International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 9, No.6 (Nov-2020)
http://dx.doi.org/10.12785/ijcds/090602

Persation: an IoT Based Personal Safety Prediction Model Aided Solution

Olasunkanmi Matthew Alofe[1], Kaniz Fatema [2], Muhammad Ajmal Azad [3] and Fatih Kurugollu [4]

[1,3,4] *Department of Electronics, Computing and Mathematics, University of Derby, Derby, UK*
[2]*Department of Computer Science, Aston University, Birmingham, UK*

Received 30 May 2020, Revised 21 Oct. 2020, Accepted 24 Oct. 2020, Published 1 Nov. 2020

Abstract: The number of attacks on innocent victims in moving vehicles, and abduction of individuals in their vehicles has risen alarmingly in the past few years. One common scenario evident from the modus operandi of this kind of attack is the random motion of these vehicles, due to the driver's unpredictable behaviours. To save the victims in such kinds of assault, it is essential to offer help promptly. An effective strategy to save victims is to predict the future location of the vehicles so that the rescue mission can be actioned at the earliest possibility. We have done a comprehensive survey of the state-of-the-art personal safety solutions and location prediction technologies and proposes an Internet of Things (IoT) based personal safety model, encompassing a prediction framework to anticipate the future vehicle locations by exploiting complex analytics of current and past data variables including the speed, direction and geolocation of the vehicles. Experiments conducted based on real-world datasets demonstrate the feasibility of our proposed framework in accurately predicting future vehicle locations. In this paper, we have a risk assessment of our safety solution model based on the OCTAVE ALLEGRO model and the implementation of our prediction model.

Keywords: IoT, Mobile Application, Vehicle Location Identification, GPS, Location Prediction

1. INTRODUCTION

Assaulting females have been frequently witnessed in the past few years around the world, with most of such incidents causing serious consequences to the victims [1, 2, 3, 4]. In most incidents, abnormal behaviour of the drivers is a commonly witnessed pattern such as diverted routes than normal, and the vehicle has been on the move while the assaults were taking place. For example, in January 2018, an abducted student forcefully driven away in her car from a car park in Atlanta, Georgia was able to prevent a more sinister ending to her ordeal by sharing her location with a third party and exchanged messages before the attacker seized the phone [5]. Furthermore, a quick response from a rescue team in San Jose, California helped to reunite a 5-year-old, who was in her father's vehicle when snatched, with her family in October 2018 [6]. In December 2018, a victim managed to escape assault in Berkshire, after she was forced into the boot of her car [7]. In February 2019, a 12-year-old girl suffered the same fate while waiting for her mother in the parking lot of a mall in Indiana [8]. Despite saving the lives of some of the victims, some had a more sinister ending, or death on some occasions. One of the survivors was saved because she was able to share her location with a third party, but not before some level of damage was done. The damage done to these victims would have been prevented earlier if there were any means to track and predict the location of the moving vehicle and inform the responder immediately. These scenarios do not share identical factors. In some cases, the speed, bearing, and location of the vehicle change rapidly and continuously while some changes were consistent. For either scenario, the delay in administering help would be damaging. Therefore, the solution should cater to the following requirements:

The helping device should be easily accessible when the attack occurs instead of attempting to unlock her phone, which might be taken away by the attacker. Therefore, the solution should cater to the following requirements.

- The device has to be lightweight for easy mobility.
- The solution should provide a way to detect and predict the location of the victim or the moving vehicle.
- It needs to inform the third party to offer the quickest help.

Fulfilling these requirements would ensure prompt actions taken. The solution needs to provide an avenue to

E-mail: o.alofe1@unimail.derby.ac.uk, k.fatema@aston.ac.uk, m.azad@derby.ac.uk, f.kurugollu @derby.ac.uk

obtain location information from the user, predict location from the obtained information, inform the third party with the predicted location, and create an intersection path for help to be offered by the third party. The contributions of this paper are:

- Proposing an IoT based personal safety solution model for ensuring prompt help for victims attacked in moving vehicle
- Reviewing solutions available for location tracking and personal safety
- Implementing the 1st and 2nd version of our location prediction model as a step towards the implementation of the safety solution model.
- Risk assessment of the safety solution model based on OCTAVE ALLEGRO.

The rest of the paper is organized as follows: Section 2 which is related works addresses the review of location tracking and personal safety solution reveals the related work and Section 3 explores the proposed personal safety solution. Section 4 is the review of the location tracking and prediction algorithm. Section 5 explores various risk assessment approach. Section 6 provides discussion about OCTAVE ALLEGRO risk assessment methodology deployed. Section 7 presents the Implementation of the location prediction model and Section 8 concludes with the results and discussion.

2. RELATED WORK

In this section, we are reviewing solutions available from the literature on providing help for victims attacked in moving vehicles and personal safety solutions for individuals. Shinde et al., [9] have presented an IoT based solution to notify some pre-saved numbers via text message in case of an accident. Although the text message provides the current location of the accident, it does not detect the location of a moving vehicle. Sharma et al., in [10] have presented an ARM7 processor-based safety device, which activates GPS location tracking and sends text messages to the responder with a single button press. Although it activates GPS tracking, it does not predict the location of a victim in a moving vehicle and does not find the nearest police station to take prompt action to save the victim. Furthermore, the device is heavy to carry and may not be suitable to carry all the time. The safety solution presented by Bhavale et al., [11] alerts pre-registered phones with the captured images. However, in a panicking attack, the biggest concern would be to send alerts in the quickest possible way. The bus- monitoring unit used will need to be pre-installed in the bus to activate tracking which may not be a practical expectation.

A wearable device was proposed by Pawar et al., [12] consisting of a microcontroller, Raspberry pi, GPS and Global System for Mobile communications (GSM) module. Readings are continually taken from the sensors and compared against assigned threshold values. Computational overhead is excessively consumed with continuous tracking and comparison of readings from sensors.

Monisha et al., [13] proposed an ARM controller incorporated with GSM, GPS, Bluetooth, and RF detector and powered by 12V for the controller. The device gets activated by pressing the emergency button and sends out messages containing location to a pre-set number. The proposed method uses a hardware device that is too heavy to carry by the individual; furthermore, access to the device is required for activation, which is disastrous in situations where the mobile is not accessible.

Choudhary et al., proposed a safety device [14] which consists of a variety of sensing units such as heartbeat sensor, temperature sensor, and a push button. It also uses ATmega8L, GPS and GSM modules, flashlight, and a taser. The device fetches heartbeat and temperature reading and compares the readings against a set threshold. If there is a variation, the device would be activated, and a message containing the location would be sent to the police alongside known personals with the help of the GSM module. The proposed method may result in false positives as the heartbeat and temperature readings may spike due to other reasons, which are not detrimental to the individual. A false message is sent to the police and known personals to initiate emergency protocols, which leads to the waste of resources.

In the bid to offer easily accessible and less cumbersome personal safety solutions, smartphone applications were developed that are capable of harnessing the modules present in the smartphone to track user's location and send alerts to third party during distress. Safetipin is a smartphone personal safety application that helps users make informed decisions about visiting an area and location tracking of the user. The app operates by providing a safety score for the intended area of visit based on disturbance and risk within the area and alternate routes are displayed to the destination. Tracking of the user can be done when the user invites a third party to track their location. Street smart is another smartphone personal safety application that allows users to make informed decisions about an area before visiting by providing articles and reviews about the safety level of an area of interest by holding the camera at the location. The safety level is determined from posted reviews and articles as positive, negative or neutral using sentimental analysis [15].

Life360 serves as a smartphone family locator application and personal safety application. The application is used for tracking the location of the user and provide the location for wellbeing centres that could be required during distress. The concern remains privacy issues regarding the location information of the user [16].

Vithu is a smartphone personal safety application that alerts selected third party when the user is in distress.

When the cycle of operation is initiated by the double press of the activation button, an SMS message with the location information of the user is sent to the third party every two minutes to track the trajectory of the user [17].

BSafe is a smartphone personal safety application that alerts selected guardians to inform them that the user is in distress. The alert contains location information of the user at the point of distress and phone call is placed to one of the selected guardians with tracking of the user possible. Guard works similarly with placing calls with the name of the user, present location and alert of help required by the user to the third party. The requirement for the use of the app is cumbersome and rigorous [18].

Streetsafe measures up as a smartphone personal safety application that operates by sending alerts for users in distress based on four features. These features are high volume alarm is initiated, the current location is uploaded to the user's Facebook account, SMS sent to preferred associates in the area and lastly call is placed to the user's chosen emergency number. Fightback is a smartphone personal safety application that is similar to Streetsafe. It allows user to send alerts for users in distress making use of features such as e-mail, GPS, SMS and GSM and track location of user on map [19].

The reviewed solutions show their ability to track the present location of devices, send alert to third parties but do not predict the future location. The proposed personal safety solution model, presented in the following section, is enriched with the capacity to predict location using acquired location information.

3. PROPOSED PERSONAL SAFETY SOLUTION (PERSATION)

This paper provides a possible solution for helping victims in violent attacks in moving vehicles, where prompt response is crucially needed for saving the victim. In the implementation of the solution, as seen in Fig. 1, the problem can be separated into two parts. The first part of the solution involves finding a wearable that can work as a panic button, when pressed it can communicate with the gateway, which could be a GPS enabled mobile device held by the victim that communicates through the internet to a cloud-based application server to send help alert and the location information to third party and nearest responder. Once the button is activated, the second part of the solution will be activated, which involves using a suitable solution for tracking and predicting the location of moving vehicles and informing the nearest first responder. To accomplish tracking and prediction, implementation of state-of-the-art location tracking and prediction technologies is required and empowered with geographical data delivery service to aid the identification of the nearest police station. This way prompt action can be taken and the victim might be saved timely. The novelty of our proposed model lies in the usage of prediction capacity for providing timely help in a crucial situation.

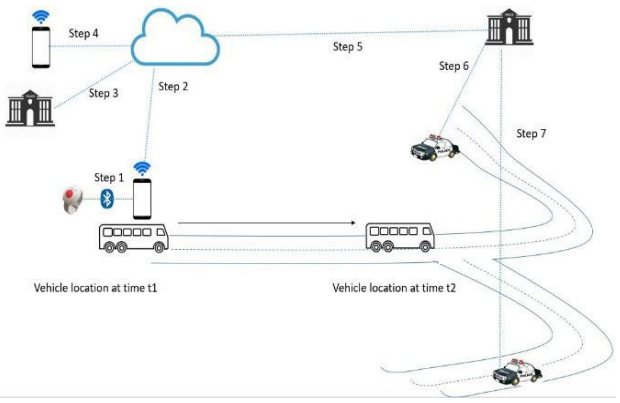

Figure 1. Personal safety solution scenario

Therefore, as a step towards the implementation of our proposed model we have first reviewed location tracking systems and location prediction algorithms in Section IV. We then have implemented the 1st version of our location prediction system by using a simple method for gathering location data and using prediction algorithm for location prediction.

4. REVIEW OF LOCATION TRACKING SERVICE AND PREDICTION MODEL

A. Selecting Location Tracking Service

Location based tracking service (LBS) offers a medium where data about a user can be collected in a coordinated and systematic manner and provides the user with the capability to find their bearing, and find other locations using semantic information about the present location and immediate environment. This service combines mobile services, location awareness, internet and GPS in the collection of a new layer of client's data and authentic data [20, 21, 22, 23, 24] Various technologies are involved in the operation of LBS with positioning technology, which handles the accurate location of the user regarded as the most important. The other technologies required by LBS are application technology that consists of two elements and deals with the presentation of information to the user. Geographic data that renders structures such as road network and manage data of the point of interest, and communication data for the transmission of user's location to the control centre for the provision of necessary service [25].

The delivery service of LBS can be categorized mainly as time- based delivery service, and distance-based delivery service. Time-based delivery service updates location information periodically to maintain high location tracking accuracy while Distance-based delivery service updates location information based on distance [26].

B. Review of Prediction Model

Time series analysis is a method used to gain insight into time series data about statistical patterns in the data and develop a suitable model for forecasting events. The most widely used linear time series models are AutoRegressive (AR) and Moving Average (MA). Autoregression is a model that relies on the dependent relationship between past period values and some degree of lagged observations to predict future values. Moving Average uses dependency within the dataset to provide output that depends linearly on the current and past values of stochastic distribution.

1) AutoRegressive Integrated Moving Average (ARIMA)

Autoregressive Moving Average (ARMA), which is suitable for univariate time series combines both AR and MA. The forecasted value by the model is a linear combination of past observations and random error with a constant term. The model is suitable for stationary time series data but comes short with data exhibiting non-stationary trends and seasonal patterns. AutoRegressive Integrated Moving Average (ARIMA) implements an integrated model that uses differencing to account for the establishment of stationarity. The suitability of the ARIMA model for the dataset is based on the exploration of trends and seasonality features in the dataset. ARIMA uses the ARIMA(p,d,q) notation based on the three models incorporated in the model (AR, integrated and MA). the p stated in the notation represents the AR that indicates the lag present in the stationarised series, the d stands for the integrated model that indicates the differencing required to attain stationarity and the q stands for the MA that indicates the lagged forecast errors in the series.

ARIMA time series model was introduced by Box and Jenkins. The model uses sets of activities to identify, estimate and diagnose the ARIMA algorithm suitable for time series data. ARIMA model forecast time series data by accounting for growth/decline pattern in the data with the Auto-Regressive part, rate of change of growth/decline with the Integrated part and the moving average to account for the noise between consecutive points in the data [27, 28]. Time series is the non-deterministic model for sequential observation of data in relation to a trend or seasonality. Time series applies a model to historic facts from data and forecast futures value of the series making use of movement along with the data over a long period of time (Trend), fluctuations available in the data over a particular period of time (Seasonality) and autocorrelation to distinguish time series operation from other types of statistical analysis. Autocorrelation (ACF), partial autocorrelation (PACF), inverse correlation and cross-correlation are used to identify and specify the form of time series model [29, 30, 31].

The appropriate model for the series is identified by initially determining the degree of differencing required to remove gross features of seasonality and non-stationarity of the series. After differencing, the next step involves checking for autocorrelated errors using the ACF to determine the order of AR required and the lagged error to determine the order of MA. The ACF displays the correlation between past values and helps in determining the term to use for the time series. Positive autocorrelation (PACF) at the first lag indicates that AR model can be used and Moving Average (MA) model indicates the random jumps for calculating error in subsequent periods within the plot.

ARIMA (p, d, q) is the standard notation used to indicate a specific model used by ARIMA where p is a number of lagged observations to be taken in, d is the degree of differencing and q is the size of the moving average window [25, 26]. The ARIMA equation after combining AR and MA becomes

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ..+ \beta_p Y_{t-p} \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1 \varepsilon_{t-1} + ..+ \phi_q \varepsilon_{t-q} \quad (1)$$

where it can be translated in words as: Predicted output Y_t is the addition of the Constant α with Linear combination Lags of input Y up to p (number of observations) lags and Linear Combination of Lagged forecast errors up to q (moving average) lags.

2) Regression Tree Ensemble

This is a predictive model composed of a weighted combination of multiple regression trees. Ensemble methods combine several base model decision trees classifier to provide an optimal predictive model and increase the accuracy of the model. The model aims at constructing a linear combination of various models and attempt using the combinations for the improvement of the predictive performance of a model fitting technique. There are two approaches for the model bagging or bootstrap aggregating and boosting.

a) BAGGing, or Bootstrap AGGregating

The model combines Bootstrapping and Aggregation into one model which works on improving unstable estimation or classification schemes. Bagging is a variance and Mean Squared Error (MSE) reduction technique that is effective for the improvement of the predictive performance of regression or classification trees. Given a sample of data, multiple bootstrapped subsamples are pulled. For each of the bootstrapped subsamples, a decision tree is formed. After Decision Trees have been formed for all the subsamples, an algorithm is used for the aggregation of the decision trees for the most efficient predictor.

b) Boosting

Boosting is a sequential process that adjusts the weight of an observation based on the last classification. The first algorithm is trained on the entire dataset and subsequent algorithms are built by fitting the residual of the previous algorithm. The principle of the model is to decrease bias

error to build strong predictive models. The prediction uses a weighted majority vote (classification) or a weighted sum (regression) to produce the final prediction.

5. RISK ASSESSMENT APPROACH

The compounding of a framework, models and methodologies knowledge for the assessment of security and privacy for the Internet of things device is crucial in integrating and examining the cyber risk standards and governance to understand the risk faced by devices and IoT networks. The adaptation of traditional cybersecurity standards indicates the need for the identification of specific IoT cyber risk vectors and integrated into a holistic cyber risk impact assessment model. To reduce cyber risk in cloud technologies, proper design of cloud architecture is maintained between the cloud services and devices connected to it [32]. This scenario involves the interaction between humans and technology in providing real-time feedback that demands the security of data in transit.

The majority of the established framework applies the quantitative approach to measuring cyber risk. One widely accepted framework is the Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) methodology. The goal of OCTAVE methodology is to help organisations with operational and strategic mediums to perform information security risk assessment. OCTAVE works to connect organisation and their operational point of view activities of information security risk management. There are three publicly available OCTAVE methodologies. The first methodology introduced is the OCTAVE-consistent methodology that is defined by the implementation guide and training. Series of workshops facilitated by an interdisciplinary analysis team from various units that are in the organisation connect the organisation and the operational point of view activities of information security risk management [33]. The method is designed for large organisations with multi layered hierarchy, independence in performing vulnerability evaluation and interpret the results when maintaining the organisation computing infrastructure. The method can be used to tailor the approach to suit the distinct environment they operate. This method is performed in three phases, the first phase is the identification of assets and protection strategies presently been implemented, the second phase evaluates the infrastructure to supplement the analysis performed in the first phase. Risk mitigation plans for critical assets are developed after performing risk identification activities. In providing OCTAVE methodologies for small organisations, Technology Insertion, Demonstration, and Evaluation (TIDE) developed OCTAVE-S. The criteria are similar to OCTAVE and operate in with the same three phases except OCTAVE-S does not depend on formal knowledge workshop to obtain information. OCTAVE-S does not require extensive examination into the organisation infrastructure and helps practitioner address a wide range of risk which they have no familiarity about them. The risk identification, assessment, and mitigation processes are developed based on the collaborative aspect from an interdisciplinary perspective. With the collaboration strengthening the quality of risk assessment and mitigation, there are limitations in the interdisciplinary collaboration such as varying levels of expertise in threat evaluation, disparity in communication channels, practices, and intended efforts [34].

With the landscape of the information security risk changing coupled with the above limitations and the change in the required capability to manage the risks, the development of a new approach was inevitable to accommodate these changes. OCTAVE ALLEGRO adopts a different approach to organisation information technology environment and information assets than the other OCTAVE methodologies [35]. This method maps information assets to all containers where they are stored, transported or processed. Unlike the other two methodologies, OCTAVE ALLEGRO streamlines and optimizes the process of assessing the security risk of an organization and eliminates the use of vulnerability tools for threat identification by introducing the concept of an information risk environment map. The map help user defines all places information has been stored, transported or processed. The map serves as baseline documentation of the risk environment for the asset and helps establish boundaries of the threat environment and scope of risk assessment. The method uses a value known as relative risk score derived from the evaluation of qualitative description of risk probability combined with the prioritization of the organizational impact of risk in terms of the organization's risk measurement criteria. Mitigation guidelines, and specific strategies are considered for each container where the asset resides [34].

The methodology has four distinct activity areas carried out in various steps. The methodology establishes quantitative measures that are used as criteria for the evaluation of risk effect and serves as the foundation of information risk assessment, identification of the location of the asset and possible situations that threatens the asset, identification of threats and risks that could impact the asset, analysing the discovered risk and selecting the mitigation approach. The methodology does not provide details of methodology implementation and does not adequately address the impact of risk on assets. The methodology can serve as the starting point for risk assessment [34].

Threat Assessment & Remediation Analysis (TARA) is another system level quantitative methodology for the identification, prioritization and response based on three activities. Cyber Threat Susceptibility Analysis (CTSA) to assess the susceptibility of the asset to threats, Cyber Risk Remediation Analysis (CRRA) to determine best-suited countermeasures, and data and tools development to deliver recommendations for informed decisions. The

methodology targets the most crucial risk and offers a complementary form of protection [36].

Common Vulnerability Scoring System (CVSS) is a combination of the qualitative and quantitative framework for providing metric groups for assigning metric values to vulnerabilities and allocating cyber risk into levels and calculate an overall risk level. The methodology applies numerical values ranging from 0 – 10 to indicate the severity of vulnerability along with 3 color-coded levels to differentiate among the actual system. During result simplification, different vulnerabilities can produce the same level and similar score values that can increase the number of the colors in the color-code to enhance the visibility of different score values. The worthiness of the score can be faulted based on the basic mathematical formalism [37].

The Capability Maturity Model Integrated (CMMI) focus on the enterprise risk and development life cycle risk by integrating five levels of the Capability Maturity Model (CMM). The methodology identifies vulnerabilities and does not indicate ways to address the identified vulnerability [38].

The National Institute of Standard and Technology (NIST) provides the combination of risk assessment and risk management with a collection of standards and guidelines when combined with automated tools, aims to improve the security infrastructure [39].

The Factor Analysis of Information Risk Institute (FAIR) approach is focused on impact assessment and complementary with existing risk frameworks. The methodology address weaknesses of ISO standardized approach and creates a standardization reference for compliance. The FAIR model enhanced the deployment of RiskLens. This is a software-as-a-Service (SaaS) platform for the management and quantification of cyber-risks. The methodology is a quantitative model for cybersecurity and technology risk with integrated advanced quantitative risk analytics, best-practice risk assessment and reporting workflows. Another quantitative approach is the Cyber VaR (CyVaR), which shares complexity similarity with RiskLens but allows for the addition of a new type of risk [40, 41, 42].

6. RISK ASSESSMENT RESULT

This section aims to collect security threats discovered from information security risk assessments with OCTAVE Allegro methodology. OCTAVE ALLEGRO focus on information asset in various contexts such as how the assets were used; their exposure to threats, vulnerability and disruptions; where they are stored, transported and processed. The approach implores the user to explicitly consider the implication of risk consequence on security requirements and risk mitigation. The requirement of the approach is to allow focus on assets by ensuring they are selected through a systematic and consistent review process. The approach streamlines and improves threat identification and risk mitigation process without extensive risk assessment knowledge required. OCTAVE ALLEGRO development minimizes certain features which contribute to the ease of use requirement, minimal resource commitment and approach usability through fewer and more focused activities directed towards risk management. The adoption of scenario questionnaires by the approach instead of threat trees used by the earlier version of OCTAVE further help user in the identification of threats associated with information assets. OCTAVE ALLEGRO uses an information asset risk worksheet to capture the relevant information regarding specific risk for an information asset. The worksheet reduces documentation, organization, and data manipulation required to perform the risk assessment and help in producing a concise view of risk. A simple quantitative analysis of risk introduces a relative risk score that is computed on the worksheet using threat and impact information associated with risks captured. The introduced relative risk score is used to compare the significance of individual risks and computed from the combination of the risk probability qualitative description and the prioritization of risk impact based on the organization's risk measurement criteria.

The worksheet is used to compute relative risk scores for the assets are based on the component of the identified assets. Primary usage of the worksheet includes the identification of assets; determining the area of concerns accompanying the identified assets; threats and risks associated with the assets. From the relative risk score shown in table 2, identity theft and privacy violation possess higher risk to the framework as indicated by their high score compared to the other risks. Based on the relative high score of these two risks, security and privacy of user's information remains of utmost importance to maintain efficiency of the framework. Privacy violation could lead to tracking and monitoring of user activities and lead to a replay of the user's previous activities when the user is in distress, the event can be masqueraded by replaying user's activities instead of alerts been sent to the third party.

Based on the identified assets and area of concern related to the assets, associated threats are identified. Table 2 shows the threats associated with the assets and area of concern. The threats associated with these assets undermines the efficiency of the framework as user impersonation could lead to prevention of mechanism activation when the user is in need, device spoofing and data spoofing could result in the system receiving fake data or device information leading to waste of resources and inability to provide help to victims. Location information plays a crucial role in the framework forming the basis required for the prediction. In providing help for individuals in distress, accurate location information is essential. Threats jeopardizing the precision of location information reduces the chances of offering help that ensures the health and safety of the user.

TABLE I. INFORMATION ASSETS AND AREA OF CONCERN

Asset ID	Asset	Area of Concern
1	Personal Information	User's privacy
		Personal Identifiable Information
2	Device Information	Device configuration
		Data stored on device
		Network Topology
3	Location information	User's behavioural pattern
		Area of interest of the user
		User's movement
4	Log Information	Device operation
		User and device activities

TABLE II. INFORMATION ASSETS AND SECURITY THREATS

Asset ID	Asset	Threats
1	Personal Information	Data disclosure
		User impersonation
2	Device Information	Device spoofing
		Device breach/Theft
		Data spoofing
3	Location information	Tracking of user
		Disruption of the user's activity
		Monitoring of user
4	Log Information	Clearing of attack traces
		Map activities of service and application in the device

From the identified assets, reviewed area of concern and threats associated with the assets, risks are computed for the scenario. The relative risk score computed from OCTAVE ALLEGRO worksheet is based on the impact of the risk on the reputation and customer confidence, financial, productivity, safety and health, fines and legal penalties, user-defined impact area. Risks related with Asset ID 2 which includes data manipulation, data leakage and service denial/starvation could lead to disruption in delivering help during distress with inaccurate data been sent or service denial to prevent alerts been sent to the third party for help and poses a high risk to the productivity of the framework and safety and health of the user. Risks related to asset ID 1 and asset ID 3 pose a high risk to the safety and health of the user while risk related to asset ID 4 poses a high risk to productivity. The relative risk score associated with risk linked with asset ID 1 is based on their impact on the user. The health and safety component has high-risk value among the other components with financial and productivity components exhibiting medium risk impact on the asset.

TABLE III. RISKS AND COMPUTED RELATIVE RISK SCORES

Asset ID	Risks	Score
1	Privacy violation	25
	Identity Theft	32
2	Denial/Starvation of service	19
	Data leakage	22
	Data manipulation	12
3	Unauthorised app execution	19
	Interruption of activity	21
	Tracking of user	22
4	Exposure to extra service	13
	Loss of information	14

Table 4 shows different mitigation for the threats identified in table 2. For the framework, the major mitigation approach is to ensures the health and safety of the user. The mitigation approach stated in the table ensure that the safety of the user is not jeopardized at any time and especially when the user is in distress. The mitigation techniques improve the efficiency of the framework and provide adequate protection for users calling out for help.

TABLE IV. POSSIBLE MITIGATION APPROACHES

Asset ID	Mitigation Approach
1	Ensuring a good understanding of user privacy concerns
	Encryption of data
2	Device hardening
	Physical security of the device
	Device firmware update
3	Monitoring device security permissions
	Secured means of communication (VPN)
	Permission restriction
4	Secured system configuration
	Device hardening

7. IMPLEMENTATION OF ARIMA PREDICTION MODEL

The tool that is used for the implementation of the first model is SAS/ETS® and the data used was obtained using the location-based tracking services of a mobile device in a moving vehicle. From the review of section IV, the essential variables latitude, longitude and time used for the evaluated algorithm are collected by LBS. The algorithm gets an error when there is a disparity in subsequent time values or an empty value for any of the

variables. For the evaluation, the data was divided into two datasets. The first dataset which is made up of 88% of the original data is the data used for training the model and the second dataset is the test data.

Mean procedure test is the first test carried out. This test result shows the mean, standard deviation, minimum value and maximum value of the acquired data. The relevance of the test is to determine whether differencing is required based on the standard deviation result. If the value for standard deviation is insignificant i.e STD < 0.05, then differencing of the data is not required but when the value is significant, the Augmented Dickey-Fuller (ADF) test is carried out. The ADF test is based on the hypothesis that time-series data is non- stationary [27, 43] or has a significant standard deviation value.

Application of ARIMA procedure for the forecast of the next values. As discussed in section IV., this test is to determine the ARIMA (p,d,q) notation to be used for forecasting. Positive autocorrelation (PACF) at the first lag indicates that AR model can be used and Moving Average (MA) model indicates the random jumps for calculating error in subsequent periods within the plot. If the AR model is required there is a negative autocorrelation at Lag-1, a sharp drop in PACF after few lags and a gradual increase in PACF [44]. Other factors which help determine the most suitable model is the relatively small value provided by (2) for Akaike Information Criterion (AIC), (3) for Schwarz Bayesian Information Criterion (SBC) and the standard error value of regression (S.E. of regression)

$$AIC(p) = n\ln(\sigma^2_e/n) + 2p \quad (2)$$

$$BIC(p) = n\ln(\sigma^2_e/n) + p + p\ln(n) \quad (3)$$

The last test is to verify the efficiency of the result produced by the ARIMA model. The test involves the comparison of the acquired data from the model forecast and the test data obtained through LBS.

8. Implementation of Ensemble Regression Tree Model

The tool used for the implementation of the second implementation model is the regression learner app of MATLAB 2019a. The app can be used to train regression models for prediction of data, perform automated training for determination of the best regression model type, select features, specify validation schemes and assess results. The model types are including linear regression models, regression trees, Gaussian process regression models, support vector machines, and ensembles of regression trees.

After loading the dataset into the app, the validation method is to be selected to examine the predictive accuracy of the fitted models. Validation helps prevent overfitting, estimate the performance of the model and choose the best model.

The first type of validation is cross-validation, this selects the number of divisions to partition the dataset. If k division is selected, then the app:

- Partitions the data into k disjoint divisions
- For each fold, out-of-fold observation is used to train the model and model performance is assessed using in-fold data.
- Calculates the average test error over all divisions.

The method makes efficient use of all data and requires multiple fits which makes it suitable for small datasets.

The second type of validation is the holdout validation. For this validation type, a percentage of the data is reserved and used as the validation set. This type of validation is appropriate for large data sets as it segregates the data efficiently based on percentage.

If no validation is selected, there is no protection against overfitting. All the data are used for training and computing the error rate on the same data. The lack of test data makes the model performance for the estimation of new inaccurate and unrealistic.

Without any test data or validation data, the model can provide unrealistic estimates when used for new data. With the dataset used for the implementation, the cross-validation was selected as it suits the dataset.

9. Results and Discussion

From the result of the means procedure test as shown in Fig 2, the standard deviation for the latitude and longitude is insignificant with a value less than 0.05. The insignificant value of the standard deviation indicates that differencing is not required for the data and the ADF test would not be carried out.

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
lat	lat	120	52.9252574	0.0043273	52.9198600	52.9295710
lon	lon	120	-1.4913512	0.0077818	-1.4990940	-1.4816560
elevation	elevation	70	109.2222736	15.4281721	43.2273404	131.0873011
accuracy	accuracy	120	19.8872000	3.6176629	6.0000000	32.7580000
bearing	bearing	8	181.6297544	73.5573774	72.7008400	301.7622000

Figure 2. Means procedure

The next test is to determine the specific ARIMA model to be used. The major factors for determining the most suitable model to used are the ACF, PACF graphs, AIC and SBC values.

As shown in Fig. 3, the ACF obtained shows a gradual decrease across the lag which indicates that AR model can be use and the PACF plot showing a sharp drop after a few lags and gradual increase across the lag indicates the MA model can be used. Based on these findings, various models of ARIMA was implemented to determine the AIC and SBC value and determine the best model for forecasting.

The values of AIC and SBC of various models of ARIMA are indicated in Table 5. The models explored have low AIC and SBC which indicates that they are all suitable for the research. The best model to use for prediction is the model with the lowest value of AIC and SBC. From Table 5, the lowest AIC and SBC value are 1735.68 and 1721.74 for latitude and 1478.09 and 1464.15 for longitude respectively, indicating the most suitable model is ARIMA (2,0,1) where 2 represent the number of lags, 0 is the degree of differencing, and 1 is the order of moving average.

TABLE V. AIC AND SBC RESULT OF ARIMA MODEL

ARIMA Model	Latitude		Longitude	
	AIC	SBC	AIC	SBC
ARIMA(1,0,0)	1627.95	1622.38	1419.60	1414.02
ARIMA(0,0,1)	1122.44	1116.87	981.49	975.92
ARIMA(1,0,1)	1696.50	1688.14	1445.14	1436.78
ARIMA(1,1,0)	1627.06	1621.49	1417.11	1411.53
ARIMA(0,1,1)	1122.44	1116.87	981.49	975.92
ARIMA(1,1,1)	1696.50	1688.14	1445.14	1436.78
ARIMA(1,1,3)	1712.58	1698.65	1455.00	1441.07
ARIMA(2,0,0)	1724.07	1715.70	1454.37	1446.01
ARIMA(2,1,3)	1459.45	1442.72	1459.45	1442.72
ARIMA(2,0,1)	1735.68	1721.74	1478.09	1464.15

For the last test, the predicted values from the model are compared with the test data from the acquired data. Fig 4 and Fig 6 shows the trend of the latitude training data for latitude and longitude respectively while Fig 5 and Fig 7 shows the difference between the test data and the predicted data. The pattern of the difference between values of the predicted data seems to be constant and swaying in a single direction, either increasing or declining gradually unlike the test data, which decreases, stagnates and increases over the trend. The model determines a pattern from the training data and uses the data to derive the predicted data.

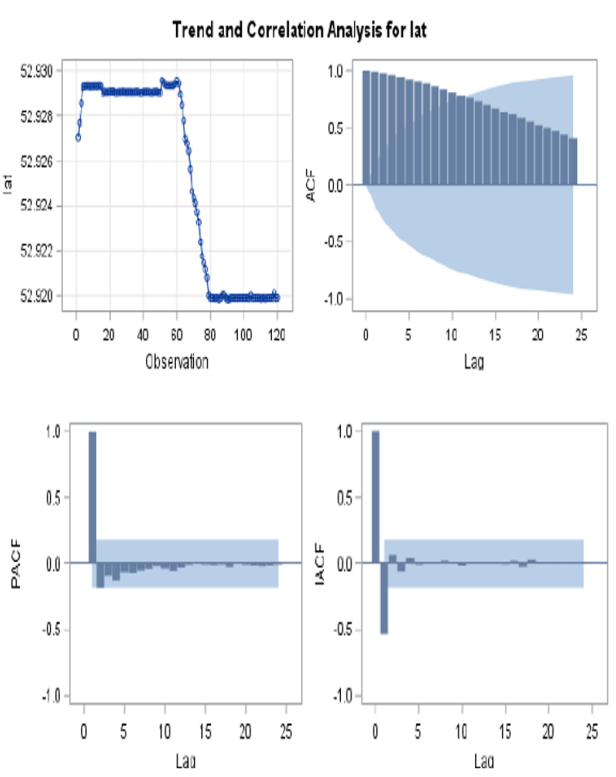

Figure 3. Correlation Analysis of training data

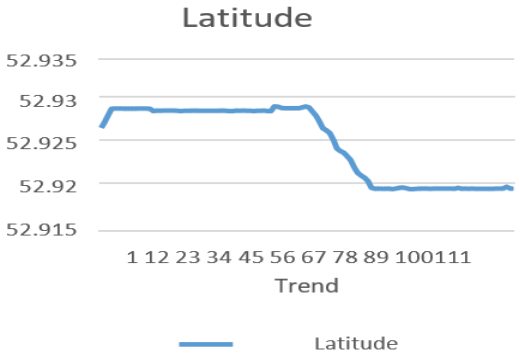

Figure 4. Latitude training data

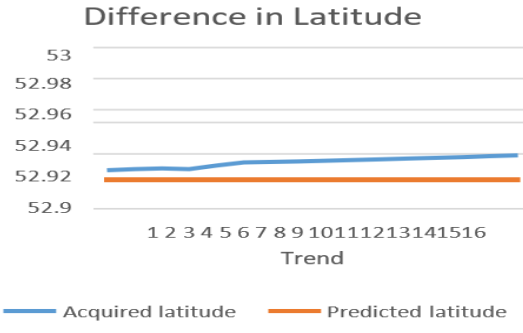

Figure 5. The difference in latitude data

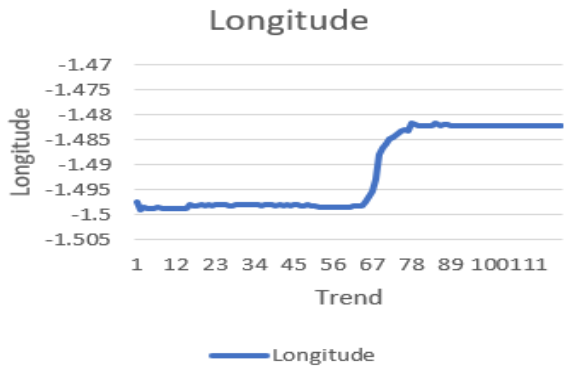

Figure 6. Longitude training data

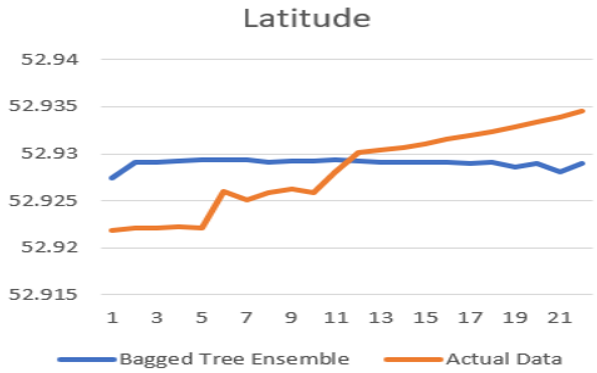

Figure 8. The difference in Bagged tree latitude data

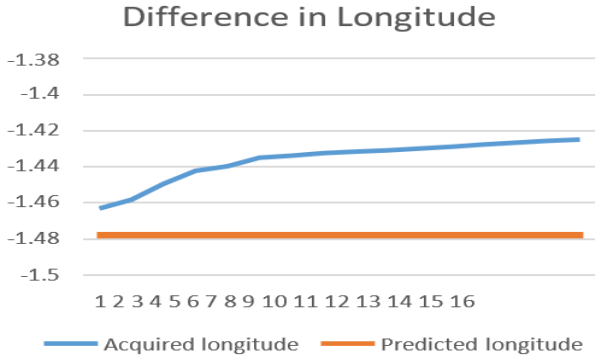

Figure 7. The difference in longitude data

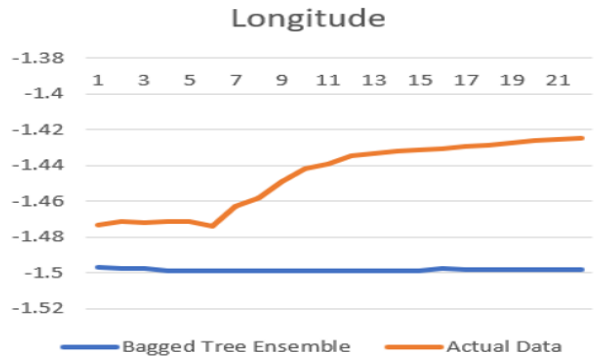

Figure 9. The difference in Bagged tree longitude data

This second implementation model uses a weighted combination of multiple regression trees to construct a linear combination that improves the predictive performance of the model.

The bagged tree ensemble aggregates the decision tree for the most efficient predictor. The figure shows the predicted values use trend which is similar to the trend of the training data. The similarity in trend shows the efficiency of the model. For the latitude in Fig 8, the trend of the values between the verification data and predicted values shows the same initial trend of an upward slope with an inconsistent upward and downward slope along with the trend. The longitude values show a great disparity along with the trend as indicated in Fig 9. Unlike the actual data that has a consistent downward slope, the predicted data displayed a conspicuous upward slope along with the trend.

The boosted tree uses a sequential process of weight adjustment and built on the fitting of the successive algorithm on the previous one. The sequential fitting can be observed in the consistent intervals shown in the figure as the predicted values seem constant across the trend. Fig 10 shows the latitude values for both the predicted value and the test data shows similar movement along with the trend but shows a significant difference between the test data value and the predicted data values. Longitude values show a significant difference in values between test data and predicted data and the movement along the trend shows significant difference along with the trend as shown in Fig 11.

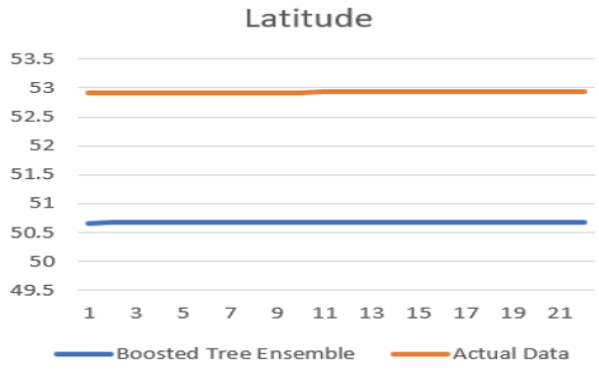

Figure 10. The difference in Boosted latitude data

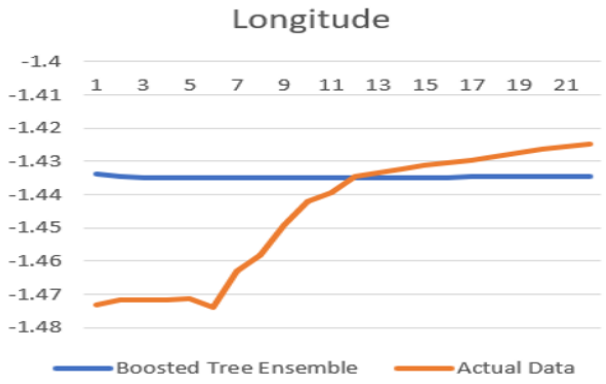

Figure 11. The difference in Boosted longitude data

10. CONCLUSION

The solution provided for our problem space is believed to help the prevention of violent attacks in high-speed vehicles and provides easily accessible devices to aid the communication of the on the move violent attacks to a third party and the nearest police station to save the lives of victims. This paper takes into consideration the prediction model for the second part of the proposed IoT solution. The first algorithm applied in this paper offers a consequential prediction based on the pattern of historical data. The results from the first implementation are predicted based on the pattern derived from the trend of the training data, the trend displayed from the result indicates the suitability of the model for forecasting slow movement along a straight path rather than random movement.

The result of the second implementation unlike the first implementation does not follow the trend of the data rather it manifested a curve based on the weighted combination of multiple regression trees.

To maximize the efficiency of the framework and prevent sabotage by insider or external vulnerabilities, risk analysis of the framework was performed using OCTAVE ALLEGRO for the identification of various assets available in the scenario, area of concerns regarding these assets, threats and mitigation approach deployable in the framework. The risk assessment looks to prevent undermining the performance of the framework.

As a next step of the research, we would look into adopting deep learning algorithms that best suit the purpose of forecasting the next location of a moving vehicle with random motion.

REFERENCES

[1] Shock and outrage over India Delhi bus gang rape, BBC, 2012.

[2] Bangladeshi law student killed after five men gang-raped her on bus, 2017.

[3] K. Lewis, Mother gang-raped on bus as two-week old baby dies in attack, Independent Digital News and Media, 2016.

[4] I. Qureshi, India woman raped in moving bus in Karnataka, BBC, 2015.

[5] A. McSorley, Who is Jastine Valdez? Wicklow woman abducted by Dublin dad-of-two Mark Hennessy, 2018.

[6] K. Utehs, ABC7 NEWS EXCLUSIVE: 5-year-old girl kidnapped during San Jose car theft, reunited with parents, 2018.

[7] Woman kidnapped in boot of her own car in Emmer Green, BBC, 2018.

[8] E. Longnecker, 12-year-old calls 911 during frightening ride in stolen car, 2019.

[9] P. Shinde, P. Taware, S. Thorat, T. Waghmare and A. Kadam, "Emergency Panic Button," International Journal of Scientific & Engineering Research, vol. 3, p. 3, 2012.

[10] S. Sharma, F. Ayaz, R. Sharma and D. Jain, "IoT Based Women Safety Device using ARM 7," 2017.

[11] M. D. M. Bhavale, M. P. S. Bhawale, M. T. Sasane and M. A. S. Bhawale, "IOT Based Unified Approach for Women and Children Security Using Wireless and GPS," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume, vol. 5, 2016.

[12] R. Pawar, M. Kulabkar, K. Pawar, A. Tambe and P. Smita Khairnar, "Smart Shield for Women Safety," International Research Journal of Engineering and Technology (IRJET) e-ISSN, vol. 05, no. 4, pp. 56-2395, 4 2018.

[13] D. G. Monisha, M. Monisha, G. Pavithra and R. Subhashini, "Women safety device and application-FEMME," Indian Journal of Science and Technology, vol. 9, no. 10, 2016.

[14] Y. Choudhary, S. Upadhyay, R. Jain and A. Chakraborty, "Women Safety Device (Safety Using GPS, GSM, Shock, Siren and LED)," International Journal of Advance Research in Science and Engineering, vol. 6, no. 5, p. 413–421, 5 2017.

[15] P. Kartik, S. Jose and G. K. MK, "Safetipin: A Mobile Application Towards Women Safety," Rajagiri Journal of Social Development, vol. 9, no. 1, pp. 5-12, 2017.

[16] Life360 - Feel free, together..

[17] M. Umar, Akash and Naveen, VithU App: A Woman Safety App by Gumrah, 2018.

[18] K. Sharma and A. More, "Android Application for women security system," International Journal of Advanced Research in Computer Engineering & Technology, vol. 5, no. 3, pp. 725-729, 2016.

[19] K. J. M. Baker, The Street Safety App for Proactive and Paranoid Woman, Jezebel, 2013.

[20] S. Khan, W. Ahmad, R. Ali and S. Saleem, "A Research on Mobile Applications for Location Tracking through Web Server and Short Messages Services (SMS)," VFAST Transactions on Software Engineering, vol. 7, no. 2, pp. 12-17, 2 2015.

[21] C. Pontikakos, M. Sambrakos, T. Glezakos and T. Tsiligiridis, "Location-based services: A framework for an architecture

design," Neural Parallel and Scientific Computatiions, vol. 14, no. 2/3, p. 273, 2006.

[22] J. O. Aasha, S. Monica and E. Brumancia, "A tracking system with high accuracy using location prediction and dynamic threshold for minimizing SMS delivery," in 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), 2015.

[23] R. Kolvoord, K. Keranen and P. Rittenhouse, "Applications of location-based services and mobile technologies in K-12 classrooms," ISPRS International Journal of Geo-Information, vol. 6, no. 7, p. 209, 2017.

[24] Y.-C. Lai, J.-W. Lin, Y.-H. Yeh, C.-N. Lai and H.-C. Weng, "A tracking system using location prediction and dynamic threshold for minimizing SMS delivery," Journal of Communications and Networks, vol. 15, no. 1, pp. 54-60, 2013.

[25] N. Chan and H. Lars, "Introduction to location-based services," Lund University GIS Centre, p. 1–12, 8 2003.

[26] P. Gupta and S. S. Sutar, "Study of Various Location Tracking Techniques for Centralized Location, Monitoring & Control System," IOSR Journal of Engineering, vol. 4, no. 03, pp. 27-30, 3 2014.

[27] M. Kumar and M. Anand, "An application of time series ARIMA forecasting model for predicting sugarcane production in India," Studies in Business and Economics, vol. 9, no. 1, pp. 81-94, 2014.

[28] P. Chen, H. Yuan and X. Shu, "Forecasting crime using the arima model," in 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008.

[29] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock price prediction using the ARIMA model," in 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014.

[30] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," arXiv preprint arXiv:1302.6613, 2013.

[31] C. Liu, S. C. H. Hoi, P. Zhao and J. Sun, "Online arima algorithms for time series prediction," in Thirtieth AAAI conference on artificial intelligence, 2016.

[32] P. Radanliev, D. C. De Roure, J. R. C. Nurse, R. M. Montalvo, S. Cannady, O. Santos, P. Burnap, C. Maple and others, "Future developments in standardisation of cyber risk in the Internet of Things (IoT)," SN Applied Sciences, vol. 2, no. 2, p. 169, 2020.

[33] C. Woody, J. Coleman, M. Fancher, C. Myers and L. Young, "Applying OCTAVE: practitioners report," 2006.

[34] R. A. Caralli, J. Stevens, L. Young and I. W. R. Wilson, "Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process," 2007.

[35] C. J. Alberts and A. Dorofee, Managing information security risks: the OCTAVE approach, Addison-Wesley Longman Publishing Co., Inc., 2002.

[36] J. Wynn, J. Whitmore, G. Upton, L. Spriggs, D. McKinnon, R. McInnes, R. Graubart and L. Clausen, "Threat assessment & remediation analysis (tara): Methodology description version 1.0," 2011.

[37] P. Mell, K. Scarfone and S. Romanosky, "A complete guide to the common vulnerability scoring system version 2.0," in Published by FIRST-forum of incident response and security teams, 2007.

[38] C. W. I. C. M. Model, "Integration (CMMI)®?| CMMI Institute," CMMI Institute, 2017.

[39] C. NIST, "Cybersecurity framework| NIST," NIST Website, 2016.

[40] N. Sanna, How FAIR Can Ensure The Success of COSO Risk Management Programs, 2017.

[41] Risk Analytics Platform | FAIR Platform Management | RiskLens.

[42] R. Shaw, V. Takanti, T. Zullo, M. Director and E. Llc, Best Practices in Cyber Supply Chain Risk Management Boeing and Exostar Cyber Security Supply Chain Risk Management Interviews, 2017.

[43] J. Panneerselvam, L. Liu and N. Antonopoulos, "InOt-RePCoN: Forecasting user behavioural trend in large-scale cloud environments," Future Generation Computer Systems, vol. 80, pp. 322-341, 2018.

[44] Sangarshanan, Time series Forecasting - ARIMA models, Towards Data Science, 2018.

[45] ISO - ISO 31000 — Risk management.

Alofe Olasunkanmi Matthew presently working towards PhD in Cybersecurity at the University of Derby having received his M.Sc in cybersecurity from the same university. His current research is focused on relating cybersecurity, Internet of Things and machine learning. His research interest includes cyber security, machine learning, Internet of Things, cryptography and encryption.

Dr. Kaniz Fatema is working as a Lecturer at Aston University, UK. Previously she worked as a Senior Lecture at the University of Derby, UK. She has many years of experience of working as a Research Fellow at Trinity College Dublin and University College Cork, Ireland. She completed her PhD in Computer Science (Information Security) from the University of Kent, UK and MSc in Data Communications from the University of Sheffield, UK. She has almost a decade of research experience in the domains of Information and Cyber Security, such as, access control, data protection, compliance assurance for data protection regulation and Cloud Computing.

Muhammad Ajmal Azad received the Ph.D. (2016) degree in Electrical and Computer Engineering from the University of Porto, Portugal, and MS (2008) in Electronics Engineering from the International Islamic University Pakistan. He is a lecturer in Cyber Security at the University of Derby UK, before joining The University of Derby, he was research fellow (an equivalence of lecturer in the UK) in the department of computer science at The University of Warwick and research associate at Newcastle University. He also spent more than 5 years in the telecommunication company. His research interests include privacy-aware collaboration, reputation aggregation, privacy protection, privacy-aware outsourcing of network logs, and spam detection in a telecommunication network.

Fatih Kurugollu (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer engineering from Istanbul Technical University, Istanbul, Turkey, in 1989, 1994, and 2000, respectively. From 1991 to 2000, he was a Research Fellow with Marmara Research Centre, Kocaeli, Turkey. In 2000, he joined the School of Computer Science, Queen's University Belfast, Belfast, U.K., as a Postdoctoral Research Assistant. He was appointed as a Lecturer with Queen's University Belfast in 2003 and was promoted to a Senior Lecturer in 2011. He is currently a Professor of cyber security and the Head of the Cyber Security Research Group, University of Derby, Derby, U.K. His research interests include cyber security, multimedia security, big data analysis and AI, image and video processing applications, biometrics, and hardware architectures for image and video applications.