Ad Hoc Networks 122 (2021) 102621

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

ECU-IoHT: A dataset for analyzing cyberattacks in Internet of Health Things

Mohiuddin Ahmed *, Surender Byreddy, Anush Nutakki, Leslie F. Sikos, Paul Haskell-Dowland

School of Science, Edith Cowan University, Perth, Australia

ARTICLE INFO

Keywords:
Cyberattacks
Healthcare
Testbed
Intrusion detection
Dataset

ABSTRACT

In recent times, cyberattacks on the Internet of Health Things (IoHT) have continuously been growing, and so it is important to develop robust countermeasures. However, there is a lack of publicly available datasets reflecting cyberattacks on IoHT, mainly due to privacy concerns. This paper showcases the development of a dataset, *ECU-IoHT*, which builds upon an IoHT environment having different attacks performed that exploit various vulnerabilities. This dataset was designed to help the healthcare security community in analyzing attack behavior and developing robust countermeasures. No other publicly available datasets have been identified for cybersecurity in this domain. Anomaly detection was performed using the most common algorithms, and showed that nearest neighbor-based algorithms can identify attacks better than clustering, statistical, and kernel-based anomaly detection algorithms.

1. Introduction

The healthcare sector is being transformed at a rapid pace due to the integration of Internet and smart devices. The Internet-enabled, interconnected objects of the IT infrastructures of today's healthcare are often referred to collectively as the *Internet of Health Things (IoHT)* [1,2]. While ongoing changes aim at advancing medical services, cybercriminals are very interested to target this sector in the hope of financial gain through ransom or the sale of sensitive data.

Recent technical advances have created a digital transformation in the medical and healthcare industries, which have the potential to improve treatment and patient care. This unprecedented growth in technology is a positive catalyst for healthcare facilities. IoHT is an outcome of such advances in digitalization of healthcare. It amalgamates a network of sensors to collect data to be analyzed for improved services [3,4]. IoHT in today's healthcare enterprises have a greater impact on a wide variety of applications [1–5]. One example is to increase the connectivity between clinical systems and medical devices. This interconnectivity makes medical facilities and devices vulnerable to network intrusions similar to other networking systems that can be vulnerable. Compared to the computer systems in networks, medical devices have a significant concern as the disruption between their connectivity would directly affect healthcare, including patient safety. The combination of healthcare devices, associated software, operating systems, and networking indicates that the cyber security of these devices are challenged. Many healthcare devices, such as automated insulin pumps, pacemakers etc. are used by individuals to be able to lead a near-normal life and have been increasing both in number and complexity [6]. These medical devices are extremely reliable, capable for operating for years while connected to the patient's body, but frequently lack security features such as authentication, encryption, authorization, and other network security features [7–9].

Security vulnerabilities are severe and widespread in both wireless and wired medical devices. Not only is the confidentiality of medical data at risk, but there is a chance of unauthorized command and control that can harm, and even kill, patients. While many pacemaker vendors try to provide certain standards of wireless security in their medical devices, they might still allow control commands to be transmitted wirelessly over the network. This means that online hackers could spoof a command in order to alter the functions of medical devices. For example, insulin pumps can be remotely controlled to manipulate dosage and other settings without the patient's knowledge. Considering the vast number of vulnerabilities and data breaches, there is no doubt that the security of communication in current IoHT devices is still inadequate. This paper frames this complex issue in order to identify the methods of attacks and vulnerabilities, based on which a realistic dataset has been created, which can be used for validating the performance via various attack detection techniques. The impact of security breaches are showcased and the paper explains both in technical and conceptual view the basic environment of Internet-connected healthcare device behavior with respect to cyberattacks.

Recently, IoHT has been globally deployed in various healthcare fields due to advantageous features such as efficient drug management, and treatment computerization that results in satisfied customers and lower financial expense. About 70% of global healthcare organizations

[*] Corresponding author.
E-mail address: m.ahmed.au@ieee.org (M. Ahmed).

https://doi.org/10.1016/j.adhoc.2021.102621
Received 31 January 2021; Received in revised form 28 June 2021; Accepted 15 July 2021
Available online 31 July 2021
1570-8705/© 2021 Published by Elsevier B.V.

have already implemented IoHT in many of their medical devices and equipment, as evidenced by having about a third of IoT devices used in the healthcare industries [10]. The majority of the devices in the IoHT ecosystem have not been designed properly for defending against online attacks, which makes them vulnerable to cyberattacks. According to a study report of Tripwire research, an average of 164 online threats are detected for each 1,000 connected medical devices every day [11]. All types of Internet-connected medical devices, from Wi-Fi-enabled controlled infusion pumps to smart MRI machines, have associated cyberthreats, because the attack surface has increased on many devices that share sensitive information. In turn, this creates security concerns, including the potential violation of medical information and privacy regulations. In addition, adopting numerous IoHT devices that transmit highly sensitive medical data to a cloud server using wireless connection would increase the scope of the cyberattack surface, and introduce many new threats to healthcare systems. A lack of cybersecurity awareness among the patients, medical staff, and professionals can contribute to vulnerabilities in IoHT environments. Notable attack types include, but are not limited to: denial-of-service (DoS) attacks, network sniffing, medical data theft, stealing of personally identifiable information (PII), and treatment manipulation. The consequence of such attacks not only causes disruption in healthcare systems, but also might put the patient's life at risk.

Adequate security standards cannot be practiced when the vulnerabilities and issues are obscure. Therefore, it is crucial to be aware of the possible cyber-vulnerabilities and threats involved in the healthcare systems' IT infrastructure. The rapid increase in the usage, and constant development, of Internet technologies in the medical domain raises an increasing number of security and privacy concerns [10].

By considering all the above issues and risks, there is a need to develop solutions to identify security risks and vulnerabilities in IoHT. The research community requires datasets to study the behavior of cyberattacks, identify patterns, and develop algorithms and countermeasures to detect such attacks and defend IoHT devices. However, publicly available datasets related to cyberattacks are not available for the healthcare domain. More specifically, the datasets do not portray attacks on healthcare devices or IoHT. In addition, due to privacy concerns, healthcare facilities generally do not provide access to such sensitive datasets. To fill this gap, an IoHT testbed was created and a dataset called *ECU-IoHT* [12] developed, which reflects different types of cyberattacks. The key contribution of this paper is to disseminate our dataset for the healthcare security research community, and show the effectiveness of a set of the most popular anomaly detection algorithms on this dataset. This contribution can also encourage the artificial intelligence research community to develop specific, robust algorithms suitable for IoHT security.

The rest of the paper is organized as follows. Section 2 contains a brief literature review about cyberattacks performed on IoHT. Section 3 discusses the development of IoHT testbed and resources used. Section 4 includes an experimental analysis on the ECU-IoHT dataset, and Section 5 concludes the paper, along with providing future research directions.

2. Cyberattacks on Internet of Health Things (IoHT)

According to Segura et al. [13], the healthcare industry was one of the primary victims of the WannaCry ransomware attacks, which affected 200,000 computers in 150 countries. One of the largest healthcare organizations affected most by WannaCry was the UK National Health Service (NHS), which was completely disrupted, having outdated computer systems and software. In addition, thousands of hospitals were affected, including equipment such as MRI scanners and X-ray machines; medical treatment procedures; patient appointments; monitoring and health check devices; and many sensitive medical records and information could not be accessed. This was not restricted to the UK with US-based healthcare organizations reporting thousands of medical treatment procedures having to be stopped due to loss of medical data — ultimately affecting the lives of nearly 300,000 patients.

Most of the healthcare industry using Internet of Things (IoT) devices do not implement standard security measures and policies. Some of them lack network segmentation and cybersecurity awareness when implementing firewalls, intrusion detection system (IDS), encryption, and backup infrastructure. These create a significant attack surface that can potentially be exploited by cybercriminals determined to steal medical and personal information of patients. As the number of connected devices increase, it is crucial to determine how to handle data storage and transfer securely. A lack of embedded security features would increase the risk of human error from poor system configuration to not maintaining audit logs, unauthorized access controls, or disruption in processes typically implemented in IoHT devices [10,11].

Another research study showed how artificial intelligence algorithms can be used to modify CT scans of the lung in real time [14]. The test was conducted in a radiology department's CT room, where a Wi-Fi port was installed that grabbed the CT datasets from the scanner to the digital picture archive. A computer was connected to the Wi-Fi network from the hospital room, and with the algorithm installed on this computer, lung metastases was added to, and deleted from, the CT datasets within seconds. This was done before any radiologists had the chance to even take a look at the images. According to the report, this test attack was successful only because there was no encryption used for the data coming from the scanner to the image archive. Encryption is offered by many medical device vendors, but it is rarely used because of the extra time needed for output, and the added inconvenience in the radiology workflow. It is the responsibility of the hospital management to allow encryption to avoid such attacks.

The possibility of hackers tampering with medical devices to harm individuals is well known [15], and these devices are potentially vulnerable to attacks stealing sensitive health information and medical records. Most healthcare organizations are vulnerable to "med-jacking" (short for medical device hijacking). An example of med-jacking is an incident that occurred in a hospital, during which a blood gas analyzer was infected with malware, which was used to steal passwords for other hospital systems. Another incident happened when hackers took advantage of a vulnerability in a drug pump to gain access to the hospital network and steal medical identities and records. Such identities are more valuable than credit card numbers. This shows how the level of security in many medical devices allows cybercriminals easy access to massive volumes of sensitive healthcare data.

Interpol has recently issued an alert warning regarding online hackers using ransomware to target healthcare systems already overwhelmed by COVID-19, with an increased detection of healthcare attacks detected after the pandemic has started [16]. Online criminals using ransomware can prevent medical personnel and healthcare workers from accessing patient records or other vital medical information until they receive a payment.

According to the literature analysis, it is evident that while Internet of Health Things (IoHT) help increase treatment efficiency, improve overall healthcare quality, and lower costs, IoHT devices also bring risks to the infrastructure. Failure to address these risks make IoHT devices vulnerable to a variety of cyberattacks. After reaching their end of life, such medical devices must be disposed, with the data on their storage media destroyed irreversibly (preventing even digital forensic software tools to restore data). There is a need for additional security to mitigate online attacks, which comes at a cost. The firmware and other software of IoHT devices must be kept up to date, with periodic patches to minimize vulnerability.

The following is a list of key statistics on smart medical devices [17]:

- By 2022, the global market for smart healthcare devices will exceed $674 billion.
- Smart healthcare devices are being incorporated at a faster rate and the average hospital room houses 15–20 devices.

Table 1
Dataset comparison.

Dataset	Simulated	NIDS[a] evaluation	Healthcare
DARPA 98[b]	✓	✓	×
KDD Cup 99[c]	✓	✓	×
NSL-KDD[d]	✓	✓	×
Moore [18]	✓	✓	×
UNSW-NB15[e] [19]	✓	✓	×
BOT-IoT[f]	✓	✓	×
ToN-IoT[g]	✓	✓	×
ISCX[h] [20]	✓	✓	×
Kyoto[i]	✓	✓	×
SCADA [21,22]	✓	✓	×
ECU-IoHT	✓	✓	✓

[a] Network Intrusion Detection System.
[b] https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset
[c] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
[d] https://www.unb.ca/cic/datasets/nsl.html
[e] https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/
[f] https://ieee-dataport.org/documents/bot-iot-dataset
[g] https://ieee-dataport.org/documents/toniot-datasets
[h] http://www.iscx.ca/datasets/
[i] http://www.takakura.com/Kyoto_data/

- It is predicted that the number of IoT devices in a healthcare facility will be more than the number of traditional computing devices.
- In general, smart medical devices have an average of 6 vulnerabilities each.
- Smart devices are being used in healthcare facilities for a long time due to the associated costs, regardless that older, obsolete devices are more vulnerable than the newest ones (and as such, are prime targets to hacking).
- 465,000 pacemakers were recalled by the FDA (Food and Drug Administration) in the United States in 2017, which is one of the many examples that clearly indicate the security issues of such devices.
- More than 25 percent of cyberattacks in healthcare will be associated with IoT by 2020.

3. ECU-IoHT dataset development

The comparison of mainstream datasets (see Table 1) highlights the need for the purposeful design of a novel dataset for device security evaluations in the healthcare domain.

During dataset development, the standard methodology for white hat penetration testing was followed, this method is used by many companies and IT organizations for executing penetration tests for vulnerability analysis [23]. This methodology typically has the life cycle shown in Fig. 1.

3.1. Environment

The computing environment used for developing the ECU-IoHT dataset constituted a Windows 10 (Build 17763) and a *Kali Linux*[1] 2020.2 instance, a mobile Wi-Fi hotspot,[2] a wireless network adaptor and a Bluetooth adaptor to connect the virtual machine to the Internet, and Argus 3.0.8.2.[3]

The Libelium MySignals healthcare kit was selected to create the IoHT testbed. MySignals provides a development platform for eHealth applications and medical devices. Users and IT developers can use this healthcare kit for developing eHealth web interfaces, and can also add their own sensors to build new medical devices. The MySignals kit includes several components and multiple sensors, which can be used to monitor various biometrics. The gathered data is stored and sent to the users' private cloud account through either Wi-Fi or Bluetooth. The data visualization is optimized for tablet, smartphone, and computers. The device and the associated cloud account are compatible with iPhone and Android applications.

3.2. Testbed setup

The MySignals device was wirelessly connected to the Internet, and a sensor was connected to the MySignals device. The data from the MySignals device was wirelessly sent to the broadcast address of the router, and then sent to the Libelium cloud server. A laptop and a smartphone with the MySignals Android application installed were connected to the router. In this setup, attacks became feasible on data transmissions from the MySignals device to the default gateway. Fig. 2 shows the conceptual view of the testbed.

From the available sensors, the following 3 were used for the investigation (selected based on initial experiments):

1. Temperature sensor: easy to use and quick to update the cloud account. It can send the sensor data every 10 s to the Libelium cloud account.
2. Blood pressure sensor: even though it takes longer to send updates to the cloud and is more difficult to set up than a temperature sensor, it is good for attack detection, because it generates large quantities of values.
3. Heart rate sensor: difficult to setup, but sends data to the cloud more frequently the other two.

3.3. Cyberattacks launched

Different types of attacks were performed on the target device from host IP, with the default gateway:

- *Nmap:*[4] *attack* attack using Nmap (also knows as Network Mapper), an open source tool to determine the vulnerabilities in a network and network discovery. Nmap is used to identify the hosts on a network, their open ports, and security risks. By using Nmap on the target IP address, it was found that the device was using Actiontec-embedded Linux. This type of operating system has known vulnerabilities, according to publicly disclosed CVE details. This attack is also known as *Reconnaissance*. *Zenmap:*[5] an open source software, which is the GUI version of Nmap. This tool is used for providing interactive and graphical results, and comparing differences between two scans (thereby pinpointing new services on the network). Zenmap was run multiple times using Zenmap command profiles.
- *ARP poisoning (ARP spoofing):* a type of cyberattack, which sends spoofed ARP messages over LAN networks. ARP poisoning redirects the traffic from the intended host to the attacker. Ettercap was used to perform ARP poisoning on the target via unified sniffing on `wlan0`. This successfully diverted the traffic from the target to a different IP address, and consequently generated a huge amount of malicious ARP packets in Wireshark.

[1] https://www.kali.org
[2] Thereby avoiding excessive data traffic as compared to a public Wi-Fi
[3] Audit Record Generation and Utilization System, https://openargus.org
[4] https://nmap.org
[5] https://nmap.org/zenmap/

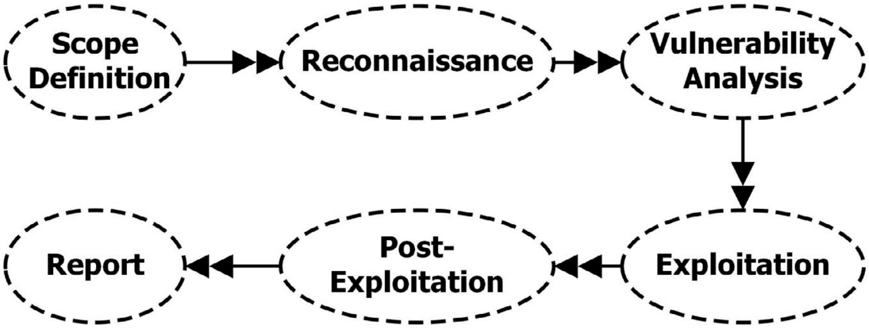


Fig. 1. Penetration testing lifecycle.

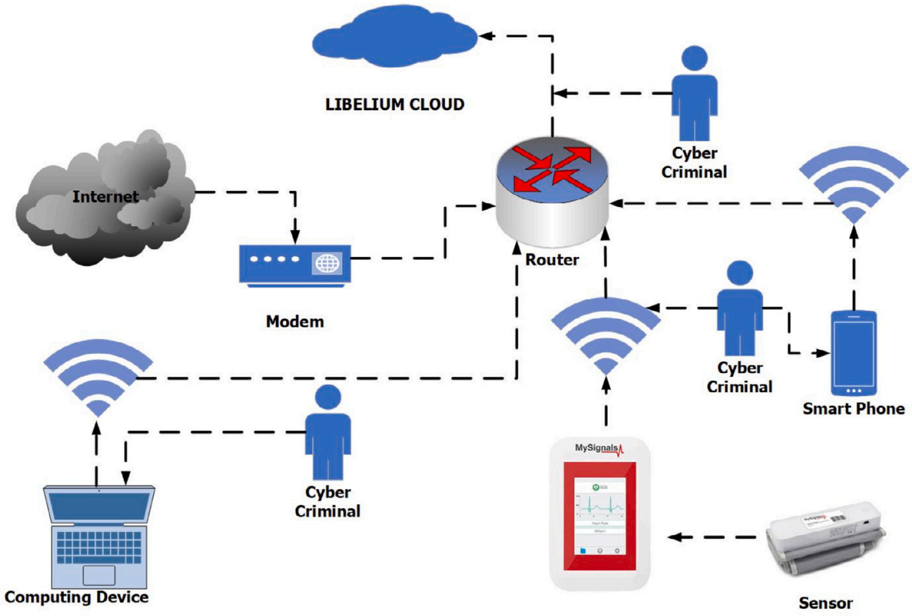


Fig. 2. Testbed design.

- *DoS attack using Ettercap:*[6] a denial-of-service (DoS) attack is an attack during which the user will not be able to access the required resources [24]. DoS attacks are easy to perform and tracing the location of the attack is difficult. Generally, during DoS attacks, large amount of data is sent to the server, potentially impacting reliability or stability. Ettercap, a free and open source network security tool for MITM attacks on LANs [25], was used to perform a DoS attack on our target, which killed the connections instantly.
- *Smurf attack:* a type of distributed denial-of-service (DDoS) attacks. In smurf attacks, a bunch of `echo` request packets are sent continuously like `ping` floods. Smurf attacks exploit the characteristics of broadcast networks to amplify the attack traffic significantly. Ettercap was used to perform a smurf attack on the broadcast address, resulting in a huge number of ICMP packets generated in wireshark.
- *Script injection using the MITMf framework*:[7] a man-in-the-middle attack is an attack, which compromises the communication between two parties. It can be used to either eavesdrop or even modify the data being transmitted. The MITMf framework was used to perform the attack on the broadcast address of the router, and injected an empty shell script into it. As a result, a huge amount of corrupted ARP packets appeared in wireshark.
- *Bettercap attack:* Bettercap is a tool available in Kali Linux, allowing Bluetooth attacks. Bluetooth low-energy device scans were performed, but were unable to identify and capture the MySignals device.

3.4. Dataset development

The configuration and generation of the IoHT dataset is presented in this section, focusing on testbed generation and the dataset development process.

For the period of simulation, the dataset statistics are provided for the cumulative flows of the traffic. In Table 2, different statistical features of the dataset are provided, such as type of protocol, simulation period, number of flows, source bytes, destination bytes, source packets, destination packets, normal and attack instances, and unique source and destination IP addresses.

Argus is the first solution of flow monitoring in a data file. It is an open source Layer 2+ auditing tool. Argus is used for things like bandwidth utilization, higher level protocol data and track performance through the stack. Argus was used to differentiate statistical features from the Pcap file from Wireshark. Racluster is used to aggregate the Argus data and TShark is used to aggregate unique source IP addresses and destination IP addresses in the pcap file. The Table 2 implicitly indicates the number of flows: it allows counting the total number of

[6] https://www.ettercap-project.org

[7] https://github.com/byt3bl33d3r/MITMf

Table 2
Dataset features.

Feature type	Feature count (in 3 h)
Number of flows	8,905
Source bytes	4,726,130
Destination bytes	7,819,570
Source packets	95,148
Destination packets	16,059
Protocol — TCP	23 494
Protocol — TLS	6 034
Protocol — ICMP	77,920
Protocol — DNS	1 241
Protocol — ARP	2 355
Protocol — Others	173
Normal instances	23,453
ARP spoofing	2,359
DoS attack	639
Nmap Port scan	6,836
Smurf attack	77,920
Unique source IP	69
Unique destination IP	71

connections in the capture file. This can be achieved by using TShark or statistics from the Wireshark capture file (by checking the conversations). Source bytes, destination bytes, source packets, and destination packets can be found using Argus; the pcap file was converted into an Argus file, and analyzed using the `Racluster` command. This way, the source bytes, source packets, destination bytes, and destination packets were found. Unique source IP addresses and destination IP addresses can be found using TShark and counting the total number of IP addresses. By using the corresponding filters in Wireshark, the packets of different protocols (TCP, UDP, etc.) have been identified. Finally, packet analysis was performed to differentiate between attack traffic and normal traffic in the capture file [26]. Since the launched attacks are conducted in a constrained environment at a specific time, the packet analysis was not tedious even if it was a manual process.

4. Anomaly detection on the ECU-IoHT dataset

Anomaly detection is an important cyber defence task to detect abnormal data from a given dataset [21,27]. Anomaly is also termed as outliers in the research community. Although there are numerous application domains [28–31], this paper considers anomaly in the context of Internet of Health Things (IoHT). Anomaly detection methods are used to detect different types of cyberattacks. Fig. 3 reflects the standard set of anomaly detection algorithms applied to the ECU-IoHT dataset.

Anomaly detection algorithms are of four different types: nearest neighbor, clustering, statistical, and kernel-based. All these algorithms are unsupervised except *LIBSVM*, which is semi-supervised. Supervised learning-based anomaly detection techniques are considered out of scope here as they are not suitable for addressing zero-day attacks. The motivation behind applying anomaly detection algorithms on the dataset is to identify the strengths and weaknesses of these algorithms in detecting cyberattacks in a state-of-the-art simulated healthcare environment.

The anomaly detection techniques used for the experimentation, along with the popular *RapidMiner*[8] tool, are the following:

- *k-Nearest Neighbor (k-NN)*: a score for being anomalous is assigned to all the data instances based on the average distance to the nearest neighbors.
- *Local Outlier Factor (LOF)*: all the data instances are assigned with an anomaly score based on local density.

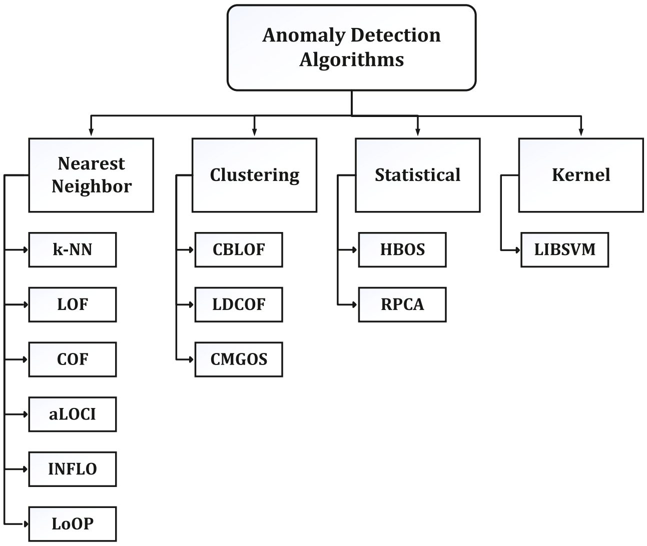


Fig. 3. Standard anomaly detection algorithms.

- *Connectivity-Based Outlier Factor (COF)*: a variant of *LOF* based on density.
- *approximate Local Correlation Integral (aLOCI)*: local correlation integral is used to assign the score for being anomalous to all data instances.
- *Local Outlier Probability (LoOP)*: a probability score for being anomalous is given based on local density.
- *Influenced Outlierness (INFLO)*: the concept of *influenced outlierness* based on neighbors are used to assign scores.
- *Cluster-Based Local Outlier Factor (CBLOF)*: the clustered data instances are assigned anomalous scores based on distances between larger and smaller clusters.
- *Clustering-based Multivariate Gaussian Outlier Score (CMGOS)*: the clustered instances are assigned scores based on their distances to the cluster center.
- *Local Density Cluster-Based Outlier Factor (LDCOF)*: based on the distance to the nearest large cluster, scores are assigned for being anomalous.
- *Robust Principal Component Analysis (RPCA)*: originated from principal component analysis.
- *Histogram-based Outlier Score (HBOS)*: a histogram based techniques that uses either fixed or dynamic bin width to assign scores to all data instances for being anomalous.
- *One Class Support Vector Machine (LIBSVM)*: anomalous score for each data instance in the dataset is calculated using support vector machine (SVM). It is a semi-supervised one-class SVM that is used for unsupervised learning.

The parameters for these algorithms are shown in Table 3.

4.1. Evaluation

While applying the anomaly detection algorithms, the dataset is converted to four different subsets where each of the set contains normal data from the original dataset and a single type of attack data. For example, a set contains all the normal instances from the original dataset and only the *ARP* labeled instances from the original dataset. Therefore, the anomaly detection algorithms will be more focused in identifying a certain type of attack instead of detection different types of attacks. Also, the effectiveness of individual anomaly detection algorithms in identifying a certain type of attack in IoHT will provide more meaningful insights. Fig. 4 shows the distribution of normal and

[8] https://rapidminer.com

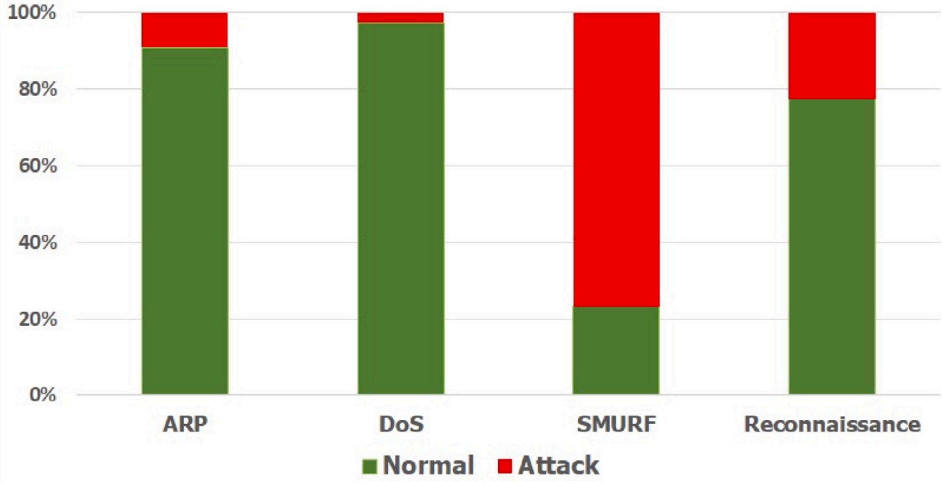
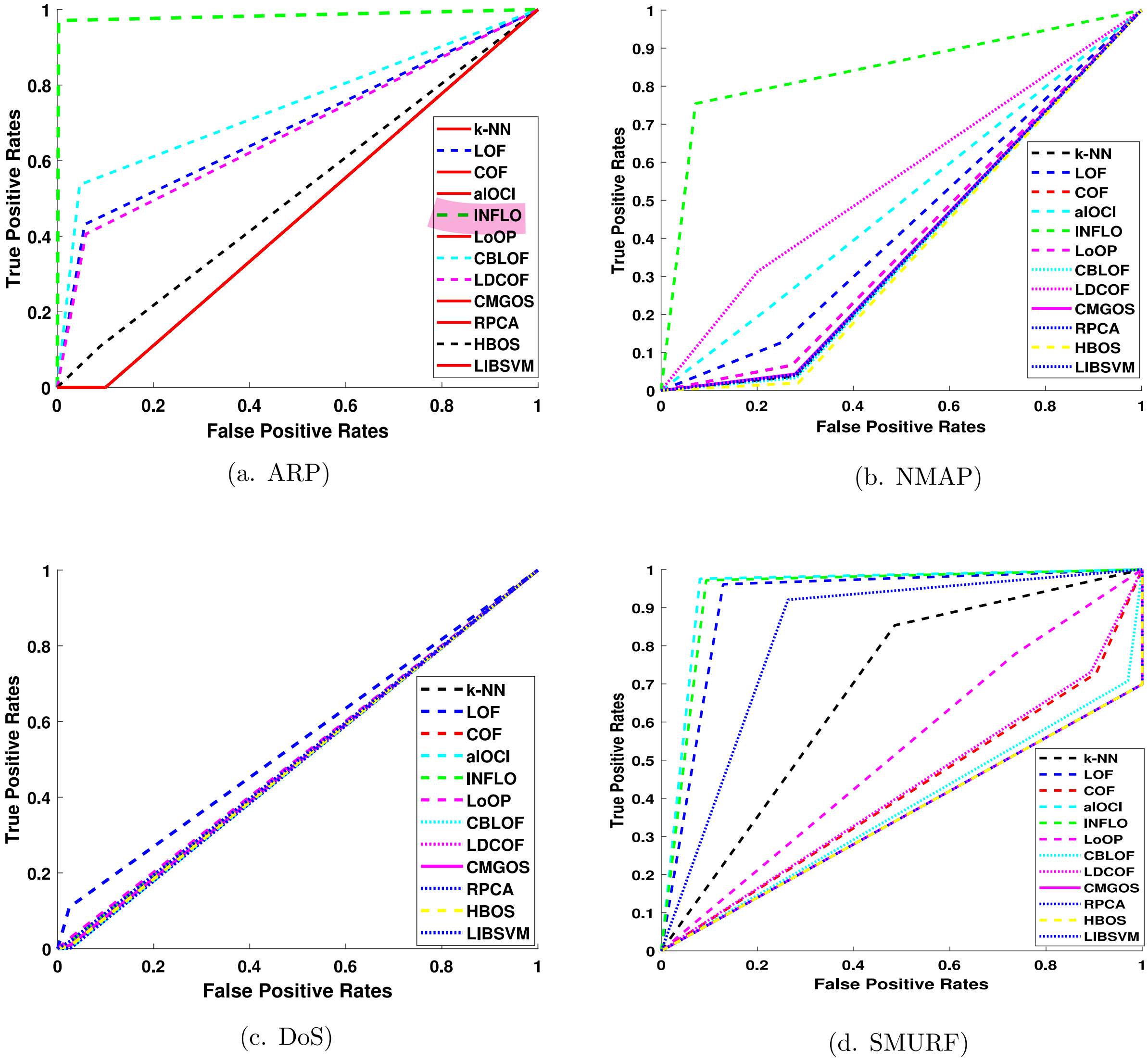


Fig. 4. Distribution of normal and anomalous instances.



(a. ARP)

(b. NMAP)

(c. DoS)

(d. SMURF)

Fig. 5. Performance of anomaly detection algorithms.

anomalous instances in the four sets of data used for experimental analysis.

- **ARP:** Fig. 5a shows the performance of anomaly detection algorithms to identify *ARP* attacks and it is clearly evident that the

Table 3
Parameters used by anomaly detection algorithms.

Algorithm	Parameter
k-NN	Number of neighbors considered (k)
LOF	Minimum and maximum number of neighbors (k_{min} and k_{max})
COF	Number of neighbors considered (k)
aLOCI	Tree depth, number of grids, minimum neighbors and difference of levels
INFLO	Number of neighbors considered (k)
LoOP	Number of neighbors considered and normalization factor (k)
CBLOF	Percentage of normal data (α), ratio between the sizes of clusters (β)
LDCOF	Ratio between sizes of clusters (γ)
CMGOS	Probability of normal class and covariance estimation (γ)
HBOS	Number of bins and bin width
RPCA	Probability of normal class
LIBSVM	Type of support vector machine and kernel

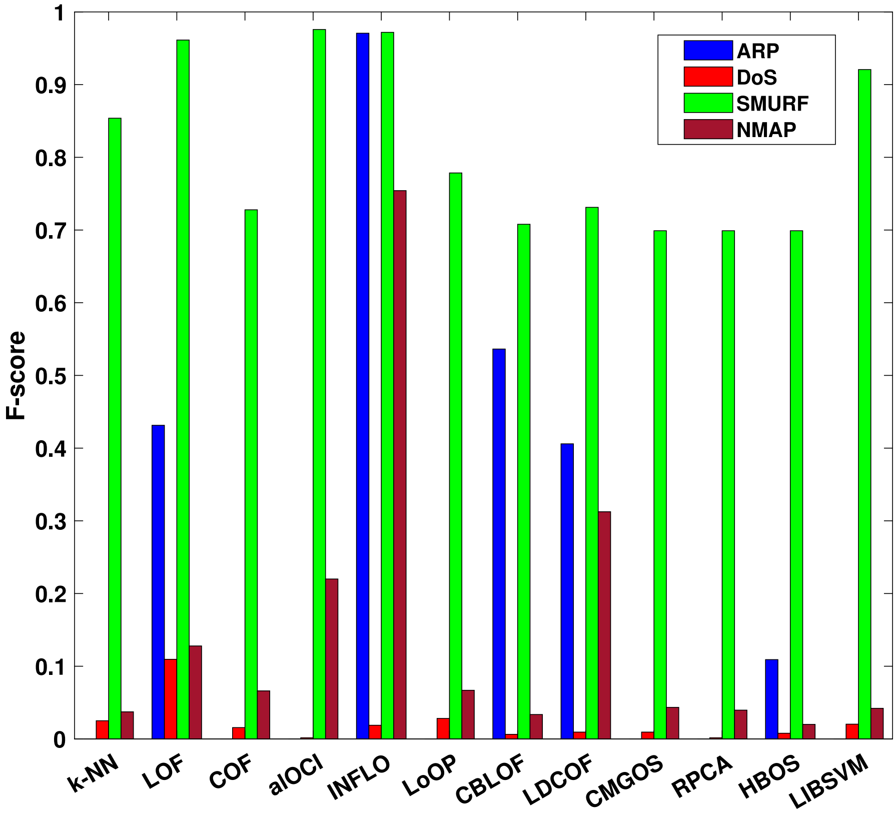


Fig. 6. F-scores of the anomaly detection algorithms on different attacks.

INFLO method is superior than others, even if the proportion of anomalous instances are not too high in the dataset. The key takeaway from this result is that *ARP* attacks are best handled by the *INFLO* algorithm and well established algorithms for rare anomaly detection such as *k-NN, LOF, COF* etc. are not suitable for identifying such attacks, especially from IoHT environment.

- **NMAP/Reconnaissance**: Fig. 5b shows that *INFLO* algorithm is performing better than others, even if the distribution of anomalous instances is relatively normal. For example, the proportion of anomalous instances are showcasing the rarity. The result indicates that, the majority of existing anomaly detection techniques are unable to identify the attacks which are used by cyber criminals for reconnaissance purpose. Therefore, this dataset and the attack patterns will be able to security researchers to devise algorithms to identify such attacks with high accuracy and low false positive rates.
- **DoS:** Fig. 5c showcases the performance of anomaly detection algorithms in identifying the *DoS* attacks from IoHT environment, even if the presence of these attacks are very low in number, the *LOF* algorithm identified few such instances and rest of the algorithms have performed poorly in detecting such attacks. It is important to note that, the denial of service attacks are considered as collective anomaly [32], however, in this dataset, the distribution of attack instances are very low and hence the true nature of collective anomalies are absent.
- **SMURF**: Fig. 5d shows that *INFLO* and *aLOCI* algorithms are superior than others in detecting *smurf* attacks. These attacks are also considered as a variant of denial of service attacks. In the data distribution, it is also shown that the number of anomalous instances are much higher than the normal instances. Hence, this set represents the true characteristics of collective anomalies and most of the algorithms are able to identify these attacks with comparatively lower false positive rates.

$$TPR = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (1)$$

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives} \quad (2)$$

Almost all the algorithms have a consistent success rate (Eqs. (1) and (2)) in identifying *smurf* attacks. Apart from the nature of the network traffic associated with this kind of attack, the high proportion of such traffic made it easier for the anomaly detection algorithms to perform. In contrast, the algorithms were unable to detect *DoS* attacks, with the exception of *LOF*. The nearest neighbor and clustering-based algorithms had moderate success in identifying *ARP* attacks, however the statistical and kernel-based techniques had less than 20% hit rate. For the *Nmap/Reconnassance* attack, only *INFLO* had significant success while the rest of the algorithms had less than 10% hit rate on average. Based on the performance of these algorithms, it is evident that the attacks from contemporary IoHT are sophisticated, and it is imperative to devise newer techniques to identify these attacks with fewer false alarms.

$$F - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

Fig. 6 shows the *F-scores* (Eq. (3)) of the anomaly detection algorithms. It is evident that the *INFLO* algorithm performs consistently in identifying the different attacks than other techniques used.

5. Conclusions and future research directions

In parallel with the global and continuously increasing deployment of Internet-supported medical devices, cyberattacks against these devices are also increasing. This urges implementations that are less prone to targeted attacks than the ones in use today, which requires the deep analysis of potential attack types. However, while general network intrusion and incident response datasets are readily available, there are no purpose-generated datasets in the healthcare domain. This paper is the first to attempt to fill this gap by providing a novel dataset that can help identify vulnerabilities of current deployments, and set directions for future research across research communities in information security, data science, and artificial intelligence.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M.B. Yassein, I. Hmeidi, M. Al-Harbi, L. Mrayan, W. Mardini, Y. Khamayseh, IoT-based healthcare systems: A survey, in: Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, ACM, New York, 2019, http://dx.doi.org/10.1145/3368691.3368721.

[2] C.A. da Costa, C. F.Pasluosta, B. Eskofier, D.B. da Silva, R. da Rosa Righi, Internet of health things: Toward intelligent vital signs monitoring in hospital wards, Artif. Intell. Med. 89 (2018) 61–69, http://dx.doi.org/10.1016/j.artmed.2018.05.005.

[3] S.A. Haque, S.M. Aziz, M. Rahman, Review of cyber-physical system in healthcare, Int. J. Distrib. Sensor Netw. (2014) http://dx.doi.org/10.1155/2014/217415.

[4] K. Saleem, Z. Tan, W. Buchanan, Security for cyber-physical systems in healthcare, in: C. Thuemmler, C. Bai (Eds.), Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare, Springer, Cham, 2017, pp. 233–251, http://dx.doi.org/10.1007/978-3-319-47617-9_12.

[5] S.M.R. Islam, D. Kwak, M.H. Kabir, M. Hossain, K.-S. Kwak, The internet of things for health care: A comprehensive survey, IEEE Access 3 (2015) 678–708, http://dx.doi.org/10.1109/ACCESS.2015.2437951.

[6] P.A.H. Williams, A.J. Woodward, Cybersecurity vulnerabilities in medical devices: A complex environment and multifaceted problem, Med. Devices: Evidence Res. (8) (2015) 305–316, http://dx.doi.org/10.2147/MDER.S50048.

[7] M. Ahmed, False image injection prevention using ichain, Appl. Sci. 9 (20) (2019) http://dx.doi.org/10.3390/app9204328.

[8] S.C. Sethuraman, V. Vijayakumar, S. Walczak, Cyber attacks on healthcare devices using unmanned aerial vehicles, J. Med. Syst. 44 (2020) http://dx.doi.org/10.1007/s10916-019-1489-9.

[9] M. Ahmed, A.S.S.M.B. Ullah, False data injection attacks in healthcare, in: Y.L. Boo, D. Stirling, L. Chi, L. Liu, K.-L. Ong, G. Williams (Eds.), Data Mining, Springer, Singapore, 2018, pp. 192–202, http://dx.doi.org/10.1007/978-981-13-0292-3_12.

[10] F. Alsubaei, A. Abuhussein, S. Shiva, Ontology-based security recommendation for the internet of medical things, IEEE Access 7 (2019) 48948–48960, http://dx.doi.org/10.1109/ACCESS.2019.2910087.

[11] A. Arampatzis, Protecting modern IoMT against cybersecurity challenges, URL: https://www.tripwire.com/modern-iomt-cybersecurity-challenges/.

[12] M. Ahmed, S. Byreddy, A. Nutakki, L. Sikos, P. Haskell-Dowland, ECU-IoHT, 2020, http://dx.doi.org/10.25958/5f1f97b837aca, URL: https://ro.ecu.edu.au/datasets/48/.

[13] M. Segura, C. Butler, F. Tabibkhoei, The internet of medical things raises novel compliance challenges, 2018, URL: https://www.meddeviceonline.com/doc/the-internet-of-medical-things-raises-novel-compliance-challenges-0001.

[14] P.G. von Grätz, Dealing with an internet of medical threats, 2019, URL: https://www.mobihealthnews.com/dealing-internet-medical-threats.

[15] A. Chacko, T. Hayajneh, Security and privacy issues with IoT in healthcare, EAI Endorsed Trans. Pervasive Health Technol. 4 (14) (2018) http://dx.doi.org/10.4108/eai.13-7-2018.155079.

[16] M. Garde, Interpol chief: 'be vigilant, be sceptical, and protect your computers', 2020, URL: https://www.euractiv.com/interpol-chief-be-vigilant-be-skeptical-and-protect-your-computers/.

[17] S. Morgan, How vulnerable is the iomt to cyber threats?, 2019, URL: https://cybersecurityventures.com/patient-insecurity-explosion-of-the-internet-of-medical-things/.

[18] A.W. Moore, D. Zuev, Internet traffic classification using Bayesian analysis techniques, in: Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, ACM, New York, 2005, pp. 50–60, http://dx.doi.org/10.1145/1064212.1064220.

[19] N. Moustafa, J. Slay, UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), in: 2015 Military Communications and Information Systems Conference, IEEE, 2015, http://dx.doi.org/10.1109/MilCIS.2015.7348942.

[20] A. Shiravi, H. Shiravi, M. Tavallaee, A.A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Comput. Secur. 31 (3) (2012) 357–374, http://dx.doi.org/10.1016/j.cose.2011.12.012.

[21] M. Ahmed, A. Anwar, A.N. Mahmood, Z. Shah, M.J. Maher, An investigation of performance analysis of anomaly detection techniques for big data in SCADA systems, EAI Endorsed Trans. Ind. Netw. Intell. Syst. 15 (3) (2015) http://dx.doi.org/10.4108/inis.2.3.e5.

[22] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, M. Palaniswami, Labelled data collection for anomaly detection in wireless sensor networks, in: S. Marusic, M. Palaniswami, J. Gubbi, P. Corke (Eds.), Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, IEEE, 2010, http://dx.doi.org/10.1109/ISSNIP.2010.5706782.

[23] P. Engebretson, The Basics of Hacking and Penetration Testing: Ethical Hacking and Penetration Testing Made Easy, Syngress, Waltham, MA, USA, 2013.

[24] M. Ahmed, Thwarting DoS attacks: A framework for detection based on collective anomalies and clustering, Computer 50 (9) (2017) 76–82, http://dx.doi.org/10.1109/MC.2017.3571051.

[25] L.F. Sikos, Packet analysis for network forensics: A comprehensive survey, Forensic Sci. Int.: Digit. Investigation 32 (2020) http://dx.doi.org/10.1016/j.fsidi.2019.200892, Article 200892.

[26] K. Alexis Fidele, Suryono, W. Amien Syafei, Denial of service (DoS) attack identification and analyse using sniffing technique in the network environment, in: E3S Web of Conferences, in: E3S Web of Conferences, vol. 202, 2020, p. 15003.

[27] M. Ahmed, A.N. Mahmood, J. Hu, A survey of network anomaly detection techniques, J. Netw. Comput. Appl. 60 (2016) 19–31, http://dx.doi.org/10.1016/j.jnca.2015.11.016.

[28] M. Ahmed, A. Barkat, Performance analysis of hard clustering techniques for big IoT data analytics, in: 2019 Cybersecurity and Cyberforensics Conference, IEEE, 2019, pp. 62–66, http://dx.doi.org/10.1109/CCC.2019.000-8.

[29] M. Ahmed, N. Choudhury, S. Uddin, Anomaly detection on big data in financial markets, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, New York, ISBN: 9781450349932, 2017, pp. 998–1001, http://dx.doi.org/10.1145/3110025.3119402.

[30] M. Ahmed, A.N. Mahmood, M.R. Islam, A survey of anomaly detection techniques in financial domain, Future Gener. Comput. Syst. 55 (2016) 278–288, http://dx.doi.org/10.1016/j.future.2015.01.001.

[31] S.M. Khandaker, A. Hussain, M. Ahmed, Effectiveness of hard clustering algorithms for securing cyber space, in: A.-S.K. Pathan, Z.M. Fadlullah, M. Guerroumi (Eds.), Smart Grid and Internet of Things, Springer, Cham, 2019, pp. 113–120, http://dx.doi.org/10.1007/978-3-030-05928-6_11.

[32] M. Ahmed, Thwarting DoS attacks: A framework for detection based on collective anomalies and clustering, Computer 50 (9) (2017) 76–82.

Mohiuddin Ahmed is currently working as Lecturer of Computing and Security in the School of Science at Edith Cowan University. Mohiuddin has been working in the areas of data analytic and cyber security, in particular false data injection attacks in Internet of Health Things (IoHT) and Internet of Flying Things (IoFT). His research projects are funded by different external agencies. He has edited books on data analytics, security analytics, blockchain and other contemporary issues. He has also engaged with media outlets such as newspaper, magazine, The Conversation etc. He is also an ACM Distinguished Speaker, Australian Computer Society Certified Professional and a Senior Member of IEEE.

Surender Byreddy received a Masters degree specialising in Cyber Security at the Edith Cowan University, Australia. He is an experienced programmer and worked on several projects. He is currently working as a cyber security engineer providing security services for a large mining company. He is interested in developing robust countermeasures for cyber attacks in the healthcare sector.

Anush Nutakki received his postgraduate masters from Edith Cowan university in the field of cyber security specialization. He has experience background as a User Interface (UI) designer, PowerApps developer and network analyst. He is interested in working on different aspects of cyber security like network security monitoring, exploring vulnerabilities and analysing them.

Leslie Sikos Ph.D. is a computer scientist specializing in network forensics and cybersecurity applications powered by artificial intelligence and data science. He has industry experience in data center and cloud infrastructures, cyberthreat prevention and mitigation, and firewall management. He regularly works on cybersecurity research projects, and collaborates with the Defence Science and Technology Group of the Australian Government, CSIRO's Data61, and the Cyber Security Collaborative Research Centre. He is a reviewer of academic journals such as Computers & Security and IEEE Transactions on Dependable and Secure Computing, and chairs sessions at international conferences, and regularly edits books, on AI in cybersecurity. Dr. Sikos holds professional certificates, and is a member of the IEEE Computer Society Technical Committee on Security and Privacy, and a founding member of the IEEE Special Interest Group on Big Data for Cybersecurity and Privacy.

Associate Professor Paul Haskell-Dowland is the Associate Dean for Computing and Security in the School of Science at Edith Cowan University, Perth, Australia. Paul is the Working Group Coordinator and the ACS/Australian Country Member Representative to the International Federation for Information Processing (IFIP) Technical Committee 11; secretary to IFIP Working Group 11.1; a member of the ACS Cyber Security Committee; a Senior Member of the IEEE and the ACS(Certified Professional); and, a Fellow of the Higher Education Authority, BCS and the Australian Information Security Association. He is the author of over 90 papers in refereed international journals and conference proceedings. Paul has more than 20 years of experience in cyber security research and education in both the UK and Australia.