Using pattern-of-life as contextual information for anomaly-based intrusion detection systems

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 7, 2017, accepted September 16, 2017, date of publication October 20, 2017, date of current version November 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2762162

Using Pattern-of-Life as Contextual Information for Anomaly-Based Intrusion Detection Systems

FRANCISCO J. APARICIO-NAVARRO[1], KONSTANTINOS G. KYRIAKOPOULOS[2,3], (Member, IEEE), YU GONG[2], (Member, IEEE), DAVID J. PARISH[2], (Member, IEEE), AND JONATHON A. CHAMBERS[1], (Fellow, IEEE)

[1]School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.
[2]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, U.K.
[3]Institute for Digital Technologies, Loughborough University London, London E20 3BS, U.K.

Corresponding author: Francisco J. Aparicio-Navarro (francisco.aparicio-navarro@ncl.ac.uk)

This work was supported in part by the Engineering and Physical Sciences Research Council under Grant EP/K014307/2 and in part by the MOD University Defence Research Collaboration in Signal Processing.

ABSTRACT As the complexity of cyber-attacks keeps increasing, new robust detection mechanisms need to be developed. The next generation of Intrusion Detection Systems (IDSs) should be able to adapt their detection characteristics based not only on the measureable network traffic, but also on the available high-level information related to the protected network. To this end, we make use of the Pattern-of-Life (PoL) of a computer network as the main source of high-level information. We propose two novel approaches that make use of a Fuzzy Cognitive Map (FCM) to incorporate the PoL into the detection process. There are four main aims of the work. First, to evaluate the efficiency of the proposed approaches in identifying the presence of attacks. Second, to identify which of the proposed approaches to integrate an FCM into the IDS framework produces the best results. Third, to identify which of the metrics used in the design of the FCM produces the best detection results. Fourth, to evidence the improved detection performance that contextual information can offer in IDSs. The results that we present verify that the proposed approaches improve the effectiveness of our IDS by reducing the total number of false alarms; providing almost perfect detection rate (i.e., 99.76%) and only 6.33% false positive rate, depending on the particular metric combination.

INDEX TERMS Basic probability assignment, contextual information, Dempster-Shafer theory, Fuzzy cognitive maps, intrusion detection systems, network security, pattern-of-life, port scanning attack.

I. INTRODUCTION

Cyber-security has increasing importance to Internet users. Providing strong and reliable security mechanisms has become vital in all areas of society. The implementation of Intrusion Detection Systems (IDSs) is fundamental in security infrastructures in order to provide extra level of assurance, identifying evidence of attacks or intrusion attempts. As the complexity of these attacks keeps increasing, new and more robust detection mechanisms need to be developed.

As we previously discussed in [1] and [2], the next generation of IDSs should be designed to include reasoning engines supported by modules that could assess the quality of the analysed datasets [3], manage contextual and non-contextual information about the network, handle uncertainty or deal with incongruent decisions between different IDSs. In order to accommodate all these functionalities, the architecture of advanced IDSs would be noticeably different from the design of more conventional IDSs. The design of the domain anomaly detection system presented in [4] can be considered as a reference model.

Current IDSs use measurable network traffic information from the protected system or signatures of already known cyber-attacks during the intrusion detection process. However, these systems do not generally take into account available high-level information (i.e. above the network operation) regarding the protected system [5]. Ideally, the available high-level information (i.e. contextual information, situational awareness and cognitive information, pertaining to the experts' judgment on the network behaviour) should be incorporated within the intrusion detection process. IDSs should be able to adapt their detection characteristics based not only on the measureable network traffic

information, but also on the context in which these systems operate, and the information provided by the network users or administrators.

In the experiments that we previously presented in [2], we made use of the Pattern-of-Life (PoL) of the network usage as the main source of high-level information. In particular, we correlated the number of network users with the time of the day and the usage of the network resources to characterise the PoL of the network usage and to generate useful contextual information. The results that we presented in [2] evidenced that this available high-level information can be used to improve efficiency of an IDS. In order to incorporate the PoL into the detection process, a Fuzzy Cognitive Map (FCM) [6] can be used in conjunction with an anomaly-based IDS. The FCM is used to fine-tune the techniques used by the anomaly-based IDS to assign evidence of attack. The use of contextual information enhances the generation of a reference of normal network traffic behaviour, making the detection of malicious data more accurate, thereby improving the detection results.

Nonetheless, the work in [2] left a number of areas open for research, which are addressed in this paper. One of these open areas is to identify and evaluate how and at which stage of the data fusion process, augmenting the contextual information is most beneficial. Another open area is the selection of the metric used to represent the PoL of the network usage in the design of the FCM. In our prior work, we made use of the throughput in the construction of the FCM design [2]. However, we did not explore the use of alternative metrics in the design of the FCM, or assess how the use of these metrics may affect the detection accuracy. Therefore, in this work, we extend the analysis of our previously proposed approach for using FCMs to augment the detection process by adding contextual information.

Our contribution can be summarised as follows:

First, we propose two novel approaches that employ an FCM to incorporate the contextual information from the PoL into the detection process. These two approaches exploit the alternative stages at which this information can be added to the detection process. The first approach is based on the use of the output of the FCM to construct an additional metric to be fused by Dempster-Shafer (D-S) Theory of Evidence [7], and the second approach is based on the adjustment of the values resulting from the D-S data fusion process. The performance of these two approaches is evaluated and compared against the framework previously proposed in [2], which is based on the adjustment of certain values involved in the data fusion process, as well as the performance of the D-S based IDS without the use of the FCM. Additional description of the proposed approaches is presented in Section IV.

Second, we have designed three different FCMs by using three different metrics that characterise the PoL of the network traffic. These metrics are Throughput (THR), the number of transmitted bytes per second; Communication Rate (COM), the number of frames transmitted per second; and Destination Port Distribution (DPD), the number of unique destination ports per second. The metric Source Port Distribution (SPD), the number of unique source ports per second, was also extracted from the analysed dataset and used by the IDS during the detection process. However, since the metrics DPD and SPD have very similar profiles, the FCM construction based on SPD is not presented in this work. Further details about these metrics will be presented in Section VII.C.

Third, we have significantly extended the results initially presented in prior work [2]. We have evaluated and compared the performance of our proposed approaches using the different FCM designs. These results provide substantial insight about the behaviour of the IDS when the contextual information is taken into account. Similarly, we have presented extensive analysis of the presented results.

Finally, the method implemented to construct the FCM extracts high-level information from the network users, with a process that is completely transparent to them. In detail, we have correlated the number of researchers present in the monitored offices with the time of the day and the usage of the network resources. Additionally, the network administrator may also contribute to the FCM design by providing prior knowledge about the expected usage of the network resources. This knowledge is provided in the form of different thresholds for each of the metrics, which represent the expected usage of the network resources at a particular time of the day.

The main aims of the experiments that we present in this work are summarised as follows:

- To demonstrate the improved detection performance of our IDS using an FCM to include contextual information from the PoL of the network usage into the detection process.
- To identify which of the proposed approaches to integrate the FCM into the D-S based IDS framework produces the best results.
- To indicate which of the network metrics used to design the different FCMs produce the best detection results.
- To evaluate the efficiency of the proposed IDS in identifying the presence of probing attacks, and reducing the number of false alarms.

The remainder of the paper is organised as follows. In Section II, the most relevant previous work is reviewed. An explanation of the detection methodology used by our IDS is presented in Section III. The proposed use of an FCM within an IDS is introduced in Section IV. In Section V, a description of the FCM is provided. The process of characterising and designing the PoL with an FCM is presented in Section VI. The network testbed, the implemented attacks, and the evaluated dataset are described in Section VII. Section VIII describes the results and provides an analysis of the different findings. Finally, conclusions are given in Section IX.

II. RELATED WORK

With the increasing complexity of cyber-attacks, the next generation of IDSs need to detect network attacks, not only by using measurable information from the network, but also

by integrating human cognition and contextual information into the detection process to improve their effectiveness.

According to [8], contextual information could be defined as any information that surrounds a situation of interest, which helps to understand and to characterise the situation. Snidaro *et al.* [8] present an extensive and very detailed survey about current research on context-based information fusion systems. This work explains that data fusion systems that use contextual information to improve the quality of the fused output have gained importance in the last few years. It also emphasises that contextual information should be an important asset at any level of modern fusion systems.

In [9], the authors proposed an IDS that relies on contextual information to classify the alerts as relevant or irrelevant. The alerts generated by the IDS are processed along with high-level information about hosts present in the network and known vulnerabilities to generate a relevance score about the alerts. Then, a threshold is used to classify alerts as relevant or irrelevant according to the relevance score. Their results demonstrate the effectiveness of using contextual information in the detection process to increase the efficiency of the IDSs.

Xu *et al.* [10] present a context-sensitive detection system based on the use of Hidden Markov Models (HMMs). The host-based system models the system call sequences of a program to detect anomalous patterns. This work uses the caller function of each library or system call as context. The HMM technique can compute the likelihood of occurrences of the observed call sequences. However, this is achieved after a training process using only normal program traces.

An ontology is another technique used to provide contextual information to intelligent systems. Ontologies have proven to be powerful tools to specify and structure knowledge, or to provide formal specification of different entities in a system and their relationships. For instance, Sadighian *et al.* [5] propose a security approach based on the use of ontologies to add context information into a process that fuses the outcome of heterogeneous distributed IDSs. By using this high-level information, the authors reduce the false positive alerts.

A technique that also provides the capability of integrating contextual information from the network user to the detection process is the FCM. FCMs have been previously described and used in [6], [11], and [12] to model human knowledge. Stylios and Groumpos [6] provide a detailed description of the FCM and its mathematical foundation. Although the work presented in [12] does not focus on network security, it comprehensively describes the FCM concept with clear examples. Similarly, Ndousse and Okuda [11] provide a detailed description of an FCM and examples that use an FCM to model fault propagation in interconnected systems.

The work presented in [13] focuses on developing an actionable model of situation awareness for army infantry platoon leaders that could replicate human cognition using FCMs. Their FCM design is based on a goals submap, a tree-like diagram that structures the goals and subgoals of the platoon, and the relationships between these goals. One of the characteristics of the FCM presented in [13] is that the people responsible for designing the FCM do not provide weight values to the concepts, but rank the importance of each modelled concept. A similar approach is presented in [14], in which situation awareness is represented using an FCM. Also, the authors of [14] use ontologies to replicate situations.

In [15], the authors compare the roles of an FCM and another graphical knowledge representation technique, namely a Bayesian belief network, from the perspective of knowledge engineering and representation. This work also describes a knowledge acquisition system that systematically acquires design knowledge from multiple experts from which the FCM is constructed. The authors of [16] highlight the time consuming issues related to the manual construction of large FCMs. In order to solve these issues, the authors present a framework to semi-automate the construction of FCMs, extracting information from a database and other sources of information.

In [17], the authors use an FCM to model causal knowledge within network data. Based on this knowledge, their system calculates the severity/relevance of the modelled network data to attacks. This approach would allow their IDS to discard irrelevant events and focus only on important ones. However, in contrast to the approach that we propose, this research does not use an FCM to modify parameters in the detection process, but as an events filtering process prior to the actual detection.

III. INTRUSION DETECTION METHODOLOGY

The methodology that we present in this work builds upon the design of an unsupervised anomaly-based IDS that we previously presented in [18]. This IDS is based on the combined use of multiple metrics from multiple layers of the network stack to carry out the detection. It uses D-S [7] as a data fusion technique, and is able to detect different types of cyber-attacks in real-time. The goal is to create an overall belief on whether there is an attack in the network traffic.

As many researchers have previously shown [19], [20], the combined use of multiple metrics from the same or different network stack layers may result in higher Detection Rate (DR) with lower numbers of false alarms for an IDS. Each metric provides different levels of evidence about the real nature of the network traffic. Hence, the higher the number of metrics used, the greater the chances to identify the presence of attack.

A. DEMPSTER-SHAFER THEORY OF EVIDENCE

D-S is a data fusion technique that combines evidence of information from multiple and heterogeneous events in order to calculate the belief of occurrence of another event. D-S theory starts by defining a frame of discernment $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$, which is the finite set of all possible mutually exclusive outcomes of a particular problem domain. With regards to this work, we want to identify whether the analysed network traffic is malicious or non-malicious. Therefore, the frame of discernment is comprised of

$A = Attack$ and $N = Normal$. Assuming Θ has two outcomes $\{A, N\}$, the possible hypotheses are $\{A, N, \{A|N\}, \emptyset\} \triangleq 2^{\Theta}$. In the case of $\{A|N\}$, this subset corresponds to *Uncertainty* (either A or N). In addition, $\emptyset$ is the empty set. Each hypothesis is assigned a belief value within the range [0, 1], also known as a Basic Probability Assignment (BPA), through the mass probability function m, which expresses the evidence attributed directly to the hypothesis. This is:

$$m{:}2^{\Theta} \rightarrow [0,1] \quad if \begin{cases} m(\emptyset) = 0 \\ m(H) \geq 0, \quad \forall H \subseteq \Theta \\ \sum_{H \subseteq \Theta} m(H) = 1 \end{cases} \tag{1}$$

Then, D-S uses Dempster's rule of combination to calculate the orthogonal summation of the belief values from two different sensors or observers, and fuses this information into a single belief. This rule is defined in (2), where $m_1(H)$ and $m_2(H)$ are the beliefs in the hypothesis H, from observers 1 and 2, respectively. Similarly, $X \cap Y = H$ refers to all combinations of evidence which yield H; whereas $X \cap Y = \emptyset$ refers to the mutually exclusive subsets of the hypothesis H, thus their intersection is the empty set.

$$m(H) = \frac{\sum_{X \cap Y = H} m_1(X) * m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) * m_2(Y)} \quad \forall H \neq \emptyset \tag{2}$$

Dempster's rule allows the combination of evidence from two observers at a time. In order to combine evidence from more observers, Dempster's rule can be used in consecutive iterations. The output of the initial combination process is used as input evidence in the next iteration, along with the evidence of information from a third observer. Dempster's rule satisfices the associative property, thus the order in which the belief values are fused does not affect the final combined belief values. To better understand how Dempster's rule of combination is implemented, the reader is referred to the practical example presented in our previous work [18].

B. AUTOMATIC BPA METHODOLOGY

There exist multiple ways of assigning BPA values to each of the hypotheses in D-S theory, ranging from data mining techniques to empirical approaches. However, few of them could be used without a prior thorough training or a fine tuning period. In [18], we proposed a novel BPA methodology able to automatically adapt the assignment of its evidence to the current characteristics of the network traffic, without intervention from an IDS administrator.

The proposed BPA methodology exploits a Sliding Window (SW) scheme to compute statistical parameters from the data, used to generate the different BPA values. Our system has one SW for each metric used in the detection process. Although each SW is independent from each other, each metric is extracted from a common piece of information (i.e. the same network frame). All the statistical parameters are computed from the content of the whole SW. However, only the last metric measurement to enter the SW is analysed at a time. The SW slides one slot at a time only if the final decision indicates that the analysed data are normal. Otherwise, the SW does not slide and the data identified as malicious are discarded.

The main benefit of using this scheme is that the SW provides a countermeasure against attackers trying to skew the statistical parameters within the SW. However, the efficiency of this methodology requires a period of non-malicious traffic when the initial SW is filled. Only if the majority of the data within the initial SW are non-malicious can the effective operation of the IDS be expected. This scheme also creates situations in which the computed statistical parameters change substantially as different metrics are fused when the variability of the analysed data is high. In addition, the length of the SW will generally affect the final detection results. The analysis of the optimum SW length has been previously investigated in [21] and it is beyond the scope of this work. In the experiments conducted for this work, the SW length has been empirically set to 50 slots, based on previous experience. This SW length has previously been found to be an appropriate length for our IDS to provide accurate detection results [21].

The BPA methodology that we proposed in [18] uses three independent statistical approaches to provide the BPA values for each analysed metric. The approach that assigns BPA values to the hypothesis *Normal* uses the distribution of the network traffic within the SW. The BPA in *Normal* indicates how strong the belief is that the current analysed data are non-malicious. The content of the SW is divided in sections using the median (Me) and the first and third quartiles (Q_1 and Q_3). Then, the parameters *Max* and *Min* are computed using Inter Quartile Range (IQR):

$$IQR = Q_3 - Q_1 \tag{3}$$

$$Min = Q_1 - 1.5 \times \text{IQR} \tag{4}$$

$$Max = Q_3 + 1.5 \times \text{IQR} \tag{5}$$

A particular BPA value is empirically assigned to each of the portions of the SW as is represented in Fig. 1. The more distributed the data are within each of the portions of the SW, the wider the portion. The analysed data receive the BPA assigned to the portion that it falls in.

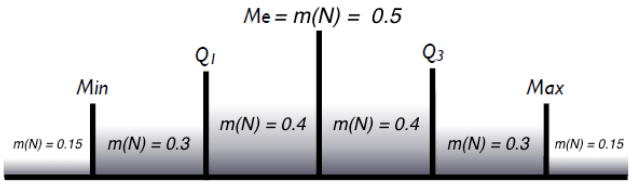

FIGURE 1. BPA Scale for Belief in *Normal* Based on the Distribution of Data.

The approach that assigns BPA values to the hypothesis *Attack* uses the Euclidean distance from a defined reference of normality (i.e. the mean of information within the SW). The BPA in *Attack* indicates how strong the belief is that the current analysed data are malicious. The Euclidean distance

from the mean to the most distant value in the SW is defined as the Maximum Distance (D_{max}), which defines the upper limit for the BPA value. Next, the distance from the mean to the currently analysed data (D) is also calculated. This is represented in Fig. 2. Finally, the BPA in *Attack* is assigned according to equation (6).

$$m(A) = \frac{|D| * 0.5}{|D_{max}|} \tag{6}$$

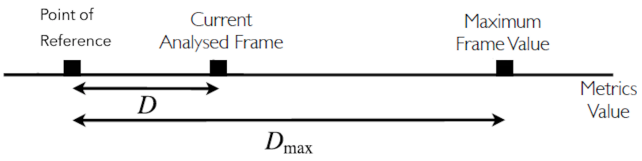

FIGURE 2. BPA scale for belief in *Attack* based on the distance.

Lastly, the BPA in the hypothesis *Uncertainty* is assigned based on the belief values assigned in *Normal* and *Attack* in the current SW, as described in (7).

$$m(N|A) = \frac{min\,(m\,(N)\,,m\,(A))}{max\,(m\,(N)\,,m\,(A))} \tag{7}$$

The BPA in *Uncertainty* indicates how doubtful the system is regarding whether the current analysed data are malicious or non-malicious. The numerator is the smallest of the two hypotheses, whereas the denominator is the largest one.

C. FINAL DECISION ASSESSMENT

Once the BPA values have been assigned, it is required that all the three conditions in (1) are assured. However, it is unlikely that the summation of the belief values assigned by the three previously described approaches add to 1. In order to guarantee that the third condition in (1) is assured, we compute an adjustment factor φ, as described in (8), that will be subtracted from each of the three BPA values, where Z is the number of different hypotheses initially considered within 2^{Θ}.

$$\varphi = \frac{\sum_{x=1}^{Z} m\,(x) - 1}{Z} \tag{8}$$

Then, the BPA values assigned by all the observers are adjusted and fused. The outcome of the D-S theory is a complete set of BPA values (i.e. one for each hypothesis initially considered). The analysed information is classified according to the hypothesis with the highest BPA, which is considered to be the correct decision. There may be cases in which both $m(N)$ and $m(A)$ receive the same final BPA values, or in which the belief in *Uncertainty* is larger than the other hypotheses. In the former case, the hypothesis *Normal* is considered to be the correct decision, whereas in the latter case, the hypothesis with the highest BPA between *Normal* and *Attack* is selected.

IV. PROPOSED USE OF AN FCM WITHIN AN IDS

This section describes the approaches that we propose by which an FCM could be integrated within our unsupervised anomaly-based IDS. These approaches are all based on the generation or modification of the BPA values used in a D-S formulation as described in Section III, by using the outcome of the FCM. Fig. 3 shows the schematic representation of the structure of the IDS, including the extraction of the different metrics, the automatic generation of the BPA values and the data fusion process. Additionally, the figure also indicates the different stages in the IDS detection process at which each of the proposed approaches adds the contribution of the FCM.

A. BPA ADJUSTMENT USING THE FCM PRIOR TO DATA FUSION

The first approach represented by the channel (a) in Fig. 3 to incorporate contextual information was initially proposed and evaluated in [2]. It is based on the adjustment of the BPA values assigned prior to the data fusion process, by using the outcome of the FCM. Once the BPA values have been computed as explained in Section III.B, the outcome of the FCM is used to adjust these accordingly. This is done by adding the outcome of the FCM to $m(N)$ and $m(A)$. Then, $m(U)$ is calculated using the newly computed values of $m(N)$ and $m(A)$. Finally, the BPAs are adjusted as described in (8).

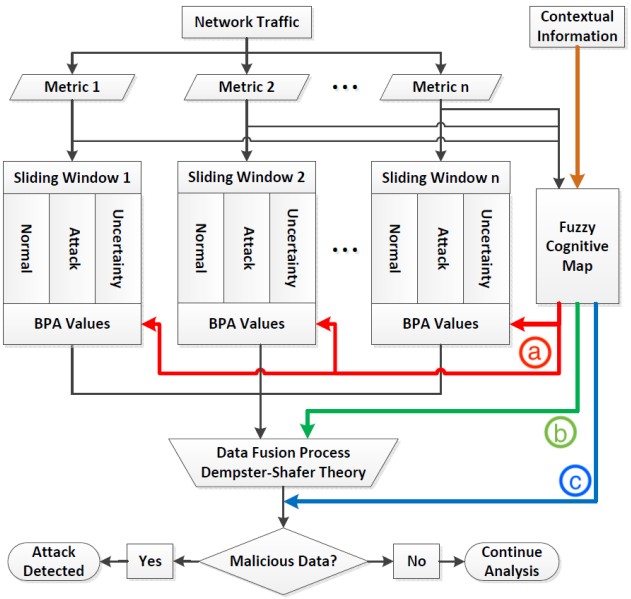

FIGURE 3. Schematic structure of the IDS, including the extraction of the metrics, the generation of the BPAs, the data fusion process and the addition of contextual information into the detection process by using an FCM.

After all the BPA values have been adjusted, the data fusion process is carried out using Dempster's rule defined in (2) and the final decision is taken as described in Section III.C. It is worth noting that, although the same weight values are used to adjust all of the considered metrics, it is unlikely for these metrics to have the same BPA value. Therefore, the adjustments would impact each of the metrics differently.

B. USING THE FCM AS ADDITIONAL BPA VALUES

The first of the two novel approaches that we propose to incorporate contextual information into the detection process is based on the use of the output of an FCM to construct an additional metric to be fused by D-S. This is represented by the green channel (b) in Fig. 3. We propose to use the outcome of the FCM to assign belief values to the hypotheses *Normal* and *Abnormal*, to yield an extra set of BPAs to be fused. These values are then used to infer the BPA in the hypothesis *Uncertainty*. Once these three values have been computed, the newly generated BPAs are merged with D-S, along with the rest of the considered network traffic metrics, using Dempster's rule defined in (2). For clarification, the BPA values computed from the network traffic, as described in Section III, remain unchanged.

It is worth noting that for this approach, in contrast to the one that adjusts the BPAs prior to the fusion process, the contextual information might have less influence over the final IDS decision, as its contribution is reduced to one set of BPAs to be fused. In Section VIII, we will discuss in further detail the contribution of each of the approaches to the final results.

C. BPA ADJUSTMENT AFTER DATA FUSION PROCESS

The second approach that we propose is also based on the adjustment of the BPA values. This is represented by the blue channel (c) in Fig. 3. However, in contrast to the approach in the red channel (a) that adjusts the BPA values prior to the fusion process, the outcome of the FCM will be used to adjust the resulting BPAs, after the D-S data fusion process.

The IDS carries out the detection process using solely the measurable information as described in Section III. The different metrics are extracted from the network traffic and the diverse BPA values are computed. Then, all the BPA values are fused using Dempster's rule defined in (2). Only after the data fusion process has ended, the contextual information is used to adjust the resulting BPA values, by adding the outcome of the FCM. The adjustment is implemented over the final outcome of the IDS, hence the addition of the outcome of the FCM is prone to dominate the entire detection decision.

V. FUZZY COGNITIVE MAP

An FCM is a technique used for prediction and decision making, which can be applied to model human knowledge, and to represent the behaviour of a system as perceived by human experts. The main goal of modelling a decision problem using an FCM is to predict the outcome of the evaluated problem by letting the relevant events interact, and to calculate the actual degree of influence that one event may have upon the system [6].

A. MOTIVATION

An FCM is an efficient soft computing tool that supports adaptive behaviour in complex and dynamic systems, and provides significant support for decision-makers [22]. We have made use of an FCM to model the PoL because this technique provides a number of advantages when compared against several probabilistic algorithms, (e.g. dealing with contradictory or conflicting pieces of information [22]).

An FCM provides a useful framework to calculate the degree of influence that one event or action may have upon the whole system or upon parts of the system. Also, an FCM is able to represent dynamic systems that evolve over time, supporting dynamic timeline structures [22], and to model new and unseen behaviours of particular scenarios.

Another characteristic is that an FCM provides the potential to make changes easily and intuitively, and allows additional pieces of information to be combined at a later time instance. Moreover, it supports memberships of more than one set of events and allows the overlapping of different FCM models [23]. Another important characteristic of this technique is that the fuzzy degrees of influence in an FCM are initially assigned using qualitative linguistic variables instead of hard numerical values. This makes an FCM an excellent solution for agile Command and Control (CC) and/or Human-Machine Interaction (HMI).

One of the most important characteristics of the FCM is its capability to combine multiple, incomplete, contradictory or conflicting pieces of information. An FCM handles conflicting or competing information better than probabilistic systems, which are regulated by the additivity rule. Probabilistic systems have difficulties managing situations that occur when competing statements are both true [22]. Also, according to [22], this type of system does not often handle all forms of uncertainty well, especially when information is conflicting. As part of the design of an FCM, it is not necessary that all human experts involved in the process agree on which events should compose the FCM or what weight value should be given to each link.

The use of an FCM in this work is motivated by all the advantages that this technique provides. Nonetheless, an FCM is not exempt from drawbacks. The design of an FCM relies not only upon the human's understanding of the work domain and knowledge, but also their ignorance, prejudice, or bias [22]. Also, the design of an FCM is very context-specific, and may not be easily generalised. In order for the model to be applied to other situations, a new FCM design should be constructed. Additionally, as will be briefly described next in Section V.B, the convergence of the fixed-point attractor is an open issue in the research community [24].

B. FCM DESIGN DESCRIPTION

The graphical design of FCMs is characterised by a set of nodes interconnected by causal connections. An example of an FCM model is presented in Fig. 4. The nodes in the FCM represent causal and time-varying concepts, events, actions or goals that describe the behaviour of the system. The definition of the main concepts relevant to the system is the initial step in the process of creating the FCM. In our experiments, the nodes have been defined based on the

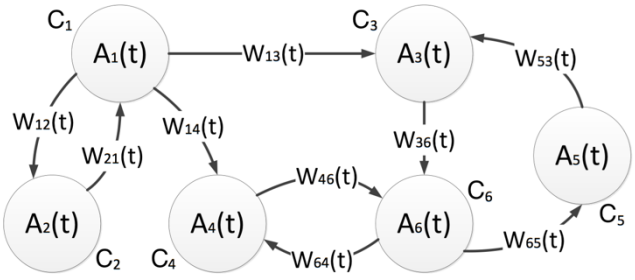

FIGURE 4. Simple FCM model in which nodes represent changes in the modelled system and connections denote relationships between concepts.

events that characterise the PoL of the network and concepts that represent prior knowledge of the network administrator. A very detailed description of the design of FCM models can be found in [6].

Each node C carries a weight A(t) in the fuzzy range [0, 1], which indicates the quantitative measure of the importance that each concept has in the system, at time t. The connections between nodes represent the causal relationship between the defined concepts. Each link is assigned a weight value $w_{ij}(t)$ in the fuzzy interval [−1, 1], which indicates the relationship and degree of influence from the nodes C_i to C_j.

There are three possible relationships between concepts:

1) Positive relationship $w_{ij} > 0$, indicates that $A_j(t)$ increases as $A_i(t)$ also increases.
2) Negative relationship $w_{ij} < 0$, indicates that $A_j(t)$ increases as $A_i(t)$ decreases.
3) No relationship $w_{ij} = 0$, indicates that there is no correlation between $A_j(t)$ and $A_i(t)$.

The fuzzy degrees of influence $w_{ij}(t)$ are initially assigned using qualitative linguistic variables by the network users or administrator, as described in [12], but then transformed into numerical values. In our experiments, the fuzzy degrees are assigned by the network administrator. We consider five linguistic variables, {*very low*, *low*, *medium*, *high*, and *very high*}. In order to transform the variables to numerical degree of influence values, the five linguistic variables are sorted in an ascendant order of importance and represented by $\{\mu_{vl}, \mu_l, \mu_m, \mu_h, \mu_{vh}\}$. These variables are transformed to the numerical values $w_{ij}(t)$ associated with each link using (9), as explained in [15], where n is the total number of variables, and p is the ordinal number that represents the position of the respective linguistic variable in the list. In our experiments, in which $n = 5$, the weights assigned to the variables $\{\mu_{vl}, \mu_l, \mu_m, \mu_h, \mu_{vh}\}$ would be {0.1, 0.3, 0.5, 0.7, 0.9}, respectively.

$$\mu_p = \frac{p}{n} - \frac{1}{2n} \tag{9}$$

An FCM can be represented by an $[m \times m]$ matrix M, where $[M(t)]_{ij} = |w_{ij}(t)|$ which is also known as an adjacency matrix, and m is the number of nodes in the modelled FCM. The matrix M describes the relationship between the nodes and the weight values $w_{ij}(t)$ associated with each link. An FCM allows different adjacency matrices to be combined in an adjacency matrix M as follows:

$$[M]_{ij} = \frac{1}{k} \sum_{m=1}^{k} \left| w_{ij}(t) \right|_m \quad \forall i, j \tag{10}$$

where k is the number of adjacency matrices to be merged (i.e. number of network users in the monitored network) and m is the number of nodes (i.e. concepts) in the modelled FCM.

As an example, Fig. 5 comprises the $[6\times6]$ adjacency matrix of the FCM in Fig. 4.

$$M = \begin{bmatrix} 0 & w_{12}(t) & w_{13}(t) & w_{14}(t) & 0 & 0 \\ w_{21}(t) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{36}(t) \\ 0 & 0 & 0 & 0 & 0 & w_{46}(t) \\ 0 & 0 & w_{53}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{64}(t) & w_{65}(t) & 0 \end{bmatrix}$$

FIGURE 5. Simple $[6\times6]$ adjacency matrix of the FCM represented in Fig. 4.

The initial weight value A(t) of all the nodes in a model at time t = 0 can be represented by the initial vector state A, where $A(0) = (A_1(0), A_2(0), \ldots, A_m(0))$. $A_i(0)$ is the weight value of node i at time t = 0. Then, the FCMs evolve via an iterative process in which, at each future time step, the weight value of each concept A(t) is computed by aggregating the influence of the interconnected concepts, using an activation function f. The value of $A_i(t)$ changes at each iteration as described in (11):

$$A_i(t+1) = f(K) = f\left(A_i(t) + \sum_{j=1 j\neq i}^{m} w_{ji}(t) * A_j(t) \right) \tag{11}$$

where $A_i(t+1)$ is the weight value of node C_i at time t+1, $A_j(t)$ is the weight value of node C_j at time t, and $w_{ji}(t)$ is the degree of influence of node C_j on node C_i.

Bueno and Salmeron [25] describe four activation functions f; these are the sigmoid, hyperbolic tangent, linear threshold, and step functions. Among the four, the hyperbolic tangent activation function, described in (12), is the one used in our experiments. This is because the hyperbolic tangent activation function produces weight values $A_i(t)$ normalised in the range [−1, 1]. Hence, this activation function complies to one of the requirements of the D-S theory, which requires that the BPA values assigned to each hypothesis could be any value up to 1.

$$f(K) = \frac{e^K - e^{-K}}{e^K + e^{-K}} \tag{12}$$

This process continues for a number of iterations until the FCM reaches one final fixed model, known as a hidden pattern or fixed-point attractor. This is when the weight

values A(t) in all the nodes do not change in successive iterations. It is also possible that an FCM keeps cycling between several fixed models, known as a limit cycle, or it may continue generating different models indefinitely. Nápoles *et al.* [24] indicate that in non-stable FCMs, a stopping criterion can be set to overcome the convergence problem of these last two situations. In our experiments, we have empirically set the stopping criterion at $t = 60$. However, the authors also highlight that this approach may be unreliable due to the lack of convergence.

The problem of the fixed model convergence in non-stable FCMs has been previously investigated by other researchers [24], [26]. Nápoles *et al.* [24] describe that non-stable FCMs are mostly related with antisymmetric adjacency matrices, which lead the system to a periodic behaviour. Another factor is the used activation function f. Continuous functions such as sigmoid and hyperbolic tangent can result in chaotic behaviours since the FCM could produce infinite different states [24], [26]. Also, according to [26], a small change in the initial vector state A can drastically change the fixed model convergence. Hence, it is clear that the development of more efficient strategies to improve the fixed model convergence of FCM is still required [24], but this is beyond the scope of this work.

VI. FUZZY COGNITIVE MAP CONSTRUCTION

A. CHARACTERISING THE PATTEN-OF-LIFE

In this work, we have made use of the PoL of the network usage as the main source of contextual information. Moreover, the network administrator has also contributed to the design of the FCM by providing its knowledge in the form of expected network usage levels. The concept of PoL refers to the information generated by observing repeated behaviours over an extended period of time. According to Craddock *et al.* [27], PoL analysis typically involves the surveillance of a group of people over a period of time to characterise their behaviours and habits, and determine if their behaviour is suspicious.

Generally, the PoL of the network usage is directly associated with the number of users accessing the Internet. In order to characterise the PoL of the network usage and to generate useful contextual information, we have correlated the number of researchers present in the monitored offices in the Wolfson School at Loughborough University, UK, with the time of the day and the usage of the network resources. These are the three parameters used to characterise the PoL of the network usage and to extract the contextual information.

Additionally, the design process of the FCM for this work is also based on three predefined assumptions, which help to represent the PoL in the design of the FCM:

- An increase in the network usage is expected to be seen during common office hours (i.e. from 9am to 5pm, weekdays) when most of the network users are expected to use the network, and a decrease outside this timeframe.
- Legitimate high network usage outside common office hours is also feasible because University staff have unrestricted access to their labs at any time of the day.
- Lastly, it is expected to see an illegitimate increase in the network usage during the implementation of the evaluated attack/threat, which could occur at any time of the day.

B. DESIGN OF THE FCM USING THE PATTEN-OF-LIFE

In order to characterise the time of the day in the PoL, four timeframes per day have been defined. These are 00-09h, 09-17h, 17-19h, and 19-24h, distinguishing between weekdays and weekends. As can be seen in Table 1, these timeframes have been used to define eight of the concepts that compose the modelled FCM (i.e. C_{1-8}). These timeframes have been defined after monitoring, for an extended period, the time when the researchers are more frequently present in the monitored offices. This pattern can change for multiple reasons (e.g. bank holidays or festive periods). Therefore, a more comprehensive methodology to characterise the time of the day in the FCM, which could seamlessly adapt to any pattern change, could be developed in future work.

TABLE 1. List of concepts that compose the FCM, building upon the throughput of the network traffic.

FCM Concepts	Concepts Definition
C1	00 – 09 h / Weekday
C2	09 – 17 h / Weekday
C3	17 – 19 h / Weekday
C4	19 – 24 h / Weekday
C5	00 – 09 h / Weekend
C6	09 – 17 h / Weekend
C7	17 – 19 h / Weekend
C8	19 – 24 h / Weekend
C9	Throughput < 4 Mbps
C10	4 Mbps < Throughput < 12 Mbps
C11	12 Mbps < Throughput < 40 Mbps
C12	Throughput > 40 Mbps
C13	Normal
C14	Abnormal

Similarly, in order to characterise the network usage in the design of the FCM, we require a number of metrics that could represent the amount of network resources utilised, such as the throughput of the network traffic. From the gathered network traffic dataset, four different metrics have been identified as the most appropriate metrics. These are THR, the number of transmitted bytes per second; COM, the number of frames transmitted per second; SPD, the number of unique source ports per second; and DPD, the number of unique destination ports per second. Further details about these metrics will be presented in Section VII.C.

Since the network traffic would present variable levels of usage depending on the PoL (i.e. the cycles of the PoL), the

network administrator used each of these metrics to define a number of thresholds based on its prior knowledge. These thresholds represent the maximum expected network usage at given time of the day, according to the network administrator. Therefore, each of these thresholds allows the system to characterise the PoL of the network usage. For instance, if we consider the metric THR to design the FCM, the thresholds 4 Mbps, 12 Mbps, and 40 Mbps have been defined by the network administrator to characterise the PoL of the network usage. These thresholds are used, in turn, to define four additional concepts (i.e. C_{9-12}) in the FCM design. As another example, if the metric DPD was considered to design the FCM, the network administrator defined the thresholds 100 and 250 for unique destination ports per second to characterise the PoL.

Although any of the available metrics can be used in the construction of the FCM to characterise the network usage, we decided to utilise only one of the metrics at a time to reduce the complexity of the FCM. The use of fixed thresholds may not provide a flexible framework that captures the stochastic nature of the network traffic. Therefore, in future work, alternative methodologies to dynamically characterise the PoL of the network usage in the FCM could be also proposed.

Finally, two additional concepts are defined as the two possible outcomes of the FCM (i.e. C_{13} = *Normal* and C_{14} = *Abnormal*). The weights A(t) associated with these two concepts are used to incorporate the contextual information extracted from the PoL into the detection process of our IDS.

The next step in the design process of the FCM is to define the relationships between concepts and the positive or negative influence. In our experiments, the FCM concepts defining the time of the day have direct influence upon the number of network users in the monitored research office and, in turn, upon the concepts defining the usage of the network resources. Thus, the concepts C_{1-8} present positive relationship over C_{9-12}. Similarly, the FCM concepts defining the usage of the network resources (i.e. C_{9-12}) have direct influence upon the possible outcomes of the FCM. Then, the concepts C_{9-12} also present positive relationship over the concepts C_{13-14}. Finally, the two possible FCM outcomes cannot occur at the same time. Hence, the concepts C_{13} and C_{14} present negative relationship with each other.

One adjacency matrix is defined for each network user, considering the time of the day and the usage of the network resources. Then, all the adjacency matrices are averaged out and merged in an adjacency matrix M, as described in (10). Following upon the concepts described in Table 1, in which the throughput is used to characterise the PoL of the network usage, the final [*14*×*14*] adjacency matrix M would be as shown in Fig. 6. The weight values in Fig. 6 are the average values resulting after the different adjacency matrices, one for each network user, are merged as described in (10).

For instance, the weight value $w_{19} = 0.9$ represents a very high positive relationship from C_1 to C_9. In other words, this

$$M = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.9} & \mathbf{0.1} & \mathbf{0.1} & \mathbf{0.1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.12} & \mathbf{0.9} & \mathbf{0.74} & \mathbf{0.52} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.56} & \mathbf{0.56} & \mathbf{0.34} & \mathbf{0.22} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.88} & \mathbf{0.14} & \mathbf{0.12} & \mathbf{0.1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.9} & \mathbf{0.1} & \mathbf{0.1} & \mathbf{0.1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.76} & \mathbf{0.28} & \mathbf{0.2} & \mathbf{0.14} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.84} & \mathbf{0.18} & \mathbf{0.14} & \mathbf{0.1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.9} & \mathbf{0.1} & \mathbf{0.1} & \mathbf{0.1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.9} & \mathbf{0.1} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.7} & \mathbf{0.3} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.5} & \mathbf{0.5} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0.1} & \mathbf{0.9} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-0.9} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-0.9} & 0
\end{bmatrix}$$

FIGURE 6. Merged [*14*×*14*] adjacency matrix of the FCM composed using the number of network users, time of the day and the throughput.

weight value indicates that it is very likely to measure from the network traffic a THR lower than 4 Mbps on weekdays, between midnight and 9 am. Similarly, the weight value $w_{1014} = 0.3$ represents a low positive relationship from C_{10} to C_{14}. This weight value indicates that THR measurement between 4 and 12 Mbps is unlikely to correspond to an attack.

In order to understand how the BPAs are adjusted using the outcome of the FCM, consider the concepts listed in Table 1 and the adjacency matrix M shown in Fig. 6. Consider the situation in which the network traffic is evaluated at 12 pm on a weekday (C_2), and the throughput reaches 10 Mbps (C_{10}). In this situation the initial state vector would be defined as A(0) = [0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]. As we previously described in Section V.B, we have empirically set the stopping criterion at t = 60. The resulting state vector is A(60) = [0, 0.155, 0, 0, 0, 0, 0, 0, 0.379, 0.672, 0.639, 0.581, 0.907, 0.782]. Once the process has ended, the weights associated with both concepts C_{13}, $A_{13}(60) = 0.907$, and C_{14}, $A_{14}(60) = 0.782$, are used to adjust the BPA values assigned to the D-S hypotheses, or to construct the new additional metric to be fused by D-S.

VII. TESTBED AND NETWORK TRAFFIC MEASUREMENTS

A. TESTBED LAN

One drawback associated with publicly available datasets is that the underlying network usage dynamics are not described in detail and high-level information such as PoL becomes very difficult to derive. The KDD99 dataset [28] is arguably the most frequently used publicly available network traffic dataset in the area of network security. Nevertheless, this dataset does not include precise timestamps nor the source address of the network traffic [29]. This makes the extraction of contextual information from the network traffic almost impossible.

The analysed data traffic, previously described in [2], has been gathered from a Local Area Network (LAN) testbed in a research office environment, in the Wolfson School at Loughborough University. The PCs in two distinct labs are connected to the same office LAN, and these PCs are used by

researchers daily for Internet access. One Cisco switch aggregates the traffic from all the PCs in the office LAN, which is used as background traffic. In addition, two additional PCs have been connected through to a testbed LAN in order to implement both the attack and the detection process. Fig. 7 shows the logical topology of the testbed LAN.

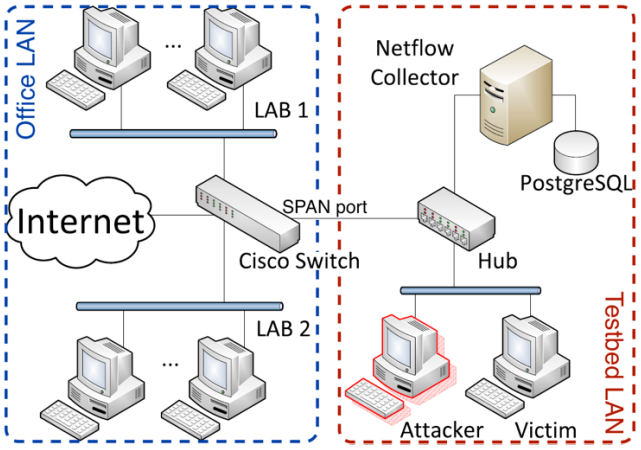

FIGURE 7. Logical topology of the testbed LAN; PCs on the left generate the background traffic, whereas those on the right are involved in the port scanning attacks implementation, and detection process.

B. EVALUATED NETWORK SCENARIO

In this paper, we use the technique port scanning to evaluate the performance of our proposed approaches. Port scanning, also known as probing, is used by network administrators to discover possible vulnerabilities in the network through the probing of open ports. However, port scanning is also used by attackers to discover active services that may have vulnerabilities that could be exploited [30]. The real threat of port scanning resides in the fact that this technique often precedes the execution of multi-stage attacks and more elaborate intrusion attempts [30]. Hence, similar to KDD99 [28] in which probing is one of the main attack categories, we consider port scanning as a test case scenario to be detected.

Port scanning has been researched for years. Christopher [31] describes different modes of port scanning attacks. One of the most common methodologies to identify port scanning is based on monitoring if there is an increase in the number of network connections in a short period of time that exceeds a predefined threshold. Another common methodology is based on monitoring the number of different destination ports per source IP address within a given period of time. However, both methodologies require predefined thresholds to be able to detect the port scanning attack.

To carry out the port scanning attacks we have used the network mapper Nmap GUI, Zenmap [32]. Nmap is a popular open source tool that provides a variety of probing techniques for network exploitation and security auditing. Nmap and Zenmap allow the implementation of different modes or profiles of the port scanning attack (e.g. intensive, quick, or slow comprehensive scan). Each of these modes of attack will manifest itself differently in the network as the intensity of the attack varies from one another. Throughout the experiments, we have implemented a number of these different modes.

As we will explain when presenting the results, the fact that we have combined different profiles of the attack makes the detection process challenging. The intensive scan profiles standout from the normal traffic and are easily differentiable, whereas the stealth scan profile is not clearly distinguishable from the normal traffic. This may cause a large number of misclassifications during the detection process.

C. NETWORK TRAFFIC MEASUREMENTS

All the network traffic from the testbed and office LANs has been gathered by the victim using the network packets analyser Tcpdump [33] in pcap format. In total, 160 GBytes of network traffic has been gathered during the 9 days that the experiment lasted. This traffic dataset comprises 99.4% of non-malicious traffic (i.e. 696638 instances) and 0.6% of malicious traffic (i.e. 4220 instances). The small proportion of malicious traffic in the dataset makes the detection process even more challenging. The dataset is available from https://figshare.com/s/4bd0fe2dab7e09ce61dc.

Multiple parameters were extracted from the network dataset, and aggregated in a per second manner (e.g. bytes per packet, source and destination IP address, and source and destination port). These parameters have been used to compute the four metrics introduced in Section VI.B that we use to carry out the intrusion detection of the port scanning attacks (i.e. COM, THR, DPD, and SPD). The metrics COM and THR are represented in Figs. 8-9. The figures present cyclic patterns in the metric measurements, which correspond to the PoL and the time of the day when the network is being utilised by more users. The section in blue corresponds to the non-malicious traffic, whereas the section in red corresponds to the traces of port scanning attacks. Fig. 8 includes extra annotations to help identify the cycles of the PoL and the traces of port scanning attacks. Additionally, a zoomed in representation of the day 1 of the THR is shown in Fig. 10.

As we can see in Figs. 8-10, there are some attack instances that standout from the normal traffic, and are easily differentiable by using a simple signature or threshold. These instances coincide with the implementation of the intensive scan profile of the port scanning attack. In contrast, when the port scanning attack is implemented in a stealthy manner, the attack is not clearly distinguishable from the normal traffic. This will arguably cause a higher number of misclassifications during the detection process. A more detailed description of the dataset is presented in [2].

VIII. RESULTS AND ANALYSIS

A. PERFORMANCE METRICS

This section describes the detection results, and compares the results generated by our anomaly-based IDS with and

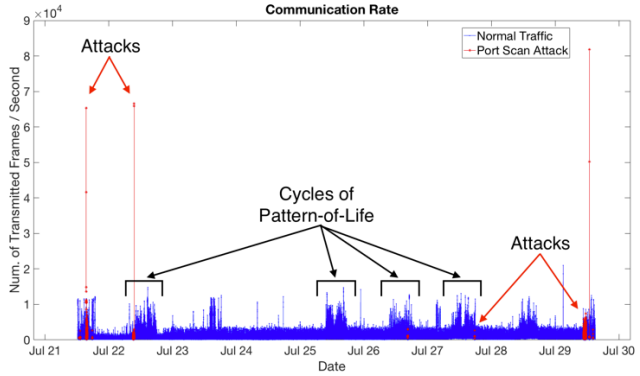

FIGURE 8. COM - Communication Rate (number of transmitted frames per second) collected over 9 days.

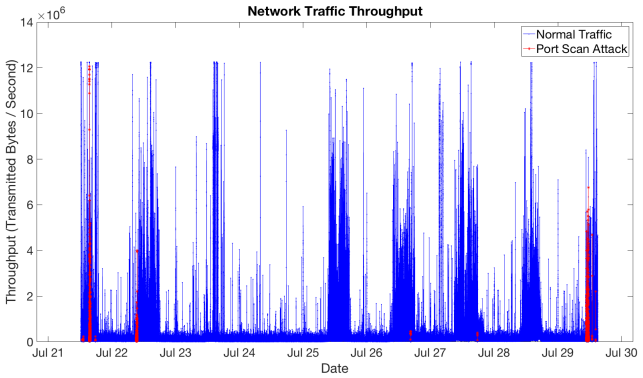

FIGURE 9. THR - Throughput (bytes per second) collected over 9 days.

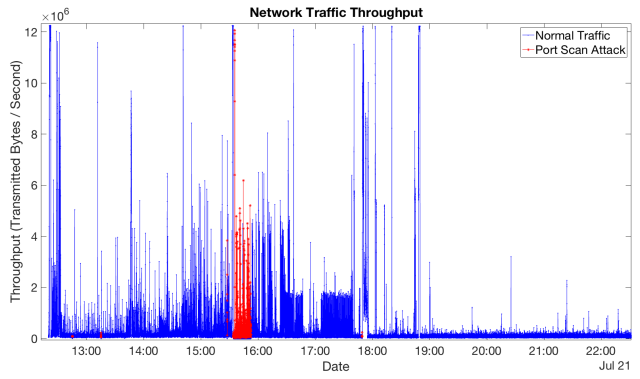

FIGURE 10. THR - Throughput gathered over 1 day, showing a zoomed in representation of the normal traffic and malicious traffic, as well as the PoL.

without the use of an FCM. There are three main purposes of these results. First, to evaluate the efficiency of the proposed approaches in identifying the presence of attacks, and reducing the number of false alarms generated by our IDS. Second, to identify which of the proposed approaches produces the best detection results. Third, to corroborate the improved detection performance of our IDS using an FCM.

The effectiveness of the IDS has been evaluated using the following performance metrics, which provide evidence of how effective the IDSs are at making correct detections:

TABLE 2. Index of the combination of metrics.

1 – DPD	6 – THR-DPD	11 – THR-SPD-DPD
2 – SPD	7 – THR-SPD	12 – CON-SPD-DPD
3 – THR	8 – COM-DPD	13 – COM-THR-DPD
4 – COM	9 – COM-SPD	14 – COM-THR-SPD
5 – SPD-DPD	10 – COM-THR	15 – COM-THR-SPD-DPD

TABLE 3. List of FCM approaches compared in the results.

Approach	Description
No FCM	*IDS Without an FCM*
FCM01	*FCM Adjustment Prior to Data Fusion Process*
FCM02	*FCM Used as an Extra Metric*
FCM03	*FCM Adjustment After Data Fusion Process*

- Detection Rate (DR) - Proportion of anomalies correctly classified as anomalous among all the anomalous data: $DR = TP/(FN+TP)$
- False Positive Rate (FPr) - Proportion of non-malicious data misclassified as anomalous among all the data: $FPr = FP/(TP+FP+TN+FN)$
- Overall Success Rate (OSR) - Proportion of all the data correctly classified among all the data: $OSR = (TN+TP)/(TP+FP+TN+FN)$

where TP represents anomalies classified as malicious; TN represents normal instances classified as normal; FP represents normal instances misclassified as attack; and FN represents anomalies misclassified as normal.

The network traffic dataset has been analysed for all the possible combinations of metrics. The Y-axis of the graphs represents the results in percentage, whereas the X-axis of the graphs represents the index of the used metrics. Each index corresponds to one possible combination of metrics, with #1 being a single metric set and #15 the set that combines all the considered metrics. Therefore, the best results overall are to be expected from the set index #15. The indexes of all the possible combinations of metrics are presented in Table 2.

Figs. 11-13 present the respective results in four graphs in each combination of metrics; one for each proposed approach. Despite the values being discrete, the line style has been used for clarity in the presentation of the results. These graphs correspond to the channels shown in Fig. 3. In these experiments, the metric used to construct the FCM was the THR; using the values in the matrix M presented in Fig. 6.

B. EVALUATION OF PROPOSED APPROACHES

The results of the proposed approaches are compared against the framework that we previously proposed in [2] and the performance of the IDS without the use of FCM (i.e. *No FCM*). Table 3 lists the different approaches that have been compared in this section.

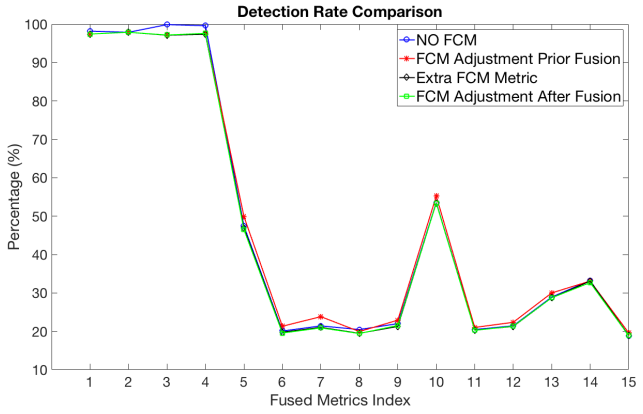

FIGURE 11. DR comparison: Three approaches that use an FCM (designed based on the THR) in conjunction with IDS, and the IDS without an FCM.

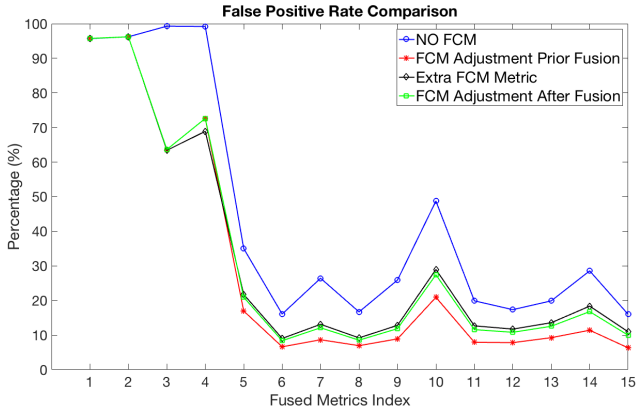

FIGURE 12. FPr comparison: Three approaches that use an FCM (designed based on the THR) in conjunction with IDS, and the IDS without an FCM.

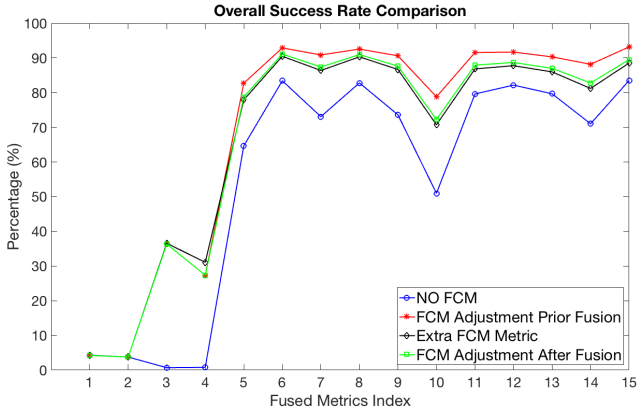

FIGURE 13. OSR comparison: Three approaches that use an FCM (designed based on the THR) in conjunction with IDS, and the IDS without an FCM.

1) DETECTION RATE RESULTS

The DR results of the IDS are compared in Fig. 11. As we can see, with regards to the DR, there is no evident difference between all the proposed approaches, using similar combination of metrics. It is worth noting that the approach that adjusts the BPA values prior to the fusion process using the FCM (i.e. *FCM01*) produces the highest DR in most cases. However, the maximum difference with the rest of DR results is only $\sim$2%. One undesirable phenomenon that is shown in Fig. 11 is that the DR decreases as the number of fused metrics increases, which is in contrast to what is expected from cross-layer IDSs. In our experiments, this phenomenon is caused by the automatic BPA methodology and the way the SW slides. When multiple metrics are used, the reference of normality in the BPA methodology becomes wider over time. Hence, the IDS becomes less sensitive and more malicious instances are misclassified as non-malicious.

2) FALSE POSITIVE RATE RESULTS

The FPr results of our IDS with and without the use of an FCM are compared in Fig. 12. Again, the metric used to construct the FCM for these results was the THR. Similarly to the DR results, we can see that the use of an FCM outperforms the FPr results produced by the IDS without the use of an FCM, for all the evaluated approaches. Also, the difference in the performance is evident. The largest difference in the FPr results occurs between *FCM01* and *No FCM*, for the set #3 (THR), which is over 35%. The largest difference for all the sets that combine two metrics is obtained in #10 (COM-THR), where the difference is 27.82%. Among the sets that combine three metrics, the largest FPr difference is obtained in #14 (COM-THR-SPD), where the difference is 17.15%. When all the metrics are combined, the difference between *FCM01* and *No FCM* is 9.68%. Also, the set of metrics #15 is the one that produces the best FPr (i.e. the lowest FPr) results, only 6.33%.

Among the three approaches that use an FCM, *FCM01* is the one that always produces the lowest FPr, when two or more metrics are combined. *FCM01* outperforms the FPr results generated by the other two approaches in $\sim$5% for all the combination of metrics; and a peak improvement of up to 8.05%, for the set #10 (COM-THR). Also, in contrast to the DR results, the FPr decreases as the number of fused metrics increases. Ideally, we would prefer to see a decrease in the FPr and an increase in the DR, at the same time. This behaviour could be associated with the tradeoff in network intrusion detection described in [34]. The authors explain that there exists a tradeoff between the reduction of the FPs by decreasing the sensitivity of the IDS and the increase of the number of misclassifications.

3) OVERALL SUCCESS RATE RESULTS

The final performance metric that we have used is the OSR. In contrast to the DR that is only based on the malicious content of the analysed data, the OSR usefully represents all the instances that have been correctly classified, regardless of whether these are malicious or not. Therefore, the OSR provides a more representative understanding of the efficiency of the IDS. Fig. 13 presents the OSR results comparison between all the approaches. Similar to the DR and FPr, the use of an FCM in the IDS outperforms the OSR results

produced by the IDS without an FCM, for all the evaluated approaches. Additionally, once again, *FCM01* is the one that always produces the best results among all the approaches.

Focusing upon the evaluation of the two approaches, *FCM01* and *No FCM*, for the set #3 (THR) the improvement in the OSR between the two methods is 35.64%. The largest improvement in all the sets that combine two metrics is obtained in #10 (COM-THR), where the difference between the two approaches is 27.38%. Among the sets that combine three metrics, the largest OSR improvement is obtained in #14 (COM-THR-SPD), where the difference is 17.15%. And finally, when all the metrics are combined, the difference between the two approaches is over 9.68% of improvement. This improvement is constant for all the combinations of metrics, and shows once more that the use of contextual information improves the detection capabilities of our anomaly-based IDS. Again, the set of metrics #15 is the one that produces the best OSR (i.e. the highest OSR) results, 93.19%. The results show that, more instances are correctly classified as more metrics are combined. With respect to the three approaches that use an FCM, once again, *FCM01* outperforms the OSR results generated by the other two approaches. The average improvement is ∼6% for all the combination of metrics; and a peak improvement of up to 8.06%, for the set #10 (COM-THR). All the detection results plotted in Figs. 11-13 have been tabulated in Table 4.

TABLE 4. Detection results - proposed approaches.

Fused Metrics Index	DR (%)				FPr (%)				OSR (%)			
	No FCM	*FCM 01*	*FCM 02*	*FCM 03*	*No FCM*	*FCM 01*	*FCM 02*	*FCM 03*	*No FCM*	*FCM 01*	*FCM 02*	*FCM 03*
1	**98.2**	97.4	97.4	97.4	95.7	95.7	95.7	95.7	**4.3**	4.2	4.2	4.2
2	97.9	97.9	97.9	97.9	96.2	96.2	96.2	96.2	3.8	3.8	3.8	3.8
3	**99.9**	97.2	97.1	97.2	99.3	63.6	63.4	63.6	0.7	36.3	**36.6**	36.3
4	**99.6**	97.6	97.4	97.6	99.2	72.6	**68.9**	72.6	0.8	27.3	**31.1**	27.3
5	47.5	**49.9**	46.9	46.5	35	**17**	21.7	20.9	64.7	**82.7**	77.9	78.7
6	20.1	**21.3**	19.8	19.1	16.1	**6.6**	9	8.4	83.4	**92.9**	90.5	91.1
7	21.4	**23.8**	21	20.9	26.4	**8.7**	13.1	12.1	73.1	**90.9**	86.4	87.4
8	**20.5**	19.9	19.5	19.4	16.7	**7**	9.2	8.5	82.8	**92.6**	90.3	91
9	22	**22.9**	21.3	21.5	25.9	**8.9**	12.8	11.9	73.6	**90.6**	86.7	87.6
10	53.5	**55.2**	53.4	53.4	48.7	**20.9**	29	27.4	51	**78.8**	70.7	72.3
11	20.5	**21**	20.3	20.3	19.9	**7.9**	12.7	11.6	79.6	**91.6**	86.8	87.9
12	21.4	**22.3**	21.3	21.3	17.4	**7.8**	11.7	10.8	82.2	**91.7**	87.8	88.7
13	28.9	**30**	28.7	28.7	19.9	**9.3**	13.6	12.6	79.7	**90.3**	86	87
14	**33.2**	33.1	33.1	32.7	28.6	**11.4**	18.4	16.8	71.0	**88.2**	81.2	82.8
15	18.8	**19.6**	19	18.9	16	**6.3**	11	10	83.5	**93.2**	88.5	89.5

*Values in bold represent the best results for each performance metric.

4) RESULTS ANALYSIS

These results indicate that by utilising only measureable information from the network without considering the available contextual information, the IDS may reach a wrong conclusion, leading to an overall low accuracy. Also, from the presented results, we can infer that the most efficient approach is to adjust the BPA values prior to the data fusion process. This is because the *FCM01* approach adjusts all of the considered metrics individually. Hence, the contribution of the contextual information adapts according to the BPA values given by the IDS for each individual metric, and would impact each of the metrics differently. Also, the contribution of the contextual information through the approach that constructs an additional metric, *FCM02*, decreases as the number of metrics being fused increases. This is because generally, after a number of consecutive D-S fusions, the BPA value given to one of the hypotheses will be largely higher than the rest of the BPAs. Therefore, the fusion of the metric computed from the FCM may not reverse the decision of the IDS. Only in cases in which the BPA values of both hypotheses, *Normal* and *Attack*, are close to each other, could the addition of the new metric have an evident effect on the final IDS decision. In the case of the approach that adjusts the BPA values after the data fusion process, *FCM03*, is prone to dominate the entire decision. On the one hand, in cases in which the BPA values of both hypotheses, *Normal* and *Attack*, are close to each other, the addition of the contextual information could greatly influence the final IDS decision. On the other hand, even if one of the hypotheses receives a largely higher BPA value than the rest, the addition of the contextual information could overturn the final IDS decision if the outcome of the FCM is larger than the resulting BPA values given by the IDS.

The most important findings can be summarised as follows:

- The use of an FCM in the detection process outperforms the efficiency of an IDS without considering the use of contextual information.
- The *FCM01* approach performs generally better than the other proposed approaches.
- The contribution of the contextual information is more evident in the case of FPr and OSR.
- The contribution of the *FCM02* approach decreases as the number of metrics being fused increases.
- The *FCM03* approach is prone to dominate the decision.

C. METRICS EVALUATION FOR FCM DESIGN

As it was explained in Section VI.B, in order to characterise the PoL of the network usage in the FCM, the network administrator defined a number of thresholds for each of the considered metrics (i.e. THR, COM, and DPD). These thresholds define expected levels of the normal usage of the network resources according to the a priori knowledge of the network administrator. One FCM model is designed for each of the considered metrics. As described in Section VI.B, any of the available metrics can be used in the construction of the FCM to characterise the network usage. However, we decided to utilise only one of the metrics at a time to reduce the complexity of the model.

This section compares the detection results generated by our IDS when the metrics THR, COM, and DPD are used to model the FCM, as well as the results of the IDS without an FCM. Since the addition of the FCM contribution prior to the data fusion process is the most efficient approach of the

three proposed, only the *FCM01* approach is considered in the results presented in Figs. 14-16. For clarity, Table 5 lists the different approaches that have been compared in this section. Again, the effectiveness of the IDS has been evaluated in terms of DR, FPr, and OSR, presented in four graphs for all the possible combinations of metrics described in Table 2.

TABLE 5. List of FCM models compared in the results.

Approach	Description
No FCM	*IDS Without an FCM*
FCM-THR	*FCM Modelled Using Throughput*
FCM-COM	*FCM Modelled Using Communication Rate*
FCM-DPD	*FCM Modelled Using Destination Port Distribution*

1) DETECTION RATE RESULTS

The DR results of the IDS are compared in Fig. 14. The most noticeable characteristic that we can see in the results is the drastic improvement provided by the *FCM-COM* and *FCM-DPD* approaches. For almost all the possible combination of metrics, both approaches provide over 99% of DR. In particular, for the set #15 (COM-THR-SPD-DPD) the DR reaches 99.76%, which improves the DR results provided by the IDS to 80.94% without the use of an FCM. Also, in contrast to the results generated by *FCM-THR*, the DR does not decrease as the number of fused metrics increases. This phenomenon manifests that the contribution of the contextual information tends to dominate the detection.

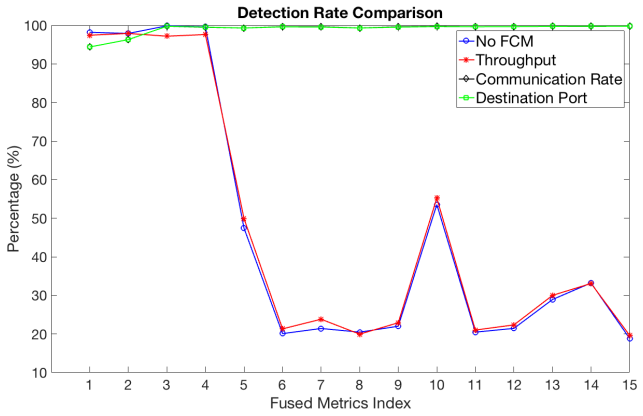

FIGURE 14. DR comparison: Adjustment of BPAs prior to the fusion process; FCM designed based on the THR, COM or DPD, and the IDS without an FCM.

2) FALSE POSITIVE RATE RESULTS

The FPr results of our IDS are compared in Fig. 15. In contrast to the DR results, we can see that the use of the *FCM-COM* and *FCM-DPD* produces a slightly higher number of false alarms, in comparison with the use of the *FCM-THR*. For the set #15 (COM-THR-SPD-DPD) the FPr reaches 26.42% and 25.42% for the *FCM-COM* and *FCM-DPD*, respectively. This represents an increase in the number of false alarms of approximately 20% (i.e. ∼140800 normal data instances misclassified). In comparison with the *No FCM* approach, the increase in the number of false alarms reaches approximately 10% (i.e. ∼73000 normal data instances misclassified). As these results suggest, the use of the *FCM-THR* approach produces the best detection results overall in terms of FPr. Also, similarly to the DR results presented in Fig. 14, the FPr does not decrease as the number of fused metrics increases.

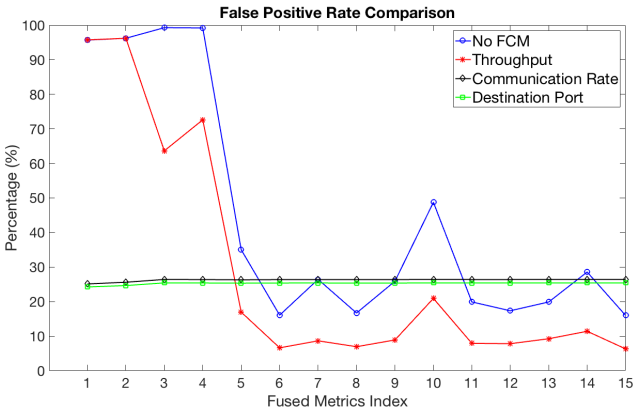

FIGURE 15. FPr comparison: Adjustment of BPAs prior to the fusion process; FCM designed based on the THR, COM or DPD, and the IDS without an FCM.

3) OVERALL SUCCESS RATE RESULTS

Fig. 16 presents the OSR results comparison of our IDS. Generally, the OSR results produced by the *FCM-COM* and *FCM-DPD* approaches remain almost unchanged regardless of whether the number of fused metrics increases or not. As in the previous results (i.e. DR and FPr) when the *FCM-COM* and *FCM-DPD* are used, the contribution of the FCM tends to dominate the intrusion detection process. Both approaches *FCM-COM* and *FCM-DPD* produce lower OSR overall than the *FCM-THR* approach for almost all the combinations of metrics. For the set #15 (COM-THR-SPD-DPD), which is the set index expected to produce the best results overall, the OSR reaches 73.58% and 74.58% for the *FCM-COM* and *FCM-DPD*, respectively. This represents a decrease of approximately 20% of OSR with respect to the use of *FCM-THR*, and a decrease of approximately 10% of OSR with respect to the IDS without an FCM. As these results suggest, once again, the use of the metric THR to construct the FCM produces the best detection results overall.

4) RESULTS ANALYSIS

From the presented results, the important role played by the metric selection in the design of the modelled FCM (i.e. THR, COM and DPD) in the effectiveness of the IDS is clear. Based on the 99% of DR obtained when either the approach *FCM-COM* or *FCM-DPD* is used, we might incorrectly assume that these are the best selection

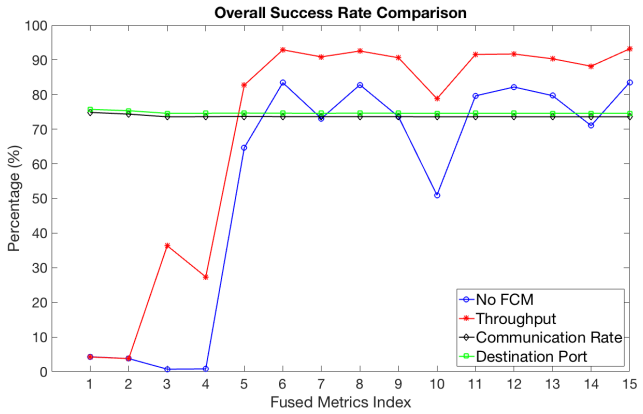

FIGURE 16. OSR comparison: Adjustment of BPAs prior to the fusion process; FCM designed based on the THR, COM or DPD, and the IDS without an FCM.

TABLE 6. Detection results - FCM design.

Fused Metrics Index	DR (%)				FPr (%)				OSR (%)			
	No FCM	*FCM THR*	*FCM COM*	*FCM DPD*	*No FCM*	*FCM THR*	*FCM COM*	*FCM DPD*	*No FCM*	*FCM THR*	*FCM COM*	*FCM DPD*
1	**98.2**	97.4	94.4	94.4	95.7	95.7	25.1	**24.2**	4.3	4.2	74.8	**75.7**
2	97.9	97.9	96.3	96.3	96.2	96.2	25.6	**24.6**	3.8	3.8	74.3	**75.3**
3	**99.9**	97.2	99.8	99.8	99.3	63.6	26.4	**25.4**	0.7	36.3	73.6	**74.6**
4	**99.6**	97.6	99.5	99.5	99.2	72.6	26.4	**25.4**	0.8	27.3	73.6	**74.6**
5	47.5	49.9	99.3	99.3	35	**17**	26.3	25.3	64.7	**82.7**	73.7	74.6
6	20.1	21.3	99.6	99.6	16.1	**6.6**	26.4	25.4	83.4	**92.9**	73.6	74.6
7	21.4	23.8	**99.6**	99.5	26.4	**8.7**	26.4	25.4	73.1	**90.9**	73.6	74.6
8	20.5	19.9	99.3	99.3	16.7	**7**	26.4	25.4	82.8	**92.6**	73.6	74.6
9	22	22.9	**99.6**	99.5	25.9	**8.9**	26.4	25.4	73.6	**90.6**	73.6	74.6
10	53.5	55.2	99.7	99.7	48.7	**20.9**	26.4	25.4	51	**78.8**	73.6	74.6
11	20.5	21	99.6	99.6	19.9	**7.9**	26.4	25.4	79.6	**91.6**	73.6	74.6
12	21.4	22.3	99.7	99.7	17.4	**7.8**	26.4	25.4	82.2	**91.7**	73.6	74.6
13	28.9	30	99.7	99.7	19.9	**9.3**	26.4	25.4	79.7	**90.3**	73.6	74.6
14	33.2	33.1	99.7	**99.8**	28.6	**11.4**	26.4	25.4	71.0	**88.2**	73.6	74.6
15	18.8	19.6	99.7	99.7	16	**6.3**	26.4	25.4	83.5	**93.2**	73.6	74.6

*Values in bold represent the best results for each performance metric.

of metrics. An IDS that triggers ∼140800 false alarms during the 9 days that the experiment lasted would make the network administrator ignore the generated alarms. A tradeoff between the DR and the false alarms should be found, based on the needs of the protected network. The metric selection should be based on whether we prioritise a system that detects most of the attacks regardless of the number of false alarms, or whether we prioritise reducing the number of misclassifications. All the results plotted in Figs. 14-16 have been tabulated in Table 6.

The most important findings can be summarised as follows:

- The *FCM-COM* and *FCM-DPD* approaches produce the best detection results overall in terms of DR.
- The *FCM-THR* approach produces the best detection results overall in terms of FPr and OSR.
- The use of contextual information in the detection process, regardless of the metric used in the modelling of the FCM, outperforms the efficiency of an IDS without an FCM.

IX. CONCLUSIONS AND FUTURE WORK

In this paper we have advocated incorporating high-level information regarding a monitored network, in the form of the PoL and the network administrator prior knowledge, when taking decisions on whether an attack is present in the network traffic. We have proposed two additional approaches to the one previously presented in [2], that use an FCM in conjunction with an IDS to add the contextual information into the detection process. The analysed dataset was gathered from a real LAN in a research office during 9 days, and comprises normal traffic and traces of port scanning attacks.

Initially, we have compared the results generated by our IDS without the use of an FCM and with the application of all the proposed approaches, using the metric THR to design the FCM. In terms of DR results, the proposed approaches provide an improvement to the DR results of only ∼2%. Nonetheless, with regards to the FPr and the OSR, it is evident that the use of PoL constantly improves the detection capabilities of the IDS, for all the possible combinations of metrics. Among all the FCM approaches, *FCM01*, the approach that adjusts the BPA values prior to the fusion process using the FCM, is the one that always produces the lowest FPr and higher OSR. The *FCM01* approach is able to produce only 6.33% of FPr, and up to 93.19% of OSR, in the best-case scenario.

During a second set of experiments, we compared how the selection of metrics used by the network administrator to represent the expected normal usage of the network resources may influence the detection results. From the presented results, we can see that there are significantly distinct results, especially in terms of DR. The use of the metric COM and DPD in the design of the FCMs drastically improves the DR results produced when the THR is utilised. For almost all the possible combination of metrics, both approaches provide over 99% of DR. This represents an improvement of almost 80% of DR in comparison to the use of the THR. However, with regards to the FPr and the OSR, the use of the metrics COM and DPD produced slightly worse detection results than the use of the THR. By using all the considered metrics, the FPr reaches 26.42% and 25.42% for the COM and DPD, respectively. This represents an increase in the number of false alarms of approximately 140800 data instances misclassified.

Different conclusions could be extracted from the presented results. First, these results empirically confirm that the use of the FCM provides improvement to the effectiveness of the IDS. Also, the presented results ratify that adjusting the BPAs prior to the data fusion provides the best use of the PoL in the detection process. The number of false alarms may not be low enough to be acceptable and make the IDS usable in practice in multiple scenarios. Nonetheless, it is important to reiterate and emphasise that the aim of this work is to

investigate the best way to use an FCM in conjunction with an IDS. Similarly, as we previously explained, in this work we have chosen this type of attack mainly as a means to evidence the benefit of including high-level information as part of the detection process.

Additionally, the selection of the metrics used to design the modelled FCM plays an important role in the effectiveness of the IDS. The metric selection should be based on whether a system that detects most of the attacks or a system that reduces the number of misclassifications should be prioritised. A tradeoff between the DR and the false alarms should be found, based on the needs of the protected network.

As for future work, we wish to research novel methods to characterise the time of the day in the FCM, to extract the available high-level information used to construct the FCM, and to refine the setting of the different thresholds that characterise the various concepts in the FCM model. Also, we wish to implement novel techniques to address the problem of fixed model convergence in non-stable FCMs, and evaluate the time complexity of the presented system. Finally, we wish to focus on the use of an FCM in conjunction with the IDS to enhance the detection capabilities of multi-stage attacks.

REFERENCES

[1] F. J. Aparicio-Navarro, K. G. Kyriakopoulos, D. J. Parish, and J. A. Chambers, "Adding contextual information to intrusion detection systems using fuzzy cognitive maps," in *Proc. IEEE Int. Multi-Disciplinary Conf. Cognit. Methods Situation Awareness Decision Support (CogSIMA)*, Mar. 2016, pp. 187–193.

[2] F. J. Aparicio-Navarro, J. A. Chambers, K. G. Kyriakopoulos, Y. Gong, and D. J. Parish, "Using the pattern-of-life in networks to improve the effectiveness of intrusion detection systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.

[3] F. J. Aparicio-Navarro, K. G. Kyriakopoulos, and D. J. Parish, "Automatic dataset labelling and feature selection for intrusion detection systems," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Oct. 2014, pp. 46–51.

[4] J. Kittler *et al.*, "Domain anomaly detection in machine perception: A system architecture and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 845–859, May 2014.

[5] A. Sadighian, S. T. Zargar, J. M. Fernandez, and A. Lemay, "Semantic-based context-aware alert fusion for distributed intrusion detection systems," in *Proc. Int. Conf. Risks Secur. Internet Syst. (CRiSIS)*, Oct. 2013, pp. 1–6.

[6] C. D. Stylios and P. P. Groumpos, "Modeling complex systems using fuzzy cognitive maps," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 34, no. 1, pp. 155–162, Jan. 2004.

[7] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[8] L. Snidaro, J. García, and J. Llinas, "Context-based information fusion: A survey and discussion," *Inf. Fusion*, vol. 25, pp. 16–31, Sep. 2015.

[9] D. Gupta, P. S. Joshi, A. K. Bhattacharjee, and R. S. Mundada, "IDS alerts classification using knowledge-based evaluation," in *Proc. 4th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2012, pp. 1–8.

[10] K. Xu, K. Tian, D. Yao, and B. G. Ryder, "A sharper sense of self: Probabilistic reasoning of program behaviors for anomaly detection with context sensitivity," in *Proc. 46th Int. Conf. Dependable Syst. Netw. (DSN)*, Jun./Jul. 2016, pp. 467–478.

[11] T. D. Ndousse and T. Okuda, "Computational intelligence for distributed fault management in networks using fuzzy cognitive maps," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 3. Jun. 1996, pp. 1558–1562.

[12] E. I. Papageorgiou *et al.*, "Brain tumor characterization using the soft computing technique of fuzzy cognitive maps," *Appl. Soft Comput.*, vol. 8, no. 1, pp. 820–828, 2008.

[13] R. E. T. Jones, E. S. Connors, M. E. Mossey, J. R. Hyatt, N. J. Hansen, and M. R. Endsley, "Modeling situation awareness for Army infantry platoon leaders using fuzzy cognitive mapping techniques," in *Proc. Behavior Represent. Modelling Simulation Conf. (BRiMS)*, Mar. 2010, pp. 216–223.

[14] M. M. Kokar and M. R. Endsley, "Situation awareness and cognitive modeling," *IEEE Intell. Syst.*, vol. 27, no. 3, pp. 91–96, May/Jun. 2012.

[15] W. P. Cheah, Y. S. Kim, K.-Y. Kim, and H.-J. Yang, "Systematic causal knowledge acquisition using FCM constructor for product design decision support," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 15316–15331, 2011.

[16] S. Limon, O. P. Yadav, and B. Nepal, "Modeling cognitive network of a physical system using design knowledge base," in *Proc. IEEE Int. Conf. Ind. Eng., Eng. Manage. (IEEM)*, Dec. 2014, pp. 238–242.

[17] M. Jazzar and A. Jantan, "Towards real-time intrusion detection using fuzzy cognitive maps modeling and simulation," in *Proc. Int. Symp. Inf. Technol. (ITSim)*, vol. 2, Aug. 2008, pp. 1–6.

[18] K. G. Kyriakopoulos, F. J. Aparicio-Navarro, and D. J. Parish, "Manual and automatic assigned thresholds in multi-layer data fusion intrusion detection system for 802.11 attacks," *IET Inf. Secur.*, vol. 8, no. 1, pp. 42–50, 2014.

[19] G. Thamilarasu, S. Mishra, and R. Sridhar, "A cross-layer approach to detect jamming attacks in wireless ad hoc networks," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Oct. 2006, pp. 1–7.

[20] X. Wang, J. S. Wong, F. Stanley, and S. Basu, "Cross-layer based anomaly detection in wireless mesh networks," in *Proc. Int. Symp. Appl. Internet (SAINT)*, Jul. 2009, pp. 9–15.

[21] F. J. Aparicio-Navarro, K. G. Kyriakopoulos, and D. J. Parish, "An automatic and self-adaptive multi-layer data fusion system for WiFi attack detection," *Int. J. Internet Technol. Secured Trans.*, vol. 5, no. 1, pp. 42–62, 2013.

[22] R. E. T. Jones, E. S. Connors, and M. R. Endsley, "Incorporating the human analyst into the data fusion process by modeling situation awareness using fuzzy cognitive maps," in *Proc. Int. Conf. Inf. Fusion (FUSION)*, Jul. 2009, pp. 1265–1271.

[23] L. Rodriguez-Repiso, R. Setchi, and J. L. Salmeron, "Modelling IT projects success with fuzzy cognitive maps," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 543–559, 2007.

[24] G. Nápoles, E. Papageorgiou, R. Bello, and K. Vanhoof, "On the convergence of sigmoid fuzzy cognitive maps," *Inf. Sci.*, vols. 349–350, pp. 154–171, Jul. 2016.

[25] S. Bueno and J. L. Salmeron, "Benchmarking main activation functions in fuzzy cognitive maps," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5221–5229, 2009.

[26] A. K. Tsadiras, "Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps," *Inf. Sci.*, vol. 178, no. 20, pp. 3880–3894, 2008.

[27] R. Craddock, D. Watson, and W. Saunders, "Generic Pattern of Life and behaviour analysis," in *Proc. IEEE Int. Multi-Disciplinary Conf. Cognit. Methods Situation Awareness Decision Support (CogSIMA)*, Mar. 2016, pp. 152–158.

[28] UCI. (1999). *KDD Cup 1999 Data*. Accessed: May 5, 2017. [Online]. Available: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[29] B. Portier and J. Froment, "Data mining techniques for intrusion detection," Univ. Texas, Austin, TX, USA, 2000.

[30] E. Bou-Harb, M. Debbabi, and C. Assi, "Cyber scanning: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1496–1519, 3rd Quart., 2014.

[31] R. Christopher, "Port scanning techniques and the defense against them," SANS Inst. Tech. Rep., 2002, pp. 1–6. [Online]. Available: https://uk.sans.org/reading-room/whitepapers/auditing/port-scanning-techniques-defense-70

[32] G. Lyon. *NMAP: The Network Mapper—Free Security Scanner*. Accessed: Jun. 21, 2016. [Online]. Available: http://nmap.org/

[33] V. Jacobson, C. Leres, and S. McCanne. (1987). *TCPDUMP/LIBPCAP*. Accessed: Jun. 23, 2016. [Online]. Available: http://www.tcpdump.org

[34] T. Alpcan and T. Başar, *Network Security: A Decision and Game-Theoretic Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

FRANCISCO J. APARICIO-NAVARRO received the B.Eng. degree in telecommunications engineering, specialized in computer networks from the Technical University of Cartagena, Spain, in 2009, and the Ph.D. degree in computer network security from Loughborough University, Loughborough, U.K., in 2014.

From 2013 to 2016, he was a Research Associate with the School of Electronic, Electrical and Systems Engineering, Loughborough University, U.K. Since 2016, he has been with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, U.K., where he is currently a Research Associate with the Communications, Sensors, Signal, and Information Processing Group. He is also an Academic Visitor with the Wolfson School of Mechanical, Electronic and Manufacturing Engineering, Loughborough University, Loughborough, U.K.

KONSTANTINOS G. KYRIAKOPOULOS (M'07) received the B.Sc. degree in electrical engineering from the Technological Education Institute of Larisa, Greece, in 2003, the M.Sc. degree in digital communication systems and the Ph.D. degree in computer networks from Loughborough University, Loughborough, U.K., in 2004 and 2008, respectively.

From 2008 to 2016, he was a Research Associate with the School of Electronic, Electrical, and Systems Engineering, Loughborough University, U.K., involved mainly in EPSRC projects and successfully licensing research output from his work. Since 2016, he has been an Academic Member with the Wolfson School of Mechanical, Electronic and Manufacturing Engineering, Loughborough University, Loughborough, U.K., and the Institute of Digital Technologies, Loughborough University London, London, U.K. His research interests are in the areas of computer networks, including network security, intrusion detection, vehicular communications, intelligent decision making based on network situational awareness and network performance measurements in emerging network paradigms and their applications.

YU GONG (M'07) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1992 and 1995, respectively, and the Ph.D. degree in communications from the National University of Singapore, in 2002.

He has had several research positions with the Institute of Infocomm Research, Singapore, and Queen's University, Belfast, U.K., respectively. From 2006 to 2012, he was an Academic Member with the School of Systems Engineering, University of Reading, Reading, U.K. Since 2012, he has been with the School of Electronic, Electrical, and Systems Engineering, Loughborough University, Loughborough, U.K. His research interests are in the area of signal processing and communications, including wireless communications, cooperative networks, nonlinear and nonstationary system identification, and adaptive filters.

DAVID J. PARISH received the B.Sc. degree (Hons.) in physics with electronics and the Ph.D. degree in electronic engineering from the University of Liverpool, Liverpool, U.K., in 1979 and 1983, respectively.

From 1984 to 2014, he researched communication networks with the School of Electronic, Electrical, and Systems Engineering, Loughborough University, U.K., becoming a Professor. In 2012, he became the Dean of the School and currently an Emeritus Professor. He has over 30 years of experience of communication networks research, with particular emphasis on network performance measurement, network abuse detection, and network simulation. He has established an award-winning M.Eng. programme in Systems Engineering and has graduated over 25 Ph.D. students. He has also been the Treasurer and the Chair of the IEEE U.K. and RI Communications Society (COMSOC) Section.

JONATHON A. CHAMBERS (S'83–M'90–SM'98–F'11) received the Ph.D. and D.Sc. degrees in signal processing from the Imperial College of Science, Technology, and Medicine, Imperial College London, London, U.K., in 1990 and 2014, respectively.

From 1991 to 1994, he was a Research Scientist with the Schlumberger Cambridge Research Centre, Cambridge, U.K. In 1994, he returned to Imperial College London as a Lecturer in signal processing and was promoted to a Reader (Associate Professor) in 1998. From 2001 to 2004, he was the Director of the Centre for Digital Signal Processing and a Professor of signal processing with the Division of Engineering, King's College London. From 2004 to 2007, he was a Cardiff Professorial Research Fellow with the School of Engineering, Cardiff University, Cardiff, U.K. From 2007 to 2014, he led the Advanced Signal Processing Group, School of Electronic, Electrical, and Systems Engineering, Loughborough University, U.K., where he is currently a Visiting Professor. Since 2015, he has been with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, U.K., where he is currently a Professor of signal and information processing and heads the Communications, Sensors, Signal, and Information Processing Group. He is also the International Honorary Dean with the School of Automation, Harbin Engineering University, Harbin, China. He has advised over 70 researchers through to Ph.D. graduation and authored over 500 conference proceedings and journal articles, many of which are in IEEE journals. His research interests include adaptive signal processing, machine learning and their applications.

Prof. Chambers is a fellow of the Royal Academy of Engineering, U.K., and the Institution of Electrical Engineers. He has served on the IEEE Signal Processing Theory and Methods Technical Committee for six years, the IEEE Signal Processing Society Awards Board, the IEEE Signal Processing Conference Board and the European Signal Processing Society Best Paper Awards Selection Panel. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING for three terms over the periods 1997–1999, 2004–2007, and between 2011 and 2015 as a Senior Area Editor. He received the first QinetiQ Visiting Fellowship in 2007 for his outstanding contributions to adaptive signal processing and his successful industrial collaboration with the international defence systems company QinetiQ.