and roll sensors to make direction corrections and inform the system of environment volatility. Also, the system uses an informative meta-data sensor that measures the temperature of the heat sink connected to the on-board voltage regulators which supply power for the camera and robotic arm. In this way, the system is able to determine if these heavy-current-drawing system components are in use. These sensors and additional meta-data extraction allow automatic characterisation of the behaviour of a robotic testbed vehicle whilst in operation.

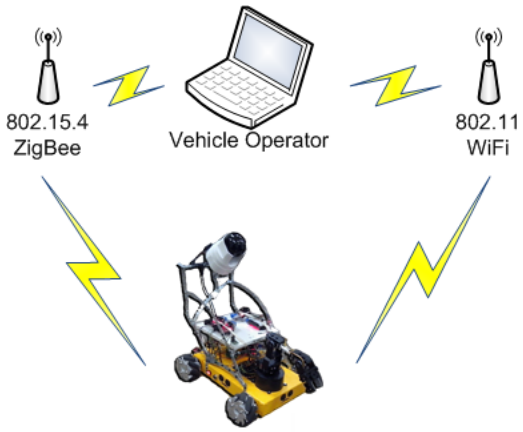

Figure 1. High-level communication

To gather the data for off-line analysis, we use an external workstation. Sensor data from the testbed is collected and stored in a knowledge base. Communication between the workstation and the robotic testbed vehicle is achieved using a dedicated ZigBee network. The ZigBee connection also enables us to transmit commands to the testbed (e.g. to initiate missions). The camera is a self-contained unit; its audio and video feeds are streamed using a standard WiFi protocol. An overview of high-level communication architecture between workstation and robotic testbed vehicle can be seen in Figure 1.

A variety of commands can be sent to the robotic vehicle as simple navigation commands, camera or robotic arm control commands. Additionally, the vehicle supports complex mission task uploads. The command transmission is one-way communication functionality; commands are only executed if they are received from verified ZigBee network nodes and the command is in the correct format. The robotic vehicle testbed does not send any commands to any external nodes within the ZigBee network. The testbed will only periodically report its instrumentation data to a verified connected workstation. The instrumentation report periodicity is one second, due to the low bandwidth ZigBee protocol and unique ZigBee ZE10 module behaviour. Therefore higher-rate sample aggregation is performed on-platform on the sensor hosting nodes.

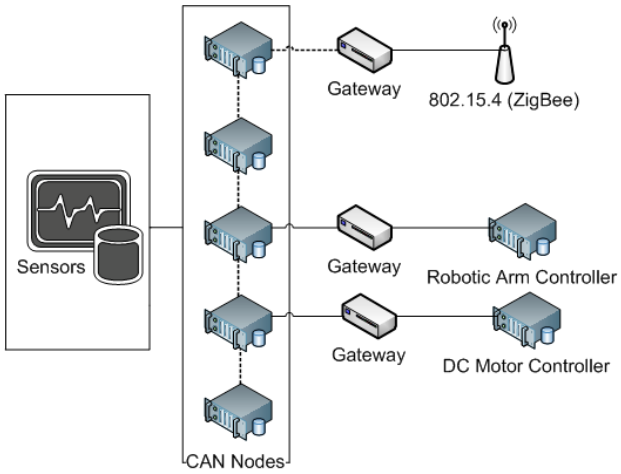

Figure 2. Internal Communication: gateways connect different subsystems

For communication between system components, the testbed uses a CAN bus. This bus is used to share overall sensor data from data sources, including additional metadata extracted during data analysis by the processing nodes. Internal communication architecture is shown in Figure 2. This data is retransmitted to other nodes through gateways and is collected at the reporting node which will transmit data to the workstation when appropriate.

The software structure of the robotic vehicle testbed uses a layered architecture, which separates the different levels of reasoning from the lowest physical sensor level, represented by individual embedded nodes performing analog to digital conversions interpreting signals into an understandable software language. The next-higher level is the classification layer where data is analysed using statistical analysis approaches, such as exponential smoothing to determine the trends in the data. A level higher, we have an autonomous module controller layer which controls actuating capabilities based on the data received from the lower layers of the model. The autonomic module controller layer is a set of autonomic controllers that are carrying out their defined tasks, such as robotic arm movement or navigational control. A mission layer then collects knowledge from autonomic controllers and evaluates if the expected mission goal has been achieved. The layered software approach improves flexibility and maintainability in terms of a robotic vehicle testbed programming, as all these layers are implemented as a set of libraries that can be extended further.

In a real-world environment, there are a variety of physical threats to the system that can be caused by unknown factors, such as rain which affects the grip on a road, windy weather that can affect vehicle movement etc. An issue can arise when an attacker targets a specific sensor to disrupt its activity during the learning process, as it will affect overall operation of the vehicle at later operational stages. One of the examples would be to disrupt a compass intentionally during learning so that the robotic vehicle will learn the disrupted pattern as being 'normal'. We eliminate this risk by securing our vehicle from attacks during the learning process.

Figure 3. Robotic vehicle testbed

To summarise, our robotic testbed vehicle has been designed to facilitate a variety of experiments targeting different data sources and identifying behavioural abnormalities. Our goal is to develop a methodology that will improve robustness of autonomous vehicles using a sensor-agnostic learning approach where the type of data source does not matter, as the requirement is to learn the "normal" signal characteristics, including noise characteristics, generated by the data sources. This robotic vehicle testbed has been built to conduct experiments for a variety of navigational tasks combined with robotic arm actuation. Additional sensors can be added to extend evaluation of the behavioural model.

III. Experiment Environment

Initial experiments that were used for behaviour definition are conducted in the Queen Mary Building at Greenwich University. The environment provides an area with stochastic elements for the data sources, such as old uneven stone flooring with an irregular surface as can be seen in Figure 4. The space is a controlled environment that will not change overtime. The flooring has a variety of dents and lumps that affect the testbed movement throughout the experiment and introduce a stochastic randomness that is used to learn normal deviations. The distance between walls is constant. This allows us to identify the behavioural profile of an environment based on the data source information. The corridor has a set of inset door openings on either side which allows observation of periodic behaviour in the ultrasonic distance sensor signals as the vehicle passes by.

The corridor is 28 m long and the distance from wall to wall is 2 m. The experiments were repeated five times to ensure that the collected data set is representative and these

Figure 4. Experiment environment

were used for the creation of the behavioural profile. Two further experiments were used for evaluation of the behavioural profile. The behavioural profile is built using patterns of the variation and background noise in data sources; mainly we are looking at the spikiness of the data variations and the variety of deviations. The experimental environment facilitates repeatability and contains static elements that can be used as guideline features during analysis of gained data, but it also introduces significant stochastic elements which are essential for understanding the normal levels of noise and variability in sensor signals.

The experimental scenario evaluated in this paper is a mission in which the robotic vehicle testbed has to reach the end of the corridor using its own sensing capabilities. The complexity of such a mission is not obvious. The uniqueness of the flooring surface disrupts direction of the vehicle, forcing it to continuously adapt the speed of its motors and its direction and ensure that it maintains a safe distance from the walls during operation. The scenario was chosen due to the structural uniqueness of the vehicle, and as such the scenario exercises all sensor capabilities. The experiment is organised in two steps. The first is a training step, where over several runs we collect a learning data set that will allow us to create a "normal" behavioural profile. The second step is being conducted to evaluate the recognition of "normal" behaviour profile, as well as we evaluate the representational normality value of a signature that was obtained during the learning step. This value will be used to monitor normality at the higher level observation, if an anomalous representational value has been identified, the system will examine the amount of anomalies and relationship between them, thus reducing computational power requirement of the system.

IV. Normal Behaviour Definition

Our behavioural model uses a sensor-independent approach, in the sense that the sensor-signal characterisation is performed without any additional contextual information to indicate the type of the sensor. Each different type of sensor has its unique output, but we are not interested in determining the type of the sensor, but instead we are interested in learning the signal characteristics under normal operating conditions and thus being able to automatically determine when an anomalous condition occurs by monitoring data source signature.

A compass sensor provides a valuable example: due to the limited speed at which the vehicle can turn, there is a corresponding limit to the rate at which the compass bearing can be expected to change. The compass bearing will also contain a certain amount of noise as the vehicle travels over non-perfect surfaces and does not track in a perfect straight line (there is a detectable "wobble" of typically one to two degrees). These characteristics can be learnt by examining the signal over a series of test missions, without having to explicitly know that the sensor is a compass. For simplicity, we demonstrate the impact on the compass of a cyber-physical disruption using a magnet-based sensory channel attack. By placing a magnet in the vicinity of the sensor, we cause a variable disruption of the vehicle's navigational ability. By so doing, the data stream from the sensor is affected in two detectable ways. Firstly the sharp change in bearing when the magnet is applied (or removed), and secondly, in the reduced noise levels since the magnet causes the sensor to read near-static values (which are anomalous because they are suspiciously "clean"). The proposed approach enables attack detection without prior knowledge of the attack type. The compass example is a part of the experimental set used in our evaluation.

We represent the characteristics of sensor signals in a signature format that can be used to compare expected and actual behaviour in order to detect anomalous events. The signature comprises a number of metrics whose values are learnt during the mission experiments described earlier. The metrics describe characteristics such as the signal-to-noise levels, maximum and minimum sensor readings detected, size and frequency of spike values and rate of change of sensor values. The signature approach facilitates evaluation of the enviroconsistency of a particular trace. For example, the system may learn that a particular data source generates data values distributed in the range 100 to 400 with a mean of 200 during normal operation. The new trace can be compared against the expected behaviour based on these specific characteristics. There is no need to compare the raw data directly. The model will determine whether a particular trace represents normal or abnormal behaviour based on the distance between the trace characteristics and the corresponding values in the signature.

TABLE II. SIGNATURE CHARACTERISTICS

Value Type	Characteristic
Raw	Minimum Maximum
Exponential Smoothing	Minimum Maximum Lowest Difference Highest Difference
Deviation	Standard Deviation
Spike Areas	50% - 100% 100% - 150% 150% - 200% Over 200%

Our signature format contains various characteristics as shown in Table II. Values are exponentially smoothed to provide a basis for comparing instantaneous values with the recent trend, thus detecting noise levels and abrupt changes in values which are short lived are categorised as spikes. Such concept has been used in a dynamic system in [18].

V. Identification of anomalous signals and behaviour

The signatures are constructed during the learning stage to define normal behaviour on a per-sensor signal stream basis. To capture the range of normal behaviour the experiments were

repeated five times to identify the domain of values where the data sources operate and their normal deviations. Using such an approach it is possible to classify normality when the system operates within the normal experiment environment. Currently, we evaluate the results of test runs off-line after each run, however the learnt-signature based approach has the potential to be used in real-time, for self-protection of the autonomous vehicle. In this publication, we review multiple results from seven experiments and different scenarios which are: learning stage, evaluation stage, and physical-cyber compass attack scenario.

To smooth data we are using exponential smoothing in our model as it enables dynamic smoothing to be more reactive or passive by changing the α value. It is a simple and efficient means by which to follow an unfolding trend in sample values. The technique if very efficient in regards to memory and processing and so is well suited for use in embedded systems. Each element of a signature comprises of characteristics that may indicate an anomaly. An operational signature is applied to a learnt "normal" signature, and this facilitates observation of a data source anomalous behaviour and reasoning about component behaviour at the higher layers of our software stack and evaluate the deviations from normality. By observing deviation coefficients (from the learnt normal characteristics), we form a dynamic behaviour score for the data source.

The behavioural score (from the signature) can be used to evaluate the level of threat to the system i.e. the higher the deviation from the learnt normality, the higher the likelihood of an attack. Table III shows an example analysis of data from a compass bearing sensor. By comparing signature elements we identify those elements which indicate that an attack might be present. The deviation extents are weighted and combined to determine the likelihood of an actual attack, i.e. co-deviation on multiple elements reinforces the attack risk.

TABLE III. COMPASS BEARING BEHAVIOUR SIGNATURE DATA

		Deviation Coefficient		
Characteristic	Value	Learnt	Test	Attack
Min	167.8	0.0292	0.0148	0.4388(A)
Max	194.7	0.0151	0.0128	0.3180(A)
Exp. Min.	170.9	0.0145	0.0035	0.0995(A)
Exp. Max.	188.9	0.0083	0.0171(A)	0.3269(A)
Exp. Diff. Min.	0.0	0.0000	0.0000	0.0000
Exp. Diff. Max.	10.1	0.2362	0.1289	>1.0000(A)
Std. Deviation.	5.3	0.1094	0.0582	>1.0000(A)
Spikes >50%	55	0.2435	0.3043(A)	0.9348(A)
Spikes >100%	25	0.5632	0.4079	0.8684(A)
Spikes >150%	10	0.5576	0.5455	0.8485(A)
Spikes >200%	5	0.4462	0.2308	0.6154(A)
Threat	**Summary**	2.2231	1.7237	17.921

Table III shows the deviations from a variety of data sets which are a **Learning scenario**, **Test scenario** and **Compass bearing attack scenario**. By learning we mean that the vehicle is operated in a series of known missions which exercise the sensor signals across their normal value ranges. For example, following the earlier discussion concerning the compass sensor, the vehicle can be operated moving over various types of surfaces to determine the levels of noise in the compass sensor signal, and also can be made to turn at various angular rates-of-change in compass sensor values.

By test scenario, we mean that the vehicle is operated (post learning) in a variety of normal scenarios, with the objective of testing the vehicle's ability to detect the abnormalities solely on the basis of finding anomalous conditions where the sensor signals do not conform to expected learnt behaviour.

By attack scenario, we mean that the vehicle is operated using the same conditions as in the learning and test scenarios, but during the experiment we place the magnet near the compass sensor for forty seconds. This is to validate the behavioural profile approach and determine if the vehicle is able to identify anomalous conditions.

To calculate the deviation coefficient reference for each element in the signature, we have used results of five learning stage experiments. The data set is normalised in the following way, firstly for each signature characteristic the deviation of its current value from the mean is calculated. This deviation is then divided by the mean value of that corresponding signature characteristic, the result is an absolute value. Then the knowledge base is updated with the highest deviations from all learning stage experiment data sets and forms the "Learnt" knowledge which will be used as a reference for anomaly detection. Test scenario is then compared to evaluate the normality behavioural profile and identify the quantity of allowed anomalies. To demonstrate the ability to identify anomalous behaviour the data set of an "Attack" scenario was used.

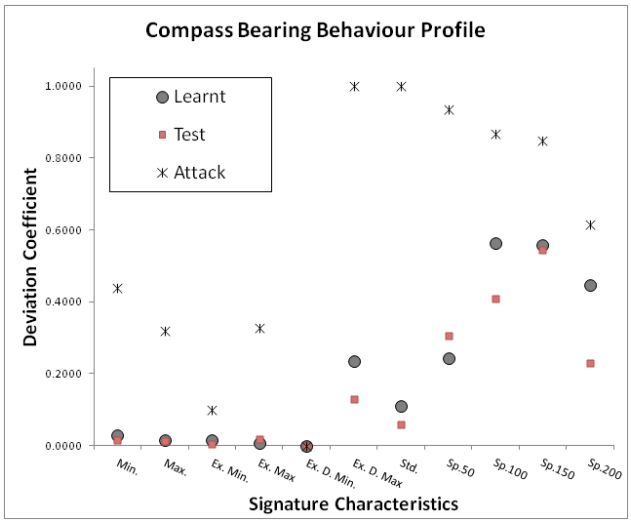

Figure 5. Compass bearing behaviour analysis

We detect an anomaly in terms of sensor signal values as a situation where the signal deviates significantly from expected (learnt) mean behaviour, as held in the particular sensor's signature. We investigate the automated detection of anomalies, based on our signature approach using, initially a single sensor. The corresponding data values are shown in Table III for one data source using two different scenarios. The learnt behaviour signature is based on running the identical normal scenario experiment five times. The absolute value of the registered maximums of all learning data sets is used as the anomaly limit. To identify the number of acceptable anomalies we have used our test scenario experiment runs. The data set from these runs is evaluated in regards to the "Learnt" knowledge to identify the amount of anomalies that exceed the learnt threshold. This procedure has shown that two anomalies have been observed and the threat summary score has not exceed the learnt score, therefore these anomalies were classified as

acceptable.

The number of anomalies identifiable from an attack scenario data are shown (A) in the Table III. The summary score (behavioural score) is a sum of all deviation coefficients of signature characteristics, and is used as an indicator of a threat to the system. Such score explains the deviation of the data source saying; the higher the score, the higher the deviation, therefore the higher threat risk to the system. This is also shown graphically in the Figure 5.

Initially all signature characteristics have equal weights, and due to the weighing scheme we use, signature elements are significantly out of line with expected values when an attack occurs and so the robot is able to autonomously identify an attack based on the detection of anomalies in the sensor data using our methodology. At a level higher, if we will take into account other data sources we can form a system behavioural profile and identify if the system is exhibiting normal or abnormal behaviour.

VI. Evaluation

Earlier, we have demonstrated how a single data source is analyzed producing the behavioural score that can be used at the level higher for surface analysis of the data source. If the behavioural score is exceeding the normality score, the system will investigate the lower layer and will identify what is the cause of such a high score. All system data source signatures operate in the same data domain allowing the system to produce a behavioural score by combining these signatures together.

In this publication, we have reviewed a single data source from multiple experiment scenarios. At the level higher observation, the system uses the overall behavioural score produced by all available data sources. Such approach can decrease the computational power requirement, however potentially a situation can arise where multiple signature characteristic readings are abnormal, but cancel each other out. Leading to a threat summary score which does not indicate a threat. This could mask an actual threat. This can be avoided by an occasional low-level analysis and generating an interrupt-based procedures when anomaly has been identified within the signature.

As for the system's final evaluation, it combines the threat summary scores of all data sources to classify the behaviour profile of a vehicle. The system has access to 17 instrumentation channels from the internal components, which include physical and cyber metrics, such as internal communication utilisation, or sensing the physical environment such as a compass bearing. For each data source a signature is automatically generated. These signatures can be used in isolation or in combination to determine the presence of anomalies and thus determine the level of threat.

To summarize our experimental setup, we combine signatures together to form a behavioural profile score of the system which can represent level of threat to the system during a mission. All data sources have equal weights when combined together producing a sum of all available data source signature scores. For current experimental setup that is reviewed in this publication, we have learnt that the overall behavioural score from the learning scenarios was **46.323**. This score has been produced by a combination of signature summary scores from all available data sources during the learning stage. In this we publication, the key aspect was made to demonstrate the conceptual idea of a data source signature approach and it would be thoroughly explained, therefore the higher level of anomaly detection approach will be investigated further and published in the future. The test scenario produced a score of **54.237**, we can notice that the score for the test scenario is higher than the "Learnt" behavioural score which was learnt using the learning scenarios, through a thorough investigation of the results we have identified that the amount of allowed anomalies has not been exceeded, and overall higher behavioural score was produced by accumulated anomalies that were classified as allowed, resulting in a higher behavioural score. The attack scenario has produced a score of **113.6568**, which is considerably higher than the score produced by the learnt and test scenarios. This shows that the deviations from normality were highly exceeding the threshold allowance on multiple data source signatures.

Further improvements have to be made to increase robustness of the described methodology. Currently, we are using sensor characteristic weights that are equally distributed, thus affecting an overall threat score of a sensor and system itself. Also to make system more robust it would be necessary to investigate how correlation affects an overall behaviour score, as some data sources may have dependencies and these dependencies would result in an anomalous accumulative behaviour score that was described earlier. Weighing system has to be enhanced on a data source and signature characteristic level. One of the solutions is the examination of the spikiness level that can be used to implement dynamic weighing. One of the examples would be that the values from a sensor that are continually volatile (e.g., an accelerometer reading when travelling over a bumpy surface). In such cases, a lower weight would be assigned to the particular sensor characteristic or a signature characteristic. In this way the system can adapt to changing environmental contexts. It is less sensitive to noise or spikes when the ambient noise level or spike frequency is higher.

VII. Conclusion

The work presented here forms part of a wider project to develop techniques for autonomous systems to self-detect attacks. In this paper we have presented a sensor-agnostic learning technique in which a set of sensor-signal characteristics are collectively represented in a signature for each particular sensor. In terms of detecting attacks, the system need not know the type of the sensor, but instead looks at characteristics such as the typical noise levels, the range of data values, the rate of change of data values, the occurrence of spike values, etc.

The initial signatures are generated in experimental mission scenarios but in the absence of attacks, the data signals from sensors are therefore realistic in terms of data values, noise levels, etc. An attack is subsequently detected by observing significant deviations in one or more signature elements for a specific sensor or across several sensors. This approach lends itself to dynamic adaptation which enables the anomaly detection thresholds to be adjusted in line with the environmental volatility, although to date, we have only addressed this step at the concept level (using static signatures for detection).

The main strengths of our approach are that it can be applied universally across a wide range of sensor types without needing manual configuration and that multiple signatures can be used to enhance the attack discrimination accuracy

(facilitated by the standardised signature representation). In addition it has the potential to operate in a continuous learning mode in which it will adapt to its environmental conditions over short to medium time spans, but it will always be sensitive to abrupt changes.

We have developed a custom testbed vehicle in order to evaluate the approach. The experimental method and some initial results are presented above and illustrate how the vehicle was able to successfully identify anomalous events which were part of an attack. Our current findings are very encouraging. Due to the weighting scheme we use, the effects of significant differences between expected and actual sensor data are amplified and thus we have achieved a high true-positive rate and simultaneously a low false-positive rate.

The current implementation requires a training phase, during which it builds up behavioural signatures based on the sensed data signals. These signatures then form the basis on which reasoning is performed at several layers in our software stack. The first layer is concerned with anomaly detection at the level of a sensor, whereas at higher levels it is possible to gain a picture of the attack status across the entire vehicle.

Further work includes dynamic adjustment of the anomaly threshold, as discussed above with the intention of removing the need to retrain the vehicle for use in different environments, as well as further evaluation on the training algorithm itself to understand the optimal level of training and to avoid over-training or under-training issues.

Our cyber-security approach is to consider the robotic system from the perspective that the system initially has no knowledge about itself i.e. a box-in-a-box concept where the perception of the robotic system is stored in a box with several doors and the outer box represents the operating environment. The robotic system only observes values coming in or out and is not able to directly observe the true outside environment. In such a case the robotic system's perception has to make sense to itself, without knowing what is outside and is entirely based on sensor data and patterns within. In such a scenario the autonomous system is sensitive to manipulated sensor data and therefore the signature based approach has been devised specifically to facilitate discrimination between normal and abnormal situations, using a combination of learnt mean behaviour, current data signals and trends in data signals.

ACKNOWLEDGEMENT

This research has been funded and supported by the Defence Science and Technology Laboratory. We thank Robert Sayers for his invaluable assistance and feedback.

REFERENCES

[1] T. Vuong, A. Filippoupolitis, G. Loukas, and D. Gan, "Physical indicators of cyber attacks against a rescue robot," in IEEE International Conference on Pervasive Computing and Communications. IEEE, 2014, pp. 338–343.

[2] R. Mitchell and I.-R. Chen, "Adaptive Intrusion Detection of Malicious Unmanned Air Vehicles Using Behavior Rule Specifications," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. PP, no. 99, 2013, p. 1.

[3] G. Loukas, Cyber-Physical Attacks: A Growing Invisible Threat. Butterworth-Heinemann, 2015.

[4] J. Petit and S. E. Shladover, "Potential cyberattacks on automated vehicles," Intelligent Transportation Systems, IEEE Transactions on, vol. 16, no. 2, 2015, pp. 546–556.

[5] K. Koscher et al., "Experimental security analysis of a modern automobile," in Security and Privacy (SP), 2010 IEEE Symposium on. IEEE, 2010, pp. 447–462.

[6] M. Wolf, A. Weimerskirch, and C. Paar, "Secure in-vehicle communication," in Embedded Security in Cars. Springer, 2006, pp. 95–109.

[7] R. Mitchell and I.-R. Chen, "Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems," Dependable and Secure Computing, IEEE Transactions on, vol. 12, no. 1, 2015, pp. 16–30.

[8] T. P. Vuong, G. Loukas, D. Gan, and A. Bezemskij, "Decision tree-based detection of denial of service and command injection attacks on robotic vehicles," in Information Forensics and Security (WIFS), 2015 IEEE International Workshop on. IEEE, 2015, pp. 1–6.

[9] T. P. Vuong, G. Loukas, and D. Gan, "Performance evaluation of cyber-physical intrusion detection on a robotic vehicle," in Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on. IEEE, 2015, pp. 2106–2113.

[10] A. Fagiolini, M. Pellinacci, G. Valenti, G. Dini, and A. Bicchi, "Consensus-based distributed intrusion detection for multi-robot systems," in Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on. IEEE, 2008, pp. 120–127.

[11] T. Bonaci, J. Herron, T. Yusuf, J. Yan, T. Kohno, and H. J. Chizeck, "To make a robot secure: an experimental analysis of cyber security threats against teleoperated surgical robots," arXiv preprint arXiv:1504.04339, 2015.

[12] H. Orojloo and M. A. Azgomi, "A method for modeling and evaluation of the security of cyber-physical systems," in Information Security and Cryptology (ISCISC), 2014 11th International ISC Conference on. IEEE, 2014, pp. 131–136.

[13] S. Majed, S. Ibrahim, and M. Shaaban, "Energy smart grid cyber-threat exposure analysis and evaluation framework," in Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services. ACM, 2014, pp. 163–169.

[14] A. Hahn and M. Govindarasu, "Cyber attack exposure evaluation framework for the smart grid," Smart Grid, IEEE Transactions on, vol. 2, no. 4, 2011, pp. 835–843.

[15] R. Mitchell and I.-R. Chen, "On survivability of mobile cyber physical systems with intrusion detection," Wireless Personal Communications, vol. 68, 2013, pp. 1377–1391.

[16] M. Yampolskiy, P. Horváth, X. D. Koutsoukos, Y. Xue, and J. Sztipanovits, "A language for describing attacks on cyber-physical systems," International Journal of Critical Infrastructure Protection, vol. 8, 2015, pp. 40–52.

[17] M. Mateski et al., Cyber threat metrics. Sandia National Laboratories, 2012.

[18] R. J. Anthony, "Load sharing in loosely-coupled distributed systems: A rich-information approach," Ph.D. dissertation, University of York, 2000.