INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

*ISSN 2278 – 0211 (Online)*

# Intrusion Detection System on Android Phones Using Filter Base Feature Selection Algorithm

**Dr. Okoronkwo Matthew C.**
Senior Lecturer, Department of Computer Science
University of Nigeria, Nsukka, Nigeria
**Onyedeke Obinna C.**
Masters Student, Department of Computer Science
University of Nigeria, Nsukka, Nigeria

***Abstract:***
*With every day increase in the number of android phone applications, there is a huge increment in malicious activities in android phone. Intrusion detection system (IDS) is an effort gear towards monitoring and controlling such intrusion activities. This paper proposes a system that detects any illegal/malicious intrusions in android phone using filter based feature selections. The filter based feature selection approach uses the characteristics of data to select and evaluate the features thus; eliminating variables that are not in the dataset and ensures that the selection processes and results are statistically good. This paper adopted the object oriented analysis and design method (OOADM), and uses the approach to model real world processes, operations and data in a more flexibly, efficiently and realistically manner. The paper goal is to raise user awareness of the permission-based system of Android phones and the threat it pose and provide counter measure system for more secured operations. The proposed Android Intrusion Detection System (AIDS) will be downloaded from Google play store by users of Android phones to help in monitor detecting, filtering, authenticating, logging and mail alert of intrusion attempt to the user's phone.*

***Keywords:*** *Intrusion Detection System (IDS), Feature selection, filter base, AIDS*

## 1. Introduction

Intrusion is any action that endeavours to unsettle the confidentially, availability or integrity of a resource or the controlling apps. Because of high prevalence, IDS are provided to checkmate intrusions. IDS is a sort of security programming intended to provide caution when a person or thing is attempting to bargain data framework through pernicious exercises or security infringement.

IDS, works by checking apps framework; looking at vulnerabilities in the construction, the trustworthiness of records and directing investigations on known assaults. It frequently scans the Internet for most recent dangers which could result in attack. There are a number of approaches in IDS including wrapper approaches, filter-based approach among others. This project adopted the filter-base feature selection approach on android app.

The filter-based approach uses feature selection. In this approach selected metric are used to distinguish superfluous qualities, and channel out excess segments from a model. At that point factual measure that suits your information is connected to figure scores (result) for each component section. The segments are returned according to their component scores. By picking the correct features, you can conceivably improve the exactness and productivity of grouping.

By applying factual measures (detection rate, false alarm, accuracy rate), you can figure out which sections don't add to the precision of the prescient model (or may diminish the exactness of the model) and expel them before structure your model.

Filter based approaches are computed very quickly, and works by computing a mathematically interpretable quantity such as an information theoretic based metric. It explains exactly why a given feature is selected or not selected, and a physical reason for that selection. The wrapper approach is just a brute force approach, and you really do not know exactly why a feature is selected - other than that it helped the classification performance of some classifier on some dataset when used with some arbitrarily chosen parameters. Hence, the adoption of filter based approach.

Android was launched by Google and Open Handset Alliance in September 23, 2005, and has experience vast growth since inception because of its user friendliness, open source, ease of developing and publishing applications. It has become the most widely used operating system on Smart phones with an estimated market share of 81% in 2015. According to report about 432 million smart phones were sold with Googles Android OS making up 81.7% of the market followed by Apples iOS with 17.9% of the overall market share. The ubiquitous usage of Android OS has induced the burst of mobile application market. But frustrating malware assaults in Android phones is a flourishing exploration region with a lot of unsolved issues [1].

This application performs three principle capacities: Detection rate, Accuracy rate and false alert (false positive and false negative). Detection rate of the app signifies how effective the app can detect a malicious attack, Accuracy rate tells how the app can accurately identify a malicious activity and False *alarm* is a result that indicates there is an intrusion, when there is not. False Positive is the interruption recognized when there is no interruption and False Negative is unidentified intrusion when there is an intrusion.

## 2. Related Works

### *2.1. Feature Selection*

Feature selection has its root in factual learning hypothesis, which guarantees that the choice procedure and results are measurably stable. Feature selection is of vital significance in AI, because it improves the forecast presentation of AI models by wiping out boisterous factors, provides less difficult models that encourage better elucidation of the complex stochastic procedure, removes the expense of huge measure of trials, and recognize factors that can be examined intently for stimulation [2].

### *2.2. General Approach for Feature Selection*

The general methodologies for feature choice are classified into three: filter technique, wrapper technique, and embedded technique. Each component calculation utilizes any of the three element choice systems [3]. The Embedded methodology is incorporated with versatile frameworks for information investigation, The Wrapper approach is folded over indicators that provide subsets of highlights and getting their input. The methodology is for improving after effects of the particular indicators they work with. The Filter approach incorporates features choice methods that are free of any indicators and features that have minimal opportunity to be helpful in investigation of information. Its strategies depend on execution assessment metric determined legitimately from the information, without direct criticism from indicators that will at long last be utilized on information with decreased number of features.

Adebayo Olawale, et al [4] this article aims at identifying the malware as one of the most dreaded threats to an emerging computer and communication technology. It identified the category of malware, malware classification algorithms, malwares activities and ways of preventing and removing malware if it eventually infects system. Once a system has been compromised by a malware, an attacker can then launch their attack through several tools like packet sniffer, port scanner, vulnerability scanner, password crackers among others.

Safaa O, et al [5] reviewed calculations in information mining utilizing (knowledge discovery in database) KDD-cup 99 (this is a sort of dataset use to assess the performance of an intrusion detection framework precedent least square-bolster vector machine (LSSVM IDS) in the order of assaults and analyzed their outcomes which have been come to. A lot of calculation was assessed on KDD dataset and it was utilized for recognizing the classes of assaults. The shortcoming AI calculations utilized as classifiers for KDD cup information collection is that it doesn't offer much guarantee for recognizing client to root (U2R) and remote to nearby (R2L) assaults.

Mohammed A, et al [6] proposed a mutual data based calculation that logically chooses the ideal element for grouping. The assessment results demonstrates that the feature selection calculation contributes progressively basic highlights (Logs records, hot logins, number of compromised condition) for least square help vector machine-based interruption discovery framework for a better precision and lower computational expense. The shortcoming is that "huge information" hinders the whole identification process and may prompt inadmissible grouping precision because of the computational challenges.

Adetunmbi A, et al [7] manages the importance of each component in KDD 99 intrusion recognition dataset to the discovery of each class. Their exact outcomes uncovered that a few highlights (hot Login, number of Compromised conditions, number of record creation tasks, visitor login) have no pertinence in interruption identification. Harsh set (Rough set) produces a set of compact rules made up of relevant features only suitable for misuse and anomalous detection.

A.A.Waskita, et al [8] proposed a novel way to deal with break down factually the system traffic crude information. The enormous measure of crude information of real system traffic from the IDS is investigated to decide whether traffic is an ordinary or hurtful one. The issue (problem) is now turned into the sensor system to build the precision recognition rate, on the grounds that no hunt spaces are diminished.

Yousef Farhaoui, et al [9] presented and characterized various structures to determine level of adequacy of IDS and the ongoing work of institutionalization and homogenization of IDS.

The system enables updating the analyzer to find new or varieties of assaults. Their restrictions (limitations) don't guarantee 100% security, and the disservice of this arrangement is the rate of false positives because of strange or unordinary conduct of clients, who are not really hurtful.

Heba Fathy, et al [10] proposed NID models utilize diverse keen calculations and feature selection and extraction methods and approve another component choice methodology "Bi-Layer social based element choice methodology", which relies upon the conduct of the order precision as indicated by positioned include. The test results on four proposed NID models, demonstrates the models points of interest of upgrading the identification exactness and testing speed by diminishing the component measurement space. One of the significant research difficulties for building elite NIDS is managing information containing huge number of features.

Akhilesh Kumar, et al [11] proposed an outfit model that is blend of C4.5 and Classification and Regression Tree (CART) as hearty classifier for arrangement of assaults. The proposed ensemble model gives tasteful precision as 99.67% and 99.53% if there should be an occurrence of double class and multiclass NSL-KDD informational index separately. A standout amongst the most significant efficiencies in the KDD informational collection is the colossal number of repetitive records, which makes the learning calculations be one-sided towards the incessant records, and in this way keep them from learning rare records which are normally increasingly destructive to systems, for example, U2R and R2L assaults.

Akash J, et al [12] proposed a framework so as to improve the security of the portable applications which will assess the versatile applications security dependent on the distributed computing stage and information mining. The assessment results demonstrate that it is reasonable to utilize distributed computing stage and information mining to confirm all put away applications routinely to sift through malware applications from versatile application markets. The moving of the security usefulness into the cloud could likewise be dangerous, if not all pieces of the phone can be imitated into the cloud.

Abdulla Amin, et al [13] evaluated data in respect to classifiers configuration, utilized dataset, feature extraction, clustering strategies, exactness location measures and so on. The work of numerous and cross breed classifiers, improves the precision of the grouping and encourages understanding troublesome issues. The shortcoming is that binomial or typical (measurable circulations) can't delineate example acknowledgment conduct, which implies that customary systems of parametric techniques may not work.

Mehrnaz Mazini, et al [14] proposed another solid half breed technique for an oddity system based IDS (NIDS) utilizing artificial bee colony (ABC) and Adaptive Boosting calculations (ADA Boost) so as to pick up a high recognition rate with low false positive rate. The exactness and identification rate of this technique has been improved in correlation with unbelievable strategies. The shortcoming is the bogus alert report of interruption to the system and interruption identification precision that occurs because of the high volume of system information.

## 3. The Filter Approach

This approach utilizes a free measure, (for instance, information measures, separate measures, or consistency measures) as an establishment for surveying the association of a great deal of features [15]. The features are situated by the score and either picked or ousted from the dataset. The strategy is much of the time univariate and considers the component unreservedly, or with regards to the dependent variable. A couple of models consolidate the chi square test, information expansion and relationship coefficient score.

Filter approach is categorized into two sorts: The methodologies are ordered dependent on whether they rate the pertinence of individual features or feature subsets.

- Attribute evaluation approach: The trait evaluate methodologies rank the features independently and allocate a weight to each element as per each component's level of importance to the objective element. The characteristic assessment strategies are probably going to yield subsets with repetitive features since these techniques don't quantify the connection between features
- Subset evaluation approach: The subset evaluation strategies, interestingly, select element subsets and rank them dependent on certain assessment criteria and subsequently are increasingly proficient in expelling repetitive features. The primary hindrance of the channel technique is it overlooks the conditions among the highlights and treats the features independently.

Filter approach is independent of learning figuring which uses the credits of data to pick and survey the features. It is progressively wide, speedier, requires low computational multifaceted nature.

### *3.1. Proposed Work*

The proposed Android intrusion detection system (AIDS) is an application that runs on an Android OS, and can detects malicious acts accurately. The main Objectives are as follows:

- Develop an app that will scan to ascertain, legal or illegal users through password validity.
- Develop a new filter based feature selection method to evaluate the dependence between features and output classes.
- Create log information on both legal and illegal user activities, block or stop and report it.

### *3.4. System Architecture*

The architectural design of the Proposed system (AIDS) is of 3 (three) tiers as shown in figure 3.9. Android Intrusion detection system (AIDS) was designed based on four layers that manage the activities on the system starting from; Event Boxes are sensors in charge of information accumulation and are in this manner the data wellsprings of IDS. This data is drawn from different sources, for example, enlisted information and log documents. The event boxes forward the information gathered to the analyzer to decide if an intrusion has happened or not. [16]An analysis box is utilized to dissect occasions and distinguish unfriendly conduct. It gets contribution from the event boxes and chooses whether an intrusion has happened or not. They can likewise give proof to help their decisions. The outcomes are sent back to the framework, if the IDS utilize a functioning methodology. The database box stores the event created by event or investigation boxes, permitting after death examination and ensuring ingenuity.

Reaction box execute a reaction, to impede the danger, if an intrusion is recognized. It takes the information and contrast and the prepared dataset and match on the off chance that the information is assaulted or typical, on the off chance that the information is assault, at that point an alarm will be sent to the phone number of the client (showing intrusion day, month, time and year).

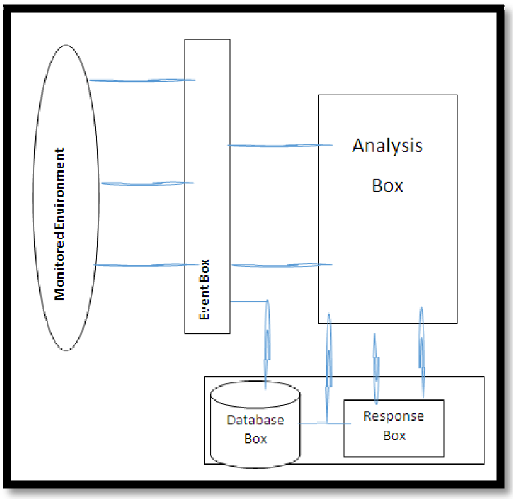


*Figure 1*

## 4. System Implementation

The system implementation is the development of the new system or application following the laid plans from analysis and design stage. This chapter depicts how the plan from the previous section is executed with the aim of providing a proficient system to detection of intrusion on Android phone. Apparatuses and techniques used to actualize are presented in this section.

### *4.1. Choice of Development Environment*

The integrated development environment (IDE) used in the development of this work is the
Android studio v2.2.0, on which the source codes are written, compiled and uploaded on Google Play Store. Android Studio offers numerous features that improves profitability when building Android applications, for example, Gradle-based system which is use to manage all dependencies ( to build, test, run and package your app), Android Virtual Device (Emulator) also helps run and debug apps in the Android studio. The programming languages employed in this project are Java while Realm database management system was used.

### *4.2. Implementation Architecture*

The implementation architecture of the AIDS is represented in figure 4.2 below. It is made up of the various components of the software modules and their linkages.

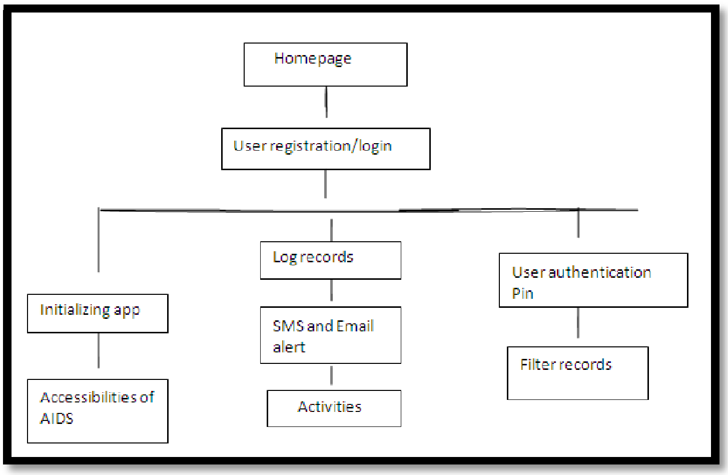
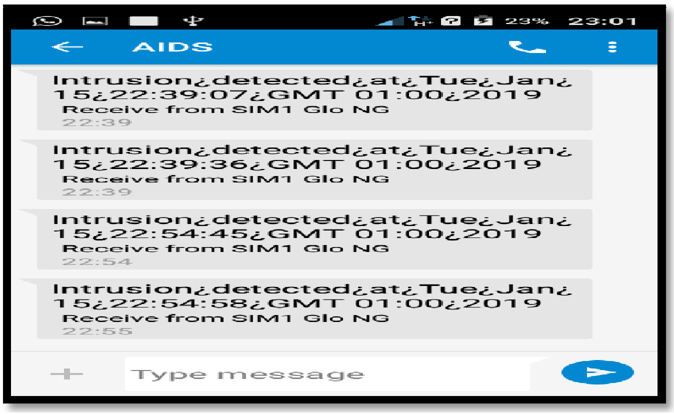


*Figure 2: The Implementation Architecture*

## 5. Result and Discussions

The intrusion detection system (AIDS) clearly shows that the performance of the system is enhanced by the feature selection and it shows promising results in terms of low computational cost and high result. AIDS is a mobile application developed using Java. After the application has been deployed, a pin will be required as the main authentication method. If an intrusion is detected, immediately the alert agent sends an SMS and email to the phone number and email of the user, also all log records of successful or failed attempt will be kept for the user. The problem of false alarm is avoided because the proposed system major alert agent is through SMS and not only email that requires ICMP (which sends error messages to email indicating service is not available or not reachable).

Results are presented in terms of class that achieves smooth levels of discrimination from others in the dataset and the analysis of feature selection in the dataset. The figure's below is an illustration of the output (result) displayed of an intrusion attempt.



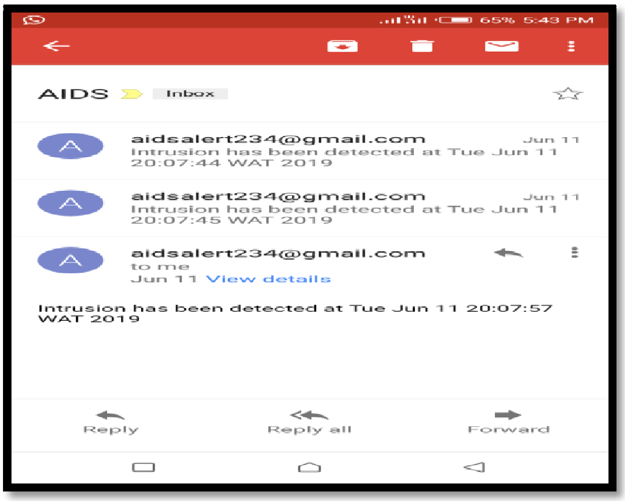
Intrusion¿detected¿at¿Tue¿Jan¿15¿22:39:07¿GMT 01:00¿2019
Receive from SIM1 Glo NG
22:39

Intrusion¿detected¿at¿Tue¿Jan¿15¿22:39:36¿GMT 01:00¿2019
Receive from SIM1 Glo NG
22:39

Intrusion¿detected¿at¿Tue¿Jan¿15¿22:54:45¿GMT 01:00¿2019
Receive from SIM1 Glo NG
22:54

Intrusion¿detected¿at¿Tue¿Jan¿15¿22:54:58¿GMT 01:00¿2019
Receive from SIM1 Glo NG
22:55

*Figure 3: The Screen Shot of the SMS Alert of an Intrusion*



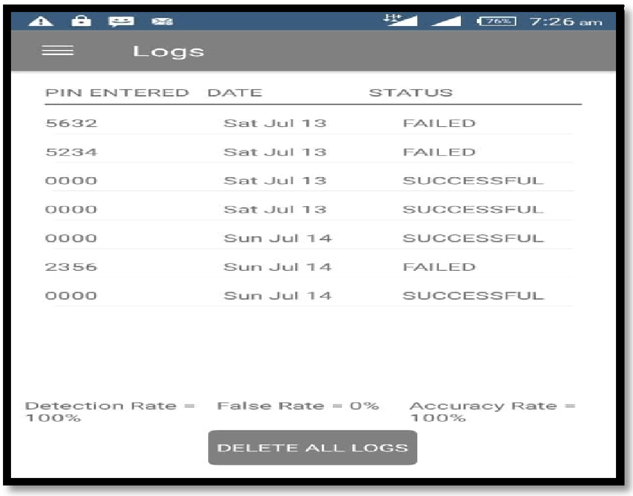
aidsalert234@gmail.com Jun 11
Intrusion has been detected at Tue Jun 11 20:07:44 WAT 2019

aidsalert234@gmail.com Jun 11
Intrusion has been detected at Tue Jun 11 20:07:45 WAT 2019

aidsalert234@gmail.com
to me
Jun 11 View details

Intrusion has been detected at Tue Jun 11 20:07:57 WAT 2019

*Figure 4: The Screen Shot of the Email Alert of an Intrusion*



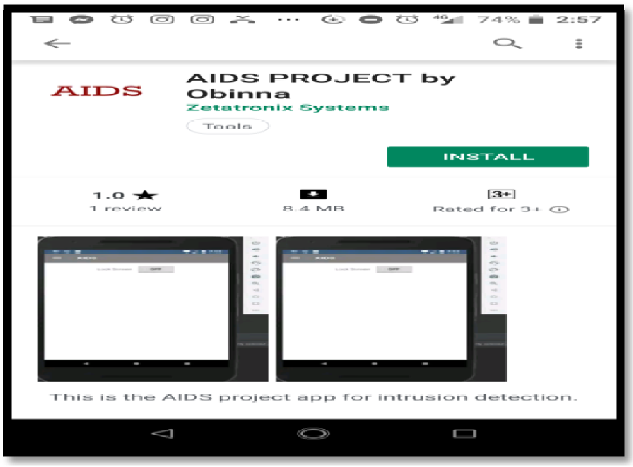
| PIN ENTERED | DATE | STATUS |
|---|---|---|
| 5632 | Sat Jul 13 | FAILED |
| 5234 | Sat Jul 13 | FAILED |
| 0000 | Sat Jul 13 | SUCCESSFUL |
| 0000 | Sat Jul 13 | SUCCESSFUL |
| 0000 | Sun Jul 14 | SUCCESSFUL |
| 2356 | Sun Jul 14 | FAILED |
| 0000 | Sun Jul 14 | SUCCESSFUL |

Detection Rate = 100% False Rate = 0% Accuracy Rate = 100%

*Figure 5: the Screen Shot Login Attempts*

### *5.1. Conclusion*

In conclusion, Android intrusion detection system has been developed and tried, and it is working in its full limit which meets the point and goals of this research work. The proposed system, Android intrusion detection system (AIDS) is an application that keeps running on an Android OS. We propose an Android Intrusion detection system that recognizes noxious acts precisely and the application is utilized by Android phone users. A few works has been investigated which present various strategies of IDS on Android phone. Below is a screen shot of AIDS about to be installed.



This is the AIDS project app for intrusion detection.

*Figure 6: The Screen Shot of AIDS (The Proposed System) About to Be Installed*

### *5.2. Recommendation*

The accomplishment made in the advancement of this project could profit the immense Users of Android phones. To ensure a safe community, it is recommended that this application is adopted for use by Android phone users.

## 6. References

i. K. Chinetha, J. D. Aphney Joann, A.shalins, 2015. “An Evolution of Android Operating System and its version”. International journal of Engineering and applied science (IJEAS), Volume 2, issues 2, pg. 30 – 31.

ii. Yubin Kuang, 2009 “A Comparative Study on Feature Selection Methods and Their Applications in Causal Inference”. Department of Computer Science, Faculty of Science, Lund University, pg. 1-3.

iii. Saranya.k1, Prabhu.r2, Dr. Ramesh Kumar.m3, Preethi.p4, 2017 “Network based Intrusion Detection System using Filter based Feature Selection Algorithm”. International Research Journal of Engineering and Technology (IRJET), Volume: 04, no.10,pg 1271**.**

iv. Adebayo Olawale, Surajudeen, M. A. Mabayoje, Amit Mishra, Osho Oluwafemi, 2012.“Malware Detection, Supportive Software Agents and Its Classification Schemes”. International Journal of Network Security & Its Applications (IJNSA), Vol.4 no.6, pg. 33,

v. Adetunmbi A. Olusola, Adeola S. Oladele, Daramola O. Abosede, 2010. “Analysis of KDD 99 Intrusion Detection Dataset for Selection of Relevance Features”. Proceedings of the World Congress on Engineering and Computer Science, Vol 1, pg 2663-2664.

vi. Mohammed A, Ambusaidi, Xiangjian He, Priyadarsi Nanda, Zhiyuan Tan, 2014. “Building an intrusion detection system using a filter-based feature selection algorithm”. IEEE transactions on computers, vol 1, no 1 pg 1-3.

vii. Jasmina Novakovic, PericaStrbac, DusanBulatović, 2011. “Toward optimal feature selection using ranking methods and classification Algorithm”. Yagoslav journal of operation Research, pg 119 -135.

viii. A.A. Waskita, H. Suhartantoy, P.D. Persadhazy, L.T. Handoko, 2014. “A simple statistical analysis approach for Intrusion Detection System”. Center for Development of Nuclear Informatics-National Nuclear Energy Agency, pg 1.

ix. Yousef Farhaoui, Ahmed Asimi, 2012. “Creating a Complete Model of an Intrusion Detection System effective on the LAN”. International Journal of Advanced Computer Science and Applications *(IJACSA)*, Vol. 3, No. 5, pg 1-2.

x. Heba Fathy, Ahmed Mohamed Eid, Prof. Afaf Abo, El Ftouh Saleh, Prof. Aboul-Ella Hassanien, 2013. “Computational Intelligence in Intrusion Detection System”. Faculty of Science, Al-Azhar University for Obtaining the Degree of Doctor of Philosophy in Computer. pg 147-150.

xi. Akhilesh Kumar Shrivas, Prabhat Kumar Mishra, 2016“Intrusion Detection System for Classification of Attacks with Cross Validation”. International Journal of Engineering Science Invention, Vol 5, Issue 9, pg 5.

xii. Mr. Akash J Wadate, Prof. N. R Chopde, Prof. D. R. Datar, 2016 “Malware Detection System for Android Mobile Applications”. International Journal of Engineering Research and General Science, Vol 4, Issue 1, pg 21-22.

xiii. Abdulla Amin Aburomman, Mamun Bin IbneReaz, 2013. “Evolution of Intrusion Detection Systems Based on Machine Learning Methods”**.** Australian Journal of Basic and Applied Sciences, Vol 7, no 7, pg. 46.

xiv. Mehrnaz Mazini[a], BabakShirazi[b], IrajMahdavi[b].2013“Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and Ada Boost algorithms”. Journal of King Saud University, pg. 799-806.

xv. MuthaiyanMadiajagan, Pragya Garg,2015“Prototype of Intrusion Detection Model using UML 5.0 and Forward Engineering”. International Journal of Security, Privacy and Trust Management (IJSPTM), Vol 4, No 1, pg. 31-34.

xvi. K.Mani, P.Kalpama, 2016 “A review on filter-based feature selection”. International Journey of Innovation Research in Computer and Communication Engineering, Vol 4, issue 5, pg 9147.

xvii. 1 Chani Jindal, 2 Mukti Chowkwale, 3 Rohan Shethia, 4 Sohail Ahmed Shaikh.2014"A Survey on Intrusion Detection Systems for Android Smartphones."IJCSN International Journal of Computer Science and Network, Volume 3, Issue 5, pg. 3.