This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

Hybrid Deep Learning-based Anomaly Detection Scheme for Suspicious Flow Detection in SDN: A Social Multimedia Perspective

Sahil Garg, *Member, IEEE,* Kuljeet Kaur, *Member, IEEE,* Neeraj Kumar, *Senior Member, IEEE*, and Joel J. P. C. Rodrigues, *Senior Member, IEEE*

***Abstract*—The continuous development and usage of multimedia-based applications and services have contributed to the exponential growth of social multimedia traffic. In this context, secure transmission of data plays a critical role in realizing all the key requirements of social multimedia networks such as reliability, scalability, Quality of Information (QoI) and Quality of Service (QoS). Thus, a trust-based paradigm for multimedia analytics is highly desired to meet the increasing user requirements and deliver more timely and actionable insights. In this regard, Software Defined Networks play a vital role; however, several factors such as as-runtime security, and energy-aware networking limit its capabilities to facilitate efficient network control and management. Thus, with the view to enhance the reliability of SDN, a hybrid deep learning based anomaly detection scheme for suspicious flow detection in the context of social multimedia is proposed. It comprises of two modules: (1) An anomaly detection module which leverages improved Restricted Boltzmann Machine and Gradient Descent-based Support Vector Machine to detect the abnormal activities, and (2) End-to-end data delivery module to satisfy strict QoS requirements of SDN, *i.e.*, high bandwidth, and low latency. Finally, the proposed scheme has been experimentally evaluated on both real-time and benchmark datasets to prove its effectiveness and efficiency in terms of anomaly detection and data delivery essential for social multimedia. Further, a large-scale analysis over CMU-based insider threat dataset has been conducted to identify its performance in terms of detecting malicious events such as-Identity theft, profile cloning, confidential data collection, etc.**

***Index Terms*—Anomaly detection, Deep learning, Flow routing, Software defined networks, and Social multimedia.**

I. Introduction

THE Web expansion from Web of Things to Web of Thoughts has made social networking more popular. It is the largest, richest and most dynamic evidence base of human behavior that brings new opportunities for understanding individuals, groups and societies. Recent insights into the world of social media suggest that currently there are around 4 billion Internet users worldwide, out of which more than 3 billion are active social-media users [1]. Due to this proliferation of social networks, multimedia content is growing at an unprecedented rate [2]. However, the content-driven and object-oriented nature of social multimedia pose difficulties in gaining meaningful insights from this vibrant and ever-growing pool of data. Moreover, the substantial multimedia content offered on social networks includes sensitive and private information about users and their interactions. Such an abundance of readily available personal information makes it highly vulnerable to threats which often result in information and identity theft. Therefore, interoperability and security are the two biggest challenges in the underlying architecture [3]. Hence, a scalable and pervasive communication paradigm is required for data analytics and management of social multimedia data while maintaining an adequate level of security.

In recent years, Cybersecurity researchers have designed many anomaly detection models to protect the network against attacks perpetrated by malicious users against different multimedia applications such as remote video-on-demand, video conferencing, real-time content delivery, online gaming, etc. In this direction, deep learning architectures such as Convolution Neural Network (CNN), Deep Belief Network (DBN), Restricted Boltzmann Machine (RBM), Stacked AutoEncoders, Recurrent Neural Networks (RNN), etc., are widely used. For example, Chu *et al.* [4] devised an abnormal event detection scheme for videos where they used 3-dimensional CNN to extract the spatiotemporal information of the inputs. Xu *et al.* [5] proposed a method for detection of unusual events in videos via Stacked Sparse Coding and intra-frame classification strategies based on the probabilistic outputs of SVM. Sabokrou *et al.* [6] used a fully CNN to detect anomalies in crowded activities.

Similarly, Ribeiro *et al.* [7] presented an anomaly detection approach using Convolutional Autoencoder (CAE) where aggregation of high-level features was done with the input frames to analyze their effect on the performance of CAE. Xu *et al.* [8] proposed the Appearance and Motion DeepNet model for anomaly detection in videos using deep neural networks and multiple one-class SVM models. Similarly, Feng *et al.* [9] devised a deep learning based model for abnormal event detection where PCANet was used for feature learning, and a deep Gaussian mixture model was proposed to explore the video event patterns. Sun *et al.* [10] propounded a hybrid neural network model for abnormal emotion detection on social media by integrating CNN and Long-Short Term Memory

S. Garg and K. Kaur are with the Department of Electrical Engineering, École de technologie supérieure, Université du Québec, Montréal, QC H3C 1K3, Canada. (e-mail: garg.sahil1990@gmail.com and kuljeet0389@gmail.com).

N. Kumar is with the Department of Computer Science & Engineering, Thapar Institute of Engineering & Technology (Deemed to be University), Patiala (Punjab), India (E-mail: neeraj.kumar@thapar.edu).

J. J. P.C. Rodrigues is with National Institute of Telecommunications (Inatel), Brazil; Instituto de Telecomunicações, Portugal; and University of Fortaleza (UNIFOR), Brazil. (e-mail: joeljr@ieee.org).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

scheme. Since these methods are based on end-to-end training and representation learning, they are widely used in pattern recognition as compared to the traditional machine learning approaches. Although, training of deep learning methods is computationally expensive and requires a massive amount of data, the combination of these approaches with reinforcement learning could be potentially useful [11].

Further, to handle the ever-increasing demands of social users, multimedia analytics requires high processing, continuous data acquisition, huge bandwidth, and computationally less complex encoding techniques [12]. Thus, it has become a trend to incorporate next-generation networking technologies for providing a high quality of experience (QoE) to its intended users. However, due to the complexity and openness of social networks, it becomes essential to filter the inherent security and privacy concerns before deploying new technologies because it results in high capital and maintenance costs. As a result, the communication network needs to be customized to meet the challenges posed by social multimedia networks. One means to manage and control this communication network could be through the use of Software-Defined Networking (SDN) [13].

SDN, a critical enabling technology, separates the network control plane from the forwarding plane to provide several characteristics such as scalability, privacy, fault tolerance, distributed routing control, incremental deployment, network programmability, runtime security policies, and procedures, etc., without compromising the QoS. Literature suggests several approaches wherein different anomaly detection schemes have been employed on top of the SDN controller to provide scalable and resilient communication. For example, He *et al.* [14] designed an SDN-based traffic anomaly detection model to identify attacks based on significant deviations from the established standard usage profiles. Two refined algorithms namely density peak based clustering algorithm with sampling adaptation and unsupervised cluster-based feature selection mechanism were also proposed to handle large-scale, high dimensional and unlabeled network data.

Similarly, Peng *et al.* [15] presented an SDN-based flow detection method using K-nearest neighbors algorithm where double P-value of transductive confidence machines was used for the classification of SDN flows. In [16], Ha *et al.* introduced a traffic sampling strategy for software-defined networking (SDN). Instead of analyzing all the packets, the sampling of suspicious traffic was performed which minimizes the capture-failure rate of the malicious flow. Carvalho *et al.* [17] presented an SDN-based ecosystem to detect unusual network traffic patterns where a multi-feature analysis was employed to profile the normal traffic usage. In all these methods SDN has been proven to deliver promising results.

A. Motivation

With the increasing popularity of social networks, there has been an unprecedented growth in the multimedia content across the Internet. This prolific development in the Internet industry has opened doors for intruders to launch wide variety of attack vectors (such as impersonation attacks, phishing attacks, account hijacking, malware distribution, obfuscated malicious URL, etc.) by exploiting the widespread popularity of social networks. Thus, interoperability and security of social multimedia content are the biggest challenges for the research fraternity. Hence, designing a pervasive and scalable communicating paradigm is need of the hour.

In this direction, SDN is expected to play a pivotal role with its ability to centralize its controlling action using its dedicated control plane. SDN's central controller can dynamically control the functions of the forwarding devices. Although, with an increase in network traffic, the load on controllers increase rapidly which often leads to security and latency issues. Thus, to improve the network security, anomaly detection models are generally implemented on the top of the SDN controller. However, the existing SDN-based anomaly detection models pose difficulties in identifying the suspicious flows due to complex, heterogeneous and multi-dimensional nature of data. Moreover, factors like unknown data distribution, the imprecise boundary between normal and anomalous events, low accuracy, occasionally emerging anomalies, and unavailability of labeled data makes this task more challenging [18], [19], [20]. Considering all these prerequisites, a real-time anomaly detection scheme on the top of the SDN controller is required to provide scalable and resilient communication and enhance its adaptability to large-scale networks [21], [22].

Another significant challenge associated with the implementation of SDN in social networking domain is the provision for end-to-end data delivery. However, the existing research proposals fall short to address the QoS requirements pertaining to delivery of social multimedia content. Additionally, energy consumption initiatives in this context are still in their infancy. Thus, it is essential to secure the end-to-end data delivery of social multimedia data in the current scenario.

B. Contributions

The key contributions of this research work are:

- We present a consolidated solution for network anomaly detection for social multimedia domain powered with end-to-end delivery assisted with SDN platform.
- For anomaly detection, an ensemble approach based on the Restricted Boltzmann machine (RBM) and Support Vector Machine (SVM) is presented; wherein the performance of RBM has been revamped by incorporating dropout functionality and SVM has been improved with encapsulation of mixed kernel function and gradient descent approach.
- For end-to-end delivery of social media traffic, an SDN-assisted multi-objective flow routing scheme has been designed; which obtains the trade-off between latency, bandwidth, and energy consumption utilization.
- Lastly, we evaluate the performance of the designed model on both real-time and benchmark datasets.

C. Organization

The rest of the paper is organized as per the following sequence. The system of the proposed SDN-based anomaly detection framework is presented in Section II; followed by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

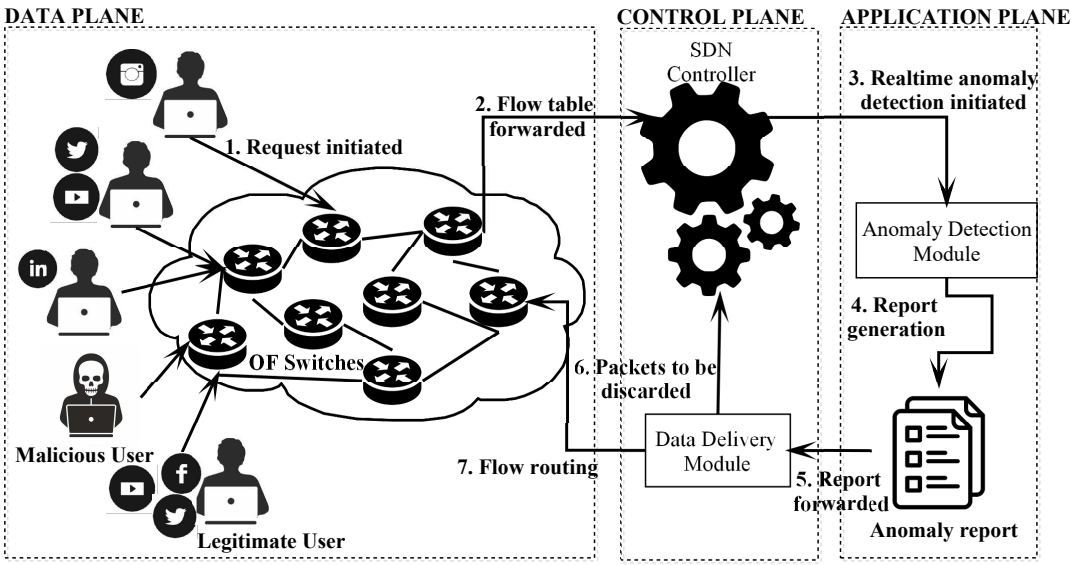


Fig. 1: System model of the proposed SDN-based anomaly detection framework.

detailed description of the *anomaly detection module* and *communication module* in Section III and Section IV respectively. The experimental results on real-time and benchmark datasets are drawn in Section V. Finally, the paper is concluded in Section VI.

II. SDN-based Anomaly Detection Framework

This section provides the preliminary details about SDN, followed by the working of the proposed SDN-based anomaly detection framework for suspicious flow detection in social multimedia.

A. Glimpse of Software Defined Networks

SDN extracts control flow management capabilities from the forwarding elements (FEs) and consolidates them into a logically centralized controller to get a unified global view of the network. SDN controller manages the functions of all the FEs such as switches, routers, gateways, or any access point by facilitating standardized network device programming. The communication architecture of SDN follows OpenFlow (OF) protocol, a standardized communication protocol, which enables the controller to perform flow level controls. The SDN framework consists of three decoupled planes: a data plane, control plane, and application plane. These are described below:

1) Data Plane: The data plane comprises of FEs which act according to the instruction set (contains packet forwarding instructions) provided by the controller. With OF-enabled FEs, data can be acquired from network traffic at a required level of granularity. Here, the traffic is abstracted as flows (a group of packets that share some common features such as addresses, transport protocol, and both source and destination ports). The instruction set configured on the FEs is mapped with the list of flow tables that are associated with each other by a pipeline. Accordingly, the transmission of packets is done by accessing the flow tables, configuration, and statistics of the data plane, which is controlled through OF protocol. FEs in data plane serve as data forwarding devices which regularly send messages to controllers to notify them of the ongoing flow status. Whenever a packet arrives at FE, the corresponding flow entry is updated. Since this module is responsible for the collection of traffic data, it can be considered as a prerequisite to anomaly detection.

2) Control Plane: This is the decision making plane of SDN which regulates the overall functions of controller such as system configuration, management, and exchange of routing table information. All the control commands and logic that are used to program the functions of FEs, reside in this plane. Here, controller solely manages the complete traffic flow through an open interface and provides consolidated decisions on flow forwarding, routing, and packet dropping. The basic purpose of control plane is to decide how the flow tables in data plane should be configured. This is where routing and switching protocols are executed to synchronize the distributed flow tables.

3) Application Plane: It mainly consists of end-user applications such as video streaming, web browsing, security implementation, network virtualization, mobility management, load balancing, etc. It also contains SDN-based applications such as policy implementation, network management, and security services. These applications are controlled by the SDN controller.

B. Working of the proposed framework

In the considered setup, SDN provides a scalable and dynamic reconfigurable architecture to the network by decoupling the network control plane from the physical network topology. By integrating SDN, network control decisions could be implemented by the centralized controller, SDN controller, via standardized interfaces, without modifying the underlying physical network components [23]. Due to its programmable ability, different nodes can interact with each other to maintain load-balancing and multimedia service operations among

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

multiple servers [24]. Since multimedia devices are considered dumb and all the intelligence is kept on the control centers. Thus, embedding SDN for multimedia analytics will not only provide faster and reliable communication but will also lead to convenient control [25]. In our proposed framework, we consider two sets of modules namely-anomaly detection module and data delivery module to provide Quality of Experience (QoE) to its intended users; as shown in Fig. 1. Their detail description is provided in the subsequent section with steps of execution outlined in Algorithm 1.

Algorithm 1 Working methodology of the proposed scheme

procedure SUSPICIOUSFLOWDETECTION
    A user initiates a social multimedia request over Internet
    Flow controller captures flow statistics
    Flow features are extracted
    Dimensionality reduction using improved RBM
    Classification using Gradient based SVM
    Anomaly Detection Module generates anomaly report
    Anomaly report is transmitted to the SDN controller over secure channel
    for $\forall$ flows **do**
        if flow is anomalous **then**
            Discard the associated packet
        else
            Apply MoFR to establish an optimal route
            Controller makes flow table updates

III. ANOMALY DETECTION MODULE

This module consists of two phases: feature selection and classification as depicted in Fig. 2. Initially, flow controller requests the FEs via OF protocol to provide flow statistics. Then, the flow collection module of the controllers collects this information to extract the flow features. Based on this, dimensionality reduction is performed using improved RBM algorithm. Finally, the extracted features are passed to the next phase which pre-processes the flow features and performs classification on the network flows with the proposed Gradient-based SVM algorithm. The anomaly detection scheme then generates an anomaly report and sends the same to the SDN controller with the help of a secure channel(s). As a result, SDN controller updates the flow table in accordance with the report received and configures OF-enabled FEs for further treatment. Details of the employed RBM and SVM are described as under.

A. Dimensionality Reduction: Restricted Boltzmann Machine

RBM is a stochastic approach which learns a probability distribution over the input. It consists of two layers of binary units: one visible, to represent the data, and one hidden, to increase learning capacity. In the proposed approach, RBM is used for dimensionality reduction. If some object is represented as a vector of n elements in n-dimensional space, then dimensionality reduction is defined as the process of refining data in a manner that every individual data vector x is transformed into different vector $x^{'}$ in an m-dimensional space; such that $m < n$. By doing this, RBM tries to keep all the important information by removing the noisy elements. However, due to the large flow of network traffic, it suffers from the problems of over-fitting. In such cases, the concept of "Dropout" is generally used which refers to a way of regularizing a neural network by adding noise to its hidden units.

Let us consider a RBM with m visible units (denoted by $v_i : v = [v_1, \cdots, v_m]$) and n number of hidden units (denoted by $h_j : h = [h_1, \cdots, h_n]$). The standard probability distribution of the RBM over visible and hidden layers is expressed as follows [26]:

$$P(h, v; \theta) = \frac{1}{Z(\theta)} exp(a^T h + b^T v + v^T W h) \tag{1}$$

In the above equation, Z depicts the partitioning function, 'W' denotes the weight from the visible units to hidden units, variable a and b are the bias values associated with the visible and hidden layers respectively while $\theta = (W, a, b)$ is equivalent to model parameters.

In order to further improve the performance of RBM such that it can quickly learn the important features from the training dataset, dropout functionality can play a pivotal role. Dropout helps to overcome the over-fitting problem in large networks; wherein it helps to randomly drop a few units and their corresponding connections from the network. Dropout of a hidden layer in RBM is achieved by encapsulating a vector of binary random variables $r \in \{0, 1\}^n$ within the RBM. Each r_j attains the value of 1 with the probability p and vice-versa. In case, r_j assumes the value of 1, then the associated hidden layer is retained or else is dropped. The joint probability distribution of the RBM with dropout is expressed as follows [27]:

$$P(r, h, v; p, \theta) = P(r; p)\mathcal{P}(h, v|r; \theta) \tag{2}$$

where,

$$P(r; p) = \sum_{j=1}^{n} p^{r_j}(1-p)^{1-r_j} \tag{3}$$

$$\mathcal{P}(h, v|r; \theta) = \frac{1}{Z^{'}(\theta, r)} exp(a^T h + b^T v + v^T W h) \times \sum_{j=1}^{n} g(h_j, r_j) \tag{4}$$

$$g(h_j, r_j) = \begin{cases} h_j = 1; \text{If } r_j = 1 \\ h_j = 0; \text{If } r_j = 0 \end{cases} \tag{5}$$

In the above equations, $Z^{'}(\theta, r)$ depicts the normalization constant. Further function $g(h_j, r_j)$ imposes restriction on the value of h_j against r_j.

Next, we present the modified conditional probability of h over v and r; followed by the corresponding activation probability.

$$P(h|r, v) = \sum_{j=1}^{n} P(h_j|r, v) \tag{6}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

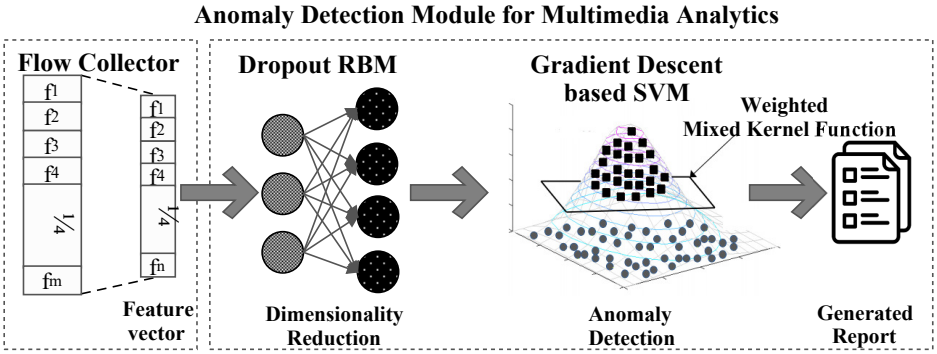


Fig. 2: Anomaly detection module for multimedia analytics.

$$P(h_j = 1|r, v) = \sigma(b_j + \sum_{i=1}^{m} w_{i,j} v_i); \text{If } r_j = 1 \quad (7)$$

In contrast, the conditional and activation probability of the visible layer over hidden layer remains unaffected by the dropout scheme, as evident from the following equations:

$$P(v|h) = \sum_{i=1}^{m} P(v_i|h) \quad (8)$$

$$P(v_i = 1|h) = \sigma(a_i + \sum_{j=1}^{n} w_{i,j} h_j) \quad (9)$$

Since, Gibbs sampling results in a satisfactory trade-off between the estimation speed and accuracy; thus input expectation $E[v]$ is computed using it. Gibbs sampling is then employed for simultaneous adjustment of the weights; which result in considerable minimization of the reconstruction error [28].

Further, contrastive divergence has been used for enabling learning in the designed dropout RBM. During the learning phase, r is initially sampled followed by the training of RBM using the preserved hidden units. The moment RBM is trained, it then transforms the input vectors into resulting vectors for a better representation; which are finally passed to SVM for further classification.

In a nutshell, it can summarized the improved version of RBM helps in extraction of features from the flow statistics training data acquired from FEs. It is achieved by using a non-linear transformation. The observed data is encoded in visible layer and the outputs in hidden layer are considered as selected features. The step by step execution of RBM is detailed using Algorithm 2.

Algorithm 2 Dimensionality Reduction using RBM

input: Training dataset with flow features received from FEs
output: Extracted features $F = \{f_1, f_2, \cdots, f_k\}$

1: Load training dataset
2: Sample training vector from training dataset
3: Initialize weights W and bias a and b
4: Set m visible units (v)
5: Set n hidden units (h)
6: Compute conditional probability P for all v
7: Compute conditional probability P for all h using dropout
8: Initialize target class $c = \{c_1, c_2, \cdots, c_t\}$
9: Set training objective as $O_T = -\sum_{t=1}^{T} \log P(c_t, v_t)$
10: To deal with the computational problem, compute gradient of $\log P(c_t, v_t)$, i.e., $\partial \log P(c_t, v_t)/\partial\theta$
11: Compute contrastive divergence from gradient to yield features from T
12: Compute Expectation with respect to the data distribution, i.e, $E[v]$ using Gibbs sampling
13: Repeat the procedure G times to obtain a G-steps Gibbs samples
14: Return extracted feature set F

B. Classification: Support Vector Machines

The proposed anomaly detection scheme employs SVM for the classification of network traffic flows. It is a supervised learning approach which maximizes the geometric margin between two classes in n-dimensional space. To minimize the generalization error, the training phase of SVM constructs a hyperplane and extends this hyperplane to non-linear boundaries [29]. The training set comprises of linearly separable labeled vectors $T = (x_i, y_i)|1 \leq i \leq k$, where $x_i \in \Re^d$ with d as the dimensionality of feature-space and $y_i \in \{-1, +1\}$ implies two different categories [30]. For non-linear classification, a kernel trick is used, which implicitly maps the given inputs to multi-dimensional feature space. Now, the classifier can be learnt for high dimensional feature space without explicitly computing $\phi(x)$. There are different types of kernel functions such as linear, polynomial, radical basis function (RBF), Gaussian, sigmoid, etc., that can be used to achieve the appropriate boundary function [31]. In this paper, we employed a mixed kernel function proposed by Wan *et al.* [32]. The details about this kernel function are provided in the subsequent section.

1) Weighted Mixed Kernel Function: The mixed kernel function proposed in [32] integrates Gaussian and Polynomial kernel functions in order to adapt the local performance capability of Gaussian kernel along with global performance ability of Polynomial kernel. It is expressed as follows:

$$k(x_j, x_k^{'}) = w_1 \exp\left[\frac{-||x_j - x_k^{'}||^2}{2\sigma^2}\right] + w_2(1 + x_j^{\intercal} x_k^{'})^d \quad (10)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

where p and q corresponds to the weight coefficients respectively.

2) Gradient Descent Approach: The proposed anomaly detection scheme aims to build and update the hyperplane (decision function) dynamically. To build such a scheme, the gradient descent algorithm is used which computes minimum of a function iteratively. In this method, iterative update mechanism is used to minimize a cost function $C(\omega)$ [33].

$$\omega_{t+1} \leftarrow \omega_t - \eta_t \nabla_\omega C(\omega_t) \tag{11}$$

where η corresponds to the learning rate of the classifier. Further, we iteratively update the ω_t coefficients at interval t as:

$$\omega_{t+1} = \begin{cases} \omega_t - \eta(\lambda\omega_t - y_i x_i) & if\ y_i f(x_i) < 1 \\ \omega_t - \eta\lambda\omega_t & otherwise \end{cases} \tag{12}$$

which helps in building and updating the decision function at run-time.

3) Modelling Parameters: The classification accuracy of the proposed gradient based SVM approach depends on four parameters, i.e., penalty parameter C, the kernel parameter σ, and weight coefficients of weighted mixed kernel functions, i.e., w_1 and w_2. Thus, changing the values of these parameters can disrupt the accuracy of classification. If C is too large, classification rate is very high, but if C is too small, classification accuracy is very low which will result in huge number of false positives. Similar is the case with kernel parameter σ. Higher values of 'σ' results in low classification accuracy whereas its smaller values lead to over-fitting problems. Nevertheless, the value of the weight coefficients w_1 and w_2 can be adjusted for better classification performance. In order to maintain an optimal trade-off between empirical risk and model complexity, all the aforementioned parameters are optimized. For this purpose, we adopted the Chaotic Differential Evolution approach proposed by Yi *et al.* in [34]. Since Differential Evolution suffers from the risk of premature and Chaotic Optimization has the feature of ergodicity and randomness, both are integrated in their work to improve the global optimization capability of the algorithm.

IV. Data Delivery Module

In addition to the robust anomaly detection mechanism explained above, the proposed scheme also ensures end-to-end delivery of the multimedia content over the SDN platform. This is achieved by amalgamating the anomaly detection model and efficient flow routing model at the heart of the SDN's control plane. The execution flow works in accordance with the following sequence: i) Classification stage, ii) Reporting Stage, iii) Migration Stage and iv) Relay Stage. In the first stage, the control plane encapsulated with the anomaly detection model classifies the incoming traffic into anomalous and benign. The results of the classification are then communicated to the Control Plane during the second stage. In case, the incoming traffic is classified as anomalous, the control plane then drops the packet and the communication with the requesting host is discarded immediately. This helps to protect

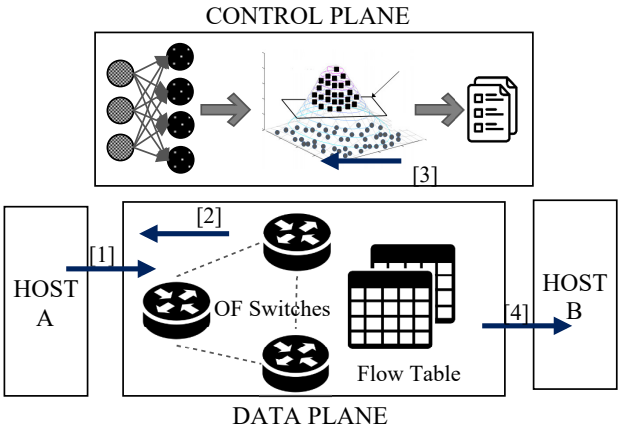


Fig. 3: SDN module for multimedia delivery. [1] Data Extraction [2] Classification [3] Reporting [4] Migration and Relay

the underlying network with the malicious content and hinders its further propagation across the network. However, if the incoming traffic is deemed benign during the classification stage, then it's passed on to the next stage. In the migration stage, the control plane updates the flow table entry of the forwarding devices and finally relays the incoming traffic to the destination host. This relaying is further accomplished using the proposed Multi-objective Flow Routing (MoFR) Scheme as shown in Fig. 3. The details of the proposed MoFR scheme are described in the following subsections.

A. Multi-objective Flow Routing (MoFR) Scheme

Multimedia applications range from delay-sensitive interactive and streaming applications to delay-tolerant object retrieval and messaging applications. Certain class types (such as object retrieval) demand higher bandwidth availability, while others (such as interactive and streaming applications) require reduced latency or a trade-off of the two (such as messaging applications). Given the complexity, it is essential to reconfigure the network resources by the multimedia class type being transmitted over the network. This can be best addressed through SDN's promising platform which can reconfigure its control logic on a real-time basis. So, the idea is to design a MoFR Scheme for the end-to-end relay of multimedia applications. The designed MoFR not only guarantees optimal trade-off between latency and bandwidth distribution for end-to-end delivery of multi-media applications but also addresses an important issue of energy consumption utilization by the underlying data plane devices. In a nutshell, the design MoFR provides an optimal trade-off against resource distribution (regarding bandwidth and latency) and energy minimization.

The overall objective is to attain efficient flow routing in SDN setup while attaining the above objectives. For this, let us assume that the problem at hand requires routing of the k^{th} flow over the SDN platform at time t. Further, consider the SDN setup to be equipped with w number of switches; wherein the switches are indexed using i and referred using v_i. Further, the adjacency between the switches, say i^{th} and l^{th} switch is denoted using $a_{i,l}$. The proposed scheme attains

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

the efficient flow routing in SDN setup in accordance with the following objective functions.

1) Objective function for Latency Minimization: Latency of the channel depends on the propagation delay, transmission delay, queuing delay and processing delay respectively; and the same is illustrated using the below mentioned objective function $\mathbb{L}(\delta_{k,v}(t))$ with the decision variable $\delta_{k,v}(t)$.

$$\mathbb{L}(\delta_{k,v}(t)) = \begin{cases} \left(\sum_j \frac{d_{vi,vl} \times \delta_{k,vi}(t) \times \delta_{k,vl}(t) \times a_{i,l}}{\mathbb{P}r(t)} \right) + \\ \left(\sum_{i \in w} \sum_{j \in |P(i)|} \frac{\mathcal{P}_{i,j}(t)}{B_{i,j} \times O_{i,j}(t)} \right) + \\ \left(\sum_{i \in w} \sum_{j \in |P(i)|} \frac{|\mathbb{Q}_{ready}(t)|}{B_{i,j} \times O_{i,j}(t)} \right) + \\ \left(\sum_{i \in w} \mathbb{P}_i(t) \times \sum_k (t_k^{end} - t_k^{start}) \times \delta_{k,vi}(t) \right) \end{cases}$$

here, $\delta_{k,v}(t)$ is a binary decision variable which denotes whether the switch v is employed in routing the l^{th} flow ($\delta_{k,v}(t) = 1$) or not ($\delta_{k,v}(t) = 0$). The first part of the Eq. (13) denotes the propagation delay; wherein the variables $d_{vi,vl}$ and $Pr(t)$ denote the distance between the i^{th} and l^{th} switch, and the medium propagation delay respectively. On the contrary, the second part of Eq. (13) depicts the transmission delay with variables $\mathcal{P}_{i,j}(t)$ and $O_{i,j}(t)$. Here, the former represents the packet size while the latter denotes the occupancy ratio. The third part of Eq. (13) depicts the queuing delay; with variable $|\mathbb{Q}_{ready}(t)|$ indicating the number of flows in the ready queue. Finally, the fourth part of the equation refers to the processing delay induced due to the intermediate nodes; for instance $\mathbb{P}_i(t)$ is the processing delay caused by the i^{th} node. Further, t_k^{start} and t_k^{end} refer the start and end time of the k^{th} flow respectively.

2) Objective function for Bandwidth Maximization: Bandwidth refers to the amount of the data that can be transmitted over a channel in a fixed interval of time. Mathematically, the flow routing problem can be linked with bandwidth utilization using the below mentioned objective function ($\mathbb{B}(\delta_{k,v}(t))$).

$$\mathbb{B}(\delta_{k,v}(t)) = \Bigg(\sum_{i \in w} \sum_{j \in |P(i)|} \sum_k B_{i,j} \times (t_k^{end} - t_k^{start}) \times \delta_{k,vi}(t) \times \delta_{k,vl}(t) \times a_{i,l} \Bigg) \quad (13)$$

In the above equation, the overall bandwidth over a selected route is expressed using $\mathbb{B}(\delta_{k,v}(t))$; while j^{th} port specific bandwidth of the i^{th} switch is denoted using $B_{i,j}$.

3) Objective Function for Energy Minimizations: For minimizing the energy consumption of SDN platform, it is essential to minimize the total energy consumption of the underlying forwarding devices, *i.e.*, switches. Total energy utilization of a switch is dependent on its fixed part and dynamic part. Here, the former is used to power the fixed components such as fans, chasis, etc. While, the latter is dependent on the number of active ports of the considered switch. Hence, it can be concluded that the energy utilization of a switch is variable of its dynamic part rather than the traffic load incident on it. Hence, the total energy utilization for routing l^{th} flow at time t is subjected to the following condition.

$$\mathbb{E}(\delta_{k,v}(t)) = \Bigg(F \times \sum_k (t_k^{end} - t_k^{start}) \times \delta_{k,vi}(t) \Bigg) + \Bigg(D \times \sum_k (t_k^{end} - t_k^{start}) \times \delta_{k,vi}(t) \times \delta_{k,vl}(t) \times a_{i,l} \Bigg) \quad (14)$$

where, F and D denote the fixed and dynamic portions of energy consumption of a switch respectively. $\mathbb{E}(\delta_{k,v}(t))$ refers to the objective function under consideration.

4) Overall Multi-objective Flow Routing (MoFR) Scheme: In view of the above objectives functions, the overall problem can be visualized as a multi-objective optimization expressed below:

$$\min \mathbb{F}(\delta_{k,v}(t)) = f(-\mathbb{B}(\delta_{k,i}(t)), \mathbb{L}(\delta_{k,i}(t)), \mathbb{E}(\delta_{k,i}(t))) \quad (15)$$

$$s.t.: \begin{cases} \textbf{C1:}\ F_{vi,vl} = \{f_j | \delta_{k,vi}(t) = 1\ \&\ \delta_{k,vl}(t) = 1; \\ \qquad\qquad \forall v_i, v_l \in w\} \\ \textbf{C2:}\ \delta_{k,v}(t) \in 0, 1;\ \forall k, v, t \end{cases}$$

In the above equation, $\delta_{k,v}(t)$ is the binary decision variable. Further, first and second constraint depict restriction on the number of flows and integrality restriction on the decision variable respectively. Keeping in view the complexity of the formulated problem, it is subjected to a multi-objective evolutionary algorithm based on decomposition (MOEA/D) [13]; for obtaining the optimal trade-off between the competing functions.

V. Implementation

This section presents the results of the experimental evaluation of the proposed anomaly detection model across different case studies. All the scripting for the model has been done using Matlab R2014a running on Intel i7-7500U CPU @ 2.70GHz with 8 GB RAM. The details about the datasets employed, performance parameters and the existing state-of-the-art models are given below.

A. Datasets Considered

For rigorous evaluation of the proposed anomaly detection model, it has been evaluated on both real-time and benchmark datasets, as described below.

1) Real-time dataset: In this case, the real-time data traffic from Thapar Institute of Engineering & Technology (TIET), Patiala, India has been captured [35]. The network traffic of nearly 400 users was captured using Wireshark for 1 hour. During this time, the users were made to access different social networking sites such as Facebook, Twitter, Instagram, Whatsapp, etc. A snippet of the captured traffic is depicted in Table I comprising majorly of HTTP requests.

The captured traffic was then subjected to anomalous traffic instances with different polymorphic and non-polymorphic HTTP attacks [36]. After the attack injection, the traffic instances could be broadly classified into the following classes: Generic, Shell-code and CLET Attacks. Here, the first class

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia

refers to the generic attack vectors like buffer-overflow, information leakage, and remote execution. On the contrary, Shellcode attacks are targeted to capture the shell of the remote server or machine. This is achieved by the encapsulation of executable code in the payload. This attack class comprises of 11 attacks like code-red, ddk and attacks against Windows media service. CLET attacks are an inherently polymorphic version of shell-code attacks. They further comprise of 96 attack traces generated using CLET polymorphic engine.

This dataset in total consists of 35 features; comprising of 10 basic (like arrival time, epoch time, frame number, etc.), 8 content-based (flags, fragment offset, time to live, protocol, etc.) and 6 host-based features (acknowledgment number, header length, flags, etc.).

2) Benchmark dataset: The benchmark dataset considered for evaluation of the proposed model is KDD'99 [37]. The entire dataset has around 5 million records with a total of 41 features (1-9 packet's basic features, 10-22 content-based features, 23-31 traffic features, and 32-41 host-based features). Further, it comprises of attack data instances which could further be classified into denial of service (DOS), user to root (U2R), remote to local (R2L), and probe attacks.

3) Insider threat dataset: Further, the above mentioned datasets depicted only the outsider attacks on social networking sites. However, the insider attacks such as Identity theft, profile cloning, data collection, etc. can adversely hamper the privacy, integrity and availability of online social networks. Thus, in order to cater the security objectives that could not be validated using the previously mentioned case studies, a new case study has been introduced which use the only publicly available dataset from CMU for insider threat analysis [38].

B. Case studies considered

Based on the above two datasets, the following case-studies have been considered.

1) Case study-I: In this case study, the real-time dataset has been chosen for evaluation of the proposed model on the social multimedia dataset. The dataset considered for evaluation is injected with anomalous traffic across the time horizon. The number of packets for both benign and malicious instances is depicted in Fig. 4a. It is evident from the figure that the generated dataset to be evaluated has a significant number of anomalous instances to be detected. Hence, in the first phase, the improved RBM with dropout functionality is employed to extract the relevant feature set. The idea about the performance of the designed RBM can be realized by the results illustrated using Fig. 4b. It is evident from the figure that RBM with dropout functionality considerably reduces the classification error rate with the increasing number of packets. A similar trend is observed for standard RBM without dropout functionality. However, the RBM with dropout functionality exhibits higher performance in comparison with its counterpart.

TABLE I: Characteristics of captured network traffic

File	Packet Count	HTTP Count	Rate	Burst Rate
Capture 1	324414	11867	1.2853	2.3400
Capture 2	318714	11616	1.3654	2.4900
Capture 3	320085	11212	1.3049	2.2200
Capture 4	318247	9010	1.1659	2.2300
Capture 5	312480	11436	1.2761	3.3300
Capture 6	320136	9745	1.0465	2.2800
Capture 7	316400	11150	1.2988	2.4200
Capture 8	316034	10615	1.1887	2.3900
Capture 9	319693	10849	1.2469	2.9900
Capture 10	314776	11017	1.2811	2.5200
Capture 11	319437	10446	1.1564	2.5500
Capture 12	321772	9748	1.0158	2.2200
Capture 13	321581	11074	1.1969	2.9000
Capture 14	318019	8296	1.0468	2.1800
Capture 15	316531	8826	1.0822	2.0600
Capture 16	316759	13238	1.4732	2.9900
Capture 17	320856	12109	1.5087	2.7000

The next phase of the proposed anomaly detection model is the classification of data instances using the designed SVM; powered with gradient descent and mixed kernel. For evaluating its performance it has been compared with three other variants of the proposed model namely-i)RBM coupled with dropout (DRBM) and SVM powered with gradient descent (GDSVM), ii)standard RBM and GDSVM, and iii)standard RBM and SVM. The related results in terms of ROC has been shown using Fig. 4c, 4d and 4e respectively. It is evident from the figure that the proposed model gives the superior performance in comparison with other schemes; with the highest area under the curve (AUC) value of 0.9786. Following the proposed model, DRBM+GDSVM, RBM+GDSVM, and RBM+SVM perform in descending order with AUC values of 0.9682, 0.9336 and 0.9328 respectively.

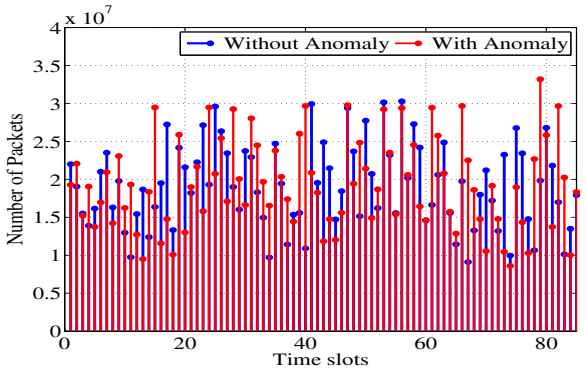
As a result of this classification, the traffic flows are separated into two types, *i.e.*, anomalous and benign. The anomalous traffic is discarded by the SDN's control plane; while the benign traffic is forwarded to the designated host using the proposed MoFR scheme. The proposed routing scheme helps to attain optimal trade-off between latency, bandwidth, and energy consumption. The related results have been highlighted using Figs. 4f, 4g and 4h respectively. For evaluating the performance of MoFR, the incoming traffic has been run on both the traditional network and SDN platforms (simulated). It is clear from the figures, the implementation of MoFR on SDN platform gives more impressive results relative to the traditional network.

Hence, it can be concluded from the obtained results that the proposed scheme achieves high performance on real-time datasets in comparison to existing schemes.

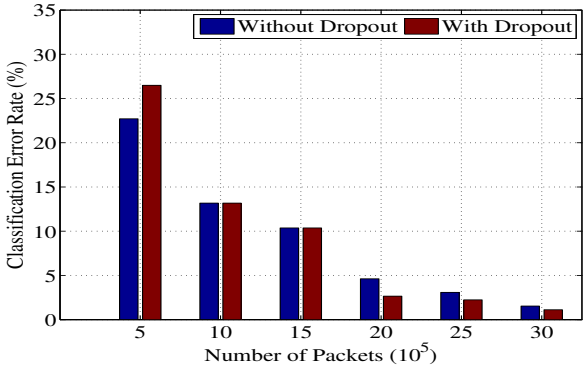
2) Case study-II: This case study presents the extensive evaluation for the proposed anomaly detection model on KDD'99 dataset against the current state-of-the-art models [39], [40], [41], [31], [42]. Fig. 5 and Table II illustrates the key findings of the proposed model on KDD'99 dataset. The related findings have been summarized in terms of different performance parameters namely-DR, FPR, accuracy, precision and F-score [43].

Fig. 5a presents the detection rate evaluation of the proposed model corresponding to successful detection of normal, DOS, U2R, R2L and Probe data instances (99.13, 99.16, 98.88, 99.01, and 99.03% respectively). It is evident from the obtained results that the maximum level of detection is attained against DOS attacks; while the least is obtained against U2R attacks. On the similar lines, FPR is minimum for DOS ($= 0.49\%$) and the highest for U2R (approx. 2%); as depicted in Fig. 5a. Further, the accuracy obtained during evaluation results are depicted using Fig. 5c; wherein the accuracy of 99.99, 99.99, 99.98, 99.97, and 99.97% was achieved in

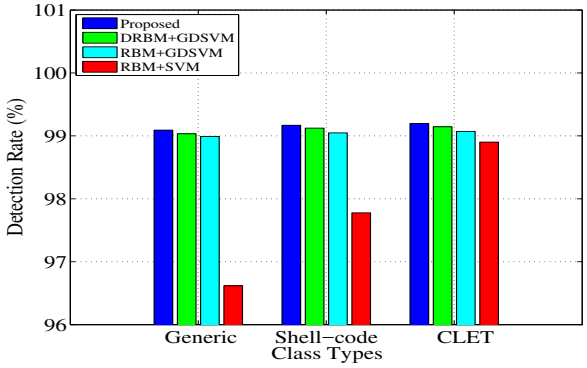
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia



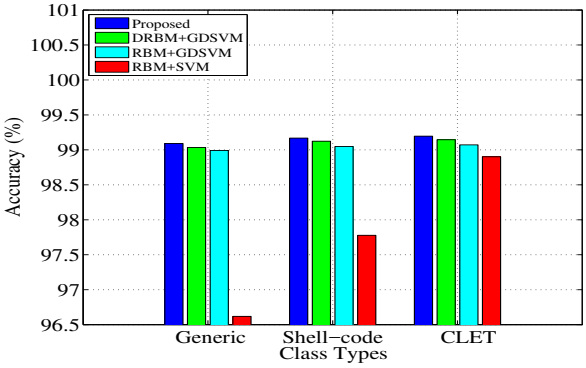
(a) Number of packets across the time horizon for the synthetic datasets.



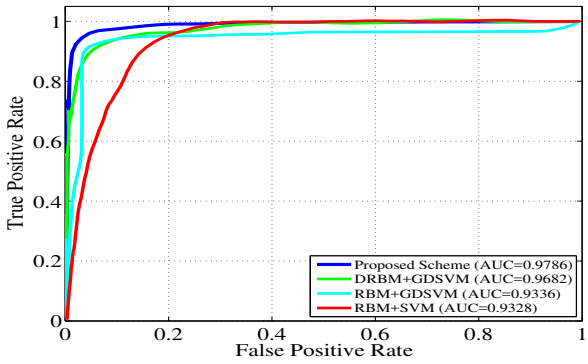
(b) Classification error rate Vs number of packets.



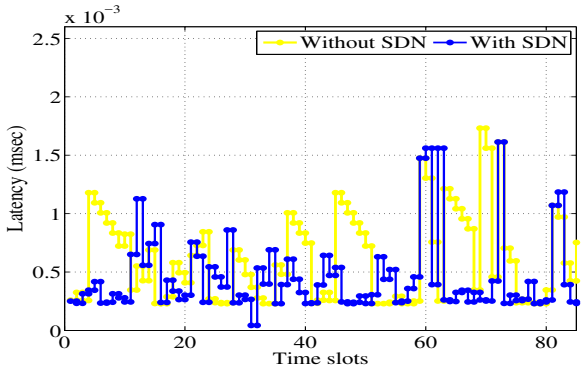
(c) Detection rate evaluation.



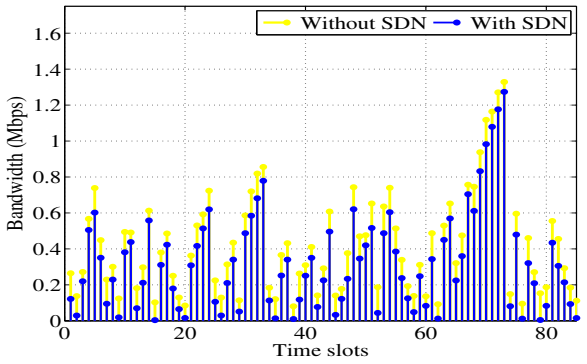
(d) Accuracy evaluation.



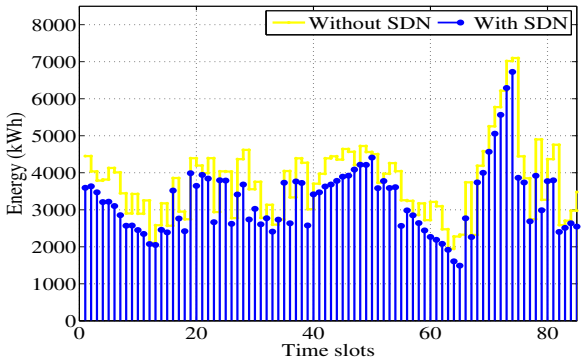
(e) ROC curve of the proposed hybrid anomaly detection model.



(f) Latency involved in end-to-end delivery of social networking data.



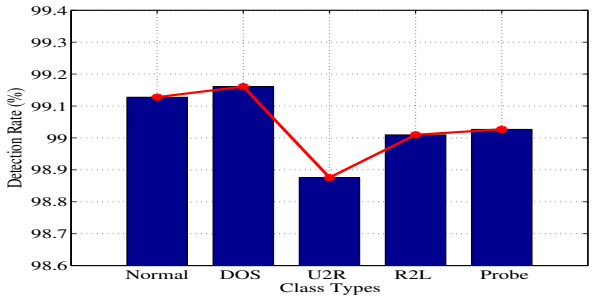
(g) Bandwidth involved in end-to-end delivery of social networking data.



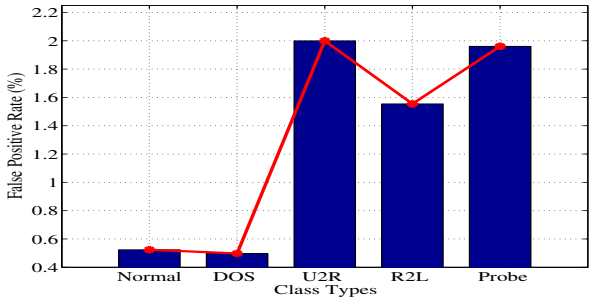
(h) Energy consumption involved in end-to-end delivery of social networking data.

Fig. 4: Experimental evaluation of the proposed model on real-time dataset.

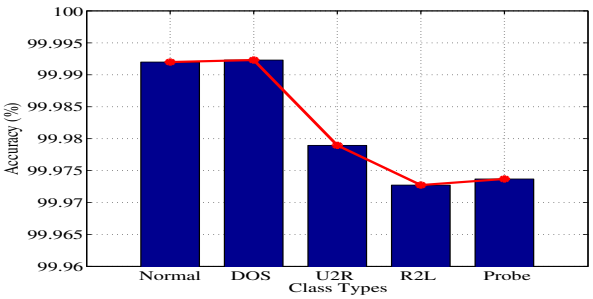
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia



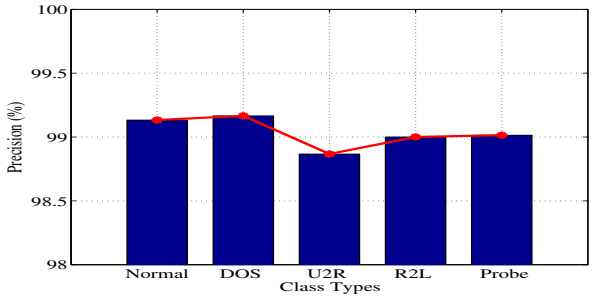
(a) Detection rate evaluation.



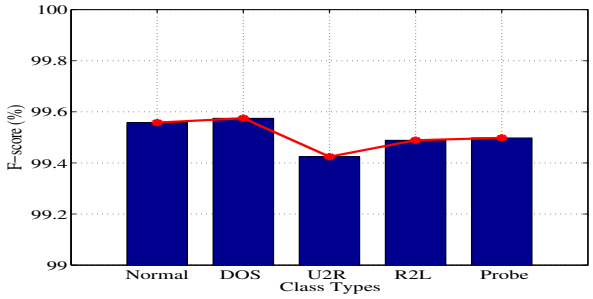
(b) False positive rate evaluation.



(c) Accuracy Evaluation.



(d) Precision evaluation.



(e) F-score evaluation.

Fig. 5: Experimental evaluation of the proposed model on benchmark KDD'99 dataset.

detecting normal, DOS, U2R, R2L, and Probe data instances. Similar performance was observed by the proposed scheme against the precision and F-score parameters as summarized in Fig. 5d and 5e respectively.

• **Comparison with the current state-of-the-art:** In addition to this, the relative comparison of the proposed anomaly detection model against the current state-of-the-art models is also illustrated in Table II. It is evident from the obtained results that the proposed model outperforms the existing schemes on the KDD'99 dataset.

TABLE II: Relative comparison of the proposed anomaly detection model with the current state-of-the-art models

Scheme	DR	FPR	Accuracy	Precision	F-score
Tan *et al.* [44]	95.11	1.26	95.20	-	-
Guo *et al.* [45]	91.86	0.78	93.29	-	-
Al-Yaseen *et al.* [42]	95.17	1.87	95.75	-	-
Bigdeli *et al.* [39]	98	2	-	-	-
Chiba *et al.* [40]	98.59	1.13	98.66	99.62	0.99
Shone *et al.* [41]	97.85	2.15	97.85	99.99	98.15
Proposed	99.04	1.31	99.98	99.03	99.50

3) Case study-III:: In order to further evaluate the effectiveness of the proposed anomaly detection module, the CMU insider threat dataset [38] has been chosen as a reference. This dataset depicts one of the crucial class of social networing attack types, *i.e.*, insider threats. Hence, it has been used for evaluation in this work.

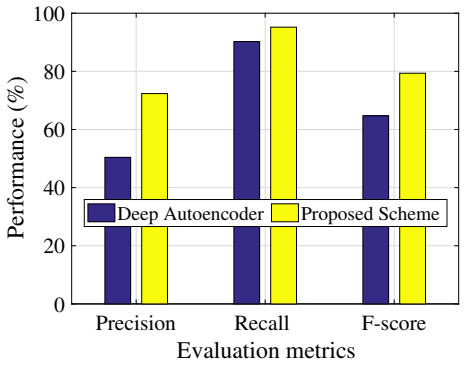
This is the only publicly available dataset for insider threat analysis and comprises of 14GB web-browsing logs in addition to device connections, email log files, and file transfers. In this work, version 4.2 of this dataset has been used for comparative analysis with an existing scheme based on the concept of deep autoencoders [46]. The dataset depicts the user-specific activity logs of almost 1000 insiders over 1.5 years. The different files related to user activitity logs and available in the dataset are namely-logon.csv, device.csv, file.csv, email.csv, http.csv, and psychometric.csv. Further, the dataset defines three different insider attack scenarios on the basis of above-mentioned user activities. Scenario 1 depicts the uncharcteristic behavior of a user, Scenario 2 represents a user who steals confidential data, and Scenario 3 denotes the case where the user installs keylogger to obtain the passwords.

The performance evaluation of the proposed anomaly detecction module on the above mentioned scenarios is depicted in Fig. 6. The obtained results depict the supremacy of the proposed scheme over the existing scheme on the basis of precision, recall and F-score.

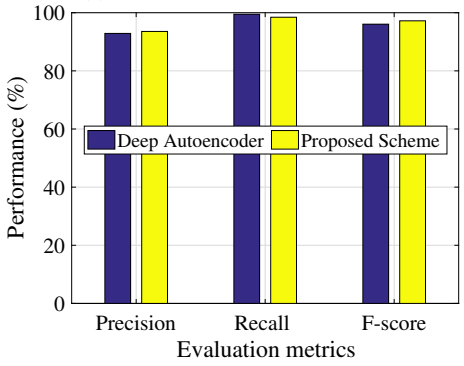
VI. Conclusion and Future Scope

With the unprecedented popularity of Social media networking sites such as Facebook, Instagram, WhatsApp, Twitter, etc., the network today has become more vulnerable to network security risks. Hence, it is essential to analyze the social multimedia traffic in real time while guaranteeing quality end-to-end delivery of the underlying data. This brings us to designing a real-time anomaly detection scheme coupled with an SDN-enabled data delivery module. The former leveraged the advantages of improved RBM supported with dropout functionality and SVM encapsulated with a mixed kernel and gradient-descent approach; for effective anomaly

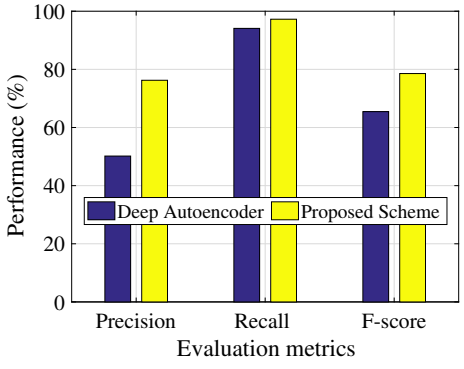
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2019.2893549, IEEE Transactions on Multimedia



(a) Performance for Scenario 1



(b) Performance for Scenario 2



(c) Performance for Scenario 3

Fig. 6: Comparative evaluation on CMU insider threat dataset [38]

detection. On the contrary, effective data delivery is realized using multi-objective flow routing scheme based on SDN. The proposed model, when evaluated on real-time and benchmark datasets, led to impressive outcomes against the current state-of-the-art models. For instance, the proposed scheme achieves significantly improved detection rate (*i.e.*, $> 99\%$) over the considered TIET, KDD'99 and CMU datasets.

In future, the proposed framework can be extended to different application domains namely Smart Grids, Intelligent Transportation Systems, Unmanned Aerial Vehicles, Smart Homes, etc.

ACKNOWLEDGMENT

The work carried out in this paper is funded by National Funding from the FCT - Fundação para a Ciência e a Tecnologia through the UID/EEA/50008/2013 Project; by Finep, with resources from Funttel, Grant No. 01.14.0231.00, under the Centro de Referência em Radiocomunicações - CRR project of the Instituto Nacional de Telecomunicações (Inatel), Brazil; and by the Brazilian National Council for Research and Development (CNPq) via Grant No. 309335/2017-5.

REFERENCES

[1] D. Chaffey, "Global Social Media Research Summary 2018," Smart Insights, Tech. Rep., 2018. [Online]. Available: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

[2] S. A. Alvi, B. Afzal, G. A. Shah, L. Atzori, and W. Mahmood, "Internet of multimedia things: Vision and challenges," *Ad Hoc Networks*, vol. 33, pp. 87 – 111, 2015.

[3] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: threats and solutions," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2019–2036, 2014.

[4] W. Chu, H. Xue, C. Yao, and D. Cai, "Sparse Coding Guided Spatiotemporal Feature Learning for Abnormal Event Detection in Large Videos," *IEEE Transactions on Multimedia*, 2018, DOI: 10.1109/TMM.2018.2846411.

[5] K. Xu, X. Jiang, and T. Sun, "Anomaly Detection Based on Stacked Sparse Coding With Intraframe Classification Strategy," *IEEE Transactions on Multimedia*, vol. 20, no. 5, pp. 1062–1074, 2018.

[6] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," *Computer Vision and Image Understanding*, 2018, DOI: https://doi.org/10.1016/j.cviu.2018.02.006.

[7] M. Ribeiro, A. E. Lazzaretti, and H. S. Lopes, "A study of deep convolutional auto-encoders for anomaly detection in videos," *Pattern Recognition Letters*, vol. 105, pp. 13 – 22, 2018.

[8] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Computer Vision and Image Understanding*, vol. 156, pp. 117 – 127, 2017.

[9] Y. Feng, Y. Yuan, and X. Lu, "Learning deep event models for crowd anomaly detection," *Neurocomputing*, vol. 219, pp. 548 – 556, 2017.

[10] X. Sun, C. Zhang, S. Ding, and C. Quan, "Detecting anomalous emotion through big data from social networks based on a deep learning method," *Multimedia Tools and Applications*, 2018, DOI: 10.1007/s11042-018-5665-6.

[11] H. Li, "Deep learning for natural language processing: advantages and challenges," *National Science Review*, 2017.

[12] S. Rathore, P. K. Sharma, V. Loia, Y.-S. Jeong, and J. H. Park, "Social network security: Issues, challenges, threats, and solutions," *Information Sciences*, vol. 421, pp. 43 – 69, 2017.

[13] K. Kaur, S. Garg, G. S. Aujla, N. Kumar, J. J. Rodrigues, and M. Guizani, "Edge computing in the industrial internet of things environment: Software-defined-networks-based edge-cloud interplay," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 44–51, 2018.

[14] D. He, S. Chan, X. Ni, and M. Guizani, "Software-Defined-Networking-Enabled Traffic Anomaly Detection and Mitigation," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1890–1898, 2017.

[15] H. Peng, Z. Sun, X. Zhao, S. Tan, and Z. Sun, "A detection method for anomaly flow in software defined network," *IEEE Access*, vol. 6, pp. 27 809 – 27 817, 2018.

[16] T. Ha, S. Kim, N. An, J. Narantuya, C. Jeong, J. Kim, and H. Lim, "Suspicious traffic sampling for intrusion detection in software-defined networks," *Computer Networks*, vol. 109, pp. 172–182, 2016.

[17] L. F. Carvalho, T. Abrão, L. de Souza Mendes, and M. L. Proença Jr, "An ecosystem for anomaly detection and mitigation in software-defined networking," *Expert Systems with Applications*, vol. 104, pp. 121–133, 2018.

[18] P. Micholia, M. Karaliopoulos, I. Koutsopoulos, L. Navarro, R. Baig, D. Boucas, M. Michalis, and P. Antoniadis, "Community Networks and Sustainability: a Survey of Perceptions, Practices, and Proposed Solutions," *IEEE Communications Surveys & Tutorials*, 2018.

[19] M. Lopez-Martin, B. Carro, J. Lloret, S. Egea, and A. Sanchez-Esguevillas, "Deep learning model for multimedia quality of experience prediction based on network flow packets," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 110–117, 2018.

[20] Z. Zhang and B. B. Gupta, "Social media security and trustworthiness: overview and new direction," *Future Generation Computer Systems*, 2016.

[21] S. Garg, A. Singh, S. Batra, N. Kumar, and M. Obaidat, "EnClass: Ensemble-based classification model for network anomaly detection in massive datasets," in *IEEE Global Communications Conference (GLOBECOM), Singapore*, Dec. 2017.

[22] D. Gupta, S. Garg, A. Singh, S. Batra, N. Kumar, and M. Obaidat, "ProIDS: Probabilistic Data Structures Based Intrusion Detection System for Network Traffic Monitoring," in *IEEE Global Communications Conference (GLOBECOM), Singapore*, Dec. 2017.

[23] P.-W. Tsai, C.-W. Tsai, C.-W. Hsu, and C.-S. Yang, "Network monitoring in software-defined networking: A review," *IEEE Systems Journal*, 2018, DOI: 10.1109/JSYST.2018.2798060.

[24] G. S. Aujla, R. Chaudhary, S. Garg, N. Kumar, and J. J. Rodrigues, "SDN-enabled Multi-Attribute-based Secure Communication for Smart Grid in IIoT Environment," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2629 – 2640, 2018.

[25] E. Molina and E. Jacob, "Software-defined networking in cyber-physical systems: A survey," *Computers & Electrical Engineering*, 2017, DOI: http://dx.doi.org/10.1016/j.compeleceng.2017.05.013.

[26] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 599–619.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[28] E. de la Rosa and W. Yu, "Data-driven fuzzy modeling using restricted boltzmann machines and probability theory," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, dOI: 10.1109/TSMC.2018.2812156.

[29] A. Zisserman, "Lecture 2 the svm classifier," C19 Machine Learning lectures Hilary 2015, [Accessed on : May 2018]. [Online]. Available: http://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf

[30] J. A. Cid-Fuentes, C. Szabo, and K. Falkner, "Adaptive Performance Anomaly Detection in Distributed Systems Using Online SVMs," *IEEE Transactions on Dependable and Secure Computing*, 2018, DOI: 10.1109/TDSC.2018.2821693.

[31] S. Garg and S. Batra, "A novel ensembled technique for anomaly detection," *International Journal of Communication Systems*, vol. 30, no. 11, p. e3248, 2017.

[32] M. Wan, W. Shang, and P. Zeng, "Double Behavior Characteristics for One-Class Classification Anomaly Detection in Networked Control Systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 3011–3023, 2017.

[33] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE transactions on signal processing*, vol. 52, no. 8, pp. 2165–2176, 2004.

[34] J. Yi, D. Jian, and S. Zhenhong, "Pattern synthesis of MIMO radar based on chaotic differential evolution algorithm," *Optik-International Journal for Light and Electron Optics*, vol. 140, pp. 794–801, 2017.

[35] S. Garg, K. Kaur, S. Batra, N. Kumar, and M. Obaidat, "HyClass: Hybrid classification model for anomaly detection in cloud environment," in *IEEE International Conference on Communications (ICC), Kansas City, USA*, May 2018.

[36] R. Perdisci, D. Ariu, P. Fogla, G. Giacinto, and W. Lee, "McPAD: A multiple classifier system for accurate payload-based anomaly detection," *Computer networks*, vol. 53, no. 6, pp. 864–881, 2009.

[37] "KDD Cup 1999 dataset," The UCI KDD Archive Information and Computer Science. University of California, Irvine, 1999, (Accessed Dec 2017). [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[38] A. P. Moore, D. M. Cappelli, T. C. Caron, E. Shaw, D. Spooner, and R. F. Trzeciak, "A preliminary model of insider theft of intellectual property," CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST, Tech. Rep., 2011.

[39] E. Bigdeli, M. Mohammadi, B. Raahemi, and S. Matwin, "Incremental anomaly detection using two-layer cluster-based structure," *Information Sciences*, vol. 429, pp. 315–331, 2018.

[40] Z. Chiba, N. Abghour, K. Moussaid, A. El Omri, and M. Rida, "A novel architecture combined with optimal parameters for back propagation neural networks applied to anomaly network intrusion detection," *Computers & Security*, vol. 75, pp. 36–58, 2018.

[41] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.

[42] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system," *Expert Systems with Applications*, vol. 67, pp. 296–303, 2017.

[43] S. Garg and S. Batra, "Fuzzified cuckoo based clustering technique for network anomaly detection," *Computers & Electrical Engineering*, vol. 71, pp. 798 – 817, 2018.

[44] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A system for denial-of-service attack detection based on multivariate correlation analysis," *IEEE transactions on parallel and distributed systems*, vol. 25, no. 2, pp. 447–456, 2014.

[45] C. Guo, Y. Ping, N. Liu, and S.-S. Luo, "A two-level hybrid approach for intrusion detection," *Neurocomputing*, vol. 214, pp. 391–400, 2016.

[46] P. Chattopadhyay, L. Wang, and Y.-P. Tan, "Scenario-Based Insider Threat Detection From Cyber Activities," *IEEE Transactions on Computational Social Systems*, no. 99, pp. 1–16, 2018.

Sahil Garg [S'15, M'18] received his B.Tech degree from Maharishi Markandeshwar University, Mullana, India, in 2012; his M.Tech degree from Punjab Technical University, Jalandhar, India in 2014; and his Ph.D. from Thapar Institute of Engineering & Technology (Deemed to be University), Patiala, India, in 2018, all in computer science and engineering. He is currently working as a Postdoctoral Research Fellow with Department of Electrical Engineering, École de technologie supérieure, Université du Québec, Montréal, Canada. His research interests include Machine Learning, Big Data Analytics, Knowledge Discovery, Cloud Computing, Internet of Things, and Vehicular Ad-hoc Networks. Some of his research findings are published in top-tier journals such as IEEE Network, IEEE TII, IEEE Communications Magazine, IEEE IoT Journal, IEEE Consumer Electronics Magazine, Elsevier FGCS, Elsevier Information Sciences, Elsevier CAEE , Wiley IJCS, and various International conferences of repute such as-IEEE Globecom, IEEE ICC, ACM Mobicom Workshops, etc. He was the recipient of prestigious Visvesvaraya PhD fellowship from the Ministry of Electronics & Information Technology under Government of India (2016-2018). For his research, he also received the best paper award at the IEEE International Conference on Communications (ICC) in 2018. He is a member of IEEE, IEEE Communications Society, IEEE Computer, ACM and IAENG.

Kuljeet Kaur [S'13, M'18] received the B.Tech degree in computer science and engineering from Punjab Technical University, Jalandhar, India, in 2011 and the M.E. (Information Security) and PhD (Computer Science and Engineering) degrees from Thapar Institute of Engineering and Technology (Deemed to be University), Patiala, India, in 2015 and 2018, respectively. She is currently working as a Postdoctoral Research Fellow with Department of Electrical Engineering, École de technologie supérieure, Université du Québec, Montréal, Canada. Her main research interests include Cloud Computing, Energy Efficiency, Smart Grid, Frequency Support, and Vehicle-to-Grid. Dr. Kaur has secured a number of research articles in top-tier journals such as IEEE Wireless Communications, IEEE TII, IEEE TVT, IEEE TSG, IEEE Sensors Journal, IEEE Communications Magazine, IEEE TPDS, IEEE PS, Springer PPNA, etc., and various International conferences including IEEE Globecom, IEEE ICC, IEEE PES GM, ACM Mobicom Workshops, etc. During her PhD, she received two prestigious fellowships, i.e., INSPIRE fellowship from Department of Science & Technology, India (in 2015) and research scholarship from Tata Consultancy Services (TCS) (from 2016-2018). Dr. Kaur also received the best paper award at IEEE International Conference on Communications (ICC) in 2018. She is a member of IEEE, IEEE Communications Society, IEEE Computer, IEEE Women in Engineering, IEEE Software Defined Networks Community, ACM and IAENG.

Neeraj Kumar [M'16, SM'17] received his Ph.D. in CSE from Shri Mata Vaishno Devi University, Katra (J & K), India, and was a postdoctoral research fellow in Coventry University, Coventry, UK. He is working as an Associate Professor in the Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology (Deemed to be University), Patiala (Pb.), India. He has published more than 150 technical research papers in leading journals and conferences from IEEE, Elsevier, Springer, John Wiley etc. Some of his research findings are published in top cited journals such as IEEE TIE, IEEE TDSC, IEEE TITS, IEEE TCE, IEEE TII, IEEE TVT, IEEE ITS, IEEE Netw., IEEE Comm., IEEE WC, IEEE IoTJ, IEEE SJ, FGCS, JNCA, and ComCom. He has guided many research scholars leading to Ph.D. and M.E./M.Tech. His research is supported by funding from UGC, DST, CSIR, and TCS. He is an Associate Technical Editor of IEEE Communication Magazine. He is an Associate Editor of IJCS, Wiley, JNCA, Elsevier, and Security & Communication, Wiley. He is a senior member of the IEEE.

Joel J. P. C. Rodrigues [S'01, M'06, SM'06] is a professor at the National Institute of Telecommunications (Inatel), Brazil and senior researcher at the Instituto de Telecomunicações, Portugal. He received the Academic Title of Aggregated Professor in informatics engineering from UBI, the Habilitation in computer science and engineering from the University of Haute Alsace, France, a PhD degree in informatics engineering and an MSc degree from the UBI, and a five-year BSc degree (licentiate) in informatics engineering from the University of Coimbra, Portugal. Prof. Rodrigues is the leader of the Internet of Things research group (CNPq), Director for Conference Development - IEEE ComSoc Board of Governors, IEEE Distinguished Lecturer, Technical Activities Committee Chair of the IEEE ComSoc Latin America Region Board, the President of the scientific council at ParkUrbis – Covilhã Science and Technology Park, the Past-Chair of the IEEE ComSoc Technical Committee on eHealth, the Past-chair of the IEEE ComSoc Technical Committee on Communications Software, Steering Committee member of the IEEE Life Sciences Technical Community and Publications co-Chair, and Member Representative of the IEEE Communications Society on the IEEE Biometrics Council. He is the editor-in-chief of the International Journal on E-Health and Medical Communications and editorial board member of several high-reputed journals. He has been general chair and TPC Chair of many international conferences, including IEEE ICC, GLOBECOM, and HEALTHCOM. He is a member of many international TPCs and participated in several international conferences organization. He has authored or coauthored over 650 papers in refereed international journals and conferences, 3 books, and 2 patents. He had been awarded several Outstanding Leadership and Outstanding Service Awards by IEEE Communications Society and several best papers awards. Prof. Rodrigues is a licensed professional engineer (as senior member), member of the Internet Society, and a senior member ACM and IEEE.