Intelligent Control and Security of Fog Resources in Healthcare Systems via a Cognitive Fog Model

MOHAMMED AL-KHAFAJIY, University of Reading, UK
SAFA OTOUM, Zayed University, UAE
THAR BAKER, Liverpool John Moores University, UK
MUHAMMAD ASIM, National University of Computer and Emerging Sciences, Pakistan
ZAKARIA MAAMAR, Zayed University, UAE
MOAYAD ALOQAILY, Al-Ain University (UAE), and xAnalytics Inc. (Canada), UAE and Canada
MARK TAYLOR, Liverpool John Moores University, UK
MARTIN RANDLES, Liverpool John Moores University, UK

There have been significant advances in the field of Internet of Things (IoT) recently, which have not always considered security or data security concerns: A high degree of security is required when considering the sharing of medical data over networks. In most IoT-based systems, especially those within smart-homes and smart-cities, there is a bridging point (fog computing) between a sensor network and the Internet which often just performs basic functions such as translating between the protocols used in the Internet and sensor networks, as well as small amounts of data processing. The fog nodes can have useful knowledge and potential for constructive security and control over both the sensor network and the data transmitted over the Internet. Smart healthcare services utilise such networks of IoT systems. It is therefore vital that medical data emanating from IoT systems is highly secure, to prevent fraudulent use, whilst maintaining quality of service providing assured, verified and complete data. In this paper, we examine the development of a Cognitive Fog (CF) model, for secure, smart healthcare services, that is able to make decisions such as opting-in and opting-out from running processes and invoking new processes when required, and providing security for the operational processes within the fog system. Overall, the proposed ensemble security model performed better in terms of Accuracy Rate, Detection Rate, and a lower False Positive Rate (standard intrusion detection measurements) than three base classifiers (K-NN, DBSCAN and DT) using a standard security dataset (NSL-KDD).

Additional Key Words and Phrases: Fog Computing, Cognitive Fog, Fog Security, Medical Data Security

Authors' addresses: Mohammed Al-khafajiy, m.d.al-khafajiy@reading.ac.uk, University of Reading, Department of Computer Science, Reading, UK; Safa Otoum, Zayed University, Abu Dhabi, UAE, Safa.Otoum@zu.ac.ae; Thar Baker, Liverpool John Moores University, Liverpool, UK, T.Baker@ljmu.ac.uk; Muhammad Asim, National University of Computer and Emerging Sciences, Pakistan, muhammad.asim@nu.edu.pk; Zakaria Maamar, Zayed University, Dubai, UAE, akaria.maamar@zu.ac.ae; Moayad Aloqaily, Al-Ain University (UAE), and xAnalytics Inc. (Canada), Al-Ain, UAE and Canada, maloqaily@ieee.org; Mark Taylor, Liverpool John Moores University, Liverpool, UK, M.J.Taylor@ljmu.ac.uk; Martin Randles, Liverpool John Moores University, Liverpool, UK, M.J.Randles@ljmu.ac.uk.

0730-0301/2020/8-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

ACM Reference Format:
Mohammed Al-khafajiy, Safa Otoum, Thar Baker, Muhammad Asim, Zakaria Maamar, Moayad Aloqaily, Mark Taylor, and Martin Randles. 2020. Intelligent Control and Security of Fog Resources in Healthcare Systems via a Cognitive Fog Model. *ACM Trans. Graph.* 37, 4, Article 111 (August 2020), 17 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The Internet of Things (IoT) is a network of smart connected objects covering electronics, software, and network connectivity. According to [Rahmani et al. 2015], IoT is a manageable set of convergent developments in sensing, identification, communication, networking, and informatics devices and systems. The IoT is increasingly being integrated into daily lives due to advances in automating daily functions and roles, such as healthcare monitoring which becomes more efficient using an IoT architecture [Wu et al. 2014]. According to a 2015 IBM white-paper [Green 2015], the IoT needs to be smarter so that better results from connected devices could be attained [Maamar et al. 2018]. In most IoT system architectures, such as healthcare monitoring systems within smart homes or hospitals, there exists a bridging point, called fog node, between the IoT devices and the Internet (i.e. the cloud). This fog layer often only performs basic functions such as translating between the protocols used in the Internet and the deployed smart objects such as Low power devices and Lossy Networks (LLN) [Agustin et al. 2017] along with the provision of basic data storage and manipulation services (e.g., data filtering and aggregation) [Rahmani et al. 2015]. Fog computing offers the ability to extend not only storage capabilities but also networking and computing capabilities of the cloud to the edge of the network. The better positioning of fog nodes within the network, in relation to fog connectivity with end-devices, can boost functionality, especially for systems that require data synchronisation with low latency (i.e. real-time healthcare service systems) [Al-khafajiy et al. 2018]. These include examples such as heart-monitors and smart-meters. Fog nodes can have the advantage of providing knowledge and constructive control over both the network of IoT devices and the data transmitted over the network. This enables fog nodes to not only act on the data but also make intelligent decisions regarding resource utilization and security.

In this paper, we propose a novel Cognitive Fog (CF) model that interacts with surrounding objects (e.g. the IoT devices and other fog nodes) and enables:

- (i) Completion of the processes of the running services and satisfying their needs and requirements, assuring correctness of medical/clinical data, for example.
- (ii) Decision making for contributing to serving other IoT devices based on their interactions and computational abilities ensuring agility and robustness within a healthcare services system.
- (iii) Security awareness for detecting malicious activities that may undermine the performance or privacy of a healthcare services fog system.

Cognitive capabilities provided by the model include reasoning, problem solving, planning, and learning from experience. The cognitive capabilities relate to monitoring performance on related fog nodes and IoT devices to determine how best to allocate resources, transmit data, and detect malicious activity based upon analysis of current and previous performance Shi et al. [2019]Balasubramanian et al. [2019b]Ali and Cheng [2019]. These are important considerations in the provision of medical and healthcare service via smart networks. The model uses monitoring data to enhance service provisioning and security within the fog system. A fog node federation gathers relevant fogs' processes and models together according to the needs and requirements of different situation that required cognitive fog decisions to handle incoming requests. The model operates in both a planned and ad-hoc mode. The planned mode is developed during the set-up stage and has the fog nodes and IoT devices already identified with respect to a situation's needs and requirements. The ad-hoc mode is activated when the existing planned federations cannot deal with a situation and, hence, necessary fog constituents that will satisfy this situation's needs and requirements, are identified. An important aspect of the Cognitive Fog model is performance free from malicious intervention. Intrusion Detection Systems (IDSs) are essential units in networks, and aim to protect the networks against malicious activities that may interfere with the network and manipulate the data traffic. The Cognitive Fog model, particularly in healthcare systems, uses an ensemble classifier based on K-Nearest Neighbor (K-NN), Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Decision Trees (DT), that enables detection of attacks that may target the CF. The goal of the research presented in this paper is a CF model that uses planned and ad-hoc federations to monitor and control resources and prevent malicious activity in a healthcare services system based upon an ensemble classifier that uses K-NN, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Decision Trees (DT) to detect malicious activity that may undermine performance and security.

To this end, the main contributions of this paper are as follow:

- Empower fog nodes with reasoning, learning, and adaptation capabilities to provide resilience to attacks, swift recovery of systems, following any security incidents and robustness to attack, though for example, early detection. These fog nodes would be incorporated into services provisioning models facilitating the proposed intelligent control and security.
- Propose an ensemble classifier based on K-NN, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Decision Tree (DT), that is able to detect attacks that may target the CF and classify these attacks with high accuracy and efficiency level.
- Implement the CFs federation based on Planned Federations ($\mathcal{PF}$) and Ad-hoc Federations ($\mathcal{AF}$). In ($\mathcal{PF}$), all CFs are known to each other from the design time to assist, or take benefit from, each other. Contrarily, in ($\mathcal{AF}$), the CFs are communicating with each other based on a need (i.e., formed on the fly). In both cases the data security benefits ensue from the robust configuration of the CFs in terms of communication, threat awareness, process recovery and, not least, because of the intelligent control of data by the CFs.

The remainder of this paper is organized as follows: In Section 2, related works to this paper are discussed, and the main challenges are described. The proposed architecture and framework presented in more details in Section 3 and 4. The case study, testing and evaluation are discussed in Section 5. Finally, the conlusions and future directions are summarised in Section 6.

2 RELATED WORK

This section discusses the state-of-the-art work in healthcare data processing over fog and cloud computing. This includes, for example, resources allocation for healthcare data processing, workload distribution, and use of machine learning to improve the performance and security of IoT-enabled applications. Despite the growing interest in fog computing for IoT-enabled applications, there does not appear to be established approaches for intelligence distribution via Cognitive Fog Computing models for applications in healthcare service systems. Previous research has discussed federated devices, Cognitive fog IoT, and load-distribution via fog cloud models. Fog entities are being used to deliver continuous/stable simple/complex services for surrounding environments [Al Ridhawi et al. 2019]. Moreover, Also, trust-based security solutions have been the focus of both industry and academia. Trust can help in detecting and isolating those malicious entities which are part of a network using legal identities. Besides, trust plays an important role in nurturing the relation between different fog nodes in term of maintaining user privacy and information security.

There are many trust-based models and resource access control across heterogeneous networks that have been reviewed thoroughly in the literature [Balasubramanian et al. 2019a; Galluccio et al. 2015; Henze et al. 2014; Wang et al. 2018b]. Kai Hwang with his team present the idea of trust in clouds, in which he suggested to combine security-based data centers, data access and virtual clusters driven by reputation systems [Hwang et al. 2009]. The work of [Henze et al. 2014] introduces a trust mechanism using a point-based technique for protecting against unauthorized entry. For securing data transmission between two devices, trust was used in the gateway devices. However, it does not guarantee the credibility of sensor data and cloud providers.

Heil et al. [Heil et al. 2007] propose a context-aware federation approach for IoT devices to support user access, connect, and locate arbitrary devices according to their functionalities. The approach utilizes the Federated Devices Assemblies (FDX) for integrating realword IoT devices into service federations. Mathlouthi et al's [Mathlouthi and Saoud 2017] present an approach which enables the composition of federated cloud based System of Systems (SoS) to work co-operatively in order to achieve common goals. A SoS constitutes several complex, heterogeneous, and autonomous system deployed on heterogeneous cloud environments. Both the functional and the non-functional requirements are considered to obtain the best SoS composition and maintain the overall Quality of Service (QoS). Heil et al's things federation and Mathlouthi et al's SoS composition approach are different from ours in the sense that we advocate and focus on the cognitive fog and fog federation in response to specific situations relating to performance and security.

Al-Turjman et al. [Al-Turjman 2017] propose a Cognitive Cashing approach for the Future Fog that focuses on data exchange in Information Centric Sensor Networks (ICSNs). It depends on functional parameters (such as age of information and data fidelity) to assign a value to the cashed data while retaining the most valuable one in the cache for prolonged time period. This enables a significant availability of the most valuable and difficult to retrieve data in the ICSNs. The work of Jalali et al. [Jalali et al. 2017] propose a cognitive IoT gateway based approach supported by cognitive analytic and machine learning to improve the performance of IoT-enabled applications. The proposed approach enables the IoT devices to automatically learn and decide whether and when to run an application on the Cloud or on the fog.

Fog-based trust management is on its inception, because there have been very few reported work on the topic of trust mechanism in fog computing. In [Alrawais et al. 2017], the authors carried out a survey for finding the current security issues and challenges in IoT and propose a fog-based security mechanism to improve the distribution of certification revocation information between IoT devices. The authors in [Wang et al. 2018a] come up with the concept of fog-based hierarchical trust-based mechanism for SDN, which has two distinctive features: trust in network structure, and the trust between cloud service providers (CSPs) and sensor service providers (SSPs). They focus on the packet loss rate, route failure rate and forwarding delay only. Elmisery et al. [Elmisery et al. 2016] propose a fog-based middleware where trust between a fog node and the cloud is calculated in a decentralized fashion using entropy definition. The authors in [Soleymani et al. 2017] proposed a fuzzy trust-based model that considers experience and plausibility for securing vehicular networks. To ensure the correctness of information collected from authorized vehicles, a series of security checks are performed. Moreover, a fog -based facility is used to evaluate the level of accuracy of event's location.

A mobile fog is proposed by Hong et al. [Hong et al. 2013], which is a high-level programming model that is geographically distributed on large scale and highly sensitive towards latency Balasubramanian et al. [2020]. A bundle of various functions and event handlers are included in the mobile fog, which an application can access whenever required. The mobile fog model is static, and therefore does not present a generic model. However, it is a model for a particular applications while excluding functions that take care of the processing primitives. Moreover, using mobile fog as a primary resource is not a good solution as the mobile fog may not be available due to signal loss. Beate et al. [Ottenwälder et al. 2013] introduce a placement and migration technique for cloud and fog resources providers. They demonstrate how the application prior knowledge of Complex Event Processing (CEP) system helps reduce the necessary bandwidth of Virtual Machines (VMs) migration. Notwithstanding, the work failed to enhance the workload mobility as fog nodes are also capable of performing computationally intensive tasks.

The authors in [Agarwal et al. 2016] focus on resource allocation. They propose a three-layer architecture: Clients, Fog, and Cloud. Then, they implement a workload distribution algorithm between the cloud and fog layers. This necessitated implementing a module that checks if enough computational resources exist in the designated fog node. Consequently, incoming tasks can be executed subject to resource availability; or otherwise postponing few tasks or dispatching them to the cloud node. Hence, the main limitation of this work is the assumption that a manager does exist between every fog and cloud node to manage the cooperation among them. This approach does not thoroughly support the proper execution of distributed tasks. Kapsalis et al. [Kapsalis et al. 2017] propose a new fog layer that incorporates the *'manager'*; thereof, allocating resources and managing tasks run in the same fog layer. It utilizes a distributed communication method based on publication/subscription pattern resource sharing between fogs. In their fog stratum, they define a so-called "utility metric" among fog nodes which identify the communication benefits when/if the fog nodes share resources. They first specify an organised list of preference pairing fog nodes for each node. Each node in the fog layer will then set a pairing request to its preferred pairing nodes. On the reception side, depending on the preference and benefits of the previously received requests, a target node decides either to accept or reject the request. A limitation of this work is that the core parameters upon which the fog nodes take decisions are the communication cost between nodes which can be affected by some factors (e.g., time and location) of the pairing. In addition, they do not consider the QoS (such as latency, bandwidth and etc.) in the resource sharing decisions.

Much research considered the challenges of intrusion detection in fog networks [Bhuyan et al. 2014], [Alom and Taha 2017], [Soheily-Khah et al. 2018], [Otoum et al. 2017b] and [Otoum et al. 2017a]. Particle Swarm Optimization (PSO) has been adopted in [Aburomman and Reaz 2016] for proposing an IDS-based on ensemble technique. Another ensemble technique has been introduced in [M.Govindarajan 2016], where the authors presented a hybrid IDS by adopting both Support Vector Machine (SVM) and the Radial Basis Function (RBF) in which they utilized various datasets and showed that heterogeneous models performed better than homogeneous models. Other ensemble-based solutions are presented in [Govindarajan 2016] and [Moustafa et al. 2019]. The authors in [Govindarajan 2016] used arcing for heterogeneous and bagging for homogeneous classifiers in which they utilized SVM and RBF as base classifiers in their proposed ensemble method. The work presented in [Moustafa et al. 2019] proposes an ensemble intrusion detection mechanism to minimize malicious activities affecting IoT protocols. The authors adopted Artificial Neural Network (ANN), Naive Bayes (NB) and Decision Tree (DT) to detect such malicious activities. Thus far, the work in the field of fog computing revolves around the on network communication and content migration.

Table 1. Notations used in the paper

Symbol	Description
$\mathcal{CF}$	cognitive fog node
ts	task to be executed by CF
$\mathcal{AF}$	ad-hoc federations
$\mathcal{PF}$	planned federations
$E(S)$	entropy for a set S
$p(x)$	probability of an event x
$E(S,T)$	Entropy with respect to feature T
$G(S,T)$	Entropy change after a decision on feature T, where S is a set

However, our work focuses on investigating a fog framework to empower optimal load on fog nodes, and provides appropriate security, facilitating protection and privacy of medical data in exchanges between IoT systems and practitioners. Abouelmehdi et al [Abouelmehdi et al. 2018] commented that concerns over healthcare data security and privacy are steadily increasing, in particular with regard to cloud based healthcare data [Sajid and Abbas 2016]. This framework aims to support fog activities and support high data traffic levels, and large volumes of data that need processing, through a cognitive fog model.

3 PROPOSED COGNITIVE FOG MODEL

The core concepts of CF and fog federations is elaborated in this section. Before we present the details, i) mostly used notations are given in Table 1, and ii) some key definitions are highlighted below:

Cognitive Fog (CF) model concerns interpreting gathered/received data from IoT devices, via pattern matching in a way that mimics the process of cognition in the human mind [Sheth 2016]. CFs can learn from their past processes according to different situations/scenarios, and improve when performing repeated processes. The Cognitive fog model employs algorithms concerning pattern recognition and data mining to boost performance and achieve better experiences on the repeated processes. In this paper, the context of CF takes the same concepts of cognitive computing which can be define according to DARPA definition of cognitive system as a system that can *"reason, use represented knowledge, learn from experience, accumulate knowledge, explain itself, accept direction, be aware of its own behavior and capabilities as well as respond in a robust manner to surprises"* [Maamar et al. 2018; Sheth 2016]

Fog Federation concerns gathering multiple fog nodes to perform/achieve a specific task in a certain situation or scenario. Fogs become members of a federation because of their capabilities that permit the satisfaction of the needs and requirements of the situation assigned to this federation for handling. Hence, fogs are to be described and discovered for federation and, then, selected for a particular federation according to *planned* and *ad-hoc* federations.

- Planned federation formed at design-time, all its fog participants are already identified and ready to act according to a task's needs and requirements.
- Ad-hoc federation formed at run-time, fogs are joined together according to certain occasions where each fog can empower the federation with various types of processing and controls that enhance performance.

3.1 Fog Architecture

We propose to adopt the general *IoT-Fog* based architecture which has been proposed in our previous work [Al-khafajiy et al. 2018] and in-line with other fog architectures in [Al-khafajiy et al. 2018; Fan and Ansari 2018; Yousefpour et al. 2018] to apply to the provision of smart medical and healthcare systems. Thus, realising the fog based architectures helps obtain a better insight into the real cognition of fog's abilities in such systems. The main layers of such CF architecture are "*Device*" layer, "*Fog*" layer, and "*Cloud*" layer as per Fig. 1.

Device Layer: Also termed the perception layer, this is the starting point where IoT data is generated. This layer contains the interconnected devices (e.g., heart-rate sensors and embedded systems) which feed the fog layer with data. Each device in this layer is facilitated with a communication protocol (e.g., IEEE 802.15.4, WiFi, Blue-tooth, and MQTT) which permits the device to transmit the generated data to the fog nodes over the network.

Fog Layer: It contains a number of decentralised nodes. This layer handles the primary refining, and processing of data generated in the devices layer. Fog nodes aim to improve the efficiency of IoT applications, in terms of the potential to reduce the amount of data transmitted to the cloud layer, and minimizing the request-response time for IoT applications, as well as performing cognitive processes and making decisions that can enhance user experiences.

Cloud Layer: is the top layer of the IoT architecture that enables convenient and appropriate network access for shared resources (e.g., storage and services) over the IoT network. The cloud performs the more processing intensive services of data analysis that the fog layer cannot perform.

3.2 Cognitive Fog Model

For the fog layer to be cognitive so that it can reason about the environment, learn from past occurrences, and adapt to changes, the fog requires components such as pattern recognition that enables the fog network to interpret the network environment. In addition, a cognitive fog requires computation/processing capability for task processing needs, resources for storage needs and communication abilities for networking and interactions. The operations over the CF run or interact with the four connected worlds as per Fig. 2. The data world featuring both row and filtered data, the process world featuring processing models, the fog world featuring the CF processes and controls, and finally the devices word which is controlled by the CF to adapt to the environment. The CF either acts upon device data or directs devices to engage in continuous interactions that should better perform certain tasks, such as directing traffic following congestion or accidents. Each CF has a number of parameters that either permit the CF to participate in decision making processes or just to step-out of the process. Such parameters can influence CF involvement in active processes in the process world due to limited availability (e.g. busy network), security restrictions (e.g., malicious processes), processing or storage limitations, and/or

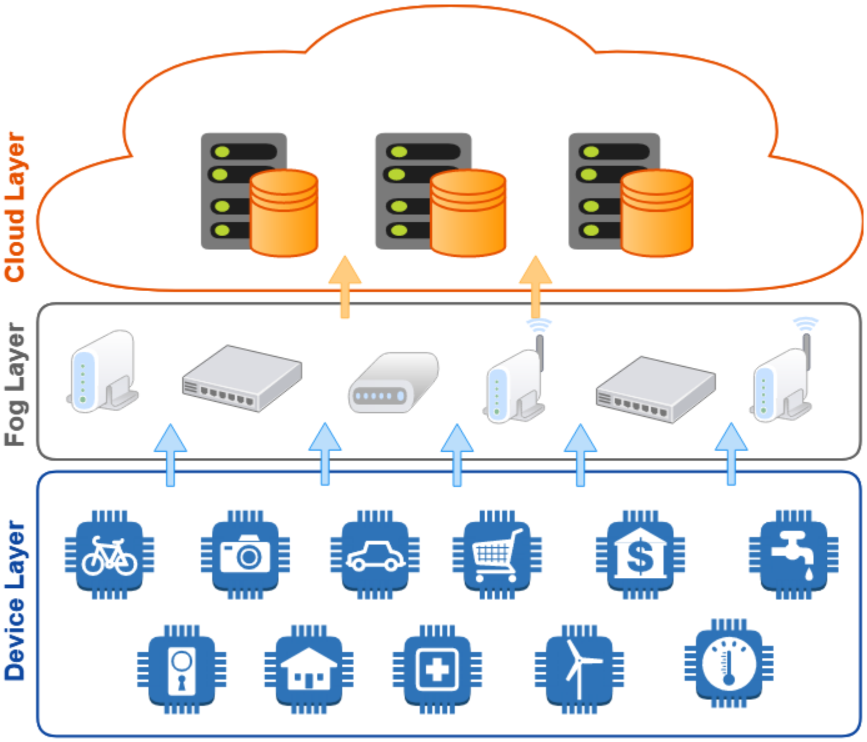

Fig. 1. IoT-Fog Architecture Layers

bandwidth limitations. Therefore, participation considers restrictions that influence CF participation with other CFs process during planned or ad-hoc federations (Section 3.3).

The cognitive model can be defined as a three stage cycle as per Fig. 3. The first stage is the reasoning stage, where all cognition activities take place to assess the surroundings according to the received data from the data and devices worlds. Prior to any decision made by the CF, it will check its parameters in terms of restrictions for any processes and/or participation within the new context that may impact the CF's performance. In the second stage, the CF relies on both the device's data and data in the data world to make decisions and reasoning that could lead to the CF participating in a new process as well as adjusting behaviours, such as executing additional processes (e.g., pattern recognition) to identify certain activities, then ideally, to redirect the connected devices accordingly. In the third stage, the lessons learned during the adaptation and participation of CF will feed into a learning process, such as making new rules/notes for its parameters. All learning outcomes will feed into the reasoning stage that applies to the CF in future interactions.

3.3 Cognitive Fog Federation

To model CF federations, understanding the insight of the CF design is essential. In our proposed model, each CF consists of set of 4-tuple CF = $\{i, t, c, l\}$. Where i refers to CF unique identifier, such as, IP address, t denotes the type of CF (e.g., type of processes or jobs that the CF is capable of). c denotes to the total capability of the CF node, such as fog hardware limitations (e.g., CPU frequency), and finally l denotes to the actual geographical location where the CF is installed. Thus, these CF's tuples are used to define each CF in the network, prior or during any federation. Fig. 4 shows both types of federations (planned and ad-hoc federations). In planned federations ($\mathcal{PF}$), all CFs are known to each other initially (i.e., during design time) and are designed to assist, or take benefit from each other. While in ad-hoc federations, the fog nodes are communicating with each other based on a need, hence they are formed and introduced to each other on the fly to perform a certain task.

Thus, the CF in a particular geographical-area, having the same t (i.e., same type of processes or jobs) and l (e.g., within the same network domain), are designed to communicate with each other to deliver a single task. We can formulate a $\mathcal{PF}$ as:

$$\mathcal{PF} = \{\mathcal{CF}_1^{ts_1}, \mathcal{CF}_2^{ts_2}, ..., \mathcal{CF}_n^{ts_n}\} \quad (1)$$

Where ts refers to the tasks required from CF during the federation. For instance, the roadside of a highway supplied with a set of CFs to perform road monitoring tasks, such as traffic and accidents (known from data provided from devices planted along the way). The CFs are connected to each other at the design time, thus in this scenario, the planned federations occurs when one or more CF

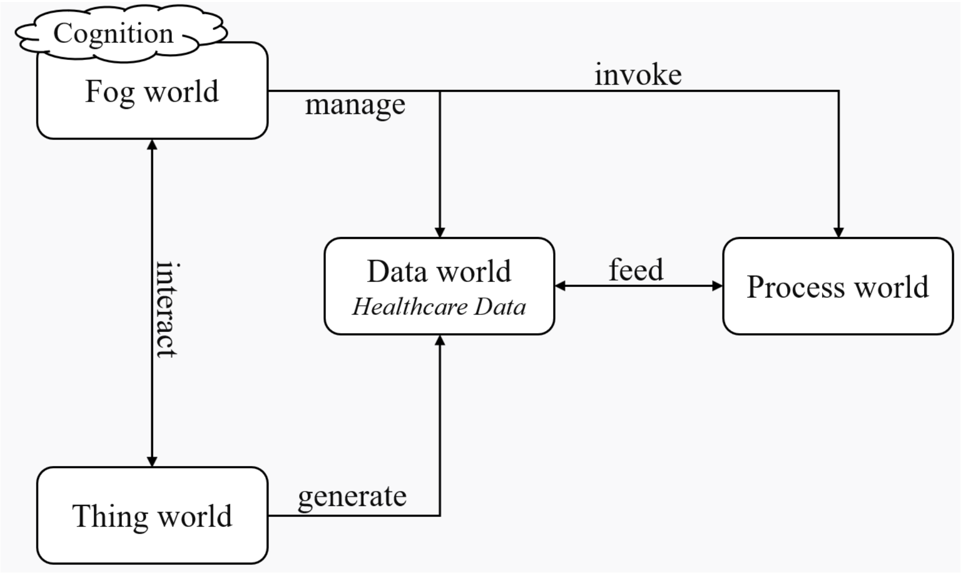

Fig. 2. Interactions of the Cognitive Fogs in healthcare

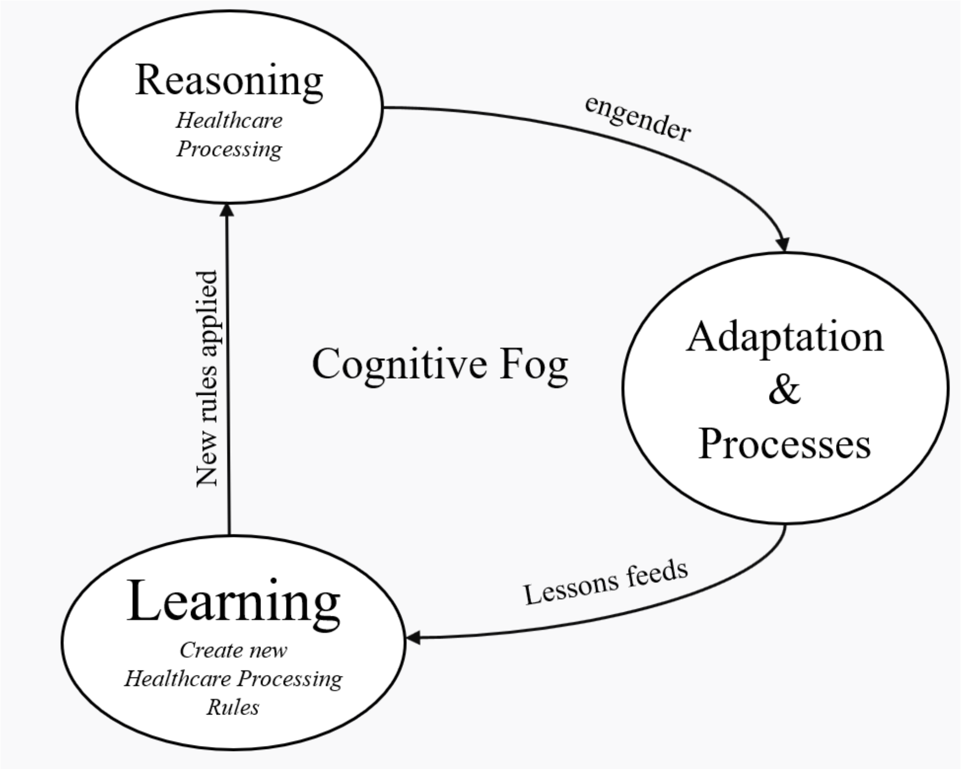

Fig. 3. Cognition of the Cognitive Fog as a 3 stage cycle based on healthcare

failed for whatever reasons, the active CF will federate to cover the failure of the CF. In $\mathcal{PF}$, CF would usually be connected to perform a specific task (e.g., road monitoring) and not multi-tasks.

In contrast, in $\mathcal{AF}$, the CFs are communicating with each other based on a need (i.e., formed on the fly) and usually perform different types of tasks (i.e., multi-tasks are achieved from the federation)

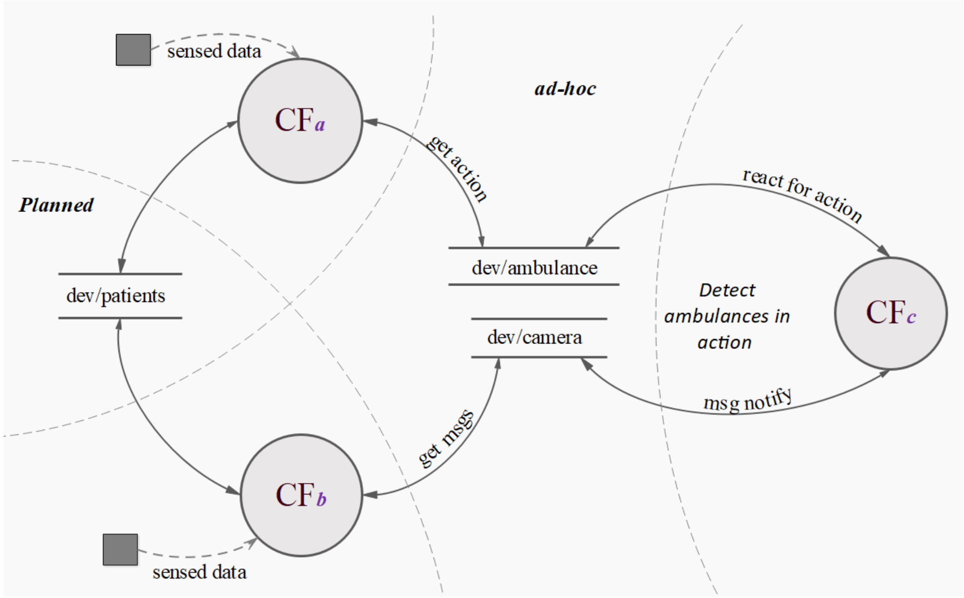

Fig. 4. Planned and Ad-hoc federations

according to a specific situation, for example, multiple CFs can form a federation to detect and react upon a patient's illness. In forming a federation the CFs create a trust environment whereby the computational power and data capabilities of the federation can be optimised for a given set of circumstances. Past occurrences of patient illness and subsequent engendered set of actions can be both utilised and reasoned upon to provide swift diagnoses and best practice responses. Hence, within the $\mathcal{AF}$, multiple CFs could perform one or more tasks according to requirements, therefore, the $\mathcal{AF}$ can be formulated as a 2-dimensional matrix of $\mathcal{CF}$ communications according to tasks, as follow:

$$\mathcal{AF} = \begin{pmatrix} \mathcal{CF}_{1,1}^{ts_1} & \mathcal{CF}_{1,2}^{ts_1} & \cdots & \mathcal{CF}_{1,n}^{ts} \\ \mathcal{CF}_{2,1}^{ts_2} & \mathcal{CF}_{2,2}^{ts_2} & \cdots & \mathcal{CF}_{2,n}^{ts} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{CF}_{x,1}^{ts} & \mathcal{CF}_{x,2}^{ts} & \cdots & \mathcal{CF}_{x,n}^{ts} \end{pmatrix} \tag{2}$$

So that, one row could refer to the multiple CFs collaborations to achieve one task ($\mathcal{CF}_{1,1}^{ts_1}$ & $\mathcal{CF}_{1,2}^{ts_1}$ & $\cdots$ & $\mathcal{CF}_{1,n}^{ts}$), while the total CFs in the federation (i.e., all rows and columns) are achieving multiple tasks.

3.4 Cognitive Fog Security

The CFs, in this area of smart healthcare systems, are proposed to both maintain policy enforcement across the various domains whilst creating, adapting and coordinating consistent security, privacy and data security through the system. The majority of cloud security options are unable to offer reliable data security solutions as they tend to be separately deployed as an overlay. The CF model allows security concerns not only to be embedded in the initial system set-up but also to be adaptable and evolvable. Upon receipt of data, reasoning takes place within the CFs, which maybe to identify malicious use or detect intrusions. At this point the current context of policies, constraints and system norms will be assessed. The CF then reasons, in this context, to reach a decision to, for example, adjust behaviour, identify an intrusion or instruct an associated device. Finally, the CF assimilates the learning into its cognitive model for use in future reasoning's.

4 PROTECTED COGNITIVE FOG

In our proposed ensemble-based intrusion detection model (Section 4.1), the gathered traffic from the CF nodes has been directed to three base classifiers namely K-Nearest Neighbor (K-NN) classifier, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) classifier, and Decision Tree (DT) classifier following a time-slotted method in a round-robin fashion as shown in Figure 5, where f_1, f_2, and f_3 refers to classifiers 1, 2 and 3 respectively, t_1, t_2, ..., t_n refers to the gathered traffic D at time slots 1, 2,..., n, respectively. The results from the three classifiers are integrated provide a conclusion from the ensemble technique.

4.1 Proposed Ensemble Method

In the proposed ensemble technique, we adopt three base classifiers, K-Nearest Neighbor (K-NN) classifier, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) classifier, and Decision Tree (DT) classifier, which together provide analysis of CF collected traffic.

The NSL-KDD dataset [for Cybersecurity 2018] has been used for training and testing the presented techniques. The NSL-KDD features records that have been forwarded to the three base classifiers. The collected decisions from the three classifiers are combined using the combiner to produce the overall conclusion of the ensemble technique as represented in Figure 6. As the combining technique, we adopt the majority voting rule. The majority voting rule is the simplest and most effective voting scheme in this instance; the class with the highest number of votes from the three classifier systems is the outcome.

4.1.1 *K-Nearest Neighbour Classifier.*

K-Nearest Neighbour (K-NN) Classifier is considered an effective classifier for classification purposes. In order to classify an input traffic, an establishment for k nearest training patterns will be done based on the Euclidean distance measurement between the gathered traffic (input traffic) and every training pattern. After that, the traffic is then assigned to the class by using the majority voting technique where the traffic is classified to the frequent class amongst the k nearest training patterns. The adopted K-NN procedure is presented in Algorithm 1 [Tay et al. 2014].

4.1.2 *Density-Based Spatial Clustering of Applications with Noise (DBSCAN).*

The DBSCAN technique has been adopted as one of the proposed ensemble technique base classifiers. DBSCAN is a density-based clustering mechanism where it formulates clusters as dense regions [Otoum et al. 2017a][Ma and Zhang 2004]. It classifies the clusters such as all adjacent clusters undergo under the same class [Jiang et al. 2001]. The adopted DBSCAN technique is shown in Algorithm 2. Where ϵ refers to the neighborhood regions' radius, $MinPts$ refers to the minimum number of gathered points together within that neighborhood and $Data$ refers to the used dataset. The algorithm starts by randomly picking up point from the adopted dataset ($Data$) till all the points have been visited. If at least $MinPts$ within ϵ radius, all these points will be within the same cluster which have been tested and achieved using the $CLUSTERING$ function as shown in Algorithm 2 .To this end, the clusters will be extended by repeating the neighborhood calculation ($Neighborpts$) for all neighboring points.

4.1.3 *Decision Tree classifier.*

Decision Tree (DT) classifier is adopted as a base classifier for our proposed ensemble-based intrusion detection. In DT, each data point can be utilized to make a choice by splitting the data points into reduced groups, made up of nodes that formulate the rooted tree. In DT, any tree contains three nodes categories, namely, the terminal, internal and root nodes [Farid et al. 2010]. Iterative Dichotomiser-3 (ID-3) is the DT which is adopted in this work.

Each node in ID-3 links to a various feature while each arc characterizes the possible value of each attribute. Entropy is utilized in ID-3 in order to calculate the event predictability which represents the amount of data uncertainty while the information gain principles are utilized to regulate the features splits goodness, such as, the feature with the maximum gain is considered as the splitting-purpose feature [Pujari 2001]. Entropy $E(S)$ for a set S is formalized as in Equation 3 [Pujari 2001],

$$E(S) = \sum_{x \varepsilon X} p(x) \log_2 \frac{1}{p(x)} \tag{3}$$

Higher $E(S)$ values specify the high uncertainty while the lower values represent the low ones. Information gains $G(S,T)$ refers to the entropy change after a decision on feature T, where S is a set. $G(S,T)$ is formalized in Equation 4.

$$G(S,T) = E(S) - E(S,T) = E(S) - \sum_{i=0}^{n} p(x)E(T) \tag{4}$$

In Equation 4, $p(x)$ refers to the probability of an event x. As a summary, ID-3 algorithm begins with the root node creation, calculating the $E(S)$, calculating the entropy with respect to feature T $E(S,T)$, choosing the feature T with maximum gain $G(S,T)$, eliminating the feature with the highest G and finally, repeating the same steps for all features [DT2 2018].

4.1.4 *The combiner technique.*

The Majority Voting technique has been adopted as the combiner technique for our proposed ensemble-based intrusion detection model. The majority voting technique forwards the gathered data into the class that has the majority among the output gathered from the base classifiers [Bouziane et al. 2011].

4.2 Cognitive Fog Ensemble for Security

Specifically for anomalous events such as intrusion detection in smart healthcare systems, the ensemble of CFs using these 3 base classifiers and majority voting effectively analyse and provide action for reasoning over the collected network traffic data. To identify anomalous events, a clear baseline for what is considered normal is required. This subjective metric is often based on the network policy that defines which users are allowed to access network resources. In the case of the CFs, this initial policy can be adapted and evolved by the CFs themselves.

5 CASE STUDY AND TESTBED SETUP

5.1 Case Study

The case study concerns improving the operation and security of a healthcare and biomedical systems by the introduction of CFs. In this paper, our case study examines a CF to monitor the health and activities of elderly people in care-home premises, including associated concerns of data security and privacy. Consider an IoT healthcare system, to monitor patients with chronic diseases symptoms data, offered by a healthcare organisation to patients in care-homes. The system supports real-time monitoring of patient activities. It consists of smart healthcare wearable devices (e.g., heart rate and fall sensors), CFs, cloud and a dashboard for the caregivers and doctors to monitor a patients symptoms. The fogs are responsible for obtaining real-time data from wearable devices, ensuring the security and privacy of that data and making primary analyses of the gathered data for healthcare purposes. While the cloud data-center is responsible for data storage and future analysis (non-real-time processes) including the machine learning (ML) segment for data training and analysis activities, both on the data itself and assessing vulnerability to attack, etc., from a data security perspective. The IT division experts install CFs according to the care-home size, with at least two CFs at any given location of their care-home premises. The reason for this is to make sure that a backup fog is always available in case of one CF's failure. Also, in case of one node being busy with processes of a planned or an ad-hoc federation. To handle the patient monitoring, we have focused on monitoring the pulse

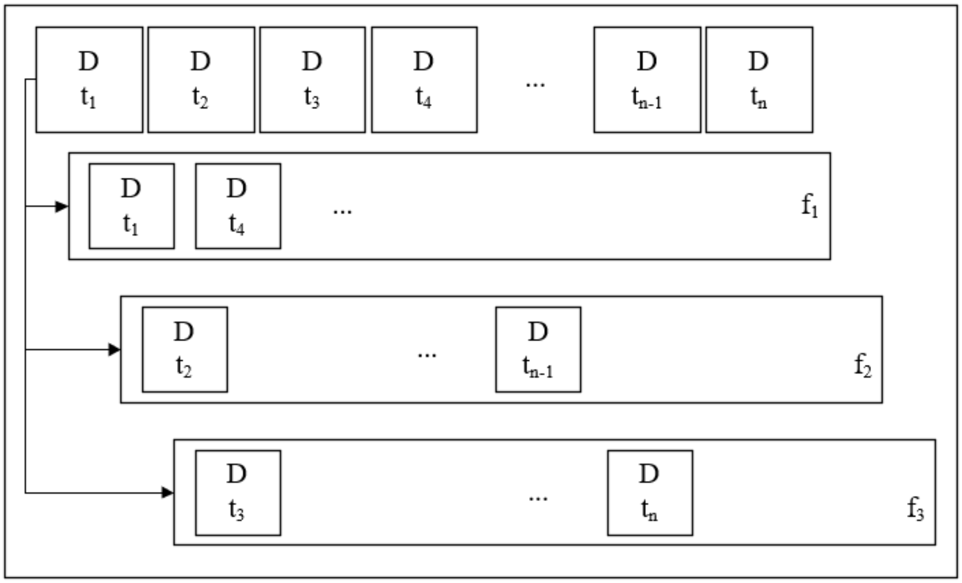

Fig. 5. Gathered traffic distribution mechanism between f_1, f_2 and f_3

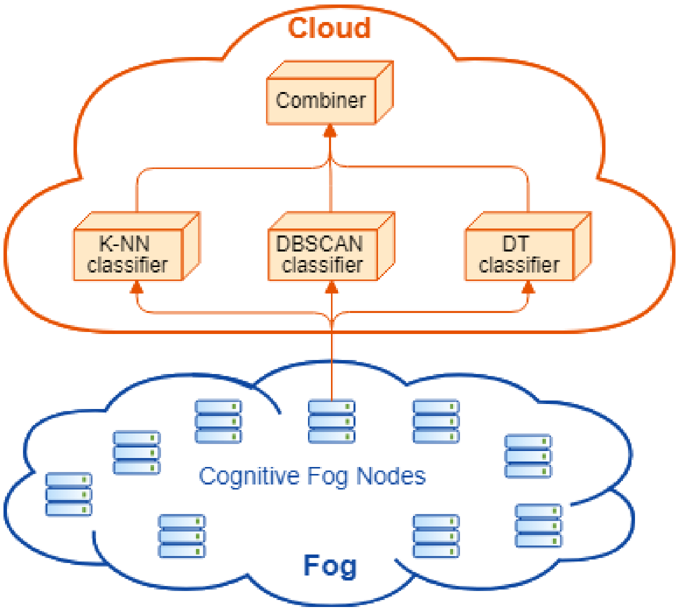

Fig. 6. The proposed Ensemble model

rate (i.e., heart-rate) with either abnormal racing or dropping in the pulse. At all stages the CFs are monitoring and reasoning on the state of the system, providing adaptable and intelligent data security services as well as high functionality. To this end, two possible cases have been considered:

(1) First time pulse rate is racing/dropping: CF will analyse the received data from the pulse sensor to detect/check for any abnormal racing or dropping in the patients pulse. For any suspicious situation, a planned CFs federation is formed, with respect to the rescue CF non-functional requirements, to investigate patient's status and make a decision based on federated CFs experiences with such situations. Once a decision is taken, the caregiver will be notified through the dashboard and every CF will make note for a set of learned lessons which could be used in the future.
(2) Recurrent abnormality detected: on a similar situation, CF will learn (i.e. make note of repeated actions) the conduct taken

Algorithm 1 K-NN pseudo-code

```
procedure KNN-LEARNING ($\mathcal{D}$ , $\mathcal{C}$ , $x$)
Input: $\mathcal{D}$ , $\mathcal{C}$ , $x$
    ▷ $\mathcal{D}$ as training data,    ▷ $\mathcal{C}$ as classes,    ▷ $x$ as data sample
Output: $dis$
Classify ($\mathcal{D}$ , $\mathcal{C}$ , $x$)

    for $i$=1 to $j$ do
Compute $dis(\mathcal{D}_i$,$x)$
     Compute set $N$    ▷ $N$ are the indices for $k$ shortest distances $dis(\mathcal{D}_i$,$x)$
Return the majority label $C_i$ such as $i \in N$
```

Algorithm 2 DBSCAN algorithm

```
procedure DBSCAN ($\epsilon$, $Minpts$, $Data$)
Input:
$\epsilon$, $Minpts$, $Data$, $Neighborpts$
Initialize: $CLUS$ = 0.
    ▷ $CLUS$ = 0 no clusters yet
    ▷ $Data$ is the dataset

    for each un-visited data point $N$ in the dataset $Data$ do
Mark $N$ as visited
$Neighborpts$= Query ($N$, $\epsilon$)

        if sizeof($Neighborpts$) < $Minpts$ then
Mark $N$ as $noise$

        else
$CLUS$= next cluster
CLUSTERING ($N$, $Neighborpts$, $CLUS$ = 0, $\epsilon$, $Minpts$)

        Query($N$, $\epsilon$)
all points in $N$ region (with $N$)

CLUSTERING ($N$, $Neighborpts$, $CLUS$ = 0, $\epsilon$, $Minpts$)
Add $N$ to cluster $CLUS$.

        for each data point $N'$ in $Neighborpts$ do

            if $N'$ is not visited then
Mark $N'$ as visited
$Neighborpts'$= Query ($N'$, $\epsilon$)

                if sizeof($Neighborpts'$) >= $Minpts$ then
$Neighborpts$= $Neighborpts$ with $Neighborpts'$

                    if $N'$ is not in cluster then
Add $N'$ to $CLUS$

```

by caregiver on such situations, so that CF can automate the processes and take the action quicker and on behalf of the caregiver, such as request ambulance and notify the doctor(s) regarding the patient's status. In such scenarios, an ad-hoc CF federation is formed after selecting the necessary CFs (i.e. according to their functional requirements) with respect to their non-functional requirements to run multiple processes.

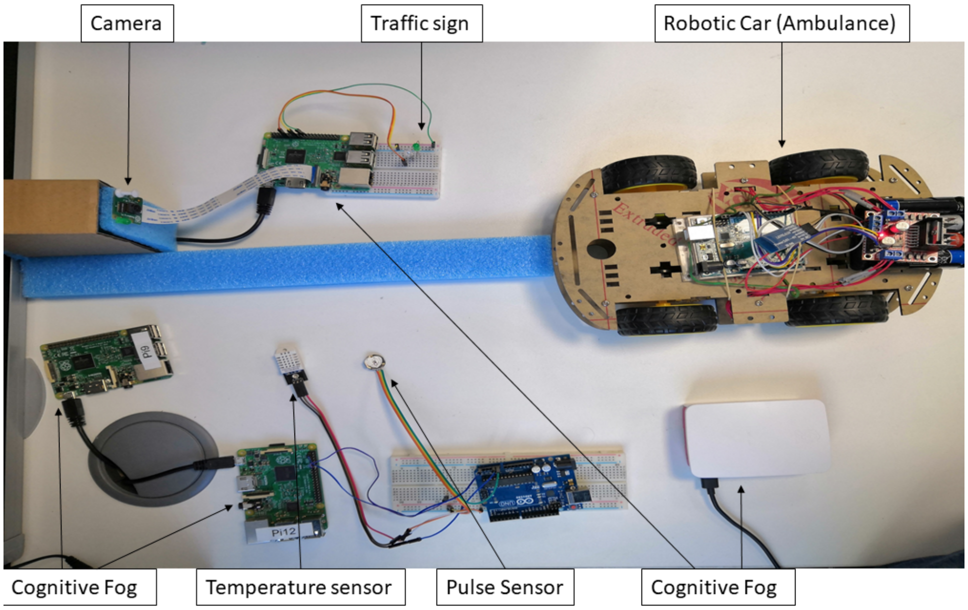

Fig. 7. Cognitive Fog Testbed

> For example, care-home CF communicating with CF of nearest hospital to send an ambulance to care-home address, also communicating with roadside CF to clear the way for the ambulance in advance to avoid traffic delays and congestion. Once the case is over, the ad-hoc CFs federation becomes a planned federation that could be initiated in the future, this should be stored under a task with a similar situation with similar non-functional requirements.

Our proposal is that CF for health monitoring would reason about sensed data, such as pulse abnormality, time detected and the circumstances of pulse racing or dropping as well as, for data security, access control and threat monitoring. Ideally, the CF will be able to act accordingly and form the appropriate federations to solve all issues. As previously stated these federations will be based on the context of functional and non-functional requirements. Functional requirement comprise the services controlled and monitored by the CF. For example, pulse monitors, camera control, sensors reading, etc. The non-functional requirements are the responses to meta-data concerned largely with security and performance of the composed CFs. For example adaptation of reasoning models, monitoring of performance, data integrity, security and privacy, etc.

5.2 Testbed Setup

Our CF testbed shown in Fig. 7. The testbed was assembled using four CFs and three device nodes. We assume that the CFs are located in these locations, in the care-home (CF_1 & CF_2) where patients are based and in the hospital (CF_3) and on-street fog (CF_4) (located on roadside between the care-home and the hospital connected to traffic-light and CCTV thing nodes). CF_1, CF_2 and CF_3 are a Raspberry Pi (RPi) (Quad Core 1.2GHz CPU, 1GB RAM). However, each with different functionality, according to our case-study, CF_1 and CF_2 are connected to a pulse sensor (SEN-11574) to measure heart-rate and temperature-humidity sensor DHT22 (AM2302) sensor, these are used as patient's sensors. While CF_3, is used for hospital processes, such as dispatching ambulance and contacting doctors according to the data received from CF_1 and/or CF_2. Finally, CF_4 is composed of a Lenovo Ideapad laptop (i5 1.8GHz CPU, 8GB RAM) connected to the Internet over Ethernet cable and fitted with an HD Lenovo EasyCamera Webcam (as a device/thing node) with 0.92MP. In addition, CF_4 is connected to a traffic light node (as a device/thing node) which has 2 LED diodes (Green and Red) wired through the breadboard to the RPi.

The interactions between CFs themselves, and CFs with IoT things are over the publish/subscript protocol, that is Message Queuing Telemetry Transport (MQTT) protocol. Thus, via the subscribed topic, which is a UTF-8469 string that MQTT broker uses to decide on which client receive which message, the subscribers of a specific topic will receive useful data in real-time. For example, the traffic light receives signals through the *"CF/traffic"* topic, upon which it changes to green or red.

During the ad-hoc federation, a CF will be responsible to communicate with the camera device and the traffic device to clear the way for an ambulance travelling from/to the hospital. Therefore, to detect ambulances, we developed an in-house Python image recognition program that processes RGB477 images using an Open Source Computer Vision (OpenCV) library. Upon ambulance detection by the CF, according to the live frames from the camera, it will send an alert to the traffic-light, to stop or redirecting the traffic, over the MQTT protocol via *"CF/traffic"* topic to set the traffic-light sign.

5.3 Experiment and Evaluation

In our experiment, we employ the four CFs as follow: CF_1 & CF_2 is for interacting with patients things (i.e., the pulse and temperature sensors) as well as interpreting the sensed data, CF_3 is for alerting the hospital's A&E about patient's situation, instruction for ambulance driver to go to patient's address (supplied from CF_1 or CF_2) and CF_4 is for interacting and controlling devices/things that are planted on the roadside (i.e., camera and traffic-light). The camera is for broadcasting live images from the way to the care-home, traffic-light is regulating the access of ambulance. For evaluation needs, two simulation scenarios were carried as follow:

Scenario $_1$: we considered a $\mathcal{PF}$ of two CFs, namely CF_1 and CF_2, upon needs after detecting abnormality in patient's pulse, thus experiences of multi CFs is required to make a decision. The $\mathcal{PF}$ evaluated in term of time-delay and efficiency in forming such federation, therefore, we measured the total time required to form $\mathcal{PF}$ between CF_1 and CF_2 when the pulse sensor provide a reading that looks abnormal (i.e., 60Bpm $\geqslant$ pulse $\geqslant$ 100Bpm as in [Al-khafajiy et al. 2019]). CF_1 interpret the sensed data from both pulse and temperature sensors to reason the measured data, thereafter, upon suspected values or abnormality, CF_1 will seek an assist from CF_2, forming a $\mathcal{PF}$ to make a decision for either alerting the caregiver or not. During the same execution life-cycle, we change the payload of sensed data and experience different set of data across number of iterations which has been grouped into 50, 100 and 150 iteration. The objective was to observe how the test-bed behaves with respect to the number of detected abnormality and the time taken to make a decision including the time required to exchange number of messages between both CFs. Fig. 8 reports the performance results of the $\mathcal{PF}$ within the three iterations. It worth noting that the aborted federation in 8 is due to some non-functional requirements.

Scenario $_2$: we expanded Scenario $_1$: to include all four CFs, namely CF_1, CF_2, CF_3 and CF_4. In this scenario both $\mathcal{PF}$ and $\mathcal{AF}$ are formed according to following: i) CF_1 detect an abnormality, in patient's pulse, and through a $\mathcal{PF}$ with CF_2 makes decision for requesting ambulance. ii) CF_1 will search for nearest hospital and communicate with its CF, in this case CF_3, and from an $\mathcal{AF}$. To this end, CF_3 will inform the doctor and send out an ambulance to the patient. iii) CF_3 will also from an $\mathcal{AF}$ with CF_4 to clear the path for the ambulance, upon detecting the ambulance via the camera thing, through controlling the traffic-light signs. The $\mathcal{AF}$ evaluated in term of time-delay and efficiency in forming the federations, thus, we measured the total time required to form an $\mathcal{AF}$ among all CFs. Fig. 9 reports the performance results of the $\mathcal{AF}$ within three iterations (50, 100 and 150 iteration). It worth noting that the time-delay (in millisecond) for $\mathcal{AF}$ is higher due to the multi-tasks required from the federation, also, the aborted federation is due to some non-functional requirements. Within this scenario, we checked how the test-bed behaves when $\mathcal{PF}$ of things (i.e., pulse sensor) are merged with an $\mathcal{AF}$ federation to evaluate the execution/process time required to perform a collaboration. Fig. 10 illustrates the results showing cases of execution time related to $\mathcal{PF}$ versus $\mathcal{AF}$ federations; it took between 85ms to 90ms to execute an $\mathcal{AF}$ federation and between 18ms to 22ms to execute $\mathcal{PF}$ federation.

5.3.1 *NSL-KDD dataset.*

NSL-KDD dataset has been used in our simulations for training and testing our proposed security model. KDD which refers to the Knowledge Discovery in Data mining is a sub-set from KDDCup99 [for Cybersecurity 2018]. We utilized NSL-KDD since it is an improved version of KDDCup99 that has been introduced to manage the KDDCup'99 problems such as the duplicate records and the huge number of records. In NSL-KDD dataset, each record consists of 41 features and attack types are categorized into 4 types: Remote to Local (R2L), Denial of Service (DoS), User to Root (U2R), and Probe.

5.3.2 *Results analysis.*

Training and testing the ensemble-based intrusion detection model has been achieved in each trial. In which each run is done for 10 trials whereas the average of the 10 trials is considered for each run. The three base classifiers (K-NN, DBSCAN and DT) are built using the KDDTrain+ dataset as the training dataset for the training phase. While KDDTest+ as the testing dataset has been used for the attacks detection and classification purposes. The models' performance evaluation is achieved using Accuracy Rate (AR), Detection Rate (DR) and False Positive Rate (FPR) which are considered as the standard intrusions detection measurements. AR refers to the ratio of truly classified behaviors, DR refers to the behaviors that accurately classified as abnormal behaviors while FPR refers to the non-malicious behaviors that inaccurately classified as malicious.

Figure 11 represents AR and DR comparison between the three base classifiers (K-NN, DBSCAN and DT) and the proposed ensemble classifier. It is clear that the ensemble-based classifier performs over the other three classifiers. False Positive Rate (FPR) has been tested for the three base classifiers and the ensemble-based classifier as presented in Figure 12. The ensemble-based classifier performs with the least *FPR* followed by the K-NN classifier.

Detection rates for NSL-KDD attacks and normal activities has been registered as shown in Figure 13.

Receiver Operating Characteristics (ROC) curve that represents the sensitivity (True Positives Rate (TPR)) versus (1-specificity) (False Positives Rate (FPR)) [Otoum et al. 2019] ratio has been traced for the proposed technique and compared with the individual classifiers ones. In Figure 14, the area under the curve considers as the sensitivity-specificity ratio where the larger area reflects the best performance. A ROC-based comparison between the base classifiers (K-NN), DBSCAN and DT along with the ensemble-based classifier to assess the system performance is shown in Figure 14. It is clear that the ensemble-based solution performs better with the largest area under the curve followed by the individual classifiers (K-NN), DBSCAN and DT.

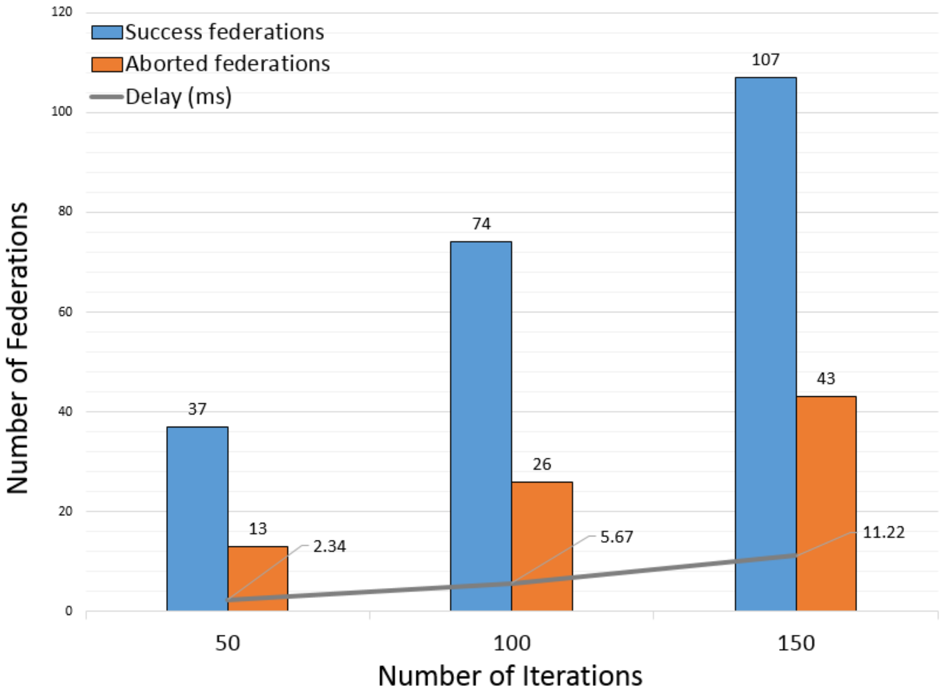

Fig. 8. The Execution time for Planned Federations

6 CONCLUSION AND FUTURE WORK

Due to the fog computing location within the network, fog nodes can have beneficial knowledge and constructive control over both the sensor network and the data transmitted over the Internet. In this paper, we utilise the strategic position of fog nodes to develop a cognitive fog for smart healthcare systems interacting with other IoT systems that is able to make decisions acting upon received data in a secure fog environment. Empowering the fog model with reasoning, adaptation, and learning capabilities allows the fog to be pro-active, i.e. to reason, learn, and adapt to different scenarios. Our CF test-bed was assembled using four CFs and three device nodes based on a healthcare scenario. The gathered traffic on the CF nodes was directed to three base classifiers namely K-NN, DB-SCAN, and DT classifiers. In future work, we will further analyze the self-learning and adaptation process. Fog nodes could be exposed to unknown situations that require new courses of action. Although intelligence in fog computing is still in its infancy, it has great potential for achieving beneficial and sustainable computing ecosystems. Further, future work involves scaling up the Case Study, described in this paper, and moving the implementation out of the testbed and into real world application. Furthermore, additional CFs are needed to be considered for realistic scenarios: A wider range of devices, both with a specialist healthcare function and with a real-world application, interacting in a smart healthcare network, are required.

REFERENCES

2018. Decision Trees for Classification: A Machine Learning Algorithm. https://www.xoriant.com/blog/product-engineering/

Karim Abouelmehdi, Abderrahim Beni Hssane, and Hayat Khaloufi. 2018. Big healthcare data: preserving security and privacy. *J. Big Data* 5 (2018), 1. https://doi.org/10.1186/s40537-017-0110-7

Abdulla Amin Aburomman and Mamun Bin Ibne Reaz. 2016. A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Applied Soft Computing* 38 (2016), 360 – 372. https://doi.org/10.1016/j.asoc.2015.10.011

Swati Agarwal, Shashank Yadav, and Arun kumar Yadav. 2016. An Efficient Architecture and Algorithm for Resource Provisioning in Fog Computing.

John Patrick C Agustin, Joshua H Jacinto, Wilbert Jethro R Limjoco, and Jhoanna Rhodette I Pedrasa. 2017. IPv6 routing protocol for low-power and lossy networks implementation in network simulator—3. In *TENCON 2017-2017 IEEE Region 10 Conference*. IEEE, 3129–3134.

Mohammed Al-khafajiy, Thar Baker, Carl Chalmers, Muhammad Asim, Hoshang Kolivand, Muhammad Fahim, and Atif Waraich. 2019. Remote health monitoring of elderly through wearable sensors. *Multimedia Tools and Applications* (24 Jan 2019). https://doi.org/10.1007/s11042-018-7134-7

M. Al-khafajiy, T. Baker, A. Waraich, D. Al-Jumeily, and A. Hussain. 2018. IoT-Fog Optimal Workload via Fog Offloading. In *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*. 359–364. https://doi.org/10.1109/UCC-Companion.2018.00081

Mohammed Al-khafajiy, Lee Webster, Thar Baker, and Atif Waraich. 2018. Towards fog driven IoT healthcare: challenges and framework of fog computing in healthcare. In *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems*. ACM, 9.

Ismaeel Al Ridhawi, Moayad Aloqaily, and Azzedine Boukerche. 2019. Comparing Fog Solutions for Energy Efficiency in Wireless Networks: Challenges and Opportunities. *IEEE Wireless Communications* 26, 6 (2019), 80–86.

Fadi Al-Turjman. 2017. Cognitive caching for the future sensors in fog networking. *Pervasive and Mobile Computing* 42 (2017), 317–334.

Bahast Ali and Xiaochun Cheng. 2019. Security Solution Based on Raspberry PI and IoT. In *Cyberspace Safety and Security*, Jaideep Vaidya, Xiao Zhang, and Jin Li (Eds.). Springer International Publishing, Cham, 162–171.

M. Z. Alom and T. M. Taha. 2017. Network intrusion detection for cyber security using unsupervised deep learning approaches. In *2017 IEEE National Aerospace and Electronics Conference (NAECON)*. 63–69. https://doi.org/10.1109/NAECON.2017.

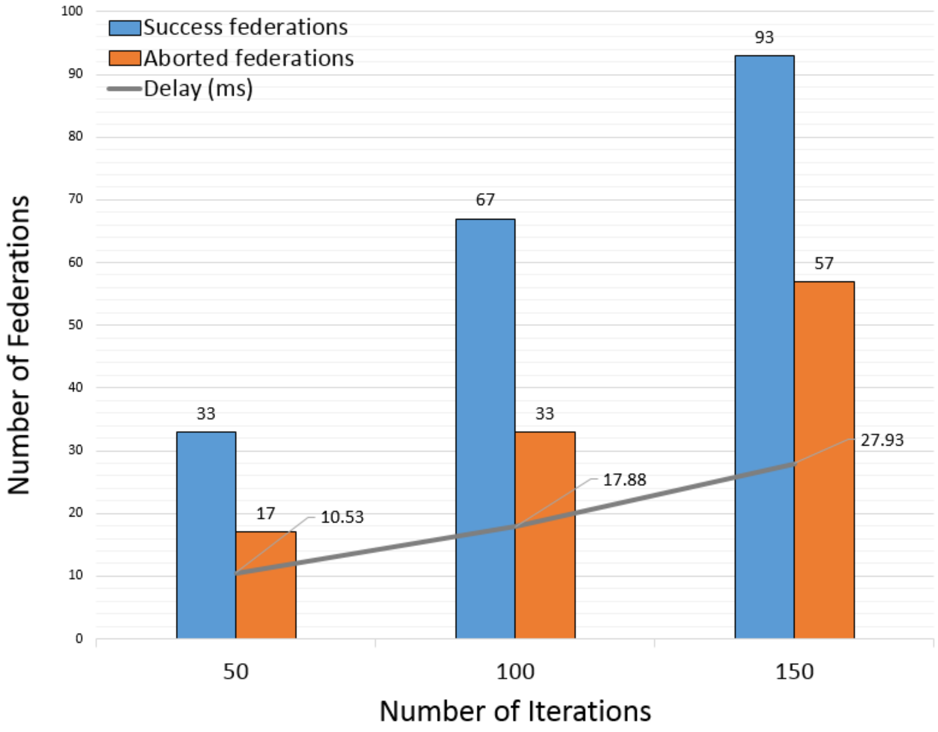

Fig. 9. The Execution time for Ad-hoc Federations

8268746

A. Alrawais, A. Alhothaily, C. Hu, and X. Cheng. 2017. Fog computing for the internet of things: Security and privacy issues. *IEEE Internet ComputingMar* 1, 21 (2017), 2.

Venkatraman Balasubramanian, Safa Otoum, Moayad Aloqaily, Ismaeel Al Ridhawi, and Yaser Jararweh. 2020. Low-latency vehicular edge: A vehicular infrastructure model for 5G. *Simulation Modelling Practice and Theory* 98 (2020), 101968.

Venkatraman Balasubramanian, Faisal Zaman, Moayad Aloqaily, Ismaeel Al Ridhawi, Yaser Jararweh, and Haythem Bany Salameh. 2019a. A mobility management architecture for seamless delivery of 5G-IoT services. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 1–7.

Venkatraman Balasubramanian, Faisal Zaman, Moayad Aloqaily, Saed Alrabaee, Maria Gorlatova, and Martin Reisslein. 2019b. Reinforcing the edge: autonomous energy management for mobile device clouds. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 44–49.

M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita. 2014. Network Anomaly Detection: Methods, Systems and Tools. *IEEE Communications Surveys Tutorials* 16, 1 (First 2014), 303–336. https://doi.org/10.1109/SURV.2013.052213.00046

Hafida Bouziane, Belhadri Messabih, and Abdallah Chouarfia. 2011. Profiles and Majority Voting-Based Ensemble Method for Protein Secondary Structure Prediction. *Evolutionary bioinformatics online* 7 (10 2011), 171–89. https://doi.org/10.4137/EBO.S7931

Ahmed M Elmisery, Seungmin Rho, and Dmitri Botvich. 2016. A fog based middleware for automated compliance with OECD privacy principles in internet of healthcare things. *IEEE Access* 4 (2016), 8418–8441.

Q. Fan and N. Ansari. 2018. Towards Workload Balancing in Fog Computing Empowered IoT. *IEEE Transactions on Network Science and Engineering* (2018), 1–1. https://doi.org/10.1109/TNSE.2018.2852762

Dewan Md. Farid, Nouria Harbi, and Mohammad Zahidur Rahman. 2010. Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection. *CoRR* abs/1005.4496 (2010). arXiv:1005.4496 http://arxiv.org/abs/1005.4496

Canadian Institute for Cybersecurity. 2018. *NSL-KDD dataset.* http://www.unb.ca/cic/datasets/nsl.html

L. Galluccio, S. Milardo, G. Morabito, and Palazzo S. Sdn wise: Design. 2015. Prototyping and Experimentation of a Stateful SDN Solution for Wireless Sensor networks. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 513–521.

T. Govindarajan. 2016. Evaluation of Ensemble Classifiers for Intrusion Detection.

Harriet Green. 2015. The Internet of Things in the Cognitive Era: Realizing the Future and Full Potential of Connected Devices. *ed: IBM Watson IoT* (2015).

Andreas Heil, Mirko Knoll, and Torben Weis. 2007. The internet of things-context-based device federations. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 58–58.

Martin Henze, René Hummen, Roman Matzutt, and Klaus Wehrle. 2014. A trust point-based security architecture for sensor data in the cloud. *In Trusted Cloud Computing* (2014), 77–106.

Kirak Hong, David Lillethun, Umakishore Ramachandran, Beate Ottenwälder, and Boris Koldehofe. 2013. Mobile Fog: A Programming Model for Large-scale Applications on the Internet of Things. In *Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing* (Hong Kong, China) *(MCC '13)*. ACM, New York, NY, USA, 15–20. https://doi.org/10.1145/2491266.2491270

Kai Hwang, Sameer Kulkareni, and Yue Hu. 2009. Cloud security with virtualized defense and reputation-based trust mangement. In *2009 Eighth IEEE International Conference on Dependable.* Autonomic and Secure Computing, IEEE, 717–722.

Fatemeh Jalali, Olivia J Smith, Timothy Lynar, and Frank Suits. 2017. Cognitive IoT gateways: automatic task sharing and switching between cloud and edge/fog computing. In *Proceedings of the SIGCOMM Posters and Demos.* ACM, 121–123.

M.f Jiang, S.s Tseng, and C.m Su. 2001. Two-phase clustering process for outliers detection. *Pattern Recognition Letters* 22/6-7 (2001), 691–700. https://doi.org/10.1016/s0167-8655(00)00131-8

A. Kapsalis, P. Kasnesis, I. S. Venieris, D. I. Kaklamani, and C. Z. Patrikakis. 2017. A Cooperative Fog Approach for Effective Workload Balancing. *IEEE Cloud Computing* 4, 2 (March 2017), 36–45. https://doi.org/10.1109/MCC.2017.25

Daoying Ma and Aidong Zhang. 2004. An Adaptive Density-Based Clustering Algorithm for Spatial Database with Noise. *IEEE Intl Conf on Data Mining (ICDM'04)* (2004). https://doi.org/10.1109/icdm.2004.10036

Zakaria Maamar, Thar Baker, Noura Faci, Emir Ugljanin, Yacine Atif, Mohammed Al-Khafajiy, and Mohamed Sellami. 2018. Cognitive computing meets the internet of things. In *13th International Conference on Software Technologies.* SciTePress.

Wided Mathlouthi and Narjes Bellamine Ben Saoud. 2017. Flexible composition of system of systems on cloud federation. In *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud).* IEEE, 358–365.

T M.Govindarajan. 2016. Evaluation of Ensemble Classifiers for Intrusion Detection.

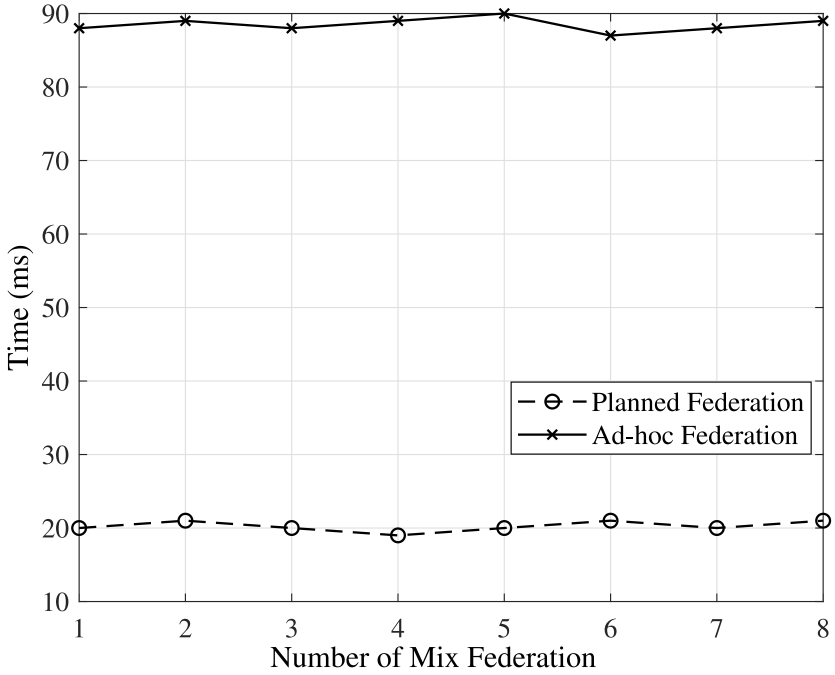

Fig. 10. Execution time related to $\mathcal{PF}$ versus $\mathcal{AF}$ federations

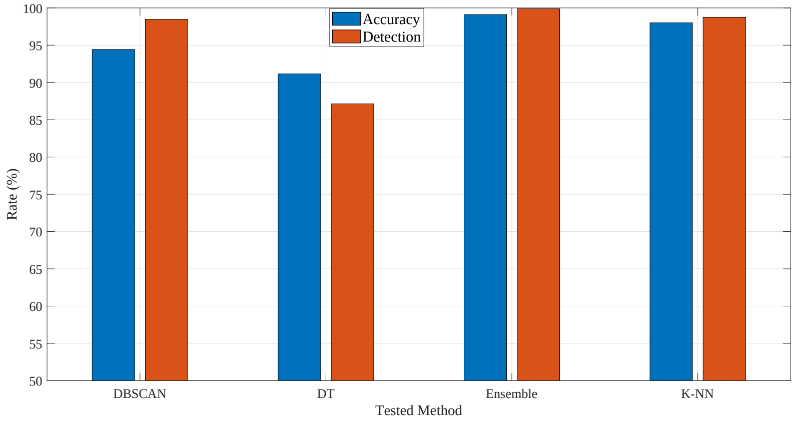

Fig. 11. AR and DR comparison between the base classifiers (K-NN), DBSCAN and DT along with the ensemble-based classifier

N. Moustafa, B. Turnbull, and K. R. Choo. 2019. An Ensemble Intrusion Detection Technique based on proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things. *IEEE Internet of Things Journal* (2019), 1–1. https://doi.org/10.1109/JIOT.2018.2871719

S. Otoum, B. Kantarci, and H. T. Mouftah. 2017a. Mitigating False Negative intruder decisions in WSN-based Smart Grid monitoring. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. 153–158. https://doi.org/10.1109/IWCMC.2017.7986278

S. Otoum, B. Kantarci, and H. T. Mouftah. 2019. On the Feasibility of Deep Learning in Sensor Network Intrusion Detection. *IEEE Networking Letters* 1, 2 (June 2019), 68–71. https://doi.org/10.1109/LNET.2019.2901792

Safa Otoum, Burak Kantraci, and Hussein T. Mouftah. 2017b. Hierarchical Trust-based Black-Hole Detection in WSN-based Smart Grid Monitoring. *IEEE International Conference on Communications(ICC'17)* (2017). https://doi.org/icc.2017

Beate Ottenwälder, Boris Koldehofe, Kurt Rothermel, and Umakishore Ramachandran. 2013. MigCEP: Operator Migration for Mobility Driven Distributed Complex Event Processing. In *Proceedings of the 7th ACM International Conference on Distributed Event-based Systems* (Arlington, Texas, USA) *(DEBS '13)*. ACM, New York, NY, USA, 183–194. https://doi.org/10.1145/2488222.2488265

A.K. Pujari. 2001. *Data Mining Techniques*. Universities Press. https://books.google.ca/books?id=dH2KQhJboSYC

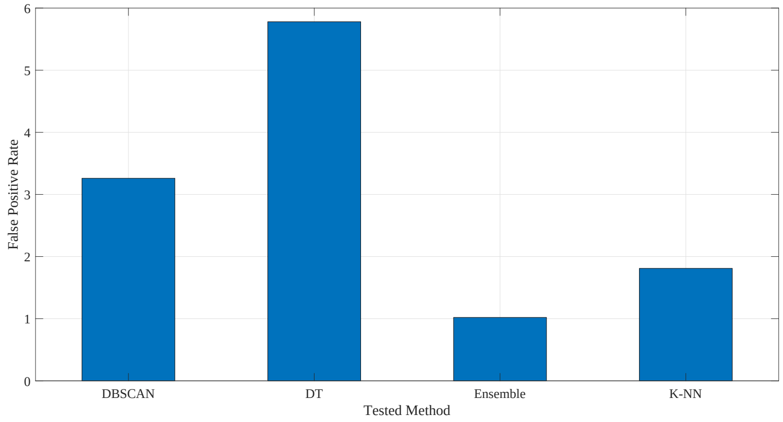

Fig. 12. FPR comparison between the base classifiers (K-NN), DBSCAN and DT along with the ensemble-based classifier

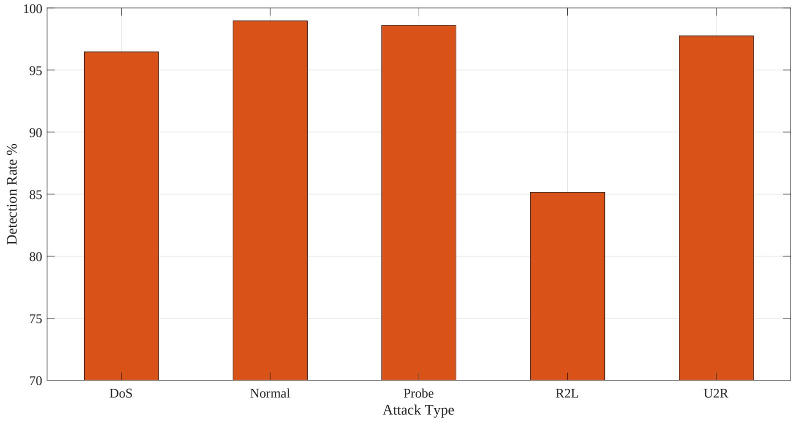

Fig. 13. DR% comparison between the detected attacks and the normal behaviors

Amir-Mohammad Rahmani, Nanda Kumar Thanigaivelan, Tuan Nguyen Gia, Jose Granados, Behailu Negash, Pasi Liljeberg, and Hannu Tenhunen. 2015. Smart e-health gateway: Bringing intelligence to internet-of-things based ubiquitous healthcare systems. In *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*. IEEE, 826–834.

Anam Sajid and Haider Abbas. 2016. Data Privacy in Cloud-assisted Healthcare Systems: State of the Art and Future Challenges. *J. Medical Systems* 40, 6 (2016), 155:1–155:16. https://doi.org/10.1007/s10916-016-0509-2

Amit Sheth. 2016. Internet of things to smart iot through semantic, cognitive, and perceptual computing. *IEEE Intelligent Systems* 31, 2 (2016), 108–112.

Fan Shi, Zhenlei Chen, and Xiaochun Cheng. 2019. Behavior Modeling and Individual Recognition of Sonar Transmitter for Secure Communication in UASNs. *IEEE Access* (2019).

S. Soheily-Khah, P. Marteau, and N. Béchet. 2018. Intrusion Detection in Network Systems Through Hybrid Supervised and Unsupervised Machine Learning Process: A Case Study on the ISCX Dataset. In *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. 219–226. https://doi.org/10.1109/ICDIS.2018.00043

Seyed Ahmad Soleymani, Abdul Hanan Abdullah, Mahdi Zareei, Mohammad Hossein Anisi, Cesar Vargas-Rosales, Muhammad Khurram Khan, and Shidrokh Goudarzi. 2017. A secure trust model based on fuzzy logic in vehicular ad hoc networks with fog computing. *IEEE Access* 5 (2017), 15619–15629.

Bunheang Tay, Jung Keun Hyun, and Sejong Oh. 2014. A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images. *Computational and mathematical methods in medicine* 2014 (01 2014), 276589. https://doi.org/10.1155/2014/276589

Tian Wang, Guangxue Zhang, M. D. Zakirul Alam Bhuiyan, Anfeng Liu, Weijia Jia, and Mande Xie. 2018a. *A novel trust mechanism based on fog computing in sensor–cloud system.* Future Generation Computer Systems.

T. Wang, G. Zhang, A. Liu, M. Z. A. Bhuiyan, and Q. Jin. 2018b. *A secure IoT service architecture with an efficient balance dynamics based on cloud and edge computing.* IEEE Internet of Things Journal.

Qihui Wu, Guoru Ding, Yuhua Xu, Shuo Feng, Zhiyong Du, Jinlong Wang, and Keping Long. 2014. Cognitive internet of things: a new paradigm beyond connection. *IEEE Internet of Things Journal* 1, 2 (2014), 129–143.

A. Yousefpour, G. Ishigaki, R. Gour, and J. P. Jue. 2018. On Reducing IoT Service Delay via Fog Offloading. *IEEE Internet of Things Journal* 5, 2 (April 2018), 998–1010. https://doi.org/10.1109/JIOT.2017.2788802

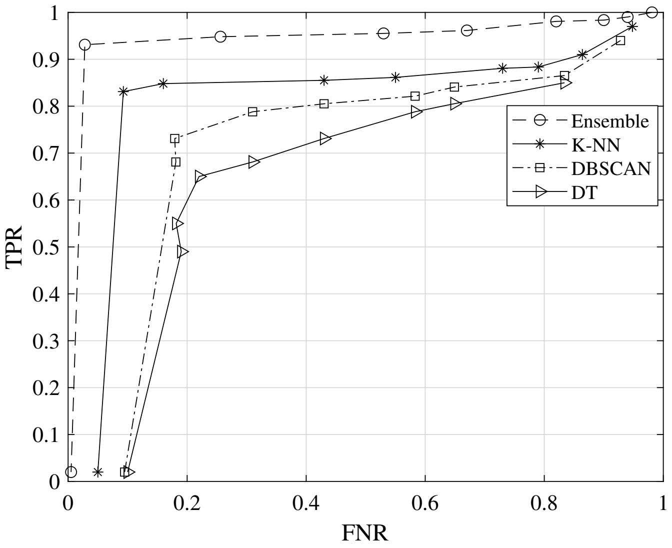

Fig. 14. ROC comparison between the proposed ensemble technique and the individual classifiers