Adding Contextual Information to Intrusion Detection Systems Using Fuzzy Cognitive Maps

Francisco J. Aparicio-Navarro, Konstantinos G. Kyriakopoulos, David J. Parish
School of Mechanical, Manufacturing and Electrical Eng.
Loughborough University
Loughborough, LE11 3TU, UK
e-mail: {elfja2, elkk, d.j.parish}@lboro.ac.uk

Jonathon A. Chambers
School of Electrical and Electronic Engineering
Newcastle University
Newcastle upon Tyne, NE1 7RU, UK
e-mail: jonathon.chambers@ncl.ac.uk

Abstract—**In the last few years there has been considerable increase in the efficiency of Intrusion Detection Systems (IDSs). However, networks are still the victim of attacks. As the complexity of these attacks keeps increasing, new and more robust detection mechanisms need to be developed. The next generation of IDSs should be designed incorporating reasoning engines supported by contextual information about the network, cognitive information and situational awareness to improve their detection results. In this paper, we propose the use of a Fuzzy Cognitive Map (FCM) in conjunction with an IDS to incorporate contextual information into the detection process. We have evaluated the use of FCMs to adjust the Basic Probability Assignment (BPA) values defined prior to the data fusion process, which is crucial for the IDS that we have developed. The experimental results that we present verify that FCMs can improve the efficiency of our IDS by reducing the number of false alarms, while not affecting the number of correct detections.**

***Keywords-** Basic Probability Assignment; Contextual Information; Dempster-Shafer Theory; Fuzzy Cognitive Maps; Intrusion Detection Systems; Network Security*

I. INTRODUCTION

There has been considerable increase in the efficiency of Intrusion Detection Systems (IDSs) in the last few years. Great effort has been made by researchers and private companies in the area of network security to increase the efficiency of the IDSs. The use of data mining techniques, data fusion approaches, and the combined use of different IDSs have contributed to this achievement. However, networks are still victims of intrusions and cyber-attacks. As the complexity of these attacks keeps increasing, new and more robust detection mechanisms need to be developed. The next generation of IDSs should be designed incorporating reasoning engines supported by modules that could assess the quality of the analysed datasets, manage contextual and non-contextual information about the network, or deal with incongruent decisions between different IDSs.

In our previous work [1], we made initial efforts to add to IDSs the capability of assessing the quality of the analysed datasets. In this paper, we argue that the next generation of IDSs should be developed with the capability of incorporating contextual information, situational awareness and cognitive information to the intrusion detection process. The network administrator and users would provide this high-level information; not always measurable within the network. Generally, the IDSs should be able to adapt their detection characteristics based on the context in which these systems operate. In addition, the information provided by the users should also be used to adjust the detection characteristics of the IDSs. Current IDSs utilise measurable network traffic information from the protected system or signatures of known attacks during the intrusion detection process, but these systems do not take into account available high-level information (i.e. above the network operation) regarding the protected system to improve their effectiveness [2].

The problem faced here is how to represent this information and how to incorporate it into the detection process. The approach that we propose in this work is the use of a Fuzzy Cognitive Map (FCM) [3] in conjunction with our anomaly-based IDS. An FCM provides a useful framework for network users to contribute their knowledge, to model new and unseen situations, unknown behaviours, and to calculate the influence that each individual event may have in the whole system and in other events. We argue that this approach may be used to fine-tune some of the techniques used by the IDSs and, as a consequence, improve their detection results.

Our unsupervised anomaly-based IDS [4] is based on the combined use of multiple metrics from multiple layers of the TCP/IP stack to carry out the intrusion detection. It uses the Dempster-Shafer (D-S) Theory of Evidence as a data fusion technique. D-S combines belief values given to different hypotheses, also known as Basic Probability Assignment (BPA) values. In this paper, we propose three possible approaches for the use of FCMs to help to generate or modify these beliefs, based on the information provided by the user. One of the approaches is to use the FCM to define the weights in Weighted D-S (WD-S) theory. Another approach is to let the FCM define the actual BPA values. The third approach is to adjust the generation of the BPA values prior to the data fusion process, by using the outcome of the FCM.

Although we propose three approaches, in this paper we have only implemented the third one, which adjusts the BPA values. The implementation of the other proposed approaches requires the adjustment of the source code of the IDS. We have conducted a series of experiments to showcase the efficiency of the proposed approach to incorporate FCM in the adjustment of the BPA values. The results obtained from these experiments empirically verify that an FCM can improve the efficiency of

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014307/1 and the MOD University Defence Research Collaboration in Signal Processing.

our IDS by greatly reducing the number of false alarms generated, while not affecting the number of correct detections.

The paper is organised as follows. In Section II, the most relevant work is reviewed. An extensive description of an FCM is presented in Section III. In Section IV, the three possible approaches for the use of an FCM that we propose, as well as a short description of D-S theory are presented. In Section V, the testbed, the dataset employed and the FCM process are described. The obtained results are explained in Section VI. Finally, conclusions are given in Section VII.

II. Related Work

Future IDSs need to intelligently tackle network attacks, not only by using measurable information from the network or past experiences identifying these attacks, but also by integrating human cognition and contextual information into the detection process. Incorporating this high-level information into security systems aims to improve their detection effectiveness.

According to [5], contextual information could be defined as any information that surrounds a situation of interest, which helps to understand and to characterise the situation. The authors of [5] 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 [6], the authors propose an IDS that relies on contextual information to classify the alerts as relevant or irrelevant. The IDS makes use of contextual information about hosts present in the network and known vulnerabilities. The alerts generated by the system are processed along with the high-level information to generate a relevance score about the alerts. Their results demonstrate the effectiveness of using contextual information in the detection process to increase the efficiency of the IDSs. However, this is a supervised system, and its performance depends on a chosen threshold, which is selected in a supervised manner after a number of runs.

The authors of [7] use finite state machines with associated performance metrics to model the behaviour of armed forces. Their system uses sensor measurements of the soldiers' actions to estimate the states that they are in, at any given time. A node is assigned to each state, along with associated attributes including classifier and the performance metrics for the state. The different states that compose the finite state machine are defined using ontologies.

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, to model behaviours schematically, or to provide formal specification of different entities in a system and their relationships. Other researchers [2, 8, 9] have used ontologies to represent contextual information. For instance, in [9], the authors tackle the problem of adding contextual information in the smart car domain. They propose a hierarchical model that defines a number of known situations using ontologies. In [8], the authors model contextual information using interrelated concepts of diverse ontologies. The authors of [2] propose a security ontologies-based approach 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.

All these techniques have proven to increase the efficiency of different systems. However, none of these techniques is able to model the influence of the different states/events in other events or the whole system. Apart from modelling influences between events, an FCM provides the capability of integrating contextual information from the user to the detection process.

FCMs have been previously described and used in [3, 10, 11] to model human knowledge, new and unseen behaviours of particular scenarios or actions. The authors of [3] provide a detailed description of the FCM and its mathematical foundation. Although the work presented in [11] does not focus on network security, it comprehensively describes the FCM concept with clear examples. Similarly, the authors in [10] provide a detailed description of an FCM and examples that use FCM to model fault propagation in interconnected systems.

The work presented in [12] focuses on developing an actionable model of situation awareness for Army infantry platoon leaders that could replicate human cognition using FCMs. Their FCM is designed based on the 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 [12] is that the people responsible for designing the FCM do not provide weight values to the concepts, but rank the importance of each concept. A similar approach is presented in [13], in which situation awareness is represented using FCM. In addition, the authors of [13] also use ontologies to replicate situations.

In [14], 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. Fuzzy Cognitive Map

An FCM is a technique that can be used to model human knowledge, to model new and unseen behaviours of a system, or to calculate the influence that one event may have in the system and in other events. This technique is a graphical representation of the modelled system, as perceived by different human experts. The main goal of modelling a decision problem using an FCM is to be able to predict the outcome of the problem by letting the relevant events interact, and to show which of the events influence the other events and the degree of influence [3].

The graphical representation of FCMs is characterised by a set of nodes interconnected by casual bidirectional connections. An example of an FCM model is presented in Fig. 1. The nodes in the FCM represent casual and time-varying concepts, events, actions or goals that describe the behaviour of the system. With regards to this work, each node represents a change in the network throughput, time and date, and each event is independent from each other. Each node C carries a weight A(t) in the fuzzy interval [0,1], which indicates the quantitative measure of the importance that each event has in the system, at time t. The connections between nodes represent the causal relationship between the modelled events. Each link is assigned a weight value $w_{ij}(t)$ in the fuzzy interval [-1, 1], which indicates the relationship and degree of influence between the nodes C_i and C_j. There are three possible relationships between nodes: 1) $w_{ij} > 0$, indicating a positive relationship; 2) $w_{ij} < 0$, negative relationship; 3) $w_{ij} = 0$, no relationship. Positive relationship indicates that $A_j(t)$ increases as $A_i(t)$ also increases, whereas a negative relationship indicates that $A_j(t)$ increases as $A_i(t)$ decreases, and $A_j(t)$ decreases as $A_i(t)$ increases.

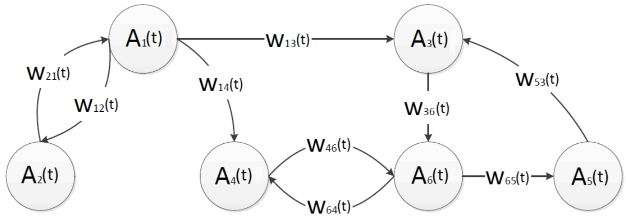


Figure 1. Simple FCM model in which nodes represent changes in modelled system and connections denote relationships between events.

The first step in the process of creating an FCM is defining the main events relevant for the considered system. This is done by a number of experts that, apart from defining the main events relevant for the considered system, also describe the relationships among these events, based on their individual knowledge. The negative or positive influence of one event on the others is defined along with the degree of influence. The fuzzy degree of influence is assigned using linguistic variables, as described in [15]. Finally, this knowledge is transformed into numerical vectors associated with each event.

An FCM can be represented by a [*n*x*n*] matrix M, where $[M]_{ij} = |w_{ij}|$ known as an adjacency matrix, describing the relationship between the nodes and the weight values $w_{ij}(t)$ associated with each link, where *n* is the number of nodes. The initial weight value $A_i(t)$ of all the nodes in a model at time t = 0 can be represented by the initial vector state A, where $A(0) = (A_1(0), A_2(0),\ldots, A_n(0))$. $A_i(0)$ is the weight value of Node i at time t = 0. At each future time step, the weight value of each event is calculated by aggregating the influence of the interconnected events on the respective weight. The new weights of the nodes are computed via an iterative process, using an activation function *f*. The value of $A_i(t)$ changes at each iteration as described below in (1), 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.

$$A_i(t+1) = f\left(A_i(t) + \sum_{j=1 j\neq i}^{n} w_{ji}(t) * A_j(t)\right) \quad (1)$$

This process continues during a number of iterations until the FCM reaches one final fixed model, known as a hidden pattern or fixed-point attractor. It is also possible that FCM keeps cycling between several fixed models, known as a limit cycle, or it may continue generating different models indefinitely. The authors of [15] describe four activation functions *f*; these are sigmoid, hyperbolic, linear threshold, and step function. The sigmoid is the most commonly used activation function in an FCM according to [15]. Therefore, by using the sigmoid function, the evolution of the weight values $A_i(t)$ would be as represented in (2), where λ is a constant that indicates the function slope (degree of normalisation) [15].

$$A_i(t+1) = \frac{1}{1+e^{-\lambda\left(A_i(t)+\sum_{j=1 j\neq i}^{n} w_{ji}(t)*A_j(t)\right)}} \quad (2)$$

One of the most important characteristics of the FCM is its capability to combine knowledge of multiple human experts. Also, it is not necessary that all human experts involved in the process agree on which events should compose the FCM or what weight should be assigned to each node. FCMs can deal with multiple, incomplete, contradictory or conflicting pieces of information. This technique also allows for different FCMs to be combined additively. The addition of k adjacency matrices can be calculated as described in (3):

$$[M]_{ij} = \frac{1}{k}\sum_{n=1}^{k} \left|w_{ij}\right|_n \quad \forall\, i, j \quad (3)$$

Another characteristic of an FCM is that the model of a system can be easily modified to incorporate new behaviours, knowledge from new human experts, or changes to the current modelled behaviour. An FCM provides the potential to make changes easily and intuitively, and allows combining additional pieces of information at a later time instance. Additionally, it supports memberships of more than one set of events and allows the overlapping of different FCM models [16]. A more detailed description of FCM models can be found in [3].

IV. Proposed Use of An FCM Within An IDS

In this section we propose three approaches by which an FCM could be integrated within our unsupervised anomaly-based IDS [4]. These approaches are based on the generation or modification of the BPA values used in a D-S formulation.

A. *Dempster-Shafer Theory*

The Dempster-Shafer theory of evidence is a data fusion technique. 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, the frame of discernment is comprised of *A* = *Attack* and *N* = *Normal*. Assuming Θ has two outcomes $\{A, N\}$, the total number of subsets of Θ, defined by the number of hypotheses that it composes, is $2^{\Theta} = \{A, N, \{A|N\}, \emptyset\}$. Each hypothesis from the power set of the frame of discernment is assigned a BPA value within the range [0, 1].

D-S uses Dempster's rule of combination to calculate the orthogonal summation of the belief values from two different sensors, and fuses this information into a single belief. This rule is defined in (4), where $m_1(H)$ and $m_2(H)$ are the beliefs in the hypothesis *H*, from observers 1 and 2, respectively.

$$m(H) = \frac{\sum_{X\cap Y=H} m_1(X) * m_2(Y)}{1-\sum_{X\cap Y=\emptyset} m_1(X) * m_2(Y)} \quad (4)$$

As we have previously described in [4], our unsupervised anomaly-based IDS is based on the combined use of multiple metrics from multiple layers of the protocol stack. It provides three levels of belief or BPA values, for each analysed network frame. These are belief in *Normal*, which indicates how strong the belief is in the hypothesis that the current analysed frame is non-malicious, belief in *Attack*, which indicates how strong the belief is in the hypothesis that the current analysed frame is malicious, and belief in *Uncertainty*, which indicates how doubtful the system is regarding whether the current analysed frame is malicious or non-malicious. The proposed system exploits a sliding window scheme to manage the statistical distribution of the data and generate the different belief values. The length of the sliding window will generally affect the final detection results. Once these values have been generated, the BPA values are fused using D-S theory.

The belief values are automatically and self-adaptively generated, based on the current characteristics of the network measurements, using three independent statistical approaches.

B. *FCM in Weighted D-S Theory*

The first approach that we propose is based on the generation of the weight values in WD-S theory using FCMs. Dempster's rule of combination (4) assumes that all the observers have a similar level of trust, however in reality some observers may be more reliable or trustworthy than others. To overcome this problem, WD-S, which extends the D-S theory, allows an independent level of trust to be assigned to each of the *n* observers. These weight values w_i, i={1, 2,…, n}, may change over time and situation. Dempster's rule for weight D-S is defined in (5) as:

$$m(H) = \frac{\sum_{X\cap Y=H}[m_1(X)]^{w_1} * [m_2(Y)]^{w_2}}{1-\sum_{X\cap Y=\emptyset}[m_1(X)]^{w_1} * [m_2(Y)]^{w_2}} \qquad (5)$$

We argue that WD-S would allow the incorporation of contextual information into the intrusion detection process via the form of individual weight values for each observer. The problem to be addressed is how to determine these weights. We propose that the single weight value $A_i(t)$, generated by (1), associated with each of the events C_i from the FCM could be used to define the weights in the weighted D-S theory. Equation (6) represents the modified Dempster's rule with two events C_1 and C_2.

$$m(H) = \frac{\sum_{X\cap Y=H}[m_1(X)]^{A_1(t)} * [m_2(Y)]^{A_2(t)}}{1-\sum_{X\cap Y=\emptyset}[m_1(X)]^{A_1(t)} * [m_2(Y)]^{A_2(t)}} \qquad (6)$$

C. *Composed FCM Weight Values*

The second approach that we propose is based on creating the FCM models including two different weights in each event, letting the FCM define the actual BPA values. As previously explained, in the FCM, the human experts define the main events relevant for the considered system and assign the importance of each event in the system. Each event C carries a single weight A(t). We propose that each expert could define the main events providing two weight values instead of one. An approach similar to the one we propose has been previously described in the literature. For instance, the authors of [17] present a Bayesian belief network model in which the nodes carry two probability outcomes (i.e. true and false outcomes). Similarly, the authors of [18] describe a Fuzzy Grey Cognitive Map that defines concepts containing two weight levels.

When designing an FCM model, we can contribute to the design process by assigning an extra weight value to each event, similar to the Bayesian network shown in [17]. Each node C would carry two weights {$N_i(t)$, $M_i(t)$} in the fuzzy interval [0, 1]. We can extend this approach by providing two weight values (e.g. *Normal* and *Malicious*), and compute the belief in *Uncertainty* as we currently do in [4]. Once these three values are computed, they could be used as an extra metric in the current IDS. This extra metric, extracted from the contextual information, would be fused using the D-S theory along with the measurable metrics from the network.

D. *BPA Adjustment Using FCM*

The third approach that we propose is based on the adjustment of the BPA values prior to the data fusion process, by using the outcome of the FCM.

Our IDS uses three independent statistical approaches to generate the BPA values based on the current characteristics of the network, and the measurable parameters of the network traffic [4]. In order to incorporate contextual information from the network users, the outcome of the FCM, i.e. the weight A(t) value that each event C carries, or the weight relationship, $w_{ij}(t)$, between events can be used prior to the data fusion process, to adjust the assigned BPA values. These weight values may be used to increase or decrease the BPA values for one particular D-S hypothesis or for all the D-S hypotheses. This approach has been empirically evaluated in Section V.

V. TESTBED AND EXPERIMENTS

A. *Testbed and Measurement Description*

We have conducted a series of experiments as a first step to confirm the efficiency of the proposed approach. These experiments use netflow measurements gathered during 168 hours (7 days) from a virtual network testbed. A detailed description of this network testbed can be found in [19].

The network traffic information is gathered from the switch of the network (Cisco Nexus) acting as a netflow exporter. For setting up the virtual network, ESXi from VMware [20] was used as a hypervisor. ESXi was installed on two distinct physical hosts composing our physical infrastructure that will host virtual machines. In total, three virtual machines were set up in the testbed to generate traffic, two clients and one server. One virtual machine, acting as a client, was installed on one physical host, and two other virtual machines, acting as secondary client and server, were installed on the second physical host. That ensures that there was traffic both internally in the physical host and between two distinct physical hosts. Fig. 2 shows the logical topology of the virtual network.

For managing the virtual network, vCenter Server was installed as a virtual machine. The administrator can access vCenter Server through a web browser and manage most properties of the virtual network. A distributed virtual switch was also deployed, namely Nexus 1000v [21]. Distributed virtual switches are particularly useful for interconnecting virtual machines hosted on distributed physical hosts. The virtual switch was set up to send netflow data regarding the network to a netflow collector.

B. *Network Traffic Measurements*

In order to create some time-based patterns on the generated network traffic data and therefore assign *time* as context to the data, three cron (built-in Linux tool) schedules were scripted: 1) connecting to a web server (port 80) and downloading a webpage (with wget); 2) securely copying a file over the network from a server over the virtual network with scp (port 22); and finally 3) streaming video from a server with VLC (port 8080) [22]. Table I presents the description of the time in which each of the services is active.

The network traffic throughput was collected from the netflow measurements to identify possible anomalies in the traffic. Each service would generate a distinctive throughput value, distinct from the other throughput values generated by the other two services. These measurements for these 168 hours have been plotted in Fig. 3. The services create periodical step-like changes in the throughput values. These changes may cause our IDS to generate False Positive (FP) alarms when analysing this data. In addition, some unexpected increases in the throughput value were seen, particularly evident between days 3 and 4. We consider these values as anomalies in the throughput measurements. A close-up of the step-like changes and the anomalies in the measurements is presented in Fig. 4. Therefore, the main purpose of the presented experiments is twofold. First, identify the anomalies in the throughput values. Second, reduce or eliminate any FP alarm that may be generated by our IDS.

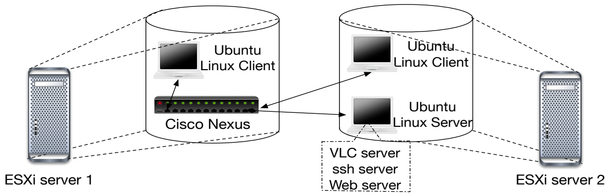


Figure 2. Logical topology of the virtual network: Three virtual machines (two clients and one server) installed on two physical hosts.

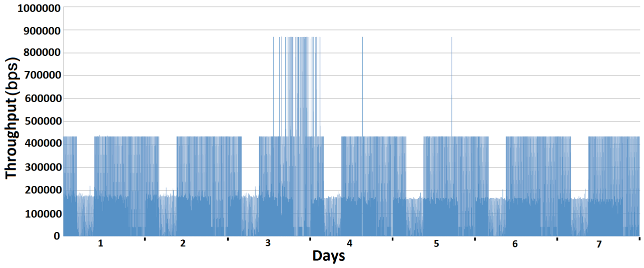


Figure 3. Throughput Measurements Over 7 Days - 168 Hours

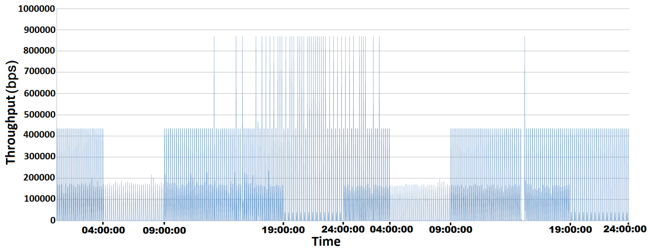


Figure 4. Throughput Measurements Over 2 Days - 48 Hours

TABLE I. TRAFFIC PATTERN DESCRIPTION

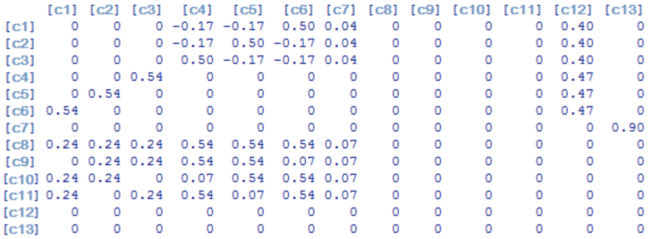
	Scheduled			
Time	00:00:00 - 03:59:00	04:00:00 - 08:59:00	09:00:00 - 18:59:00	19:00:00 - 23:59:00
Services	vlc/ssh/wget	vlc/wget	ssh/wget	vlc/ssh

C. FCM Adjustment Process Description

Three users with wide knowledge about the virtual network described in the previous section contributed to the FCM model. Each of them provided their opinion on whether the throughput measurements were normal or abnormal, based on the time and the different scheduled services. Initially, the users defined the main events relevant for the virtual network, the relationships among these events and the level of influence. This must be done without seeing the actual throughput measurements. In total, 13 events were defined, which have been sorted in Table II. Both C_{12} = *Normal* and C_{13} = *Abnormal* are the two events that would finally adjust the BPA values. Using the input from each user, three adjacency matrices were generated. The adjacency matrices were merged using (3), and the combined knowledge of the expert has been tabulated in the *13x13* adjacency matrix presented in Fig. 5.

TABLE II. LIST OF FCM EVENTS

FCM Events	Events Definition
C1	SSH
C2	WGET
C3	VLC
C4	Throughput < 50000
C5	50000 < Throughput < 200000
C6	200000 < Throughput < 450000
C7	Throughput > 450000
C8	00:00:00 - 03:59:00
C9	04:00:00 - 08:59:00
C10	09:00:00 - 18:59:00
C11	19:00:00 - 23:59:00
C12	Normal
C13	Abnormal



```
      [c1]  [c2]  [c3]   [c4]   [c5]   [c6]  [c7]  [c8]  [c9]  [c10]  [c11]  [c12]  [c13]
[c1]     0     0     0  -0.17  -0.17   0.50  0.04     0     0      0      0   0.40      0
[c2]     0     0     0  -0.17   0.50  -0.17  0.04     0     0      0      0   0.40      0
[c3]     0     0     0   0.50  -0.17  -0.17  0.04     0     0      0      0   0.40      0
[c4]     0     0  0.54      0      0      0     0     0     0      0      0   0.47      0
[c5]     0  0.54     0      0      0      0     0     0     0      0      0   0.47      0
[c6]  0.54     0     0      0      0      0     0     0     0      0      0   0.47      0
[c7]     0     0     0      0      0      0     0     0     0      0      0      0   0.90
[c8]  0.24  0.24  0.24   0.54   0.54   0.54  0.07     0     0      0      0      0      0
[c9]     0  0.24  0.24   0.54   0.54   0.07  0.07     0     0      0      0      0      0
[c10] 0.24  0.24     0   0.07   0.54   0.54  0.07     0     0      0      0      0      0
[c11] 0.24     0  0.24   0.54   0.07   0.54  0.07     0     0      0      0      0      0
[c12]    0     0     0      0      0      0     0     0     0      0      0      0      0
[c13]    0     0     0      0      0      0     0     0     0      0      0      0      0
```

Figure 5. Combined Adjacency Matrix

We can model how changes in the throughput, time and date affect the two events C_{12} and C_{13} using the initial vector state. Because the system can only measure the network traffic throughput and time, we have focused on the effect that changes in these two measurements have on the other events. For this process we have used the threshold activation function.

As an example, suppose that the time is 07:00:00 and there is a transition in the measured throughput from 29890 bps to 98500 bps, which is a normal situation. If we want to compute the weight with which the FCM would modify the BPA in this case, we would use the vector state A(0) = [0, 0, 0, -1, 0, 1, 0, 0, 1, 0, 0, 0, 0]. After a number of iterations using (1), the resulting weights would be A(6) = [0.27, 0.32, 0, -0.5, 0.38, 0.5, 0.05, 0, 0.5, 0, 0, 0.41, 0.05]. Therefore, the BPA *Normal* is adjusted by 0.41, whilst the BPA *Abnormal* is adjusted by 0.05. Similarly, suppose that the time is 16:30:00 and there is a transition in the measured throughput from 17000 bps to 450300 bps, which is an abnormal situation. The FCM process using the vector state A(0) = [0, 0, 0, 0, -1, 0, 1, 0, 0, 1, 0, 0, 0] would result in A(5) = [0.35, 0, 0, 0, -0.5, 0.44, 0.5, 0, 0, 0.5, 0, 0.12, 0.45]. Therefore, the BPA *Normal* is adjusted by 0.12 and the BPA *Abnormal* is adjusted by 0.45.

VI. Results

This section describes the detection results generated by our anomaly-based IDS. We have compared the efficiency of the detection system when contextual information is not used against detection results in which FCM is considered. The efficiency of the IDS has been evaluated using five well-known parameters, True Positive (TP), which represents anomalies correctly classified as anomalous; True Negative (TN), which represents normal instances correctly classified as non-anomalous; False Positive (FP), which represents non-anomalous instances misclassified as anomalous; and False Negative (FN), which represents anomalies misclassified as normal. These five parameters are essential to calculate the following performance metrics:

- Detection Rate (DR): Proportion of anomalies correctly classified as anomalous among all the measurements. DR (%) = TP/(FN+TP)•100
- False Positive Rate (FPr): Proportion of normal instances misclassified as anomalous among all the measurements. FPr (%) = FP/(TP+FP+TN+FN)•100
- False Negative Rate (FNr): Proportion of anomalies misclassified as normal among all the measurements. FNr (%) = FN/(FN+TP)•100
- Overall Success Rate (OSR): Proportion of all the measurements correctly classified.
  OSR (%) = (TN+TP)/(TP+FP+TN+FN)•100
- Precision: Proportion of anomalies correctly classified as anomalous among all the alarms generated.
  Precision (%) = TP/(TP+FP)•100

These parameters provide quantifiable evidence of how effective the IDSs are at making correct detections.

During the experiments we have experimentally evaluated the performance of the detection system using only measurable information from the network, as well as the performance of the system including contextual information through the use of the FCM. The length of the sliding window approach used in the D-S process to generate the different belief values has been sequentially increased from 1 to 400 slots to assess the effect that these changes may have in the final performance results.

The results without contextual information indicate that all the anomalies in the throughput measurements have been successfully identified (i.e. DR = 100% and FNr = 0%) but at the cost of a high number of false alarms. The FPr reaches up to 49% in the best-case scenario. The FPr results are presented in Fig. 6. This excessive high number of false alarms directly affects the rest of the performance measures. For instance, the Precision does not exceed 1% for any sliding window length, while the OSR result is considerably low, as can seen in Fig. 7.

As previously explained in Section V, all these false alarms generated by the IDS are caused by the periodical step-like changes in the network throughput values. These results are a clear indicator that by utilising only measureable information from the network without considering the context, IDSs may reach a wrong conclusion, leading to low accuracy.

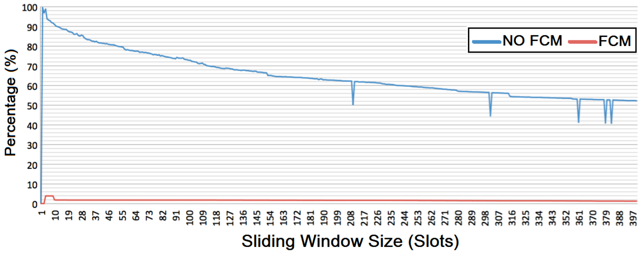


Figure 6. FPr Anomaly Detection Results

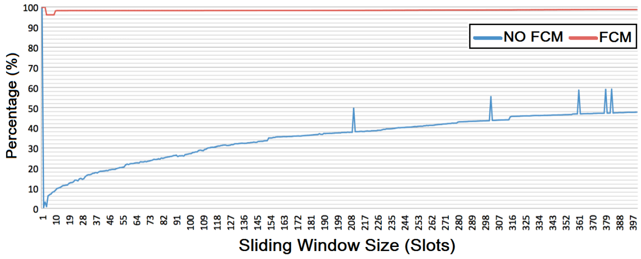


Figure 7. OSR Anomaly Detection Results

On the other hand, when contextual information is considered, the performance of the detection system can be greatly improved. The detection results including contextual information indicate that all the anomalies in the throughput measurements have been successfully identified (i.e. DR = 100% and FNr = 0%). The use of the FCM restricts the FPr to just 1.82%, in the worst-case scenario, for any evaluated sliding window length. This reduction in the number of FP alarms is a direct consequence of the adjustment induced by FCM. The FPr results when the FCM is used are also presented in Fig. 6. The precision values are also substantially higher than the results generated when the FCM is not employed; reaching up to 25% when a sliding window length 400 is used. In Fig. 7 we can also see that the OSR result is considerably higher than the results generated when the FCM is not used.

TABLE III. Traffic Pattern Description

Metrics	Result Values	
	Without FCM	*Including FCM*
Sliding Window Size	50 slots	50 slots
Detection Rate	100 %	100 %
False Positive Rate	80.61 %	1.82 %
False Negative Rate	0 %	0 %
OSR	19.39 %	98.18 %
Precision	0.54 %	19.38 %
F-Score	0.01	0.39

We have selected the sliding window length 50 as an example to compare the different performance metrics. In this particular case, both approaches generate perfect DR and FNr. However, the FPr without FCM reaches up to 80.61% whilst only up to 1.82% when FCM is used. The difference between both approaches is also evident in the OSR, F-Score and Precision results presented in Table III. These experiment results empirically confirm that the FCM can improve the efficiency of our IDS reducing the number of false alarms generated by the IDS, while not altering the number of correct detections. Therefore, we have verified improvement in the performance from the proposed approach.

VII. Conclusions And Future Work

In this paper we advocated incorporating contextual information regarding the monitored network when taking decisions on whether an anomaly in the monitored network takes place. We discussed the drawbacks of most current IDSs that only use quantifiable measurements from the network.

Our proposed contextual information methodology, presented in this paper, builds upon, expands and greatly improves the performance of our prior work on a multi-layer data fusion anomaly-based IDS, that has been briefly described in Section IV. The data fusion mechanism, based on D-S theory, combines belief values assigned to different hypotheses pertaining to whether an anomaly takes place or not.

We have proposed three possible approaches with the aim of generating or influencing these beliefs based on contextual information as assigned by multiple human experts. The incorporation of the contextual information processing method is achieved through the implementation and usage of the FCM in conjunction with our anomaly-based IDS. The FCM provides the capability of integrating contextual information from the user to the detection process in addition to modelling influences between events.

We have evaluated the use of FCMs in a virtual network topology composed of two physical hosts. The services running in the network are scripted based on the time of day, and therefore, the time becomes a contextual information input, along with the date and the measured network throughput.

We have compared the efficiency of the IDS system with and without the usage of contextual information. The experimental results strongly indicated the importance of contextual information when decisions about the behaviour of a network have to be taken. The practical results empirically confirm that by considering the contextual information, as described in the paper through the use of an FCM, the efficiency of our IDS can be improved. By using the FCM module, the number of false alarms generated by the IDS was greatly reduced, while the number of correct detections is not affected. For instance, without FCM, the FPr reaches 49% in the best-case scenario, while the use of the FCM restricts the FPr to just 1.82%, in the worst-case scenario. The results highlighted the great improvement in the performance that the incorporation of contextual information offers through the proposed approach.

As for future work, we wish to investigate the remaining two approaches that have been suggested and compare their performance. In addition, we will investigate the design of a richer scenario in terms of contextual information inputs.

References

[1] F.J. Aparicio-Navarro, K.G. Kyriakopoulos, and D.J. Parish, "Automatic dataset labelling and feature selection for intrusion detection systems," in *Proc. of the Military Communications*, 2014. pp. 46-51.

[2] A. Sadighian, S. T. Zargar, J. M. Fernandez, and A. Lemay, "Semantic-based context-aware alert fusion for distributed Intrusion Detection Systems," in *Proc. of the Int. Conf. on Risks and Security of Internet and Systems*, 2013. pp. 1-6.

[3] C. D. Stylios, and P. P. Groumpos, "Modeling complex systems using fuzzy cognitive maps," in *IEEE Transactions on Systems, Man and Cybernetics: Systems and Humans*. vol.34, no.1. 2004. pp. 155-162.

[4] 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," in *IET Information Security*, vol.8, no.1. 2014. pp. 42-50.

[5] L. Snidaro, J. García, and J. Llinas, "Context-based information fusion: A survey and discussion," in *Information Fusion*, 25. 2015. pp. 16-31.

[6] D. Gupta, P.S. Joshi, A.K. Bhattacharjee, and R.S. Mundada, "IDS alerts classification using knowledge-based evaluation," in *Proc. of the Fourth International Conference on Communication Systems and Networks*, 2012, pp. 1-8.

[7] S. Khan, H. Cheng, and R. Kumar, "A hierarchical behavior analysis approach for automated trainee performance evaluation in training ranges," in *Foundations of Augmented Cognition*, 2013, pp. 60-69.

[8] C. B. Anagnostopoulos, Y. Ntarladimas, and S. Hadjiefthymiades, "Situation awareness: Dealing with vague context," in *Proc. of the International Conference on Pervasive Services*, 2006, pp. 131-140.

[9] J. Sun, Y. Zhang, and K. He, "Providing context-awareness in the smart car environment," in *Proc. of the IEEE 10th International Conference on Computer and Information Technology*, 2010, pp. 13-19.

[10] T. D. Ndousse, and T. Okuda, "Computational intelligence for distributed fault management in networks using fuzzy cognitive maps," in *Proc. of the IEEE International Conference on Communications*, vol. 3, 1996, pp. 1558-1562.

[11] E. I. Papageorgiou, P. P. Spyridonos, D. T. Glotsos, C. D. Stylios, P. Ravazoula, G. N. Nikiforidis, and P. P. Groumpos, "Brain tumor characterization using the soft computing technique of fuzzy cognitive maps." in *Applied Soft Computing*. vol.8, no.1. 2008. pp. 820-828.

[12] 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. of the Behavior Representation in Modeling and Simulation Conference*, 2010. pp. 216-223.

[13] M. M. Kokar, and M. R. Endsley, "Situation awareness and cognitive modeling," in *IEEE Intelligent Systems*, vol.3. 2012. pp. 91-96.

[14] M. Jazzar, and A. Jantan, "Towards real-time intrusion detection using fuzzy cognitive maps modeling and simulation," in *Proc. of the International Symposium on Information Technology*, vol.2, 2008, pp. 1-6.

[15] S. Bueno, and J. L. Salmeron, "Benchmarking main activation functions in fuzzy cognitive maps," in *Expert Systems with Applications*. vol.36, no.3. 2009. pp. 5221-5229.

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

[17] C. G. Looney, and L. R. Liang, "Cognitive situation and threat assessments of ground battlespaces," in *Information Fusion*, vol.4, no.4. 2003. pp. 297-308.

[18] J.L.Salmeron, "Fuzzy cognitive maps for artificial emotions forecasting" in *Applied Soft Computing*, vol.12, no.12. 2012. pp. 3704-3710.

[19] K. G. Kyriakopoulos, D. J. Parish, and J. N. Whitley, "FlowStats: An ontology based network management tool," in *the 2nd IEEE Int. Conf. on Computing Technology and Information Management*, 2015. pp. 1-6.

[20] VMware, "vSphere Hypervisor," Available: http://www.vmware.com/products/vsphere-hypervisor (Access Date: 30 Sep, 2015).

[21] Cisco Systems, "Cisco Nexus 1000V Switch for VMware vSphere," Available: http://www.cisco.com/c/en/us/products/switches/nexus-1000v-switch-vmware-vsphere/index.htm (Access Date: 20 Sep, 2015).

[22] VideoLAN, "Vlc media player," Available: http://www.videolan.org/vlc/index.html (Access Date: 30 Sep, 2015).