2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

Analysis of Mobile P2P Malware Detection Framework through Cabir & Commwarrior Families

Muhammad Adeel, Laurissa N. Tokarchuk
School of Electronic Engineering & Computer Science, Queen Mary University of London
{muhammad.adeel, laurissa.tokarchuk}@eecs.qmul.ac.uk

Abstract— **Mobile Peer-to-Peer (P2P) malware has emerged as one of the major challenges in mobile network security in recent years. Around four hundred mobile viruses, worms, trojans and spyware, together with approximately one thousand of their variants have been discovered to-date. So far no classification of such mobile P2P security threats exists. There is no well known simulation environment to model mobile P2P network characteristics and provide a platform for the analysis of the propagation of different types of mobile malware. Therefore, our research provides a classification of mobile malware based on the behaviour of a node during infection and develops a platform to analyse malware propagation. It proposes and evaluates a novel behaviour-based approach, using AI, for the detection of various malware families. Unlike existing approaches, our approach focuses on identifying and classifying malware families rather than detecting individual malware and their variants. Adaptive detection of currently known and previously unknown mobile malware on designated mobile nodes through a deployed detection framework aided by AI classifiers enables successful detection. Although we have classified around 30% of the existing mobile P2P malware into 13 distinct malware families based on their behaviour during infection, this paper focuses on two, Cabir & Commwarrior, in order to analyse the proposed detection framework.**

Keywords—***Mobile P2P Networks, Malware Classification, MPeersim, Malware Propagation, Mobile Agents, Malware Detection, Malware Families.***

I. INTRODUCTION

Mobile malware [4] has emerged as one of the major threats for modern day mobile P2P networks. Since first mobile malware outbreak in 2004, around four hundred mobile viruses, worms, trojans & spyware and one thousand of their variants have been discovered so far [1] [4]. Mobile P2P malware is capable of propagation through mobile P2P networks using three common approaches i.e. content sharing using 2.5/3G mobile Internet & WLAN, through Bluetooth communication directly among different peers and through MMS & SMS messaging. Damages due to malware propagation through any of the means can range from loss of privacy and transfer of unsolicited information to system malfunction and failure. Most critical consequence of mobile malware infection however is it causing service disruptions and economic losses.

Mobile P2P devices are resource constrained in terms of memory and processing and malware attacks can result in overwhelming these resources. Severe nature of malware attacks like Cabir and Commwarrior family malware attacks may target bandwidth resources of the network and have consequences as dangerous as Denial of Service [15]. Mobile phone manufacturers are equipping their smartphones with security software for detection of various kinds of security threats however it is not feasible to detect evolving and ever increasing malware attacks on the terminal because of the computational cost of complex detection algorithms. Leaving malware undetected however can allow it to launch severe attacks that can potentially scale network boundaries (e.g. in case of Commwarrior family malware).

This paper is aimed at introducing an adaptive lightweight mobile P2P malware detection framework. Discussion however will focus on Cabir and Commwarrior malware families only with an aim to explore different aspect of proposed framework. With Cabir and Commwarrior families under discussion, section II briefly explores technical characteristics of these families. To make the framework lightweight in terms of detection footprint, it was ideal to detect group based behaviour instead of detecting individual malware behaviours and thus existing mobile P2P malware was classified into different groups called families. Authors in [14] emphasize that there exists no technical classification of mobile P2P malware thus it becomes a vital part of this research to classify mobile malware into families based on their behaviours and characteristic during real-time propagation. Section III elaborates on the process of classification of mobile malware into different families with a focused illustration of how different types of malware are classified into Cabir & Commwarrior families.

Although there are various simulation environments for implementing and analyzing P2P related scenario and topologies, there are literally none that could map mobile P2P malware propagation and hence it became requisite to build a simulation environment that implements propagation of various families of malware and provides a pivotal platform for mobile malware propagation analysis. Section IV briefly elaborates on characteristics and capabilities of simulation environment MPeersim (i.e. Mobile Peersim) developed under this research. As MPeersim implements various individual mobile P2P malware and malware families, it gives this research a capability of closely analyzing the propagation characteristics of mobile P2P malware. Propagation analysis of families under discussion i.e. Cabir & Commwarrior in presented in Section V which unleashes how catastrophic

978-0-7695-4578-3/11 $26.00 © 2011 IEEE
DOI

malware from these families can be in terms of their infection efficiency and battery depletion. Finally, section VI presents proposed four-layer detection framework itself and elaborates on its detection capabilities again with a focus on detection of Cabir & Commwarrior malware families. It is vital to emphasize that through adaptive self-learning capabilities employed on the designated (agent) nodes, the framework is capable of detection of previously unknown malware (i.e. capable of detection in the regions of uncertainty).

II. CABIR & COMMWARRIOR MALWARE FAMILIES

Cabir and Commwarrior are the most dangerous mobile P2P malware families discovered to-date because of their proliferation capabilities, mutational characteristics and highest number of variants. After its first instance in June 2004, 33 variants of Cabir malware family have been discovered so far [13]. It a Symbian OS worm that targets mobile phones through Bluetooth medium. Victim mobile once infected, becomes portal for further propagation of this malware to all its Bluetooth neighbours. It is a monomorphic self-carried malware with localized propagation scope [12]. Some of the important consequences of Cabir family malware are increased network throughput, denial of service, battery depletion and causing mobile failure by corrupting system binaries.

Discovered in March 2005, Commwarrior was the first mobile phone malware using MMS technology for malware propagation alongside Bluetooth. So far its 16 variants have been identified. Use of MMS gives it a global scope in terms of malware propagation capabilities. It is a polymorphic malware using self-carried approach [12] and consequence of this malware are increased network throughput, denial of service and battery depletion attack and mobile failure. Use of MMS medium for threat propagation could cause severe economic loss to the victim as well.

Figure 1 gives an analysis of threat level identification of Cabir and Commwarrior malware families by different security software vendors which is yet another evidence of the potential threats they pose to mobile networks.

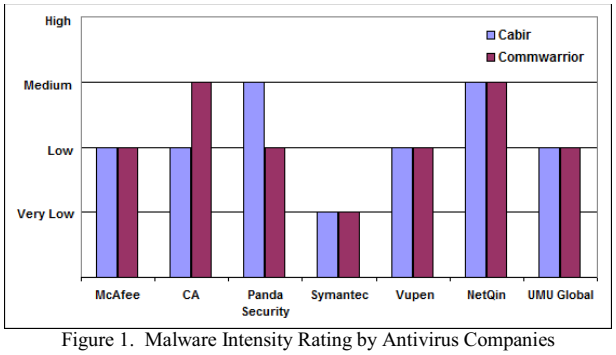

Figure 1. Malware Intensity Rating by Antivirus Companies

III. CLASSIFICATION OF MOBILE MALWARE

Realistically, it is very difficult to develop an electronic system for resources constrained mobile devices that could store updated information about all the existing mobile P2P malware and detects them when needed. Signature-based detection requires a large detection footprint, thus infeasible for mobile devices with limited memory while the same applies to some existing anomaly-detection techniques for mobile malware as they either rely on large content-based behavioural footprint [3] or record memory-intensive power signatures for abnormal behaviour detection [2]. Behaviour-based detection techniques if applied in their true essence may result in significantly reducing detection footprint on mobile devices and could prove ideal for detection on designated (agent) mobile devices. Mobile P2P malware is known to perform mutations. Hence it could not be detected through signature based detection which gives us yet another reason for employing behaviour-based detection for our framework. Adopting behaviour based detection alone was not a solution as recording behaviours of individual malware types in detection footprint could still prove memory intensive for mobile devices. Thus classification phase was a requisite to group malware based on their behaviour consequently reducing detection footprint on resource constrained mobile nodes.

Mobile viruses and worms do have commonalities in terms of their propagation behaviours. By exploiting these commonalities, we can potentially group malware into different classes or categories (called families under this research). As per authors in [2] & [14], there exists no technical categorization of mobile P2P malware hence as a novel contribution, a classification of mobile P2P malware is chalked out. In favour of space, this section demonstrates the classification mechanism only for two malware families (i.e. Cabir & Commwarrior) that are under discussion in this paper.

Figure 2 explains the classification mechanism which kicks off with rigorous analysis of individual malware in terms of their propagation characteristics. The attack strategy of a malware is then analyzed if form of sequence of operations it performs during attack. Top block in Figure 2 gives sequence of operations for five different types of malware (each presented on separate line). Of these sequences of operations and close analysis of infection characteristics of mobile P2P malware, distinct behaviours for groups of malware are identified. Individual malware exhibiting that behaviour is then added to the relevant malware category (i.e. malware family).

Based on the classification work published as part of our research under [14], around 25% of the existing mobile P2P malware was distributed across 13 malware families. In case of Cabir family, top line in the rectangular block in Figure 2 identifies the sequence of operations a Cabir-infected node undergoes during propagation. This information alone may not be sufficient to distinctly classify Cabir family. Thus from this Bluetooth transmission-specific operation set, we choose minimal distinct behaviour information (underlined in blue) and explore the infected node to acquire node-specific characteristics as well and record them in form of behaviour parameters. Cabir malware family is technically identified as Bluetooth-Propagator (BP) family as its core constituent-behaviour is *Propagator* from within Bluetooth set of behaviors.

2nd line in top block of Figure 2 explains the sequence of operations an infected node undergoes during Commwarrior family infections. Besides propagation through Bluetooth medium, Commwarrior family also uses MMS technology to replicate itself onto phonebook contacts. Thus a Commwarrior family malware not only exhibits *Propagator* behaviour under Bluetooth section of behaviors but also exhibits *N-Friends* behaviour (underlined in red) from within the MMS set of behaviors. In addition to these two transmission-specific behaviors, it obviously requires some node-specific behaviors to be acquired onto the infected nodes to distinctly identify Commwarrior malware family.

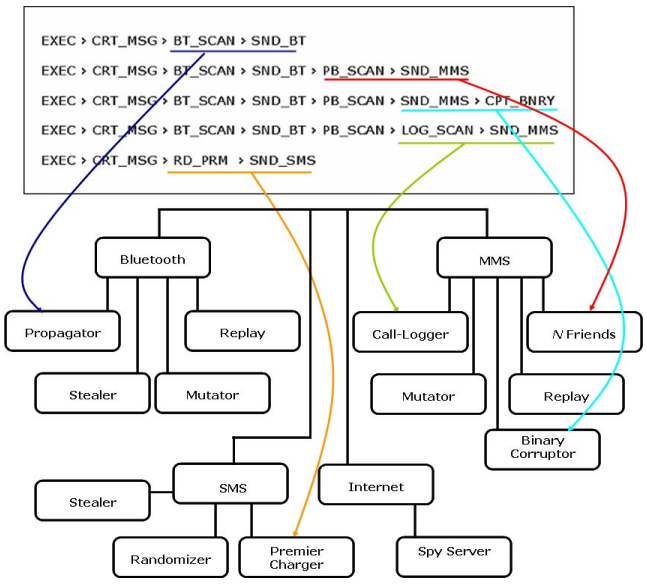

Figure 2. Classification of Mobile Malware (into families) based on Behaviour during Propagation

The classification approach under this research emphasizes on recording behavior parameters for groups (i.e. classes or families) of malware rather than storing them for individual viruses. Preferring behaviour based detection over signature based detection significantly reduces detection footprint. The same has been demonstrated in Figure 3 through analysis of memory required by the detection footprints of Cabir and Commwarrior.

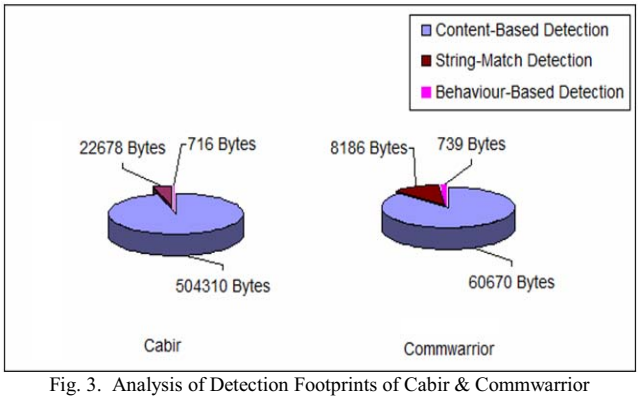

Fig. 3. Analysis of Detection Footprints of Cabir & Commwarrior

IV. Simulation Environment

Although there exist many simulation environments for P2P networks (thoroughly investigated by Stephen *et al* in [17]), there is none that could take into account the technical characteristics pertaining to mobile networks in general and mobile P2P malware propagation in particular. Thus MPeersim which is primarily a simulation environment for mobile P2P networks has been developed. Table 1 gives a brief overview of the capabilities of MPeersim in terms of what it can simulate. Different types of simulation topologies and scenarios can be constructed for a rigorous analysis of various aspects of mobile P2P networks.

TABLE 1. MPEERSIM SIMULATION CAPABILITIES

Normal File Types (Bluetooth, MMS, SMS)	Virus File Types (All families from Section 1)	Initial Virus Population (Variable)
Communication Types (Bluetooth, MMS, SMS)	Node Battery Power (Variable)	Node State (Active, Idle, Dead)
Node Associations (Bluetooth Neighbors, Phonebook Contacts)	Battery Usage Modes (Bluetooth, MMS, SMS)	Node Type (Agent, Normal Node)
Node Leave-Join Rate (High, Medium, Low)	Global Probability of Infection (0<GPI<1)	Initial Normal Files Population (Variable)
Node Leave-Join Time (Anytime)	Immunization (Bluetooth, MMS)	Mobile Node Type (Type 1, 2, & 3)
Node Mobility (Variable)	Bluetooth Neighbor Density (Variable)	AI Support (Neural Nets MLP & DTree C4.5)
Payload Type (Normal, Malware)		

Table 2 gives an overview of some of the results that could be acquired conducting MPeersim simulations. These results when analyzed could provide much needed insight into mobile P2P network communications in general and mobile malware propagation in particular.

TABLE 2. MPEERSIM RESULTS

Type	Description
Malware Prevalence	Prevalence of each family malware at any instance during simulation
Battery Power	Node, network and cumulative battery power of 2nd & 3rd degree neighbors
Throughput	Bluetooth, MMS, SMS throughput for node, network and cumulative throughput of up to 2nd & 3rd
Node Status	Number of Active, Dormant an Dead Nodes in the network at any instance during simulation
Threat Identification	Adaptive classification of activities into normal, and malicious activities
Sub-Threat Conditions Identification	Adaptive identification of flags (i.e. sub-threat conditions) on agent nodes through classification of instance/activities
Malware Family Identification	Adaptive identification of malware families based on agent nodes through classification of instance/activities
Node Relationship Diagram	Elaborates the Bluetooth associations formed by network nodes to constitute a network
Malware Propagation Modeling	Propagation modeling of generic mobile malware (i.e. Bluetooth, MMS, Hybrid) and actual mobile malware (i.e. all the families in Figure 3)

MPeersim platform provides a realistic mobile P2P environment with an emphasis on malware propagation in terms of their varying payloads, victim selection strategies and infection intensities. In MPeersim, two types of malware attacks in mobile P2P have been mapped. In first type of attacks, payload of the received transmission from infected node is a malware that infects the victim and prompt it to propagate the malware further and infect other devices. Cabir & Commwarrior belong to this category of malware. Other type of attack mapped into MPeersim is termed as repetition attack, in which a legitimate message is sent repeatedly to the victim (target) nodes. Purpose of such repetitions can be draining battery resources of the infected node or a denial of service attack on victim (target) node. Through analysis of its characteristics and capabilities, it can be safely claimed that MPeersim can prove to be a good resource in understanding and preempting about future mobile P2P malware patterns. Our research work under [18] provides a detailed discussion on MPeersim and its capabilities.

V. MOBILE MALWARE PROPAGATION

One of the novel and most useful capabilities of MPeersim is its support towards analysis of propagation characteristics of mobile P2P malware families. Before we analyze Cabir & Commwarrior families for their propagation characteristics, it is important to discuss epidemiological mobile P2P malware propagation models that have been implemented in MPeersim. The models below have been deduced from our research pertaining to P2P malware modelling [16].

A. SI Infection Model

MPeersim is capable of mapping mobile malware infection in terms of different models. The most basic is $S \rightarrow I$ model of malware propagation in which mobile nodes are divided into two classes in terms of infection i.e. susceptible and infected. When an infected node transmits to a susceptible node, susceptible node gets infected.

Given N as the total number nodes, S as the susceptible nodes, I as the infected nodes in the network, λ as Global Infection Probability (GIP) representing the rate of infection in network, ς as the average number of contacts a node has and α as the degree of immunization (or stealth against a particular malware), the rate of change of infected mobiles (or rate of infection) is be given by

$$\frac{dI}{dt} = \lambda \varsigma S \frac{I}{N}$$

It is imperative to mention that in equation above α is taken as 1 which implies that the susceptible nodes have no stealth against malware. MPeersim so far is capable of mapping malware propagating through three technologies i.e. Bluetooth, MMS and SMS. ς therefore represents mobile neighbors in Bluetooth communication while phonebook contacts in MMS and SMS communication.

With increase in I as infection spreads, S in the network goes on decreasing and hence the rate of change of susceptible nodes is given by

$$\frac{dS}{dt} = - \lambda \varsigma S \frac{I}{N}$$

Again, α in equation above is taken as 1 which means that susceptible nodes have no immunization and thus no defenses against the malware.

B. SIS Infection Model

SIS model of infection denoted by $S \rightarrow I \rightarrow S$ maps the possibility of infected node returning to a susceptible state after infection cleaning up. With infected nodes returned to the susceptible represented by $I_{I \rightarrow S}$ and rest of the parameters same as presented in section 5.1, the rate of infection is given by

$$\frac{dI}{dt} = \lambda \varsigma S \frac{I}{N} - I_{I \rightarrow S}$$

Similarly, the rate of change in susceptible nodes in the network is given by

$$\frac{dS}{dt} = - \lambda \varsigma S \frac{I}{N} + I_{I \rightarrow S}$$

C. SIR Infection Model

In SIR model of infection, an infected node can become immunized against a particular malware and could not get infected from the same malware in rest of the simulation. This model is represented by $S \rightarrow I \rightarrow R$ in which R represents immunized nodes i.e. the nodes that are immune from future infections. There are two types of immunized nodes. $R_{I \rightarrow R}$ represents the previously infected nodes turned immunized while $R_{S \rightarrow R}$ represents previously susceptible nodes turned immunized. With set of immunized nodes represented by $R_{I \rightarrow R}$ and rest of the parameters same as described in section 5.1, rate of increase in susceptible nodes is given by

$$\frac{dI}{dt} = \lambda \varsigma S \frac{I}{N} - R_{I \rightarrow R}$$

while the rate of change of susceptible nodes (getting infected) is

$$\frac{dS}{dt} = - \lambda \varsigma S \frac{I}{N} - R_{S \rightarrow R}$$

In SIR model, at any given time in simulation, there are two types of mobile nodes that could be immunized, susceptible node and infected nodes. With $R_{I \rightarrow R}$ representing infected nodes immunized in the network and $R_{S \rightarrow R}$ representing susceptible nodes the rate of increase in immunized nodes in the network is given by

$$\frac{dR}{dt} = R_{S \rightarrow R} + R_{I \rightarrow R}$$

D. SIRE Infection Model

SIRE model of communication is influenced by SIR model in which another realistic malware propagation condition has been taken into account. Where in SIR model it is possible that an infected and even susceptible node could get immunized from a specific malware forever, it is quite possible as per SIRE model that some of the infected nodes get back to susceptible status after cleaning up. Representing infected nodes returned to the susceptible represented by $I_{I\to S}$ and infected nodes represented by $R_{I\to R}$ and rest of the parameters same as described in section 5.1, rate of infection is given by

$$\frac{dI}{dt} = \lambda \varsigma S \frac{I}{N} - I_{I\to S} - R_{I\to R}$$

Similarly with $R_{S\to R}$ representing susceptible nodes that have been immunized, the rate of change of susceptible nodes is given by

$$\frac{dS}{dt} = -\lambda \varsigma S \frac{I}{N} + I_{I\to S} - R_{S\to R}$$

Just like SIR model, the rate of change of immunized nodes is dependent on $R_{I\to R}$ and $R_{S\to R}$ and is given by

$$\frac{dR}{dt} = R_{S\to R} + R_{I\to R}$$

Worms behave differently during their propagation in the network. Detection based on of such malware is highly dependent on learning of propagation behaviours of mobile malware. Hence this works analyzes the propagation characteristics of generic mobile P2P worms as well as various worm families.

As discussed, simulation environment MPeersim has been developed to analyze the propagation of different types and classes of malware and detect them in subsequent phases of the project. This environment has helped realize the extent of damage different families of malware could potentially cause.

Based on SI model of propagation, Figure 4 & 5 give propagation analysis of Cabir and Commwarrior malware. In both the propagation related simulations below, N (i.e. total nodes in the network) was set at 10000, initial population of infected nodes with Cabir and Commwarrior was set at 100 each, maximum number of Bluetooth contact a node has at 7 and average number of phonebook contacts for nodes at 83, the propagation of malware has been analyzed on varying values of global probability of infection (i.e. GPI = 0.5 & GPI = 0.8) and maximum battery power of the nodes (i.e. 900 *mAh* & 4000 *mAh*).

Comparison of propagation statistics exhibited through both the graphs below reveals that Commwarrior is more dangerous than Cabir based on the factors like propagation efficiency and battery depletion. Firstly it was observed that with infection probability of 0.8 in case of Commwarrior simulation, around 99% of the nodes got infected of Commwarrior within 82nd time unit while in case of Cabir, with the same infection probability, it took the malware around 180 time units to effect 99% of the nodes. This provides evidence of Commwarrior being almost twice as efficient as Cabir in terms of propagation. Similarly in terms of battery depletion, in the simulation with Commwarrior (with GPI as 0.8) at the end of 1000 time units' simulation, only 6572 nodes with 900*mAh* battery stood alive till the end of simulation as compared to 7040 in case of Cabir thus revealing that Commwarrior is more battery depleting compared to Cabir.

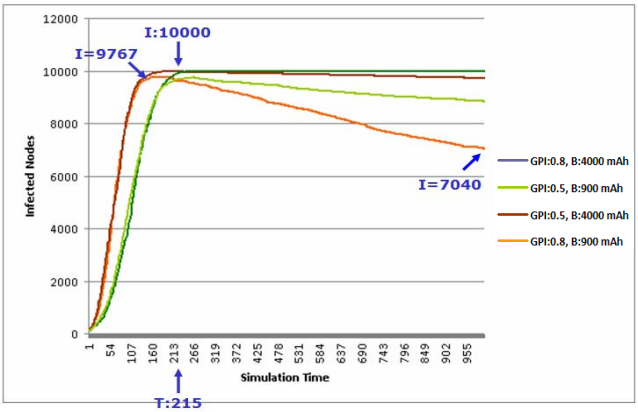

Figure 4. Propagation Analysis of Cabir

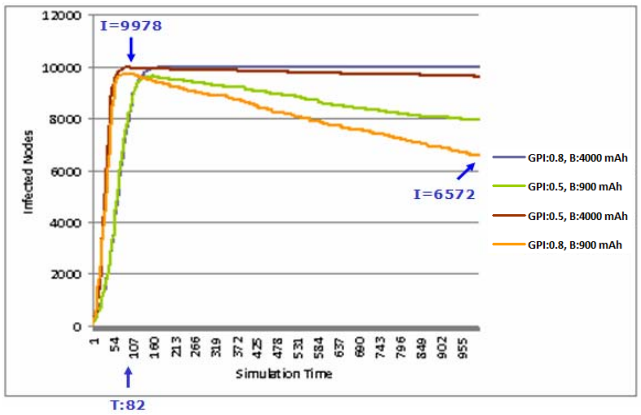

Figure 5. Propagation Analysis of Commwarrior

VI. Detection of Mobile Malware

Literature in abundance could be found for P2P malware detection [6, 7, 8, 9, 10 & 11] however very little exists in terms of mobile P2P malware detection in general and real time automated detection of mobile P2P worms in particular. Kim *et al* have come up with a ground breaking research in terms of anomaly detection in mobile P2P networks based on power signatures of devices during attack [3]. Our research has however identified critical weaknesses in Kim *et al*'s work due to which it could not be considered feasible for resources constrained mobile P2P network.

Behaviour based anomaly detections systems are known to perform better than signature based detection techniques in conditions where malware perform constant mutations [5]. The most common mobile malware families like Cabir demonstrate

mutational characteristics and are evolving at a very rapid pace. Thus there exists a growing need of a detection mechanism for mobile networks that is not only capable of detecting malware mutations but also cops-up with malware evolution. Another major contribution of our research is a behaviour based detection framework capable of (1) detecting abnormal activity in the network, (2) identifying sub-threat conditions in the network and (3) providing distinct identification of various malware families. Based on the concept of detection on designated (agent) nodes, this 4-layer detection framework records propagation behaviour of nodes against different behaviour parameters and then uses those parameters to acquire an adaptive AI based detection of known and previously unknown threats and identification of malware families. As the mobile devices are resource constrained in terms of memory and processing resources, attempt have be made to keep detection footprint of this automated AI based detection to a minimum, i.e. without compromising on detection quality. Although attempts would be made to further minimize the detection footprint, analysis in section 3 of this paper reveals that our approach is by far better than existing techniques used for mobile P2P malware detection with regards to the size of detection footprint.

Layer-1 of our four-layer mobile P2P malware detection framework records statistics from the network against various parameters. NODE_STATUS is a parameter with possible values as Active, Idle and Dead. Similarly BT_SCAN and PB_SCAN are two other parameters with Boolean values representing whether Bluetooth-neighbors or phonebook-contacts scan on a particular node has been performed. Some other parameters could be observed in Figure 6. Our framework relies on AI based classifiers for detection thus training time becomes a critical issue. It was observed that detection rules based on layer-1 parameters would require AI based prediction engine to take much longer in training. Moreover, detection accuracy of such predication engine was considerably degraded. It was thus decided to pre-process layer-1 parameters to convert them to layer-2 composite parameters and use them as an input to the classifier. Thus in layer-2 of the framework, some of the parameters are preprocessed to acquire composite parameters. Each composite parameter carries a Boolean value representing whether or not a respective condition pertaining to that composite parameter is true. Composite parameters can be considered functions that receive layer-1 parameters and return a Boolean value. Some of the composite parameters are presented in Figure 6. It is important to mention that layer-2 of the framework only focuses conversion of parameters in to composite parameters.

Detection of flags is the most vital part of the detection framework under this research and forms layer-3 of our framework. Statistics collected in terms of various composite parameters are fed to the AI based classifiers trained on detection rules to identify sub-threat conditions in terms of flags. This subsection discusses two important flags pertaining to Cabir and Commwarrior families (under discussion in this paper).

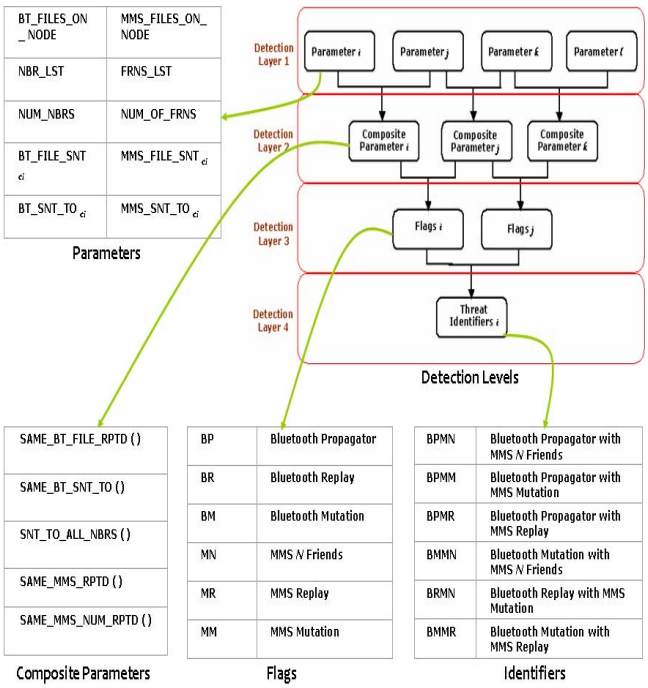

Figure 6. Multi-Layer Malware Detection on Agent Nodes

1) *MMS N Friends (MN):* MN is one of the most important MMS flags because if this flag is SET along with one more flag from Bluetooth category of flags, it can give rise to one of the deadliest attacks in mobile P2P networks. This flag is SET only if the condition below satisfies.

$$(MMS\_FILE\_SNT_{ci} == MMS\_FILE\_SNT_{ci-1}) \;\&\&$$
$$(NUM\_MMS\_SNT\_TO_{ci} == NUM\_MMS\_SNT\_TO_{ci-1})$$

MMS_FILE_SNT is the behaviour parameters identifying MMS file sent by mobile node and *NUM_MMS_SNT_TO* is the number of MMS messages mobile device sent. $_{ci-1}$ represents the previous cycle while $_{ci}$ represents current cycle. It is vital to note that *MMS_FILE_SNT* and *NUM_MMS_SNT_TO* are layer 1 parameters while the condition above in terms of composite parameters can be written as

(SAME_MMS_FILE_SNT && SAME_NUM_MMS_SNT)

2) Bluetooth Propagator (BP): BP flag in the third layer of attack detection is the most vital flag in terms of detection of mobile P2P threats. It would not be exaggerated to say that this flag helps in detection of most threats than any other flag in BT, MMS and Replay Attacks categories. This flag is SET if the following conditions holds.

$$(BT\_FILE\_SNT_{ci} \;\&\&\; BT\_SNT\_TO_{ci-1} == BT\_SNT\_TO_{ci})$$
$$\&\& (NUM\_BT\_SNT\_TO_{ci} == NUM\_NBRS)$$

BT_FILE_SNT identifies the Bluetooth file sent by the node while *BT_SNT_TO* holds the name of receiver of this file.

Total number of Bluetooth files sent is recorded against *NUM_BT_SNT_TO* while *NUM_NBRS* gives total number of neighbours this node has. Again *BT_FILE_SNT, BT_FILE_SNT_TO* and NUM_NBRS are layer-1 parameters while condition above in terms of composite parameters can be written as

SAME_BT_FILE_SNT && SAME_BT_RCVR && BT_SNT_TO_ALL_NBRS

Rather than depending on rigid rule based detection, as a novel contribution, various AI based classifiers have been introduced into the framework to classify real time instances on mobile nodes into sub-threat conditions (flags). It is important to mention that combination of various flags in next level of detection can help distinctly identify malware families. Figure 7 and 8 compare the performance of two AI based classifiers in terms of detection of flags on a designated agent node during a 100 time unit simulation. In the simulation *N* was set at 100 nodes while *I* as initial population of Cabir and Commwarrior was set at 1 each. SI propagation model was used to conduct simulation. This comparison is based on the classification of individual instance (in terms of set of composite parameters) fed to these classifiers in real time. Each instance contains composite parameters representing transmission behaviours of the selected node on a given time unit. Manual analysis of instances fed to the classifier and detection results acquired through DTree C4.5 and NN MLP classifiers reveal that 100% of the sub-threat conditions (i.e. flags) pertaining to Cabir and Commwarrior were correctly detected by both the classifiers. In presence of other malware families in the simulation though (and consequently with greater number of detectable sub-threat conditions in the network), performance of DTree C4.5 was better than NN MLPs. Although conclusions section would elaborate further on performance and characteristics of each classifier, factors like memory intensiveness, processing complexity and considerably higher training time make NN MLP not a better classifier option for resource constrained mobile devices.

Once different sub-threat conditions (in form of flags) have been identified, we go a step ahead towards fourth layer of detection in our framework (i.e. family-identifier based detection) that detects and distinctly identifies known and previously unknown malware belonging to these families based on various combinations of layer-1 parameters, layer-2 composite parameters and layer-3 flags. Cabir family can be identified if

NODE_STATUS = IDLE && BT_SCAN = TRUE && BP = TRUE

In condition above, *NODE_STATUS* and *BT_SCAN* are layer-1 parameters while *BP* is a layer 3 flag. Similarly Commwarrior family can be detected if

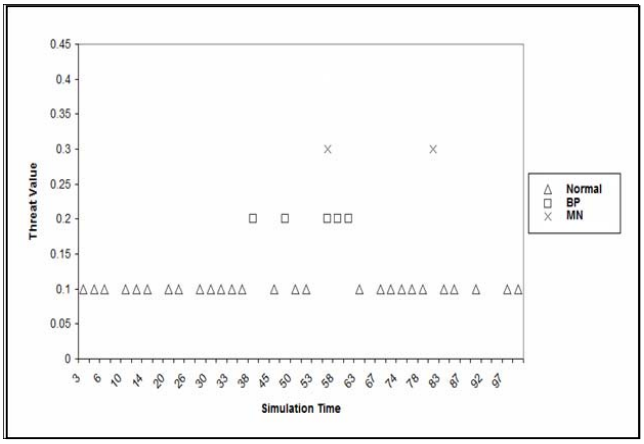

Figure 7. Detection of Flags through DTree C4.5

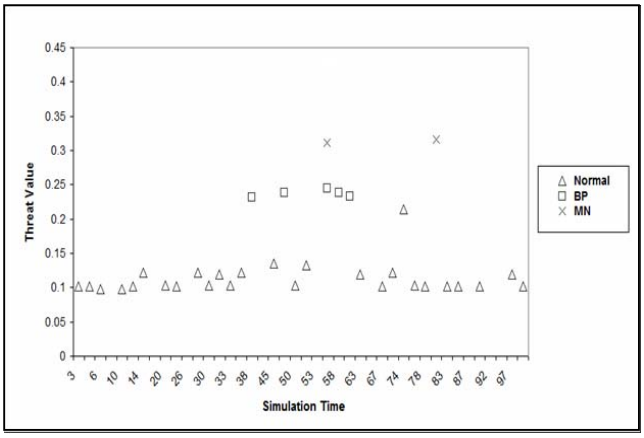

Figure 8. Detection of Flags through NN MLP

NODE_STATUS = IDLE && BT_SCAN = TRUE && BP = TRUE && PB_SCAN = TRUE && MN = TRUE

Although the rules can be hard coded as mentioned above to detect the malware, our research has incorporated AI based classifiers into the framework to classify instances into different families in real time. It not only takes care of the rigidity of rule based detection but also result in detections in region of uncertainty. It makes our approach ideal for ever evolving mobile P2P malware families and helps in detection of their unknown variants performing mutations.

With initial population of Cabir and Commwarrior family infected nodes as 1 each in the network of 100 nodes, Figure 9 & 10 give detection results of DTree C4.5 and NN MLP based prediction engine trained on layer-4 family-identifiers which comprise of flags, composite parameters and parameters. Detection again was performed at designated agent nodes and results of detection of Cabir & Commwarrior malware families detected at an agent reveal that both the classifiers were capable of detection of families with 100% accuracy. In conditions where malware from other families is propagating in the network, detection performance marginally deteriorates.

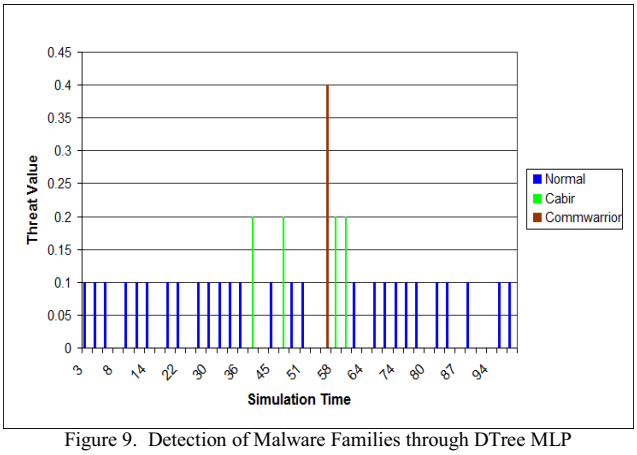

Figure 9. Detection of Malware Families through DTree MLP

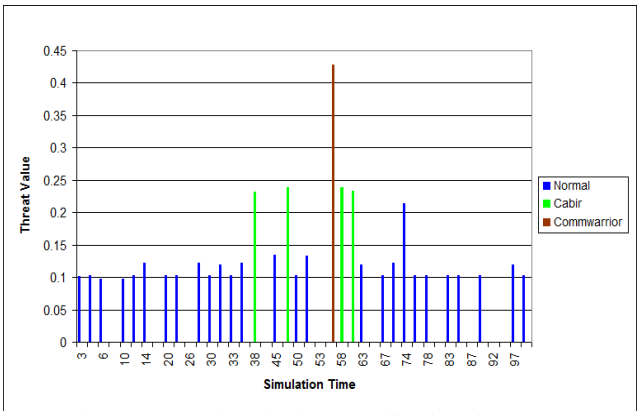

Figure 10. Detection of Malware Families through NN MLP

VI. Analysis & Conclusions

Development of mobile P2P malware detection framework can be divided into four distinct phases i.e. classification of mobile P2P malware, development of simulation environment, propagation analysis and finally the detection of mobile P2P malware. Classification phase of our research has grouped malware into 13 families that in total encompasses around 25% of the existing known malware. This paper however explains the whole framework (and its phases) in perspective of just two families i.e. Cabir and Commwarrior. Behaviour based detection adopted under this research has considerably reduced the detection footprint with literally no compromise on detection accuracy. A mobile P2P simulation environment has been implemented which not only is capable of taking various mobile P2P characteristics into account but also provides a much sough-after portal for analysis of mobile P2P malware. This portal will enable research community to preempt into the propagation behaviour of future malware. Propagation phase of the project has evaluated various kinds of viruses and results have revealed that the Commwarrior family malware are more dangerous than Cabir in terms of battery depletion, propagation speed and attack strategy.

Detection phase of this framework is capable of (1) detecting abnormal activity in the network, (2) identifying sub-threat conditions in the network and (3) distinctly identifying various malware families. Analysis has revealed that our approach of behaviour based detection requires minimal detection footprint thus highly desirable for resource constrained mobile P2P networks. Based on analysis gathered from different agent nodes in the network, use of AI based prediction engines has enabled our framework correctly classify 19% of the unknown (previously undefined) instances into malware which elaborates that we are now capable of detection in regions of uncertainty.

It was also observed that DTree based C4.5 detection module despite its minimal footprint and considerably lower training time, is equally good as Neural Nets based MLP in terms of detections accuracy. Figure 11 plots the Normal instances detected through DTree C4.5 against NN MLP for the same simulation. It can be observed that the instances classified through DTree C4.5 are uniform while detection results from NN MLP are irregular. NN MLP may thus require fuzzification of classified instances that may consume additional battery and processing resources on resources constrained mobile P2P devices.

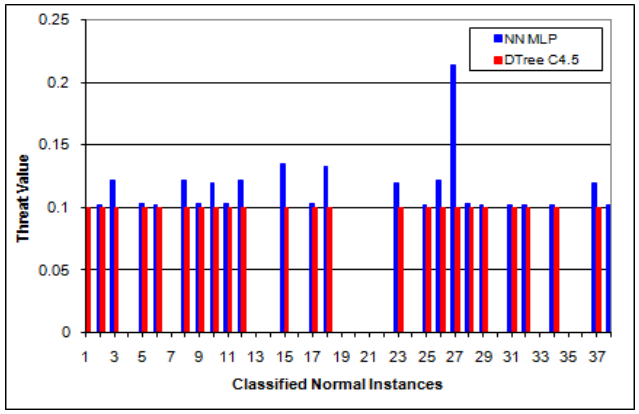

Figure 11. Comparison of Detection Accuracy

References

[1] McAfee Threats Report Fourth Quarter 2010. Available from:http://www.mcafee.com/resources/reports/

rpquarterly-threat-q4-2010.pdf

[2] Hahnsang Kim, Joshua Smith, Kang G. Shin, "Detecting Energy-Greedy Anomalies and Mobile Malware Variants", Proceedings of MobiSys'08, June 2008, Breckenridge, Colorado, USA.

[3] Pu Wang, Marta C. González, Cesar A. Hidalgo, "Understanding the Spreading

Patterns of Mobile Phone Viruses", Science, Vol. 324, No. 5930, May 2009, pp. 1071-1076.

[4] Mikko Hypponen, " Malware Goes Mobile", Proceedings of Scientific America Inc., 2006

[5] Carey Nachenberg, "Computer Virus Coevolution", Proceedings of the ACM Communications, January 1997, Vol. 40, No. 1

[6] Zhou Ruili, Pan Jianfeng, Tan Xiaobin and Xi Hongsheng, "Application of CLIPS Expert System to Malware Detection System, Proceedings of International Conference on Computational Intelligence and Security, 2008

[7] Norbik Bashah, Idris Bharanidharan Shanmugam, and Abdul Manan Ahmed, "Hybrid Intelligent Intrusion Detection System", Proceedings of World Academy of Science, Engineering and Technology, ISSN 1307-6884, Vol. 6, June 2005.

[8] Liberios Vokorokos, Anton Balaz and Martin Chovanec, "Intrusion Detection System Using Self Organizing Maps", Proceedings of Acta Electrotechnica et Informatica No. 1, Vol. 6, 2006

[9] Mehdi Moradi and Mohammad Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks" Proceedings of 2004 IEEE International Conference on Advances in Intelligent Systems Theory and Applications, Luxembourg Kirchberg, Luxembourg, 2004

[10] Hassina Bensefia1, Mohammed Ahmed-Nacer, "Towards an Adaptive Intrusion Detection System: a Critical and Comparative Study", Proceedings of 2008 International Conference on Computational Intelligence and Security

[11] C. Zhang, J. Jiang and M. Kamel, Intrusion Detection using Hierarchical neural network, Pattern Recognition Letters, Vol.26, Elsevier, 2005, pp. 779-791.

[12] Jamshed Sadiq, "Classification of Mobile Viruses", MSc Thesis Report, School of Electronic Engineering & Computer Science, Queen Mary University of London, August 2009.

[13] Online Available at: http://www.symantec.com/security_ response/writeup.jsp?docid=2004-061419-4412-99

[14] Adeel M. et al., "Classification of Mobile P2P Malware Based on Propagation Behaviour", In Proceedings of 4th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM 2010, Florence, Italy, October 2010.

[15] Dong-Her Shih, "Security Aspects of Mobile Phone Virus: A Critical Survey", Industrial Management & Data Systems, Vol. 108 Iss: 4, pp.478 – 494

[16] Chaosheng Feng *et al*, Propagation Models of Passive Worms in P2P Networks, Proceedings of 2008 IEEE Conference on Cybernetics and Intelligent Systems, Chengdu, China, September 2008

[17]Stephen Naicken Anirban Basu Barnaby Livingston Sethalat Rodhetbhai, "Towards Yet Another Peer-to-Peer Simulator", In Proceedings of The Fourth International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs), Ilkley, UK (2006)

[18] Adeel M. and Laurissa Tokarchuk, "MPeersim: Simulation Environment forMobile P2P Networks", Proceedings of 19th International Conference on Software, Telecommunication and Computer Networks (SoftCOM 2011), Split, Coratia