Evaluation of Machine Learning for Smart Phone Malware Detection

Rahul Kumar Mahato
ARKA JAIN University, Jamshedpur-831014, India

Abstract - *In the current era most of the safety downside faces the external threats and attacks. These attacks square measure destroyed the dear information and injury the growing up organization. the foremost of viruses' attacks or Malware attacks square measure pool up all the data and injury the software system and corrupt the dear information. during this on top of downside, central intelligence service (CIA triad) has organized a sorted Security system to stop the external threat and completely different Attacked like malware square measure larva, ransom ware, adware, key loggers, viruses, Trojan horses, worms et al.. The exponential growth of malware is motility a good danger to the security of counseling. This study did the performance evaluation of some classification algorithms like J45, LMT, Random Forest, Naïve Bayes, MLP Classifier, Random Tree, REP Tree, Ada Boost, Bagging, K-Star, straightforward logistical, IBK, LWL, SVM, and RBF Network. The performance of the algorithms was evaluated in terms of Accuracy, Precision, Recall, alphabetic character Statistics, F-Measure, Matthew parametric statistic, Receiver Operator Characteristics Area and Root Mean square Error victimization wood hen machine learning and data processing simulationtool.*

Keyword: - Malware, classification-algorithms, attacks, Random Forest, security.

I. INTRODUCTION

Malware is any variety of program that's anticipated to make for destruction to the pc system and network. samples of malware are larva, ransom-ware, adware, key loggers, viruses, Trojan horses, worms, and others. The exponential intensification of malware is sitting an excellent hazard to the security of to not be disclosed data. the matter with many of the present classification algorithms is their low performance in term of their ability to find and stop malware from infecting the pc system. there's an urgent got to weigh up the presentation of the present Machine Learning categorization algorithms used for malware recognition. this may facilitate in produce additional strong and strong algorithms that have the competency to beat the weakness of the present algorithms. This study did the performance analysis of some categorization algorithms such as J45, LMT, Naïve Bayes, Random Forest, MLP Classifier, Random Tree, REP Tree, Bagging, Ada-Boost, K-Star, straightforward logistical, IBK, LWL, SVM, and RBF Network. The performance of the algorithms are assess in terms of Accuracy, Precision, Recall, alphabetic character Statistics, F-Measure, Matthew correlation, Receiver Operator Characteristics space and Root Mean square Error exploitation WEKA machine learning and data processing simulation tool. Our experimentalresults showedthat Random Forest algorithmic program produced the simplest accuracy of ninety nine.2%. This absolutely indicates that the Random Forest algorithmic program achieves sensible accuracy rates in detective work malware. The breakthrough in web technology and pc networking have created high speed shared web doable. The result of this development is that the daily increase within the number of pc systems that became vulnerable to malware attacks [1, 2]. The innovation has created the net a huge repository wherever resources square measure virtualized and used to the need of users. Despite the vast edges that the internet revolution has brought, there square measure various challenges that it conjointly poses to the safety of pc systems. The conventional ADP system is entirely centered on one host machine running software package, whereas many machines connected to the host square measure running on the guest operating system [1]. The prevailing security threat braving the users is that the attack on a ADP system by malicious programs that unfold to alternative computers that haven't been infected [3]. The threat display by malware infections has become a serious challenge within the field of pc security over the years. the amount of latest malware on the net keep on increasing at AN menacing rate at the same time as anti-virus companies square measure creating effort to curtail the trend thus on create the immense range of human safe. Malware has evolved over time and is changing into a lot of refined than before. It is now harder to discover them. there's so the necessity to invent a lot of economical techniques that may discover and forestall these attacks. Malware could be a trojan horse that infringes on the safety of a ADP system in terms of privacy, reliability, and accessibility of information [3]. This trend has created academicians and trade practitioners to maneuver from the conventional static detection techniques [4, 5] to more dynamic, refined and spontaneous strategies that applies accumulated malware behavior to discover malware attacks six,[7, 8]. A malware will merely be outlined as a malicious program that the user unsuspectingly install on their machine and later these programs will begin to disrupt the proper operation of the machine or may continue unnoticed and perform malicious actions while not been detected [9]. When the wrongdoer gains management of the machine, he will then have access to any info hold on the machine. Some of the deceptive approaches accustomed install malware on the computer system through the net embrace repackaging the software, update attack [9] or want for transfer [10]. The attacker employs any of the ways mentioned before to create malicious package by inserting an explicit form of malware into it before uploading it to the net. Malware can be delineated as varied sorts of package, that have the capacity to create mayhem on a ADPS or lawlessly make use of thisinfo while not the consent of the users[6,7]. Malware are often classified in varied sorts, for instance, Botnet, Backdoor, Ransom-ware, Root-kits, Virus, Trapdoor. they're accustomed attack laptop systems and for performing criminal activities like scam, phishing, service misuse and root access thirteen.

A. Types ofMalware

For some time currently, differing types of malware are performing numerous malicious activities on laptop systems. These activities vary from just displaying undesirable subject to completely hijacking the pc system from the user and denying them access to that. the foremost well-liked and frequently noticed malware include:

Trojan Horse- may be a program that appears harmless and useful to users like all different authentic software system. However, after opening the applying, this malware distributes another malicious codes that corrupt the files and applications put in on the pc, and conjointly steal sensitive data like password. not like laptop viruses and worms, Trojans require interaction with users to breed themselves. This makes Trojans one in all the foremost damaging and unsafe types of malware as a result of it's principally discovered when it's affected the pc system [4]. consistent with [5], Trojan horse may be categorized into 2 main groups: General Trojan and Remote-Access Trojan. General Trojans: this kind of Trojans includes a big selection of malicious activities. They can threaten knowledge integrity of victim machines. they will airt victim machines to a specific computing device by exchange system files that contain URLs. they will install many malicious software on victim computers. they will even track user activities, save that data then send it to the attacker. Remote Access Trojans: we are able to claim hat they're the most dangerous form of Trojan. they need the special capability that permits the offender to remotely management the victim machine via a local area network or web. this kind of Trojan will be educated by the offender for malicious activities like harvesting counseling from the victim machine. Examples of Trojan Horses area unit Remote access Trojans (RATs), Backdoor Trojans (backdoors), IRC Trojans (IRC bots), Key workTrojans.

Virus- Virus as a malware that includes a self-replicating nature. It is constructed to change or top the functioning of a computer. It multiplies by 1st infecting one program. It is a kind of malware that may cause serious injury varied from the computer system just displaying absolute errors in making the system expertise a Denial of Service (DoS) attack. What distinguishes a virulent disease from a Trojan is that the ability of a virulent disease to duplicate itself by attaching itself to different valid software and become a locality of them. Viruses area unit sometimes propagated through repeating of files from one laptop system to a different, through websites, or e-mails that contain files that have already been contaminated with virus . Also, software system put in on the pc area unit corrupted by the viruses as a results of injecting the real software system with malicious code and because it is dead, the virus is transmitted to other programs on the pc . There are a unit several different ways for sending a virulent disease to different computers such as by causationAssociate in Nursing infected file as Associate in Nursing email attachment or by embedding copies of infected files into a removable medium such as a CD, videodisk or USB drive. Viruses will increase their probabilities of spreading to different computers by infecting files on a network classification system or a classification system that's accessed by another laptop. one in all the crucial variations between virus and worm is that the capability of worm to mechanically spread itself to different computers within the network by exploiting computer's security vulnerabilities. There area unit numerous classifications of a virulent disease, they embody Associate in Nursing encrypted, polymorphic and metamorphicvirus.

Adware- may be a malware whose solely purpose is to point out advertisements to the user. they're thought to be one amongst the least threatening classes of malware. Their intention is to display on the affected laptop commercials that the user is likely to be drawn to, it records knowledge from the pc such as browser and search engines histories nineteen. Adware is sometimes classified as spyware subject to the seriousness of the recording. Adware, or advertising-supported code, is any code package that mechanically plays, displays, or downloads advertisements to a laptop. These advertisements will be within the style of a pop-up. the item of the Adware is to come up with revenue for its author. Adware, by itself, is harmless; but, some adware might associate with integrated spyware like key loggers and different privacy invasive code. Adware is sometimes seen by the developer as a way to recover development prices, and in some cases, it may allow the code to be offered to the user freed from charge or at a reduced worth. Conversely, the advertisements is also seen by the user as interruptions or annoyances, or as distractions from the task athand.

Spyware- may be a quite self-installing malware that execute without the user’s approval. it's accustomed gather and track information regarding the person and therefore the browsing history of a computer system. it's usually prepackaged in conjunction with software that's created obtainable to users at no value. Spyware is additionally known as rootkit as a result of the packaging with freeware. Spyware may be a code that allows a 3rd party to spy on a host. Spyware has been used for a range of functions including larceny fraud and theft of private knowledge, spying on online activities of people (e.g. spouses) and look users' on-line activities. it's a sort of malware put. The presence of spyware is often hidden from the user and may be troublesome to find [2,1]. Spyware typically modifies the pc settings, leading in terribly sluggish connection speeds and/or loss of web association. Moreover, a number of the system practicality begin malfunctioning so creating the pc to be terribly slow and several strange code square measure mechanically putin.

Worm- may be a malware that doesn't attach itself to different software because it doesn't want a number code to lock itself to. This is what differentiates worm from the virus. A worm normally affects its victim through the realm of exposures that it can exploit. It employs numerous means that to propagate, and corrupt different laptop systems fourteen. Worms have the capability to wreck an equivalent extent of disturbance a pestilence can cause to associate infected ADPS. Worms don't seem to be parasitic in behavior just like the viruses. they're freelance programs that can cause hurt on their own. These worms might or might not have a payload however each sorts will be pretty harmful. Worms while not payloads don't have an effect on the system that it infects 16. Whereas the worms with payload can do hurt to the infected system also.In some cases, the payload acts as a backdoor rather than creating changes to the system. A worm may have a awfully harmful impact on systems within the network, like may consume an excessive amount of system memory or system processor (CPU) and cause several applications to prevent responding. a number of the foremost noted worms embrace the worm that has created businesses to lose upwards of five.5 billion greenbacks indamage23.

Bot- conjointly called an internet mechanism or botnet square measure application software that runs machine-controlled tasks over the web. They belong to a class of malware that permits its principal to gain access to the infected ADPS. Bots can propagate through backdoors created obtainable by a pestilence or worm on the victim laptop. Bots square measure legendary for using an application layer protocol that allows communication in the form of text with its principal. Distributed Denial of Service (DDoS) attacks that have the capability to impede the services of the target laptop by over-flooding its information measure or resources with requests will be launched victimization many bots.

Ransom ware- may be a subcategory of malware that encrypts the files on the victim's laptop or entirely fast you out. It turns your files to unintelligible data and makes them useless and payment is necessitated before the coding and returning of the ransomed files to the owner. they typically infect their victims through Trojan .

Rootkits- square measure a group of code tools utilized by hackers to induce and sustain continuous administrator-level access to a computer system therefore on camouflage the dynamic of files, or activities of the hacker to stay the user within the dark. Rootkits are usually connected with Trojans, worms, and viruses that obscure their presence and actions from users and different system processes .

Backdoor- may be a category of malware that gives a supplementary stealthy "entrance" to the system for attackers. The backdoor itself doesn't directly hurt the system however it opens the door for attackers to play disturbance. thanks to this characteristic, backdoors square measure in no means used one by one. Ordinarily, a backdoor is antecedent malware attack or different sorts of attacks24.

Key logger- conjointly called keystroke work may be a kind of surveillance malware that after the pc is plagued with it has the flexibility to record each keystroke build thereon system. The recording is saved in a very log file that is often encrypted and sent to a selected receiver. Such data will include passwords, Band Verification range, ATM card numbers and different counsel twentyfive.

II. RELATEDWORKS

With the unexampled increase within the range of malware been free on the net, several researchers have taken it upon themselves to judge the performance of classification algorithms that are used for detective work and classifying malware by employing a combination of performance metrics. We, therefore, realize it necessary to work out that algorithmic program performs best for any chosen metric to help withinthe correct classification of malware. many studies are carried out to compare the performances of some classification algorithms for malware detection. Classification algorithms whose performances are up to now compared embrace Naïve Bayes .different algorithms compared embrace call Trees, Support Vector Machine, Random Forests,J48 , C4.5, ANN , Multilayer perceptron, CART, Neural Network , IBK , theorem Network. Table one depicts the outline of the algorithms utilized in previousstudies.

III. LITERATUREREVIEWS

Sl. no.	Title	Author	Finding	Remark
1.	Machine learning techniques for the evaluation of efficiency of the software reliability growth models	Omal Sahar, Muhammad Ahsan Latif& Muhammad Imran 2017	To calculate the efficiency and achievement of proposed GA based approach, I compare outcomes of the genetic algorithm with other optimization techniques SA and MOGA.	In this paper, I suggested a very efficient and suitable technique to calculate the efficiency of the SRGMs using GA based approach. I proposed the genetic algorithm based approach to apply to the evaluation of the parameters of the SRGMs. Three operators were used in GA based approach i.e., Selection, crossover and mutation
2.	Predicting long-term mortality with first week post-operative data after Coronary Artery Bypass Grafting using Machine Learning models	J.n.alves .castela. cardoso. 2017	The accuracy of all models in predicting 5-year mortality after CABG was assessed by testing against the validation dataset, with results reported as AUROC (95%CI). Cox Regression, the most commonly used survival analysis tool in Medicine, was used as a baseline for comparison and proved the least accurate of all models with a	To my knowledge, this is the first study to describe the use of ML methods to predict long-term mortality in patients who underwent CABG. Here, I demonstrate the superiority of models developed with ML algorithms over traditional Logistic Regression for long-term mortality prediction after CABG operations. These findings are in line with the predictive capacity of ML models in other fields of Medicine

			time-dependent AUROC of 0.644 at 5 years follow-up.	
3.	A comparison of Machine Learning approaches for classifying Multiple Sclerosis courses using MRSI and brain segmentations	Adrian Ion-Margineanu, Gabriel Kocevar, Claudio Stamile et al. 2017	All performance measures can be found in Table 4. Maximum AUC values for each classification task are highlighted in gray. For CIS vs. RR i obtain a maximum AUC of 77% when combining metabolite ratios with GM, WM, and lesions percentage.	In this paper performed four binary classification tasks for discriminating between MS courses. Ireport AUC, sensitivity, and specificity values, after training simple and complex classifiers on four different types of features. I show that combining metabolic ratios with brain tissue segmentation percentages can improve classification results between CIS and RR or PP patients. This best results are always obtained with SVM-rbf, so I can safely conclude that building complex architectures of convolution neural networks do not add any improvement over classical machine learning methods.
4.	Optimization Methods for Large-Scale Machine Learning	L´eon Bottou Frank E. Curtis Jorge Nocedal 2017	This analysis of SG in 4 can be characterized as relying primarily on smoothness in the sense of Assumption 4.1. This has advantages and disadvantages. On the positive side, it allows us to prove convergence results that apply equally for the minimization of convex and non-convex functions, the latter of which has been rising in importance in machine learning; recall 2.2.	Mathematical optimization is one of the foundations of machine learning, touching almost every aspect of the discipline. In particular, numerical optimization algorithms, the main subject of this paper, have played an integral role in the transformational progress that machine learning has experienced over the past two decades. In this study, I highlight the dominant role played by the stochastic gradient method (SG) of Robbins and Monro [130], whose success derives from its superior work complexity guarantees. A concise, yet broadly applicable convergence and complexity theory for SG is presented here, providing insight into how these guarantees have translatedinto practical gains.
5.	High-Speed Tracking with KernelizedCorrelat ion Filters	Joao F. Henriques, RuiCaseiro, Pedro Martins, and Jorge Batista 2014	The explanation is that, after computing a cross-correlation between two images in the Fourier domain and converting back to the spatial domain, it is the top-left element of the result that corresponds to a shift of zero [21]. Of course, since i always deal with cyclic signals, the peak of the Gaussian function must wrap around from the top-left corner to the other corners,	In this work, I demonstrated that it is possible to analytically model natural image translations, showing that under some conditions the resulting data and kernel matrices become circulate. Their diagonalization by the DFT provides a general blueprint for creating fast algorithms that deal withtranslations.

IV. ANALYSIS

Three stages were concerned within the performance analysis of the various Machine Learning classifiers thought of during this study. The phases area unit Dataset Preparation, Pre-Processing and Application of various Machine Learning algorithms on the ClaMP (Classification of Malware with letter headers) dataset files 34. The dataset includes a total of 5184 instances, which contain 2683 Malware, and 2501 Benign. The dataset has fifty five features. The ClaMP dataset thirty six is reborn into. format(a format compatible for the file) supported by the Maori hen Machine Learning simulation setting for input file that was used for the analysis. to try to a satisfactory classification of the ClaMP dataset, J45, LMT, Naïve mathematician, Random Forest, MLP Classifier, Random Tree, REP Tree, Bagging, Ada Boost, K-Star, Simple Logistic, IBK, LWL, SVM, and RBF Network were used and a [10] folds cross-validation was used in this study. the rationale for choosing [10] folds was as a result of outputs generated from intensive tests on totally different datasets with erratic learning range of folds required to get the best estimate of error [3,5]. To carry out cross-validation, a specific range of folds is selected, the info is at random divided into [10] segments in whichthe category is denoted in virtually constant size once compared to the entire dataset. every section is control out sequentially and therefore the learning methodology trained on the nine-tenths that remain; later on, its error rate is processed on the holdout set. Consequently, the training method is dead ten times on totally different coaching sets. onceand for all, the mean of the [10] error analysis area unit hand-picked because the general permits the extraction of sure proportion of the info for assessment. A proportion split of sixty six split was used for thisstudy

VI. EXPERIMENTAL RESULTS AND DISCUSSION

Experiments were conducted exploitation the whole dataset with [10] folds cross-validation and sixty six split. The performance of every Machine Learning classifiers was evaluated in terms of Accuracy, Precision, Recall, letter of the alphabet Statistics, F-Measure, MCC, Receiver Operator Characteristics space, and Root Mean squareError.

A. Accuracy- The Accuracy is performance metrics that area unit accustomed specific the proportion of correct predictions. It doesn't take into thought actuality positives and negatives individually. this can be the essential reason why accuracy alone can't be accustomed confirm the performance of a model. different performance metrics except the accuracy area unit needed to be used. the worth of one indicates the most effective accuracy. From the experimental results of assorted classifiers during this study, the most effective Accuracy is zero.992 generated once the 10-fold cross validation was used on Random Forest classifier whereas the worst was zero.652 created once sixty six split was used on the Naïve Thomas Bayes classifier. Figure one and Table three shows the Accuracy of everyclassifier.

B. exactness and Recall- exactness, that is additionally referred to as positive prognostic worth, returns the speed of relevant results instead of inapplicable results. it's a little proportion of vital recollected instances, whereas recall is that the fraction of relevant instances that area unit recollected. The recall is that the sensitivity for the foremost relevantresult. exactness associate degreed recall rely on an understanding and live ofconnation.

$Precision = TP\ TP+FP$ ------ (1),
$Recall = TPR = TP\ TP+FN$-------(2)

The exactness and results of the various classifiers are pictured in Table three and figures a pair of and three. the very best exactness and recall values of zero.993 and 0.992 severally were created once ten -fold cross-validation was done on RandomForest.

C. Kappa Statistics- Kappa data point may be a performance metric that compares associate degree ascertained accuracy with associate degree expected accuracy (random chance). It reflects the degree of agreement between verity categories and therefore the classifications. The letter of the alphabet statistics worth of one is that the highestindicating complete agreement. during this study, the very best letter of the alphabet characteristics is zero.985 that was created once the check was conducted on Random Forest with ten folds cross-validation.

D. F-Measure- F-Measure is that the worth that estimates the whole performance of the system by uniting exactness and recall into one variety. the very best worth of one specifies the most effectiveresult.

$F - measure = 2\ x\ Recall\ x\ Precision\ Recall + Precision$ ----------(3)

ROC space- The mythical monster (AUC) Area of a classifier is that the likelihood of the classifier ranking a every which way chosen positive instance above a every which way chosen negative instance., ROC of 0.8 depicts smart prediction, ROC of 0.7 could be a mediocre prediction, whereas mythical monster of zero.6 symbolises a poor prediction. Figure half-dozen depicts the areas below mythical monster curves of classifiers utilized in this study with Random Forest achieving the most effective performance with zero.999 whereas RBF Network has the poorest performance with zero.779.

E. Matthew coefficient of correlation (MCC)- actuality and false positives cannotbe adequately represented mistreatment one indicator, the Matthews coefficient of correlation (MCC) have well-tried to be the most effective general live thirty four. MCC could be a performance metric that measures the properties of the two-class drawback. It takes into thought actuality and false positives and negatives. it's a balanced metric, even once the categories arfrom dissimilar sizes. The formula below will be wont to figure the worth forMCC:

$MCC = TPxTN - (FNxFN\ )\ (TP+FP)(TP+FN)$
$(TN+FP)(TN+FN)$ ------------(4)

when the output is +1 it represents the most effective prediction, whereas −1 signifies a whole disagreement. Table three and figure seven shows the MCC for every classifier under consideration. Random Forest classifier made the most effective MCC price of zero.985 whereas Naïve mathematician generated the worst results of zero.

VI. CONCLUSION

This paper presents a comparative study of malware detection using fifteen completely different Machine Learning algorithms. Some of the progressive models like J45, LMT, Naïve Bayes, Random Forest, MLP Classifier, Random Tree, REP Tree, Bagging, Ada Boost, K-Star, easy supplying, IBK, LWL, SVM, and RBF Network were employed in the study and their statistical results given. From the experimental results obtained from running the assorted classification exploitation 10-fold cross-validation and sixty six split check, it's been incontestable that some unpopular algorithms perform comparatively well on the ClaMP dataset thirty six on Maori hen. It becomes apparent from our study that Random Forest is that the best classifier among the fifteen (15) classifiers thought of. Experimental results indicated that even with less feature choice used, the Random Forest classifier with zero.992 performs relatively better in malware classification, far better than the favored classification algorithms like SVM with zero.956 accuracy, Ada Boost with accuracy of zero.922, sacking with zero.978, J48 with 0.978, Naïve Bayes with zero.652, and Multilayer Perceptron classifier with zero.973. we tend to suggest that additional publicly obtainable malware datasets be accustomed value the performance of different Machine Learning algorithms exploitation different data processing and Machine Learning tools like Rapid jack.

VII. REFERENCES

i. Amos B, Turner H, White J (2013) Applying Machine Learning Classifiers To Dynamic Android Malware Detection At Scale. In: Proceedings Of The 9th International Wireless Communications And Mobile Computing Conference (IWCMC), Sardinia, Italy, Pp 1666–1671

ii. Android (2013) Android 4.2, Jelly Bean. Http://Www.Android.Com/ About/Jelly-Bean/. Accessed June 2013

iii. Anuar NB, Sallehudin H, Gani A, Zakaria O (2008) Identifying False Alarm For Network Intrusion Detection System Using Hybrid Data Mining And Decision Tree. Malays J Computer Sci21(2):101–115\

iv. Anubis (2013) Anubis: Analyzing Unknown Binaries. Http://Anubis. Iseclab.Org/. Accessed Feb 2013

v. Arp D, Spreitzenbarth M, Hubner M, Gascon H, Rieck K (2014) DREBIN: Effective And Explainable Detection Of Android Malware In Your Pocket. In: Proceedings Of The 2014 Network And Distributed System Security (NDSS) Symposium, San Diego, USA (2014)

vi. Arstechnica (2013) More Bad News For Android: New Malicious Apps Found In Google Play. Http://Arstechnica.Com/Security/2013/ 04/More-Bad News-For-Android-New-Malicious-Apps-Found-In-Go Ogle-Play/. Accessed 1st Jan 2013

vii. Bradley AP (1997) The Use Of The Area Under The ROC Curve In The Evaluation Of Machine Learning Algorithms. Pattern Recognit 30(7):1145–1159

viii. Breiman L (2001) Random Forests. Mach Learn 45(1):5–32

ix. Burguera I, Zurutuza U, Nadjm-Tehrani S (2011) Crow Droid: Behavior Based Malware Detection System For Android. In: Proceedings Of The 1st ACM Workshop On Security And Privacy In Smartphones And Mobile Devices, Chicago, Pp 15–26

x. Felt AP, Finifter M, Chin E, Hanna S, Wagner D (2011) A Survey Of Mobile Malware In The Wild. In: Proceedings Of The 1st ACM Workshop On Security And Privacy In Smartphones And Mobile Devices, Chicago, Illinois, USA, Pp 3–14

xi. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian Network Classifiers. Mach Learn 29(2–3):131–163

xii. F-Secure (2013) Android Accounted For 79% Of All Mobile Malware In 2012, 96% In Q4 Alone. Http://Techcrunch.Com/2013/03/07/F-Se Cure-Android-Accounted-For-79-Of-All-Mobile-Malware-In-2012- 96-In-Q4-Alone/. Accessed 1st June 2013

xiii. García-Teodoro P, Díaz-Verdejo J, Maciá-Fernández G, Vázquez E (2009) Anomaly-Based Network Intrusion Detection: Techniques, Systems And Challenges. ComputSecur 28(1–2):18–28

xiv. Gogoi P, Bhattacharyya DK, Borah B, Kalita JK (2013) MLH-IDS: A Multi-Level Hybrid Intrusion Detection Method. Comput J 2013 Doi:10.1093/Comjnl/Bxt044. Online.Http://Comjnl.Oxfordjournals. Org/Content/Early/2013/05/12/Comjnl.Bxt044.Abstract. Accessed 12 May 2013

xv. Gribskov M, Robinson NL (1996) Use Of Receiver Operating Characteristic (ROC) Analysis To Evaluate Sequence Matching. ComputChem 20(1):25–33