Static Analysis of Android Permissions and SMS using Machine Learning Algorithms

Sonali Kothari
Comp.Sci.and Engg, Research Scholar, SGBAU Amravati, India

Pravin Karde, PhD
Information Technology, Professor, Govt. Polytechnic Amravati, India

Vilas Thakare, PhD
PG Dept. of Computer Science, Professor, SGBAU Amravati, India

ABSTRACT
Everyday user receive number of SMS messages and installs number of applications on their device. A study says that every 10 seconds Android device is facing a new attack. As user is using smartphone nowadays to store personal and professional, confidential and sensitive information on smartphone, it needs to be secured. In this paper, a static analysis model is provided for network service providers. This will allow network service providers to mark SMS as spam before sending it to user. Analyzing permissions of Android application will allow app provider to identify malicious applications. This will help in reducing attacks on smartphones using applications.

Keywords
SMS, spam, ham, Android permission, ANN

1. INTRODUCTION
In past two decades, the Internet has transformed into a big infrastructure with millions of people doing number of activities through Internet on daily basis. Thousands of businesses are dependent on Internet for daily business and providing facilities to customers. During the same period, Internet started facing threats from cyber criminals and security of Internet has become an issue to work on seriously. The features of Internet which contributed for its growth – same features like ability of every node to execute arbitrary code – became target and used for cybercrime, to perform illegal activities on massive scale. The global losses incurred by such activities are in the billions of dollars. It is necessary to protect Internet from such malicious activities.

Today numbers of researchers are working on providing security to internet. What makes internet security such an important point for researchers is its technical variety. Only technical solution is not sufficient to provide satisfactory security to internet. Considering case of DoS (Denial of Service) attack, number of solutions are provided in past years to avoid them. But none of them have managed to provide permanent security. Every time cyber criminals come with different ways of attacking. Though current solutions are not failed completely, but with new attack they need to be revised or updated. Similar is applicable to other kind of online threats like malware, phishing, spam emails.

2. BACKGROUND
As mobile based attacks mostly communicate between mobile devices (smart phones), it has the following characteristics [1]:

Limited by the power resource: The mobile devices such as smart phones are different from PC, its run time is limited due to the use of battery.

The communication costs problems: The communications of the mobile botnets will lead to the cost of the owner, and a significant rise of the phone charge will result in the investigation of the cause and thus may lead to the exposure of the cellular bot.

The connectivity changes constantly, even unstable: The connectivity may be both affected by physical environment or personal factors. It can be affected by the networks around the mobile phone owner, the action of the mobile phone owner, such as the user is in the tunnel or turn off the mobile phone during the bed time.

Lack of IP address: The lack of IP address may cause the problem of indirect connect. Due to the lack of IP address, most mobile phones are using NAT gateway and thus the devices are not directly reachable.

The diversity of operating system of smart phone: The design of mobile botnet has to consider the diversity of the OS platform of smart phone.

2.1 Various architectures used for attacking smartphones
Botnet uses four types of architectures to control network and to be invisible from detection i.e. Centralized botnet Architecture, Peer to Peer botnet Architecture (P2P), Hybrid, and Combination of Hyper Text Transfer Protocol with Peer to Peer (HttP2P). The first architecture is not very secure but easy to implement while the second architecture is hard to detect as well as hard to manage, whereas Hybrid and HttP2P are combination of first and two for bypassing firewalls and intrusion detection mechanism.

- Centralized botnet Architecture: The oldest and easiest to manage and control architecture used by the supervisor-bot is centralized. All the zombie computers or zombie army is being supervised from a center point, which makes them visible to be detected and stopped. It uses Internet Relay Chat (IRC) or HTTP protocols for its C&C. Examples of centralized models are AgoBot, SDBot, SpyBot, GTBot, and Zotob [2, 3].
- Peer to Peer (P2P) botnet Architecture: Supervisor-bot transfer command to an infected zombie peer who transfers it to other peers, acting both as Supervisor- bot and zombie army soldier.

Similarly it can transfer commands from any zombie, which lead to a slow but effective undetectable communication between zombie army. Examples of bots using P2P are Phatbot and Peacomm [4].

a) Hybrid botnet: Architecture Hybrid is similar to P2P where Supervisor-Bot maintain a P2P communication between supervisors behaving like server community. But a Supervisor-Bot breed, keep information, and prevent a robust

BOTNET able to maintain control of its remaining bots from significant exposure or making it harder through their communication traffic patterns of the network topology of its soldier zombie community. Each Supervisor- Bot has its own list of peer and does not share it with others bots for security purposes. Ping Wang *et al.*[2] designed and proposed a hybrid P2P botnet attack which is difficult to observe and even much difficult to seal.

b) Hypertext Transfer Protocol Peer to Peer (HTTP2P) botnet Architecture: In HTTP2P Supervisor-Bot cipher the message, continuously search for Soldier-Bot, and when found deliver message to it. While the Soldier-Bot does not contact dynamically to Supervisor-Bot or other soldier-bots rather it waits for a call from its supervisor [2].

3. PREVIOUS WORK BY RESEARCHERS

Mahmoud, Nir and Matrawy [5] studied about the architecture, defenses and detection of botnets in both PCs. As per their study, the most threatening attacks in mobile phones have been found to be botnet attacks and it has been established that, detecting the attacks is quite impossible even with the advanced technologies. As per the study, in-order to detect the botnet attacks, the authors developed a detection model based on identifying behavioral aspects of botnets, tracing back Bot-master and utilizing the virtual machines towards detecting botnet attacks. Thus by focusing upon these characteristics, identifying and detecting botnet attacks is a possibility.

According to the research of Abdullah [6] nowadays mobile devices namely mobile phones have been used widely. People use mobile phones not only for sending messages or phone calling but also for browsing web, online banking transaction and social networking. All confidential data are kept in their mobile phone to some extent. As an outcome mobile phones became as one of the cybercriminal major target particularly through mobile botnet installation. An example of mobile botnet is Eurograbber that being installed through affected mobile app without the knowledge of victim. It will claim as mobile banking software application and steal the financial transaction data from the users mobile phone. Eurograbber has caused huge loss of United States $47 million in 2012 all over the globe. This study provides a proof of the topic on how the botnet performs and the ongoing study to respond and detect to mobile botnet effectively. The botnet malicious activity detection is performed through an examination of Crusewind code of botnet using the static analysis technique and reverse engineering process.

According to the study of Anwar et al [7] the mobile devices usage involving tablets, mobile phones, notebooks and smart watches are developing every day in the society. They are generally linked to online and provide around similar memory, speed and functionality like a personal computer. To acquire much advantages from these mobile gadgets the applications must be installed and these applications are feasible from 3rd party websites namely Google play store. In already existing mobile devices OS, Android is simple to attack because of its open source surroundings. Android operating system utilization of open source facility attracts developers of malware to aim mobile gadgets with their new malicious app having capabilities of botnet.

In his research, Amro [8] observed the existing mobile models, OS, botnet attacks on mobiles and detection techniques. As per his findings, the malware propagation has certain techniques, such as: repackaging, dynamic payloads, drive by download and stealth malware techniques. The authors have also observed the evasion techniques of malwares and argued that: anti-security techniques, anti-analyst techniques and anti-sandbox techniques have been in use by the attackers in mobile phones which bypasses the security and safety measures initiated by application developers. Hence the study offered detection techniques, such as: a) static techniques: permission based analysis, signature-based approach and virtual machine analysis; b) dynamic techniques: anomaly based, emulation based and taint analysis. Thus the author explained about the evasion techniques and detection techniques in mobile phones.

authors insist that, there is a huge scope for further development of accuracy and detecting Botnet attacks in future.

4. EXPERIMENTAL ANALYSIS

Today most of the email providers scan email contents and its attachment for spam, SMS spam detection method still exists in offline mode. Very few applications like TrueMessanger by TrueCaller are available for blocking spam SMS. But these applications need user to specify contacts for blocking SMS. Though here static method of SMS spam-ham is explained, SMS network providers can use this and mark harmful SMS as probable spam SMS tag like email. Various datasets used during research contains different parameters. The data is converted into required format for training system. This approach includes two different category datasets. Data used for performing static analysis of SMS and permissions of applications is collected from different available sources. Though number of applications are available now a days to identify if an SMS is spam or ham, users ignore them conveniently. To identify SMS spam or ham, SMS corpus from [10,11,12] is used with various classification algorithms for training and testing model . In this static method corpus is pre-processed, stemming and tokenizing is performed and then data is split into 2 splits with 80% as training data and 20% as testing data. As different applications need different permissions and controlling permission required by application in not in the hands of user. Though developer needs to take care of permissions it is allowing to user. Figure 1 shows preparation of list of SMS ham- spam and trusted/untrusted applications on analysis server system. Here 2 different lists are maintained, one containing already known malwares while another one is list of permissions and class of application as trusted/untrusted based on allowed permissions. The first list obtained from available sandboxes is used to directly check if a file, IP or apk is already detected as harmful to device. The second list is prepared by human analyst considering number of details about applications and them marking them as trusted/untrusted based on dangerous permissions used by them. Google has implemented a service to verify applications inside the official Google Play app store as shown in Figure 2. If this serviced is on and user is installing any app, information about the app will be redirected to Google cloud server.

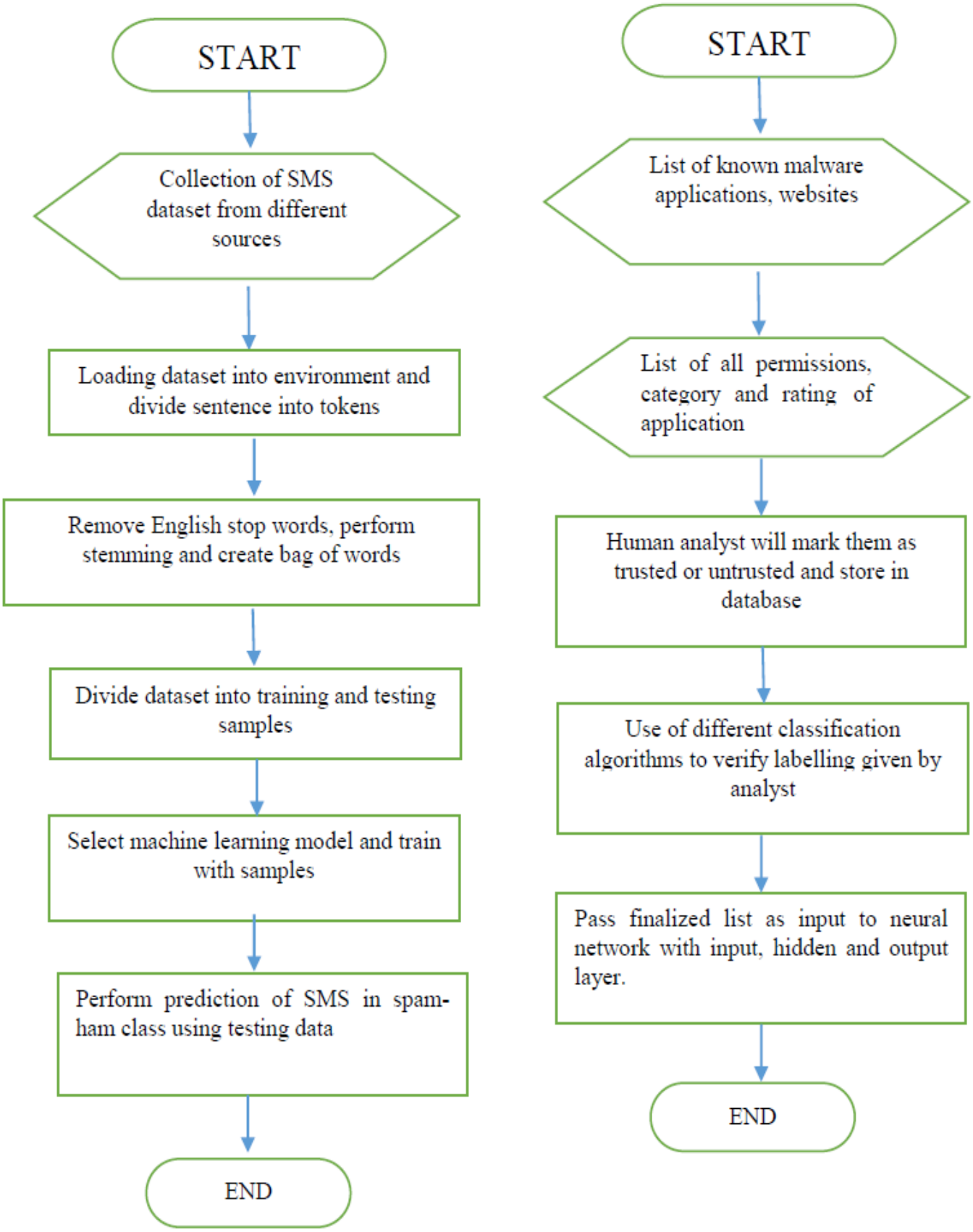

Fig 1: Flow of SMS and Android Application dataset analysis

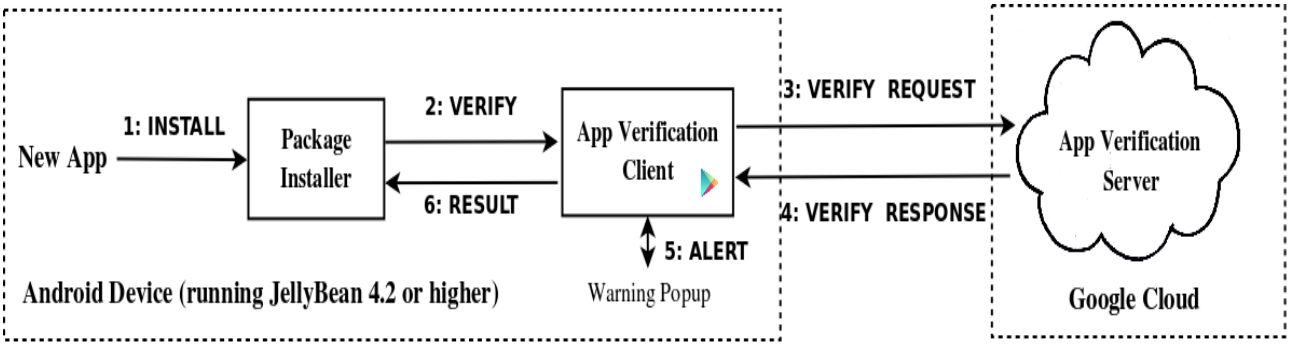

Fig 2: Google's App verification service

If Google cloud finds app is not safe, user will see warning message about app. Dangerous apps are blocked from being installed, while decision of installation of potentially dangerous apps will be taken by user [13].

Though this service is available, it only checks about already detected virus/malware/botnet applications. If any new malware is added, Google App verification server may not detect it. Applications include the preinstalled system applications and the user-installed applications. In most cases, the sources of the applications the user installed are diverse, so it is unable to verify their reliability. This makes attacks against user applications become the easiest means of attack. During the installation process of one application, users usually can not recognize the malicious program from the normal one, they often choose to trust the application, ignore safety tips, and give all the permissions they stated. This will greatly weaken the core security mechanism of application permissions control. Now some attacking codes against browsers even target on the public network. Since this ways of attacks are diverse and simple, applications become a key part of the hacker attacks [14].

5. RESULTS AND DISCUSSION

Most common method employed by bot variants is HTTP or SMS. The results shown here need a good amount of system resources in context of memory, storage space and time. It is advised to use this model on network service provider system to stop sending spam SMS or tag it with probable spam message. This will help message receiver to consider SMS before replying to it.

In table 1, details of datasets used for training models are discussed. Figure 3 shows accuracy of the output predicted by model for dataset from [10,11,12] together with various different classification algorithms. Though Logistics Regression gives maximum accuracy in predicted SMS in correct class, all classification algorithm classifies SMS with more than 86% accuracy. Similarly Figure 4 shows accuracy of various classification algorithms are used over dataset from [15]. Though output of all classification algorithms is similar, ANN gives best result with 99% accuracy with 0.02% loss at epoch 150 onwards. For ANN, input layer is having 329 neurons, first hidden layer with 25 neurons, and second hidden layer with 6 neurons while last output layer is created with single neuron. Here for first and second layer activation function RELU is used while for last layer sigmoid activation function is used.

Table 1: SMS & Permission dataset

DATASET	SAMPLES	
	TRAINING	TESTING
SPAM-HAM SMS	9962	2490
BENIGN-MALWARE PERMISSIONS	320	80

Ruchna Nigam [16] has highlighted various mobile botnets based on their timeline. Compared to PC based botnet attacks, Mobile based botnets are easy to design, propagate and take down also. Mobile botnetmaster can work from PC as well as Mobile. Condisering limited resource of mobile phone with respect to battery life, storage capacity and performance capacity, botmaster prefer to work thru PC. Besides these limitations, mobile users are accessible most of the time which is different from PC users as they are available only during device's uptime [9]. It also observerd that approximately 83% bot variants do not need administrator privileges while more than 50% bots are intended for stealing information stored on device in various formats, while other bots are focused on financial and malware propagation activity mainly.

In [17], comparison of various online sandboxes is discussed. VirusTotal by Google detects malware using supported antivirus engines. If the uploaded malware is already detected, user can understand about risk. It can be used for scanning Windows executable files, APKs, URLs, IP addresses. The other available online malware analysis tools includes Anubis which is now discontinued, Malwr, VxStream etc. Looking into the features provided by all these, only VirusTotal scans Android APKs for already available malware attacks. For identifying existing malware attacks with their package name, research work is using VirusTotal dataset available online. Besides this, admin is allowed to add new apk files, their hash codes, list of permissions which can be risky and remrks about it. Admin panel also allowes to add level of threat for existing as well as newly added APK files as shown in Figure 5.

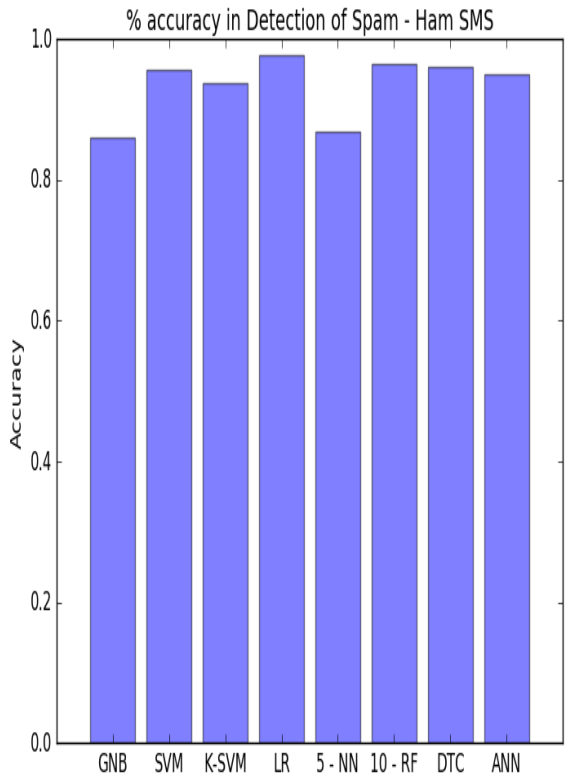

Fig 3: Accuracy ratio of various SMS datasets for classification algorithms – test data over predicted values

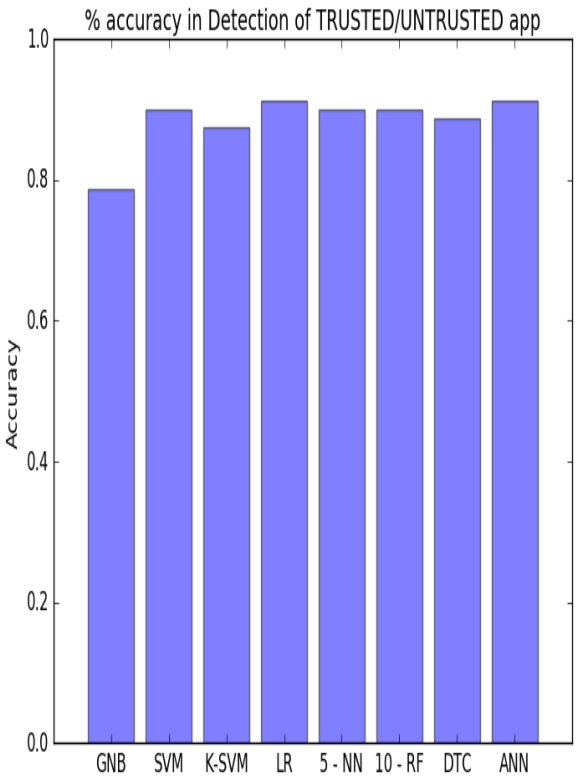

Fig 4: Accuracy ratio of various Permission datasets for classification algorithms – test data over predicted values

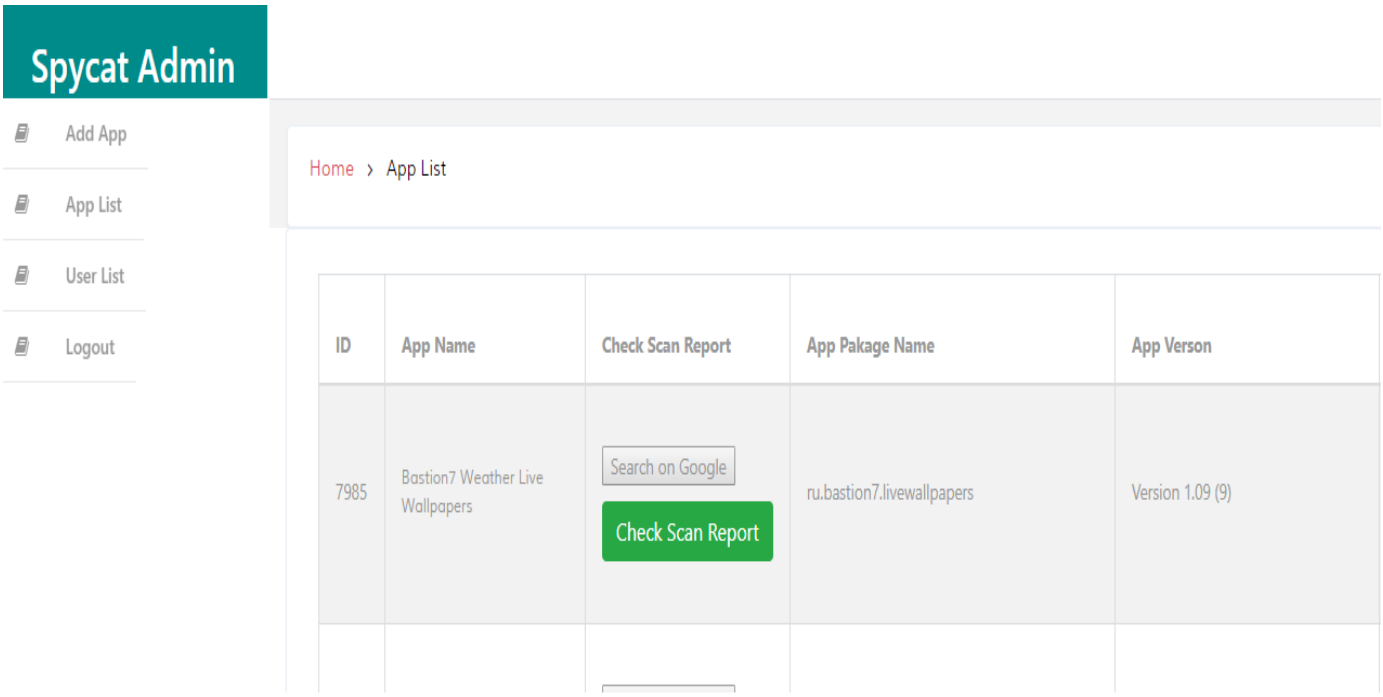

Fig 5: Admin panel for adding new malwares on server

6. CONCLUSION AND FUTURE WORK

With ever growing desire to access and control everything on a single click, smartphones manufacturers are adding number of features in devices. Similarly users are spending a good amount of their personal time using smartphone. Now a days many professional and personal activities are carried out using smartphones like internet banking, personal/professional networking, information gathering and sharing, spending leisure time etc. During all these activities most of the time safety of information available on user device is compromised either knowingly or unknowingly. Sadly manufacturers, developers, owners, users of device are not taking smartphone information security seriously. While buying a PC (Desktop/Laptop), owner demands – manufacture provides antivirus-antimalware software either free or at negligible price. But while buying/selling smartphone, this is conveniently ignored. Today smartphones store much more sensitive data than PCs, are used by greater number of people from very early age of life than any other digital device. This shows need to provide security to smartphone at high level.

Approach discussed in this paper is static analysis of SMS and Android permissions and is designed for network service providers. In dynamic analysis, prevention method for user device is provided. This will help user in identifying botnet attacks from SMS as C & C while analysing dangerous permissions of Android applications to identify probable misuse of Android applications.

7. REFERENCES

[1] Guining Geng, Guoai Xu, Miao Zhang and Yanhui Guo, Guang Yang, Wei Cui, "The Design of SMS Based Heterogeneous Mobile Botnet", Journal of Computers, Vol. 7, Issue 1, pp. 235-243, 2012

[2] 2. Jing Liu, Yang Xiao, Kaveh Ghaboosi, Hongmei Deng and Jingyuan Zhang, "Botnet: Classification, Attacks, Detection, Tracing, and Preventive Measures", Hindawi Publishing Corporation, EURASIP Journal on Wireless Communications and Networking, 19 July 2009

[3] 3. Ihsan Ullah, Naveed Khan and Hatim A. Aboalsamh, "Survey on botnet: its architecture, detection, prevention and mitigation", IEEE Transactions, pp. 660 -665, 2013

[4] 4. Mohammad Reza Faghani and Uyen Trang Nguyen, "Socellbot: a new botnet design to infect smartphones via online social Networking", 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE, 2012.

[5] 5. Mahmoud. M, Nir. M and Matrawy. A, "A Survey on Botnet Architectures, Detection and Defences", International Journal of Network Security, Vol. 0(0), 2013, pp:1-18.

[6] 6. Abdullah, Z., Saudi, M. M., & Anuar, N. B, Mobile botnet detection: Proof of concept. In Control and System Graduate Research Colloquium (ICSGRC), 2014 IEEE 5th, 2014, pp. 257-262

[7] Anwar, S., Zain, J. M., Inayat, Z., Haq, R. U., Karim, A., & Jabir, A. N,A static approach towards mobile botnet detection. In Electronic Design (ICED), 2016 3rd International Conference on IEEE, 2016, pp. 563-567.

[8] Amro.B, "Malware Detection Techniques for Mobile Devices", International Journal of Mobile Network Communications & Telematics (IJMNCT), Vol.7 (4/5/6), 2017, pp: 1-10.

[9] Karim.A et al, "On the Analysis and Detection of Mobile Botnet Applications", Journal of Universal Computer Science, Vol. 22, no. 4 (2016), pp: 567-588.

[10] SMS Corpus [online], Available: https://archive.ics.uci.edu/ml/datasets.html,[accessed on : 24/03/2018]

[11] SMS Corpus [online], Available: http://www.esp.uem.es/jmgomez/smsspamcorpus/, [accessed on : 24/03/2018]

[12] SMS Corpus [online], Available: https://www.kaggle.com /shravan3273/sms-spam, [accessed on : 24/03/2018]

[13] Xuxian Jiang , "An Evaluation of the Application ("App") Verification Service in Android 4.2" , Associate Professor, Department of Computer Science, NC State University

[14] Ting Zhao, Gang Zhang, Lei Zhang, "An Overview of Mobile Devices Security Issues and Countermeasures", 2014 International Conference on Wireless Communication and Sensor Network, pp 439-443.

[15] Urcuqui, C., & Navarro, A. (2016, April). Machine learning classifiers for android malware analysis. In Communications and Computing (COLCOM), 2016 IEEE Colombian Conference on (pp. 1-6). IEEE.

[16] Ruchna Nigam, "Time line of Mobile Botnets", FortiGuard Labs, https://www.botconf.eu/wp-content/uploads/2014/12/2014-2.2-A-Timeline-of-Mobile-Botnets-PAPER.pdf

[17] Koen Van Impe, "Comparing free online Malware analysis sandboxes", June 1 2015, https://securityintelligence.com/comparing-free-online-malware-analysis-sandboxes/, [accessed on: 30/06/2018]