Designing a Context-Aware Cyber Physical System for Detecting Security Threats in Motor Vehicles

Andrei Petrovski
Robert Gordon University
School of Computing Science and Digital Media
Aberdeen, UK
+44 (0) 1224 262788
a.petrovski@rgu.ac.uk

Prapa Rattadilok
Robert Gordon University
School of Computing Science and Digital Media
Aberdeen, UK
+44 (0) 1224 262571
p.rattadilok@rgu.ac.uk

Sergei Petrovski
Samara State Technical University
School of Electric Stations
Samara
Russian Federation
+7 846 2784493
petrovski@rambler.ru

ABSTRACT
An adaptive multi-tiered framework, which can be utilised for designing a context-aware cyber physical system is proposed in the paper and is applied within the context of providing data availability by monitoring electromagnetic interference. The adaptability is achieved through the combined use of statistical analysis and computational intelligence techniques. The proposed framework has the generality to be applied across a wide range of problem domains requiring processing, analysis and interpretation of data obtained from heterogeneous resources.

Categories and Subject Descriptors
C.4 [**Performance of Systems**]: Reliability, availability, and serviceability. I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search – *Heuristic methods.*

General Terms
Algorithms, Security, Design, Measurement.

Keywords
Context Awareness, Cyber Physical System, Data Security, Electromagnetic Interference.

1. INTRODUCTION
There exists a growing demand for intelligent and autonomous control in engineering applications, in particular related to detecting and mitigating security threats. This is especially true when some constraints are present that cannot be satisfied by human intervention with regard to decision making speed in life threatening situations (e.g. automatic collision systems, exploring hazardous environments, processing large volumes of data). Because machines are capable of processing large amounts of heterogeneous data much faster and are not subject to the same level of fatigue as humans, the use of computer-assisted threat detection in many practical situations is preferable.

Nowadays security breach events are increasing both in number and sophistication. A typical Threat Detection System (TDS) monitors the system activity and reports on observation of any security violations. Similar to Intrusion Detection Systems, TDS might be divided into two broad classes: pattern-based and anomaly-based. The former uses a database of known threat patterns and raises an alarm whenever similar pattern(s) occur, whereas the later uses a certain model of system behaviour and observes significant deviations from it [5].

Whenever a security breach event is detected, TDS generally raises an alarm that ideally contains the information describing what is detected, the most likely cause of the event, and its potential implications. The content associated with TDS alarms varies considerably depending on the nature of data and on the type of TDS mechanism (pattern- or anomaly-based).

The major concern with such systems is that they attempt to detect a very wide range of events, which often results in high false alarm rate [5]. The main cause of an excessive number of wrong detections is often attributed to a plethora of suspicious cases. However, suspected events are not necessarily security breaches to the system, which typically contains both physical and information components, leading to the concept of a cyber physical system (CPS).

Cyber physical systems are the integration of information processing, computation, sensing and networking that allows physical entities to operate various processes in dynamic environments [3]. Many of these intelligent cyber physical systems involve human intervention at some point – either during the development process by embedding expert knowledge into the systems, or during operation by requiring humans to monitor its operation, evaluate potential threats, and confirm/reject the inferences of TDS.

The latter type of intervention is often associated with another salient feature of cyber physical systems – dealing with the "big data" phenomenon. Big data has become a common research focus in the last decade due to the increasing volume, velocity, variety and veracity of data enabled by technological advancements and by a reduction in data acquisition costs. The integration of multiple data sources into a unified system leads to data heterogeneity, often resulting into difficulty, or even infeasibility, of human processing, especially in real-time environments. For example, in real-time automated process control, information about a possible failure is more useful before the failure takes place so that prevention and damage control can

be carried out in order to either completely avoid the failure, or at least alleviate its consequences.

Computational Intelligence (CI) techniques have been successfully applied to problems involving big data in various application domains [2, 9]. These techniques however require training data to provide reliable and reasonably accurate specification of the context in which a cyber physical system operates. The context enables the system to highlight potential anomalies in the data – in particular, related to security threats - so that intelligent and autonomous control of the underlying process can be carried out.

Anomalies are defined as incidences or occurrences, under a given circumstances or a set of assumptions, that are different from the expectance. By their nature, these incidences are rare and often not known in advance. This makes it difficult for the Computational Intelligence techniques to form an appropriate training dataset. Moreover, dynamic problem environments can further aggravate the lack of training data by occurrence of intermittent anomalies. Computational Intelligence techniques that are used to tackle dynamic problems should therefore be able to adapt to environmental/contextual changes.

A multi-tiered framework for cyber physical systems with heterogeneous input sources is proposed in the paper that can deal with unseen anomalies in a real-time dynamic problem environment. The goal is to develop a framework that is as generic, adaptive and versatile as possible. In order to achieve this goal both statistical and computational intelligence techniques are applied within the framework, together with the online learning capability that allows for adaptive problem solving.

2. CYBER PHYSICAL SYSTEMS (CPS)

Rapid advances in miniaturisation, speed, power and mobility have led to the pervasive use of networking and information technologies across all economic sectors. These technologies are increasingly combined with elements of the physical worlds (e.g. machines, devices) to create smart or intelligent systems that offer increased effectiveness, productivity, safety and speed [3].

Cyber physical systems are a new type of system that integrates computation with physical processes. They are similar to embedded systems but focus more on controlling the physical entities rather than simply the computational devices – embedded computers and networks monitor and control the physical processes, usually with feedback loops, where physical processes affect computations and vice versa. Components of cyber physical system (e.g., controllers, sensors, actuators, etc.) transmit the information to cyber space through sensing a real world environment; also they reflect policy of cyber space back to the real world [6]. Rather than dealing with standalone devices, cyber physical systems are designed as a network of interacting elements with physical inputs and outputs, similar to the concepts found in robotics and sensor networks.

The main challenge in developing a CPS is to create an interactive interface between the physical and cyber worlds – the role of this interface is to acquire the context information from the physical world and to implement context-aware security monitoring and threat mitigation in the cyber world. Figure 1 illustrates a conceptual framework for building context-aware cyber physical systems [8]. Each layer is dedicated to a certain context processing task, ranging from low-level context acquisition up to high level context application using either exiting or acquired knowledge.

Data acquisition (LAYER 1) covers the activities carried out on security threats, breaches, and vulnerability. The main challenge here is to regularly update TDSs against emerging threats. A good TDS must perform continuous adaptation to new vulnerabilities and changes in the system.

Data processing (LAYER 2) deals with selecting data sources and features, the ways how data is collected, logged, and formatted. One of the main problems of TDSs is analyzing and processing highly imbalanced and large amounts of acquired data efficiently. The main objective at this stage is to select appropriate features for security breach detection and to reducing the total number of features to be analysed.

The remaining layers of the proposed conceptual model operate at a higher abstraction level. The third layer (LAYER 3) is responsible for building, evaluating, and correcting (if necessary) the data-driven models based on empirical data supplied by the lower layers. The majority of research related to computational intelligence has been carried out at this layer and concerns the development of robust anomaly detection techniques [4] that are robust against unknown security threats.

The final layer (LAYER 4) purports to examine the outputs of the models built at the previous layer in order to obtain or refine knowledge about the principles or rules that govern the security of data under investigation.

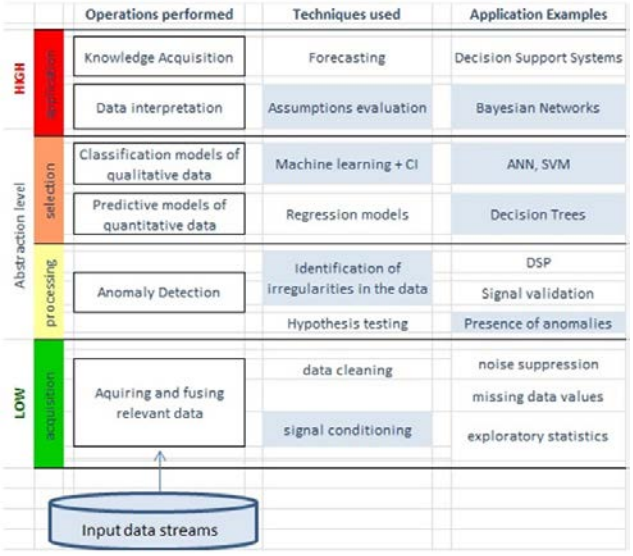

Figure 1. Framework for designing context-aware CPS

Cyber physical systems may consist of many interconnected parts that must instantaneously exchange, parse and act upon heterogeneous data in a coordinated way. This creates two major challenges when designing cyber physical systems: the amount of data available from various data sources that should be processed at any given time and the choice of process controls in response to the information obtained. An optimal balance needs to be attained between data availability and its quality in order to effectively control the underlying physical processes. Figure 2 illustrates a systematic approach to handling the challenges related to context

processing, which has been successfully applied by the authors to various real world applications [7, 8].

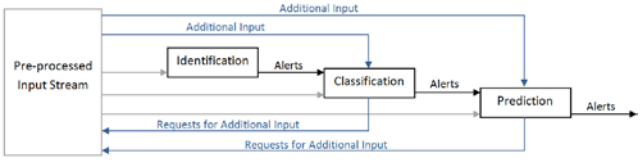

Figure 2. Systematic approach to context processing

As can be seen from Figure 2, the suggested approach segregates processing of the input stream into identification, classification and prediction phases. The identification phase minimises the volume of data and the data processing cost by analysing only inputs from easy to process data sources using context identification techniques – for instance, for finding anomalies in the acquired data. Identified potential anomalies are then passed onto the following phase, where the anomalies are classified into different types. At the end of the process, the prediction phase examines the consequences of the discovered anomalies being present in the underlying process on the operation of the cyber physical system.

Such an approach allows for the acquisition of data and/or activation of the necessary physical entities on an ad-hoc basis, depending on the outcome at each phase. Moreover, the accuracy attained at the specified phases can be enhanced by incorporating additional data from alternative sources.

Computational intelligence techniques and expert systems have been successfully applied to tackling many anomaly detection problems, where anomalies are known *a priori*. More interesting, however, is to detect previously unseen anormalies. Statistical analysis and clustering are examples of techniques that are commonly used when the characteristics of anomalies are unknown [1]. Figure 3 illustrates a more detailed process for the systematic approach (depicted in Figure 1) where statistical analysis and computational intelligence techniques are combined to tackle the unknown anomalies and learn from the experience when similar anomalies occur again.

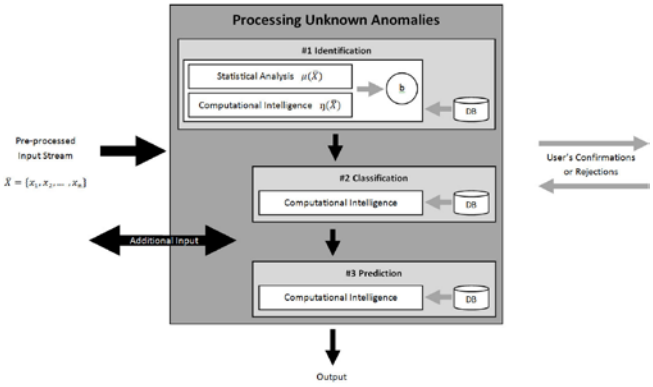

Figure 3. Context processing in a CPS

In Figure 3, 'b' represents a belief function for the output from both the statistical analysis node and computational intelligence nodes, such that

$$b(\bar{X}) = \sum_{i=1}^{n} w_i \mu_i(\bar{X}) + \sum_{j=1}^{m} w_j \eta_j(\bar{X}). \quad (1)$$

The weights (w_i and w_j) of this belief function are adaptively adjusted depending on how much knowledge related to the problem context has been obtained – the contribution of the CI nodes increases with collection of more normal and abnormal data points that can be used for training. This allows the system to run autonomously if required, and any potential anomalies are flagged for closer inspection at the second (i.e. classification) phase.

With the use of parallelisation and/or distributed systems, multiple statistical analyses, computational intelligence techniques and various belief functions can be evaluated simultaneously with their parameters being adaptively chosen. Anomaly identification using a combination of statistical analysis and computational intelligence, as described in Figure 3, has been successfully applied to a traffic surveillance application [8], robotics, a smart home environment, and automotive process control [7].

3. EXPERIMENTAL RESULTS

The proposed framework for designing context-aware CPS has been evaluated using a case study of monitoring electromagnetic interference (EMI) within a motor vehicle – in particular, with respect to the engine ignition system. The ignition system is a key part of an automobile engine - its operation, working performance and reliability have a significant influence on the automobile power and quality characteristics. Therefore, ensuring the appropriate functionality and reducing unintended effects of the ignition system on other engine subsystems become a very important problem in automotive design related to data availability and integrity due to various interferences.

An ignition spark is the radiation of waves of electromagnetic energy within the radio frequency spectrum (greater than 20 kHz). Radiated noise may be caused by any abrupt change in current flowing through a conductor, resulting in changing the magnetic field around the conductor and emanating an electro-magnetic wave that reaches a vehicle antenna. Another source of noise is arcing caused by a build-up of static electricity. Although there is no metal conductor where the arc takes place, there is a sudden change in current through space. This change of current causes a magnetic disturbance. This problem can also develop where there are poor contacts between various components of a vehicle.

In addition to appropriate antenna mounting, there are three basic methods of reducing EMI: shielding, filtering and suppression [7]. The proposed framework for designing context-aware CPS is applied to the dataset based on the effect of an interference-suppression capacitor in terms of noise at different frequencies when the capacitor is connected to the bonding or to the engine cylinders. The aim is to automatically identify the anomalies in the interference voltage, i.e. the context in which the CPS under investigation operates, and to adaptively determine the thresholds for acceptable levels of interference when various interference-suppression capacitors (0, 1, and 4.7 μF in our experiments) are used. Figure 4 illustrates the interference anomalies for frequencies above 65 MHz when an interference-suppression capacitor of 1 μF is connected to the engine. The lines of different colours represent different sized moving average windows from single sample to averaging 50 samples. The anomalies are identified when there are abrupt changes in the interference voltages.

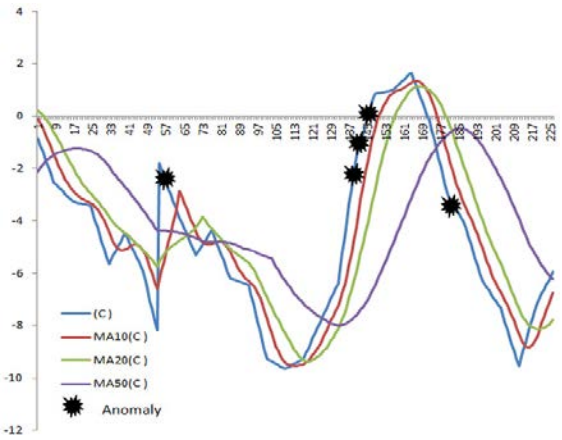

Figure 4. Identified anomalies in interference voltage

As shown in Figure 4, the anomalous frequency diapasons are at sample intervals 53-56, 133-150 and 177-180, this equates to the frequency ranges of 86.97-87.30, 96.25-98.34 and 101.76-102.15 MHz. Figure 5 illustrates the thresholds for excessive interference adaptively adjusted by the proposed framework.

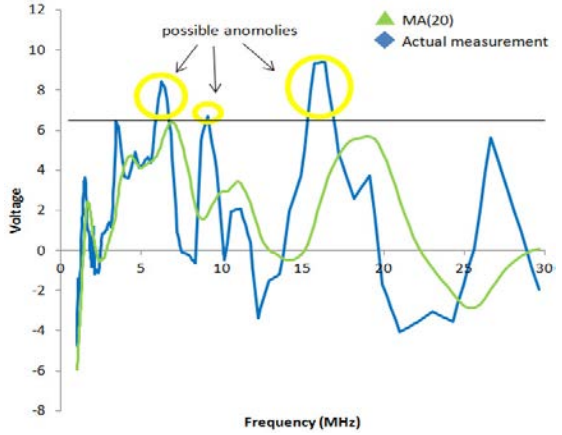

Figure 5. Adaptively adjusted interference thresholds

Figure 5 is derived using the data on the interference voltage for frequencies below 30 MHz when no interference-suppression capacitor is used. As can be seen from the figure, three anomalies are identified, which corroborates the knowledge obtained from the industry experts.

4. CONCLUSIONS

An adaptive multi-tiered framework, which can be utilised for designing a cyber physical system in general and for processing anomalous data in particular, is proposed and applied to the data related to electromagnetic interference within a motor vehicle in terms of noise at various frequencies that affect the data availability aspect of information security. Adaptability and autonomy are achieved through the combined use of statistical analysis and computational intelligence techniques. By choosing various inputs at different stages of anomaly processing, the identification and classification of several anomalies, as well as the prediction of their effects on the operation of the entire CPS, can be made more effective.

The proposed design framework has the generality to be applied across a wide range of problem domains requiring processing, analysis and interpretation of data obtained from heterogeneous resources (examples include surveillance applications, health informatics, and operational process control). There are, however, a number of further investigations to extend the functionality of the system, improve its efficiency, and enhance its scalability. One possibility for further investigation is to make the classification component (see Figure 2) more versatile. As stated previously, anomalies identified at the initial stage might be of unknown origin. The existing classification techniques commonly rely on past occurrences and historical data to build some form of expert knowledge. Classifying unknown anomalies, however, is more difficult and would require more advanced techniques.

For the case of classifying the previously unseen anomalies, a combination of clustering techniques and vector feature selections can be proposed, whereby a number of clusters for “normal” states are formed. With time the knowledge base describing the anomalies that happened in the past will grow, making the classifier component of the proposed framework more robust and reliable.

5. ACKNOWLEDGMENTS

The authors are thankful to Pavel A. Nikolaev - their collaborator at the EMC lab at the Special Testing Research Centre, JSC AutoVAZ (Togliati, Russian Federation) – for providing the experimental data and general support to the research project.

6. REFERENCES

[1] Chandola, V., Banerjee, A., and Kumar, V., 2009. Anomaly detection: A survey. *ACM Computing Surveys*, **41**(3), (September 2009), pp. 1-72.

[2] Khan, Z., Shawkat Ali, A. B. M., and Riaz, Z., Eds. Computational Intelligence for Decision Support in Cyber-Physical Systems. Springer, 2014.

[3] Lee, E., 2008. *Cyber physical systems: design challenges*. University of California, Berkeley Technical Report No. UCB/EECS-2008-8, January 2008.

[4] Sen, S., “A Survey of Intrusion Detection Systems using Evolutionary Computation”, in *Bio-Inspired Computation in Telecommunications*, Chapter 4, pp. 73-94, Elseiver, 2015.

[5] Hubballi, N., Suryanarayanan, V.,"False Alarm Minimization Techniques in Signature-Based Intrusion Detection Systems: A Survey" in *Computer Communications Journal*: Elsevier, Vol **49** (1), 2014.

[6] Park, K. J., Zheng, R., and Liu, X., 2012. Cyber-physical systems: Milestones and research challenges. *Computer Communications*. 36,1 (December 2012), 1-7.

[7] Petrovski, S., Bouchet, F., and Petrovski, A., 2013. Data-driven Modelling of Electromagnetic Interferences in Motor Vehicles Using Intelligent System Approaches. In *Proc. IEEE Symposium on Innovations in Intelligent Systems and Applications (INISTA)*. 2013, 1-7.

[8] Rattadilok, P., and Petrovski, A., 2013. Inferential measurements for situation awareness: Enhancing traffic surveillance by machine learning. In *Proc. Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. 2013, 93-98.

[9] Wu, S., and Banzhaf, W., 2010. The use of computational intelligence in intrusion detection systems: a review, *Applied Soft Computing Journal*, vol. **10**, no. 1, pp. 1–35.