This paper is published in IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/9500354 - P. Radoglou-Grammatikis, P. Sarigiannidis, G. Efstathopoulos, T. Lagkas, G. Fragulis and A. Sarigiannidis, "A Self-Learning Approach for Detecting Intrusions in Healthcare Systems," ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500354.

A Self-Learning Approach for Detecting Intrusions in Healthcare Systems

Panagiotis Radoglou-Grammatikis[†], Panagiotis Sarigiannidis[†], Georgios Efstathopoulos[‡], Thomas Lagkas[§], George Fragulis[†] and Antonios Sarigiannidis[¶]

***Abstract*—The rapid evolution of the Internet of Medical Things (IoMT) introduces the healthcare ecosystem into a new reality consisting of smart medical devices and applications that provide multiple benefits, such as remote medical assistance, timely administration of medication, real-time monitoring, preventive care and health education. However, despite the valuable advantages, this new reality increases the cybersecurity and privacy concerns since vulnerable IoMT devices can access and handle autonomously patients' data. Furthermore, the continuous evolution of cyberattacks, malware and zero-day vulnerabilities require the development of the appropriate countermeasures. In the light of the aforementioned remarks, in this paper, we present an Intrusion Detection and Prevention System (IDPS), which can protect the healthcare communications that rely on the Hypertext Transfer Protocol (HTTP) and the Modbus/Transmission Control Protocol (TCP). HTTP is commonly adopted by conventional ICT healthcare-related services, such as web-based Electronic Health Record (EHR) applications, while Modbus/TCP is an industrial protocol adopted by IoMT. Although the Machine Learning (ML) and Deep Learning (DL) methods have already demonstrated their efficacy in detecting intrusions, the rarely available intrusion detection datasets (especially in the healthcare sector) complicate their global application. The main contribution of this work lies in the fact that an active learning approach is modelled and adopted in order to re-train dynamically the supervised classifiers behind the proposed IDPS. The evaluation analysis demonstrates the efficiency of this work against HTTP and Modbus/TCP cyberattacks, showing also how the entire accuracy is increased in the various re-training phases.**

***Index Terms*—Active Learning, Cybersecurity, Intrusion Detection, Healthcare**

I. Introduction

The rapid evolution of the Information and Communication Technology (ICT) and especially of the Internet Things (IoT) [1], has led the healthcare organisations to digitise their care services, by adopting, medical telemetry and interconnected medical devices, such as wearables and medical implantables that handle and store patient data autonomously in Electronic Health Records (EHRs). Although this new reality offers multiple benefits such as remote medical assistance, timely administration of medication, real-time monitoring, preventive care, and health education, it also increases the existing security and privacy concerns, due to the heterogeneous co-existing smart and legacy nature of these entities (both hardware and software) as well as their insecure design. Moreover, among the other Critical Infrastructures (CIs), the healthcare domain is considered as the most vulnerable due to the vast amount of sensitive personal and administrative data stored and managed by the smart medical devices and EHRs software packages [2]. Based on the European Union Agency for Network and Information Security (ENISA), the healthcare sector continues to lead in the number of cybersecurity incidents (27%). The largest of such incident, 211 Los Angeles County, exposed 3,5 million records through accidental loss [3]. In particular, compared to other sectors such as government and finance, the healthcare domain lags largely behind in the cybersecurity preparedness. A recent characteristic cybersecurity incident related to the health sector was the WannaCry ransomware in May 2017, which paralysed the United Kingdom's National Health Service by encrypting multiple sensitive health data, thus locking out the legitimate users, until a specific amount in Bitcoin was paid. Furthermore, other characteristic examples in 2016 and 2017 were the cyberattacks against Princeton Community Hospital and MedStar Health Inc., a non-profit healthcare company [4]. Only during 2016 and 2017, 49 critical cybersecurity incidents were performed against healthcare organisations in the US [5]. Furthermore, in the light of many reports, such as that of Online Trust Alliance's, 2017 was the "worst year ever" for cybersecurity incidents, while healthcare seems to be one of the most targeted industries by cyberattackers. Therefore, the challenge of ensuring smart, safe, sustainable and efficient healthcare systems becomes major as the road ahead for healthcare systems is a difficult one, fact that is validated by the the European Commission (EC) decision in introduced the European Union (EU) Directive NIS 2016 enforcing, thus, all CIs as those of the healthcare domain to report any critical security incident to the Computer Security Incident Response Team (CSIRT).

Along with the various cybersecurity issues in the healthcare sector, the ICT technological achievements introduce in parallel significant privacy issues concerning the patients' data. In particular, the technological developments in the last few decades have enabled, among others the remote and collaborative healthcare services where patients' EHRs are collected in a central storage, thus giving access to many healthcare providers. Although this capability enhances clinical and other health-related decision-making processes by availing the necessary health information whenever required, at the same time, it increases the already growing privacy concerns, since multiple collaborating parties coming from a

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 787011 (SPEAR).

[†]P. Radoglou-Grammatikis, P. Sarigiannidis and G. Fragulis are with the Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani 50100, Greece - E-Mail: {pradoglou,psarigiannidis,gfragulis}@uowm.gr

[‡] G. Efstathopoulos is with the 0INF, Imperial Offices, London, UK, E6 2JG - E-Mail: george@0infinity.net

[‡] T. Lagkas is with the Department of Computer Science, International Hellenic University, Kavala Campus, 65404, Greece - E-Mail: tlagkas@cs.ihu.gr

[¶] A. Sarigiannidis is with Sidroco Holdings Ltd, 3113, Limassol, Cyprus - E-Mail: asarigia@sidroco.com

diverse range of authorities under different managements may access this data. In this kind of setup, usually, the adversaries try to access, move, or even lock EHRs. Furthermore, on the other hand, the authorised parties in a collaborative healthcare service need to trust the integrity of information displayed on their monitors. Therefore, it is clear that such data sharing and handling services should satisfy the privacy and data integrity principles. For this purpose, several existing mechanisms have been proposed that usually adopt public-key cryptosystems or pseudo anonymisation techniques [6]. Undoubtedly, these solutions provide valuable protection, especially during the data transmission processes or the storage processes in an external cloud service. However, they may not prevent sophisticated attacks such as malicious insiders or data destruction.

It is estimated that the investments of the healthcare ecosystem by the appropriate methods, tools and practices will exceed €65B over the next five years. However, the transformation of the conventional healthcare systems and services to a secure smart ecosystem utilising contemporary digital technologies, such as big data, IoT and Artificial Intelligence (AI) is not a straightforward way. Based on the aforementioned remarks, it is evident that timely and reliable intrusion detection and prevention in healthcare systems is an essential need. Although Machine Learning (ML) and Deep Learning (DL) solutions have already proved their capacity in detecting cyberthreats, the peculiarities of the healthcare sector render their adoption a challenging issue. In particular, the healthcare sector constitutes a sensitive CI, where the necessary datasets for the ML and DL solutions cannot be provided publicly. This fact complicates the cybersecurity analysts to construct appropriate intrusion detection datasets and train their models. Moreover, the heterogeneous nature of the healthcare ecosystem makes the adoption of such models more difficult since each healthcare environment is characterised by different attributes, such as medical devices and communication protocols.

In this paper, we provide an Intrusion Detection and Prevention (IDPS) system for the healthcare environments that utilise the Hypertext Transfer Protocol (HTTP) and the Modbus protocols. On the one hand, HTTP is a common ICT protocol, which is used by several computing systems, including multiple e-healthcare applications, such as EHR. On the other hand, Modbus/Transmission Control Protocol (TCP) [7] is an industrial protocol, which is widely adopted by both legacy and smart medical devices. The proposed IDPS applies an active learning approach, where first, the IDPS is trained with an initial dataset and then is re-trained continuously by its detection results in order to optimise its detection performance. The main contributions of this work are summarised in the following key-points.

- **Providing and IDPS for the healthcare ecosystem**. The proposed IDPS can recognise a plethora of HTTP and Modbus/TCP cyberattacks. In particular, regarding the HTTP protocol, four cyberattacks are detected and mitigated by the proposed IDPS, including (a) DoS, (b) Structured Query Language (SQL) injection, (c) Brute-force and (d) Cross-Site Scripting (XSS). On the other side, regarding Modbus/TCP, the proposed IDPS recognises 14 Modbus/TCP-related cyberattacks, thus solving a challenging ML classification problem.
- **Introducing an active learning approach for recognising intrusions in a healthcare ecosystem**. The proposed IDPS is re-trained continuously, thus optimising its detection efficacy by itself. Thus, the applicability of the proposed IDPS is demonstrated since it can be adapted in any healthcare ecosystem, which uses the Modbus/TCP and HTTP protocols.

The rest of this paper is organised as follows. Section II presents relevant works. Section III provides the architecture of the proposed IDPS. Section IV is focused on the active learning approach. Section V is devoted to the evaluation analysis. Finally, Section VI concludes this paper.

II. Related Work

Several papers have already studied the cybersecurity and privacy issues of the healthcare ecosystem. Some of them are listed in [8]–[11]. In particular, T. Yaqoob et al. in [8] provide a comprehensive study about the vulnerabilities of the smart medical devices and discuss relevant countermeasures. In [9]. M. Hassan et al. present a detailed analysis of the differential privacy techniques for Cyber-Physical Systems (CPS). U. Sun et al. in [10] introduce a survey regarding the cybersecurity challenges, requirements and threats related to the Internet of Medical Things (IoMT), thus identifying directions for future research works. Finally, in [11], A. Hady et al. present a thorough review about the Intrusion Detection Systems (IDS) in the healthcare area. Below, we analyse further some notable cases. Each case is analysed in a dedicated paragraph.

In [12], R. Mitchel and I. Chen provide a behaviour rule specification-based IDS for Medical CPS (MCPS). The proposed IDS is focused on operational data related to the core functionality of MCPS. In particular, they examine three cases: (a) vital sign monitor, (b) cardiac device (CD) and (c) patient-controlled analgesia. Based on the core functionality of these actuators, the authors construct behaviour-based specification rules that define the normal status and operation. Next, these rules are transformed into state machines in order to facilitate the comparison between benign and malicious states. Finally, based on an extensive threat modelling for each case, the appropriate thresholds are identified. The simulation results verify the detection performance of the proposed IDS, exceeding two similar approaches.

In [13], G. Thamilarasu et al. introduce a mobile agent-based IDS for the IoMT. Their implementation is focused on Wireless Body Area Networks (WBANs) and is capable of recognising cyberattacks in a device or network level. After introducing the necessary background about (a) IoMT, (b) WBANs, (c) security attacks and solutions and (d) mobile agent-based IDS, the authors discuss the mobile agent-based IDS requirements as well as the main threats against WBANs. In particular, the authors discriminate three threats, namely (a) DoS, (b) data fabrication and falsification and (c) privacy

data breach. However, it is worth noting that the proposed IDS cannot distinguish the aforementioned threats, but rather it identifies three classes: (a) normal, (b) malicious and (c) suspicious. Next, the architectural schema is presented, which consists of three main agents (a) sensor agents, (b) cluster agents and (c) detective agents. The sensor agents operate in device level, while the cluster agents work in the network level. The detective agents are additional nodes that support the other agents when their detection outcome is not accurate. The agents adopt regression and typical classification ML techniques, such as Support Vector Machine (SVM), Naive Bayes, Random Forest, Decision Tree and K-Nearest Neighbour (KNN). The authors evaluate the proposed IDS in a simulation environment constructed by Omnet. The simulation results demonstrate the efficiency of the proposed implementation in terms of detection accuracy and resource overhead.

M. Mohamed et al. in [14] introduce a specification-based IDS for WBANs. In particular, the authors focus on (a) jamming, (b) sinkhole and (c) flooding cyberattacks against Electrocardiogram (ECG) and Electromyogram (EMG) sensors. These cyberattacks are emulated by introducing the appropriate noise to medical signals. The architecture of the proposed IDS relies on six steps, namely (a) Data Acquisition, (b) Filtering, (c) Intrusion Detection, (d) Cancellation, (e) Anomaly Detection and (f) Diagnostic. The intrusion and anomaly detection processes rely on particular specification thresholds defined for the aforementioned medical sensors. More specifically, first, the proposed IDS adopts filters with the aim to reduce medical-based interference. Next, the intrusion detection procedure takes place based on the signal frequency and amplitude. Then, the recognised intrusions are cancelled in order to follow the medical anomaly detection that will lead to the disease diagnosis. Based on the simulation results made in Matlab, the detection performance of the proposed implementation is validated.

In [15], A. Newaz et al. present $HEKA$. $HEKA$ is an IDS especially designed to protect Personal Medical Devices (PMDs). After providing an overview of the various vulnerabilities related to PMDs, the authors first demonstrate a plethora of cyberattacks against commercial PMDs, utilising existing attacking tools. In particular, they focus on five cyberattacks, namely (a) Eavesdropping, (b) DoS, (c) Man-In-The-Middle (MITM), (d) replay attacks and (e) False Data Injection (FDI). Then, they analyse $HEKA$, which is focused on four cyberattacks: (a) MiTM, (b) Replay, (c) FDI and (d) DoS. $HEKA$ consists of four architectural modules: (a) sniffer, (b) data preprocessing, (c) n-gram generator, (d) anomaly detector and (f) notification module. The anomaly detector module applies and evaluates four typical ML methods: (a) SVM, (b) Decision Tree, (c) Random Forest and (d) KNN. To evaluate $HEKA$, the authors constructed a testbed, which is composed of eight PMDs. Based on the evaluation analysis, the accuracy of $HEKA$ reaches 0.984.

In [16], J. Stokes et al. present $ALADIN$, which stands for "Active Learning of Anomalies to Detect INtrusions". According to the authors, $ALADIN$ composes an anomaly detection and classification system, which can be adopted in both Host-based IDS (HIDS) and Network-based IDS (NIDS). $ALADIN$ applies active learning to satisfy both the improvement of the detection rate and the identification of new cybersecurity incidents. In particular, through active learning, it isolates the interesting anomalous-related network traffic items, that are given next to cybersecurity analysts in order to categorise them into predefined or new classes. After the appropriate categorisation from the cybersecurity analysts, the classification ability of $ALADIN$ is re-trained in order to include the new instances either they belong to a predefined category or a new one. For the classification process, the linear regression method is applied, while for the anomaly detection, the Naive Bayes method is used. $ALADIN$ identifies those items that present high uncertainty. These items are analysed and labelled then by the cybersecurity analysts. The evaluation analysis confirms the effectiveness of $ALADIN$. To this end, the KDD-Cup 1999 dataset [17] is used.

Undoubtedly, the works mentioned previously give significant insights and methodologies. Some of them utilise specification-based techniques, while others adopt anomaly-based techniques, such as ML solutions. On the one side, the specification-based techniques are more accurate since they define the normal state of a system and recognise potential deviations. However, they cannot easily discriminate particular cyberattack types. Moreover, they are not scalable since each healthcare device is characterised by different specifications. Therefore, the security experts need to identify and form the necessary specification rules for each of them. Also, the configuration of these devices can be changed or re-programmed, thus making it necessary to adjust the corresponding rules. On the other side, ML and DL methods can distinguish particular cyberattacks, but they rely on intrusion detection datasets that rarely are available publicly, especially for CIs. For this reason, the researchers use existing intrusion detection datasets, such as AWID [18] and KDD-Cup 1999 [17]. However, such datasets do not reflect the unique peculiarities of a healthcare environment. Moreover, it is worth mentioning that none of the previous papers investigate intrusions against healthcare communication protocols, such as HTTP and Modbus/TCP. As mentioned, HTTP is widely adopted by many healthcare computing systems, such as EHR, while Modbus/TCP is an application-layer protocol, which is adopted in IoMT. Hence, in this paper, we introduce an IDPS, which recognises efficiently HTTP and Modbus cyberattacks and adopts active learning in order to re-train itself based on its detection outcome.

III. Proposed Architecture

Fig. 1 illustrates the architecture of the proposed IDPS, which consists of three main modules, namely (a) *Network Flow Monitoring and Extraction Module*, (b) *Intrusion Detection Engine* and (c) *Notification and Response Module*. The first module undertakes to capture the monitoring network traffic and extract the corresponding Transmission Control Protocol/Internet Protocol (TCP/IP) network flows. The sec-

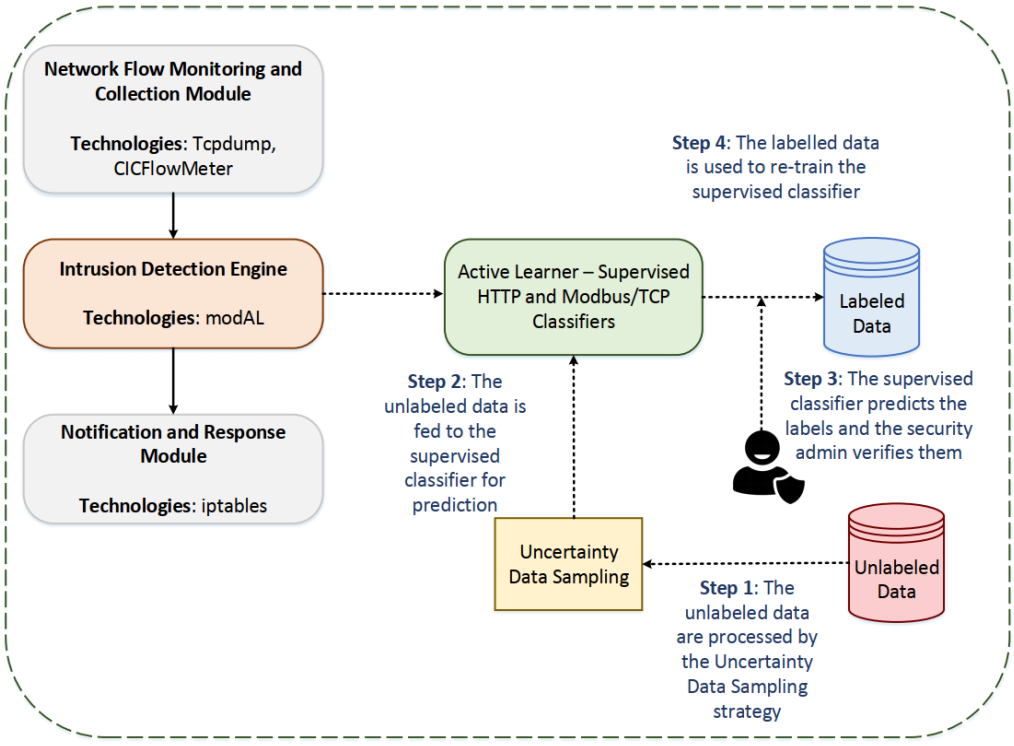

Fig. 1: Architecture of the proposed IDPS

ond module is responsible for detecting the corresponding attacks against the HTTP and Modbus/TCP protocols. Finally, the *Notification and Response Module* informs the security expert about potential intrusions and can apply some automate countermeasures that are detailed subsequently. In particular, the following subsections detail each module individually, including the respective technologies.

Moreover, Fig. 1 depicts the steps of the Active Learning methodology, which is composed of four main steps. In the first step, the unlabelled data is assessed by the query strategy, which is named Uncertainty Data Sampling. Then, the data approved by the Uncertainty Data Sampling is fed to the supervised classifiers depending on the corresponding network flows (i.e., HTTP network flows or Modbus/TCP network flows). Next, the supervised classifiers predict the labels that also are verified by a security expert. It is worth mentioning that the security expert has the ability to intervene and change the labels predicted by the supervised classifiers. Finally, the new labelled data is introduced to the new training dataset, which is used to update and re-train the supervised classifiers. The Active Learning methodology is detailed in section IV.

A. Network Flow Monitoring and Collection Module

The *Network Flow Monitoring and Collection Module* monitors the examined healthcare infrastructure through a Switch Port Analyzer (SPAN), thus receiving the overall network traffic generated by the connected healthcare devices. In particular, it applies Tcpdump [19] in order to capture the network traffic and then CICFlowMeter [20] to generate bidirectional network flow statistics. Two kinds of network flow statistics are generated related to (a) HTTP and (b) Modbus/TCP. The differentiation between these statistics is achieved through the source and destination TCP/IP ports. HTTP utilises the 80 TCP port or the 443 TCP port whether the Secure Sockets Layer/Transport Layer Security (SSL/TLS) protocol is applied. On the other side, Modbus/TCP listen to the 502 TCP port.

B. Intrusion Detection Engine

The *Intrusion Detection Engine* is the core module of the proposed IDPS. It consists of two supervised classifiers for the HTTP and Modbus/TCP, respectively. The HTPP classifier can recognise four relevant cyberattacks, namely (a) DoS, (b) SQL injection, (c) bruteforce and (d) XSS. The first HTTP-related cyberattack floods the target healthcare system with HTTP packets. The SQL injection intends to accomplish unauthorised access attacks. The bruteforce attack aims to discover the passwords of web applications by using all possible choices. Finally, XSS injects malicious scripts into the web applications. On the other side, our previous work in [7] describes the possible Modbus/TCP cyberattacks. In particular, the Modbus classifier is capable of discriminating the following cyberattacks: (a) modbus/function/readHoldingRegister, (b) modbus/scanner/uid, (c) modbus/function/readDiscreteInput, (d) modbus/dos/writeSingleCoils, (e) modbus/function/writeSingleRegister, (f) modbus/function/readInputRegister, (g) modbus/function/readCoils (DoS), (h) modbus/function/readHoldingRegister (DoS),

(i) (modbus/function/readDiscreteInputs (DoS)), (j) modbus/dos/writeSingleRegister, (k) modbus/scanner/getfunc, (l) modbus/function/writeSingleCoils and (m) modbus/function/readInputRegister (DoS). Regarding the ML and DL techniques, for the HTTP protocol, a Decision Tree classifier is utilised, while the Random Forest classifier is used for Modbus/TCP. The evaluation of these classifiers is analysed in V.

C. Notification and Response Module

The *Notification and Response Module* notifies the security expert about the possible security events via a web-based interface. The security events follow the format of the AlienVault OSSIM security events [21]. Moreover, through the aforementioned web interface, the operator has the ability to check and change the labels of the potential security events. Furthermore, the *Notification and Response Module* generates and applies some automate firewall rules that can mitigate or even prevent the various cyberattacks. For this purpose, the Linux firewall, namely iptables is adopted, utilising data from the TCP/IP network flows extracted by CICFlowMeter in the *Network Flow Monitoring and Collection Module*.

IV. Active Learning: Problem Formulation and Methodology

Active Learning is commonly adopted when there are no available labelled training datasets as in our case (i.e., intrusion detection in a healthcare ecosystem) since CIs cannot label and disclose their sensitive data. It provides an operational framework, which selects the most useful and informative data samples from a set of unlabelled data in order to optimize and construct a training dataset, which in turn will lead to producing more accurate supervised ML and DL classifiers (hypothesis). Unlike Passive Learning, which collects and feeds data samples randomly, Active Learning assesses the data samples based on particular criteria, thereby providing a training dataset with fewer data samples that include the most informative observations. These samples should be characterized by three main properties: (a) they should be represented, (b) they should be representative and (c) they should output accurate detection results. Usually, there is an external factor that annotates the samples investigated, such as a human annotator. Three main methods are utilized by an Active Learner in order to query for data samples: (a) query synthesis, (b) stream-based selective sampling and (c) pool-based sampling. The first case synthesizes the data samples *de novo*, thus producing never observed data samples. However, it does not consider the data distribution, which can be informative by the hypothesis. The other methods solve this problem. The stream-based selective sampling method receives data samples as streams continuously and decides based on a query strategy which data samples should be labelled or not. After the labelling process by the external factor (e.g., a human operator), they are moved into the training set. On the other side, the pooling method creates first a pool with unlabelled data samples and sequentially decides based on a query strategy which of them will be labelled. After the labelling process by the external factor, they are moved into the training set.

Supposing that the TCP/IP network flows from the healthcare environment flow continually and utilizing the pooling-based sampling method, let x be an unlabelled network flow from the input space X and y the respective label defined by the HTTP and Modbus/TCP threats discussed in subsection III-B, including also the normal state. Moreover, let U be a set of unlabelled TCP/IP network flows within a pool. The later is collected by the *Network Flow Monitoring and Collection Module*. Moreover, Let L be the training dataset consisting of the labelled TCP/IP network flows. We define the function $f(x) = y$ as the target function, which absolutely classifies the unlabelled TCP/IP network flows in the correct classes. On the other side, we define $h(x) = y'$ as the respective, supervised classifier, which predicts the label of an unlabelled TCP/IP network flow after the training process. Thus, the generalization error E can be expressed by equation 1.

$$E[l(h)] \int_{-\infty}^{\infty} l(h(x), f(x))\, dx \tag{1}$$

where l is the squared error function defined by equation 2.

$$l(h(x), f(x)) = (h(x) - f(x))^2 \tag{2}$$

where l is the squared error function defined by equation 2. Therefore, the Active Learning problem lies in the fact that the generalisation error should be minimised based on the new optimum training dataset L. In other words, we need to identify and label those unlabelled TCP/IP network flows in the pool that next will be used in order to re-train the supervised classifiers (hypothesis) for the HTTP and Modbus/TCP protocols with the most efficient accuracy. To this end, there are various queries strategies, namely (a) Uncertainty Sampling, (b) Query-by-Committee, (c) Expected Model Change, (d) Expected Error Reduction, (e) Variance Reduction and (f) Information Density. In this paper, we adopt the Uncertainty Sampling strategy, which takes advantage of the classifier's (hypothesis) detection uncertainty. In particular, the rationale behind the Uncertainty Sampling in the proposed IDPS is to ask the external factor about those unlabelled TCP/IP network flows for which the hypothesis is less confident. In our case, the external factor is the same hypothesis since the IDPS should be re-trained by itself. Moreover, a security expert can verify or change the labels of the selected unlabelled TCP/IP network flows from the web-based interface of the *Notification and Response Module*. The key point of the Uncertainty Sampling lies in the criterion used for calculating the uncertainty. For this purpose, various measures have been specified in the literature, such as (a) entropy, (b) least margin and (c) the least confident of prediction. In this work, we adopt the entropy criterion, which is defined by equation 3.

$$H = -\sum_{i=1}^{m} p_\theta(y_i|x) \log_2(p_\theta(y_i|x)) \tag{3}$$

where p_θ denotes the probability of class i for the observation x, while θ denotes the parameters of the supervised classifier (hypothesis). Therefore, the entropy criterion selects those TCP/IP network flows x^* from the pool U that satisfy the equation 4. In this paper, δ is defined experimentally.

$$x^* = argmax(x) + H > \delta \tag{4}$$

Based on the aforementioned remarks, Algorithm 1 defines the active learning procedure of the proposed IDPS. First, L is an initial training dataset with a few data samples that are used for training $h(x)$ for the HTTP protocol and the Modbus/TCP protocol, respectively. In particular, for the HTTP protocol, L was formed, utilising the CIC-IDS2017 dataset, while regarding the Modbus/TCP protocol, L was constructed, by emulating the cyberattacks analysed in our previous work [7]. On the other side, the *Network Flow Monitoring and Collection Module* fills U. While the size of U is greater than zero and if the entropy criterion is satisfied for each record in U, $h(x)$ predicts the label of the corresponding record and the security expert verifies or changes the outcome of this prediction via the web-based interface of the *Notification and Response Module*. Next, the specific record of U is added in L, which then is used in order to re-train $h(x)$.

Algorithm 1: Active Learner: Pooling-based Sampling and Uncertainty Sampling Strategy

```
Data: U, L, h
Result: Re-train h
initialization;
while size(U) > 0 do
    if classifier_uncertainty(U(i)) > δ then
        Predict y(i) using h;
        Verify or change the prediction of h through
          the security expert ;
        Add U(i) and y(i) in L;
        Re-train h
    end
    Remove U(i) from U;
end
```

V. Evaluation Analysis

This section is devoted to the evaluation analysis of the proposed IDPS. Before analysing the evaluation results, we need to define first the necessary terms. First, True Positives (TP) denotes the number of the correct classifications that detect the cyberattacks as intrusions. True Negatives (TN) implies the number of the correct classifications that recognise the normal network packets as normal. On the other side, False Negatives (FN) denotes the number of incorrect classifications that detect the cyberattacks as normal. Finally, False Positives (FP) indicates the number of mistaken classifications where the normal behaviours are recognised as intrusions. Based on these terms, the following metrics are defined equations 5-8.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{8}$$

Moreover, a plethora of ML supervised classifiers were tested, including (a) Decision Tree, (b) Random Forest, (c) KNN, (d) SVM, (e) Naive Bayes, (f) Multi-Layer Perceptron (ML) as well as two DL supervised classifiers, namely Dense Deep Neural network (DNN) Relu and Dense DNN Tanh originating from our previous work in [22]. After the last re-training procedure implemented by the Active Learner, Table I shows the evaluation results for the cyberattacks against the HTTP protocol. The Decision Tree achieves the best performance, where $Accuracy = 0.9644$, $TPR = 0.9111$, $FPR = 0.0222$ and $F1 = 0.9111$. In a similar manner, Table II depicts the evaluation results related to the detection of the cyberattacks against the Modbus/TCP protocol. In this case, the best performance is carried out by Random Forest, where $Accuracy = 0.94454$, $TPR = 1$, $FPR = 0.10166$ and $F1 = 0.94250$. Moreover, Fig. 2 and Fig.3 show how the accuracy is increased in each case during the re-training phases.

TABLE I: Evaluation Results related to the cyberattacks against HTTP

ML Method	Accuracy	TPR	FPR	F1
Decision Tree Classifier	**0.9644**	**0.9111**	**0.0222**	**0.9111**
Naive Bayes	0.7288	0.72888	0.27111	0.66744
SVM	0.89075	0.89075	0.10924	0.88027
Random Forest	0.92296	0.80740	0.04814	0.80740
MLP	0.90478	0.61915	0.05440	0.61915
Dense DNN Relu	0.90908	0.63633	0.05195	0.63633
Dense DNN Tanh	0.94074	0.85185	0.03703	0.85185

TABLE II: Evaluation Results related to the cyberattacks against Modbus/TCP

ML Method	Accuracy	TPR	FPR	F1
Decision Tree Classifier	0.83333	0.94827	0.28160	0.85051
Naive Bayes	0.71982	0.54597	0.10632	0.66086
SVM	0.841	0.523	0.095	0.523
Random Forest	**0.94454**	**1**	**0.10166**	**0.94250**
MLP	0.81797	0.96663	0.21924	0.68018
Dense DNN Relu	0.88357	0.87158	0.11341	0.74989
Dense DNN Tanh	0.84078	0.96122	0.18952	0.70827

VI. Conclusions

The new reality in the healthcare ecosystem introduces significant cybersecurity issues that can lead to devastating consequences or even fatal accidents. In this paper, we presented an IDPS, which is capable of detecting and mitigating cyberattacks efficiently against the HTTP and Modbus/TCP

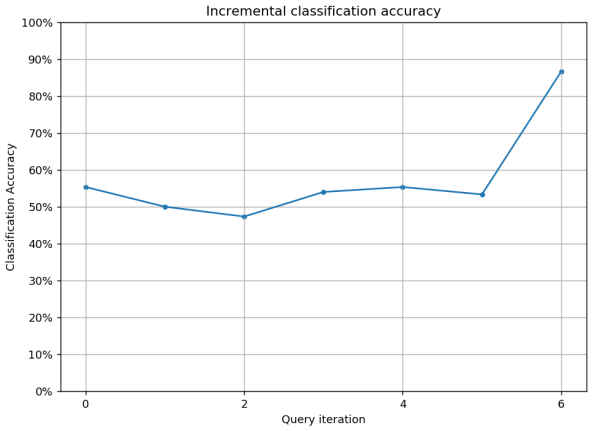

Fig. 2: Decision Tree - Accuracy Increment during the re-training phases

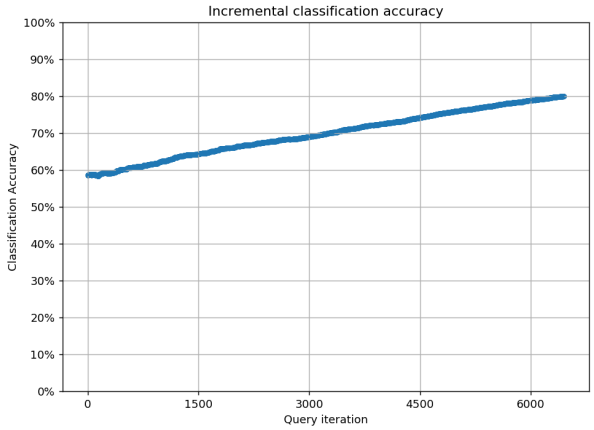

Fig. 3: Random Forest - Accuracy Increment during the re-training phases

protocols that are widely adopted in the e-healthcare services. On the one hand, HTTP is utilised by typical healthcare ICT services, such as EHR, while Modbus/TCP is used by IoMT. Given the rarely available intrusion detection datasets related to CIs and especially to the healthcare domain, the main novelty behind the proposed IDPS is its ability to re-train itself, utilising an Active Learning approach. The evaluation analysis demonstrates the efficiency of the proposed IDPS against HTTP and Modbus/TCP cyberattacks, showing additionally how the overall accuracy is increased during the re-training phases.

VII. Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 787011 (SPEAR).

References

[1] P. I. Radoglou-Grammatikis, P. G. Sarigiannidis, and I. D. Moscholios, "Securing the internet of things: Challenges, threats and solutions," *Internet of Things*, vol. 5, pp. 41–70, 2019.

[2] C. J. Dameff, J. A. Selzer, J. Fisher, J. P. Killeen, and J. L. Tully, "Clinical cybersecurity training through novel high-fidelity simulations," *The Journal of emergency medicine*, vol. 56, no. 2, pp. 233–238, 2019.

[3] A. Sfakianakis, C. Douligeris, L. Marinos, M. Lourenço, and O. Raghimi, "Enisa threat landscape report 2018: 15 top cyberthreats and trends," ENISA, Tech. Rep., 2019.

[4] X. Zhao, I. Miers, M. Green, and J. Mitrani-Reiser, "Modeling the cybersecurity of hospitals in natural and man-made hazards," *Sustainable and Resilient Infrastructure*, vol. 4, no. 1, pp. 36–49, 2019.

[5] L. E. Branch, W. S. Eller, T. K. Bias, M. A. McCawley, D. J. Myers, B. J. Gerber, and J. R. Bassler, "Trends in malware attacks against united states healthcare organizations," *Cyber Threats and Healthcare Organizations: A Public Health Preparedness Perspective*, vol. 1001, p. 29, 2018.

[6] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiquzzaman, "Privacyprotector: Privacy-protected patient data collection in iot-based healthcare systems," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 163–168, 2018.

[7] P. Radoglou-Grammatikis, I. Siniosoglou, T. Liatifis, A. Kourouniadis, K. Rompolos, and P. Sarigiannidis, "Implementation and detection of modbus cyberattacks," in *2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST)*. IEEE, 2020, pp. 1–4.

[8] T. Yaqoob, H. Abbas, and M. Atiquzzaman, "Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices—a review," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3723–3768, 2019.

[9] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for cyber physical systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 746–789, 2020.

[10] Y. Sun, F. P. . Lo, and B. Lo, "Security and privacy for the internet of medical things enabled healthcare systems: A survey," *IEEE Access*, vol. 7, pp. 183 339–183 355, 2019.

[11] A. A. Hady, A. Ghubaish, T. Salman, D. Unal, and R. Jain, "Intrusion detection system for healthcare systems using medical and network data: A comparison study," *IEEE Access*, vol. 8, pp. 106 576–106 584, 2020.

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

[13] G. Thamilarasu, A. Odesile, and A. Hoang, "An intrusion detection system for internet of medical things," *IEEE Access*, vol. 8, pp. 181 560–181 576, 2020.

[14] M. B. Mohamed, A. Meddeb-Makhlouf, and A. Fakhfakh, "Intrusion cancellation for anomaly detection in healthcare applications," in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, 2019, pp. 313–318.

[15] A. I. Newaz, A. K. Sikder, L. Babun, and A. S. Uluagac, "Heka: A novel intrusion detection system for attacks to personal medical devices," in *2020 IEEE Conference on Communications and Network Security (CNS)*, 2020, pp. 1–9.

[16] J. Stokes, J. Platt, J. Kravis, and M. Shilman, "Aladin: Active learning of anomalies to detect intrusions," Microsoft Reserach, Tech. Rep. MSR-TR-2008-24, March 2008. [Online]. Available: https://www.microsoft.com/en-us/research/publication/aladin-active-learning-of-anomalies-to-detect-intrusions/

[17] P. Aggarwal and S. K. Sharma, "Analysis of kdd dataset attributes-class wise for intrusion detection," *Procedia Computer Science*, vol. 57, pp. 842–851, 2015.

[18] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, "Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 184–208, 2016.

[19] P. Goyal and A. Goyal, "Comparative study of two most popular packet sniffing tools-tcpdump and wireshark," in *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2017, pp. 77–81.

[20] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in *ICISSP*, 2018, pp. 108–116.

[21] P. Radoglou-Grammatikis, P. Sarigiannidis, E. Iturbe, E. Rios, A. Sarigiannidis, O. Nikolis, D. Ioannidis, V. Machamint, M. Tzifas, A. Giannakoulias *et al.*, "Secure and private smart grid: The spear architecture," in *2020 6th IEEE Conference on Network Softwarization (NetSoft)*. IEEE, 2020, pp. 450–456.

[22] P. Radoglou Grammatikis, P. Sarigiannidis, G. Efstathopoulos, and E. Panaousis, "Aries: a novel multivariate intrusion detection system for smart grid," *Sensors*, vol. 20, no. 18, p. 5305, 2020.