Nasn: A Novel Approach For Securing Network From Malware Injection

Nooh Bany Muhammad

Abstract: With immense globalization and the fast growth of technologies, the world is now running with the real time technologies which means the communication made over the internet and the data is fetched simultaneously from the website. There are different websites which include the vulnerability of hosting the suspicious activities like hosting of the Malware or Worm which cause serious effect on the running system. Malware is such a threat which is injected silently and creates a massive affect on the system by creating different types of syndrome on the system like the system slow down, unexpected shut down and even the data breach. Data breach is actually the main target of the hackers through which the hacker steals data from the database and flies off. This is because of unsecured network protection and unreliable software hosting on system. So, to make the system secure over the network the protection should be upgraded with new approach so that the unauthorized access to the network can be restricted. As the hackers leave no stamp for their identification, so after the data breach takes place, the IP of the hacker cannot be recognized. This is the reason for which the hackers are hard to be recognized. There are different approaches to prevent the suspicious access but most of them are basically cracked by the data hijackers. In this paper, the discussion and the approach are made through which the unauthorized access can be obstructed and thereby the probability of hijacking of the host system can be minimized by using the Artificial Neural Network.

Index Terms: Malware detection, Malware protection, Artificial Neural Network, Networks Security, Malware prevention, Internet security, secure communications.

1. INTRODUCTION

Internet is the widely used in communication media nowadays in the world for making communications better with lesser cost. If one wants to make a call to foreign land or wants to send a message, the telephone or mobile phone service charges higher than that of the internet communication [1,4,5]. With the fame of the social networking, the communication becomes better in the terms of talking to others; make file sharing, accessing news by browsing internet, keeping documents secured by uploading on cloud and so on. As the technology progresses, the usage of the internet and its browsing is growing up with more demand [2,4]. Most of the people use to search websites randomly and are mainly searching from google or yahoo search engine. Not only from the web browser, people use the internet in different applications also like the social messaging applications like WhatsApp, Telegram etc. [1,6]. Basically, they are using the internet to access the service through the General Packer Radio Service to access the data in form of packet [2,4,6].

The use of internet is dependent upon the intention of the user like what they are browsing in which way and it may be with positive desire or may be with negative desire. In case of positive desire, internet is much helpful to the human for the development of the mankind by developing the skills and knowledge. It is also helpful to enhance the business and business opportunities as idea of such business can be gathered from the internet data [3-5]. There are many important websites which are the platforms of such works. On the basis of those, the user can develop themselves. So, it has the impact on the economic development. On the other hand, internet provides the way to communicate with the people through social media like Facebook and LinkedIn [3,6-9]. These are the websites through which the communication among the people is made better in the sense of the Communication, Product development, advertisement and interactions. Using such social media user can communicate with their friends who are far away from him and with a matter of click and enter with their desired message, it will reach to their friends just within a moment. This will not consume much time and even not make delay in conveying their messages. Again, with the use of the social media, they can develop their products and make those items to sell. As the featured utility of internet is that, once anything is uploaded publicly in it, the people around the world can view it and this is the top advantages of the use of internet. With that product development, user can make a good business over the network. So, these are the advantages of the internet for which its demand is too much increasing nowadays. This is the scenario of the people who love to browse and want to make some knowledge from internet [4,7, 9-11]. Even it helps in the use by providing some good lifestyle tips and suggestion which seems to be necessary for most of the people in the world. In the past era, people used to visit the marketplace for shopping of different goods as there was not other way to make it happen. Internet provides the world with the facility of online shopping where people can make their desired shopping by staying at home and even, they don’t need to go to the bank for payment related issue and all the work related to bank can be operated from home itself by using the internet [5,7]. Very recently, the mobile applications have developed and so that use of internet got a hike as application services

• Nooh Bany Muhammad is an Assistant Professor in the department of Computer Science and Information Systems at American University of Kuwait. He received the Bachelor degree of Computer Science and Information Systems from the Jordan University of Science and Technology, Jordan, in 2007, the M.Sc. and Ph.D. degrees (with honors) in Computer Science from the University of Southern Mississippi (USM), Mississippi, USA, in 2010 and 2015, respectively. In 2009, he joined the School of Computer Science, USM, as a graduate researcher, a project developer and system analyst for several projects at the University of Southern Mississippi. And in 2013 became a lecturer at School of Computer Science, USM. His current research interests include Security in Wireless Sensor Networks, Mobile Networks Security, Routing Protocols, and Wireless Sensor Networks Deployments. Dr. Muhammad is a Fellow in the honor society of the University of Southern Mississippi; He is a Life Member of the Jordanian Society of National Researchers, Jordan. IEEE Society.
serves the user with all kind of work like, shopping, internet banking and so on [6,8]. These are the good aspects and the positive desire to use the internet service. But the internet itself proves harmful if it is used wrongly anyhow. It may be wrong in the sense that the user is using the internet for some wrong work like they are working on something which can cause harm to the society [4]. Generally, the user does not bother about the status of a web link and they browse whatever the search engine provides them. Some of the internet source contains suspicious activities like malicious content. Sometimes, when the browser is opened with some web link, the pop-up window of that browser becomes activated by detecting the online advertisement [2,6-9]. Most of the time, the advertisement contains some malicious content and if clicked by chance, the computer of the device from where the browsing is made, may be affected. The pop-up window basically detects the suspicious activity of that website. There are different ways of attack to the user like through the browser, through the mail, through online transactions etc. So, use of internet for all the purpose is not viable for the user all the time [10, 12]. The activity of different suspicious components differs with the change in environment. Some of those attacks the Win32 to grab the control over the system as Win32 are the section of Windows from where the entire operation of that operating system commences [10,12-15]. If the Win32 is hijacked, it will be easier to grab the system through more easy way. Another type of suspicious component makes themselves replicated after entering into the system and making the operating system slower. With that case, while with such infection on the system, if the user still browse, the browser will respond slower and the hackers will get enough time to read the information through the internet [2,8,13-16]. One other kind of system threat may capture the data by reading the keys which are pressed by the user. All that the hackers need is to make money by snatching from the general and unsecured users their personal data and the transaction reports. Those data may help them to crack their system and the other systems connected to it. One of the dangerous suspicious elements is the Malware. It is one type of malicious code which is developed by the attackers who have a wish to make a penetration into the system to fetch all the important information including the login credentials and the transaction data. Malware generally observes the cookie and the cache data where all type of data is available through browsing [13, 16]. While hijacking, the attackers fetch all the information from there and store those into their database. The malware attacks basically in large organization or may be in a single computer. In case of large organization, there are several security schemes. The hackers, before the malware injection, observe the network for a long time to understand about the policy made by the CISO [15, 18-20]. After proper understanding, the hackers use to penetrate through the network flood and behave like one of the host or users of the network. So, they continue to communicate as a regular user and slowly they steal all the valuable data from the server or may be from cloud. The concept of stealing of data from local server or local database to the cloud server or cloud database is prevalent in hacking [16, 18]. For hacking of server, the hackers need to pretend themselves as one of the IP holders of the network. With the access, they behave in such a way that the server gives the access to the hackers. In the meantime, the Malware expected to be injected which searches for the data in an automated way to make the confirmation for the source device for the data storage [17,18]. After getting the access to the server through the Malware, the hackers then penetrate into the system and fetch all the data that a server holds. On the other hand, to make access to the cloud, the hackers try to hack the email id first [13,19-21]. As each cloud service is authenticated by the email id, they send mail to the user which contains the malicious code. While the user clicks the mail and the link inside, the data of that mail service will be read by the hackers. So, they can gain access to the email through the access of the mail id and the password. If they are able to achieve both these two, they will be able to access to the linked cloud administration and after that they can make the penetration through the cloud storage and then they will be able to fetch all the stored data from there [16,20]. So, there is an urge to prevent the system from the malware attack by applying some rules so that the malware can be detected while in the runtime penetration and can have the measure to block or at least obstruct those. In general, there is no way to make a full protection against the malware as it is the program code and so it can be designed in various ways to destroy the protection of the e-environment at any cost. So, better is to monitor those by vigilance of the network gateway. Presently, there are two ways of the malware detection and the making of the protection shield [20]. The first one if the Personal Vigilance over the network by network monitoring to see if there are any vulnerability caused by some suspicious activities. The second one is the use of the network monitoring tool to make detection as well as the obstruction of the malicious code. Basically, these protective measures are also built up by some code with some encryption methods like RSA, DES, and Blowfish etc. [21]. But those are not sufficient to prevent the malware attack nowadays as the technology is advanced far and punched through the Artificial intelligence [22,24]. Artificial intelligence is the technology which has the huge demand in the field of information technology but it has the dark sides also. For example, the Ransomware, which is the Artificial Intelligence enabled Malware locks the user system until the user pay for it and the payment is done by the bitcoin or some other available cryptocurrency in the black market [1,5,9]. By the time one tries to prevent the Ransomware a lump sum may be lost. Another dangerous example of the malware if the MalGEN code which is again the Artificial Intelligence influenced malware and used to make fool the malware detection tool. The available solution is the sandbox but still there are such kind of malicious coding which is so intelligent that they pass the security test and the only option left is the monitoring and the site blocking. So, there are different ways to arrange the algorithmic view to prevent the malware [2,10,11]. In this paper, the new approach to the detection algorithm will be proposed so that the malware detection can be made more efficient and the protection power will be enhanced so that the suspicious activity cannot harm the existing system. Different tools can be used for this purpose. In this paper, the network is captured with some activities and the captured file is stored in a location in the system [10,13]. This network capture file is analyzed in the Wireshark tool to detect the malware and for the suspicious activity over the network, and on the basis of that, the suitable algorithm is designed. Wireshark is also capable of visualizing in run time, no matter, as for the research purpose; this tool is used for the analysis of the suspicious activities of the network with different filters [12,14].
2. PREVIOUS WORK

Years ago, a malware was introduced in the market in 1971 which was named as Creeper. This was an experimental design to show the movement of the computer code as a virus from one machine to another by creating replica. This was one of the critical vulnerable components that were first introduced because if the replica is generated and transmitted from one computer to another, the information and the personal documents had a probability of leakage [14,15]. Day by day, the sensitivity of the malware increased because of more experiments. From that time, the hackers are used to get access of the large organizations by such attacks. But initially, there was no such system which can prevent such kind of attack [1,3,6]. But to protect the industry and other large organizations from these attacks which tend to steal the sensitive data by creating replica, Creatavin et al, 2005 have mentioned in their paper about the protective system for the worm and malware attacks. But not only the obstruction of the malware of a particular system is enough to make the system secure but also the network security is needed as stated by Bojrkman etc. al, 2014 for the security perspective of large industrial sector to prevent the data breach [4,11,18].

Generally, the malware attacks on the large network target the database and the sensitive documents. The hackers generally target the distributed networks. In the distributed network, the hackers get enough space to penetrate and if once they are capable of penetration, they get the entire access of the network. So, it will be easier for them to make steal from there [12,20,21]. So, to maintain a large network and to protect it from malware attack, the sophisticated anti-malware or the firewall will be activated. Firewall is the software which automatically monitors the suspicious activity at the gateway and if it is detected to be harmful, it has the power to obstruct those [23,24]. This technique is known by the Intrusion Detection and here the intruders are the malware which unauthorized tends to penetrate into the network [21]. The malware generally enters into the system through the IP gateway with a pretention of being a member of the system. One of the important vulnerable issue is described by the Denial or Service or DoS with which the hackers can easily penetrate through the network and is able to catch all the network document from the network database. The DoS attack is a type of technique where the attackers make the resource unavailable to the real user during the attack. So the hackers can easily grab the necessary documents and those what they want [13,16].

As the real user cannot see the network progress then, they even cannot understand what is happening as the network completely gets down. When the attack period is over, then the real user gets activated to use their system and also gets access to the network resources. This DoS attack is made through Malware and a good example is the Ransomware [25,27,28]. With an increase of the network vulnerabilities, the system architecture also changed for the protection purpose. If the previous system is considered, they were less immune to those kind of attacks as because no such anti system was developed. As for the technological progression, the system is developed and made immune to the malware attack with an introduction to the automated network. Automated network means that they can be automated such that the network monitoring can be done automatically and the filtering can be done without human intervention [26,27]. This is possible by the invocation of Artificial Intelligence which helps the system. Nowadays, the organizational system as well as the malware systems both are transferred to the newly incorporated Artificial Intelligence system and every time new algorithm needs to be developed to fight against the malware attacks [22,23]. Mostly the Neural Network takes the major algorithms for the purpose. Artificial Neural Network is a system or better it can be signified by the technology which is able to find the root of the malware by reading the network paths of the virus. It is also capable of finding of the sensitive paths of the directory where the probable attacks may take place. So, invocation of Artificial Neural Network is the efficient way to detect and to prevent the network vulnerable issues and to make an algorithm in the software to be built up for the prevention of such suspicious activities [21,23,24].

3. TYPE OF MALWARE

The introduction of malware was from the very beginning of the invention of the computer system. Basically, with the introduction of the book, “Theory and Organization of Complicated Automata” by Vann Neuman, the world becomes familiar with the self-replication of computer programs with the propagation and application of Vann Neuman Architecture [23]. This can be told as the conceptual innovation of the computer virus as the attribute of the virus is to replicate itself with the introduction into new system. This cause the system to slow down and the penetration into the system is easier for the hackers and they can steal the computer data [21,24]. The first computer virus program was written by Bob Thomas in early 1970 which is famous as the Creeper Worm. Though it was for the experimental purpose, but it was one of the important innovations in the world of the hackers which introduced the new system of the replication capacities by itself. This Creeper Virus penetrates and keeps itself alive through ARPANET and it shows the message that “I am Creeper, Catch me If you Can”. However, the term virus was not introduced then. It was introduced in the year 1986 in a PhD thesis of Fred Cohen who is known to as the Father of Virus [23,25]. In the early era of the technology, the malware was used to be spread through the floppy disk or other flash drives or even through the compact disk also. But with the advancement of the internet aspects, the developer of those virus and malware, quickly adapted the technology of the internet and used to spread the virus through the network medium [15,23,26]. The general aspects of the hacker id are to send the email link to the user which contains virus code when they will click on the link.
The general medium of spreading of such virus which can be accessed from any kind of devices and platform such as the operating systems [22]. There are different types of malwares which cause different types of vulnerability to different places and with different issues. Some of the important are discussed below:

3.1. Adware: It is a type of malware which is used to deliver advertisement automatically as the name specified by Advertisements Supported Software. The most common example the internet user can see is the pop-up window which is opened automatically when some internet websites are being opened. This is a specially designed tool which can be used by the hackers to deliver advertisement in the front end and at the back end; the required operation can be performed as because for most of the times, the adware is bundled with the spyware which thieves the data from the internet cache without being noticed by the user [28-30]. There are some inbuilt anti-adware tools that is preinstalled with the browser but most of the times it fails to stop the pop-up as because that only identifies the adware and not the back-end spyware. When the advertisement is clicked, the systems captures the spyware or more specifically, the spyware penetrates through the system and makes eyes on the system cache. They are basically entering through some remote IP which cannot be properly traced even by using the trace route technique as they are invisible [30].

3.2. Bot: This is basically a software generated robot mostly used when the internet automated service like online gaming, online suggestion etc. are used. These are mostly used for good purpose but if the same kind of bots are developed by the hackers the user will be unable to recognize those. This kind of bots can harm the user computer by misdirecting them as it is controlled by the third party. It can also be happening if the user is installing a third-party software which has a feature of chatbot or simply bot [29,31].

3.3. Bug: This is a kind of software which is precompiled and makes the false or malfunctioning outputs in the system. It may be the result of the malware injection into the system. While the malware source code is injected, it tends to do some specific work which should not obey the user desire [32]. In most cases, the user can see the pre-installed antivirus system is in crash and no virus is being detected. This is the reason of the bug which allows the malware to enter into the system with bypassing the security and authentication by the system. While doing this, most of the system hangs because then the entry of the malware is made in the system [31,39].

3.4. Ransomware: It is one of the newly introduced malwares which is capable of capturing the entire system by locking it. As the system gets locked by the remote access, the user is completely unable to access it. Ransomware demand the money in exchange of unlocking the system and for its restoration [32,34]. This amount is known as Ransom and so the name if given. Ransomware takes the ransom in form of Bitcoin or other type of cryptocurrency which is not feasible to be identified by the network security tool. In general, the Ransomware is hard to be prevented from capturing of the system but the only way out is to prevent them from entering by securing the network and by enabling the filters. While the Ransomware captures the system, it still makes the theft of the database and the documents of the system from the connected server without the authorized notice [36].

3.5. Rootkit: All the above fall into the category that will work when they are present in the system. But there are still other ways through which the malware activities can be done and that is the remote activity. Rootkit is a specially designed malware which first be installed in the user system to establish the connection in between the host system and the remote hacker computer. With that connection, the hacker will be able to make a control over the user computer to execute program and make a perfect privilege over the system [34,36]. The detection and removal of such rootkit is difficult due to their stealthy operation and can only be enabled by the regularization of patching of vulnerability of the program of the system. Otherwise the remote hacker will be able to get data in a regularized manner. Rootkit mainly enables the hacker to make some unauthorized transaction from the user computer [35,36].

3.6. Spyware: This is another specially designed malware through by which the hackers are able to monitor the activity of the user. They make their entry through spoofing or by the adware [35]. After the entrance they make their eyes on the user activities like the login information, transaction details, account details, catching the keystrokes by using the keyloggers etc. They have the additional capability of changing the user setting to make the system run by their wish. So, it is one of the dangerous ones where the user system may not be affected with respect to the speed factor but their sensitive information can be stolen [37].

4. TYPE OF ATTACKS AND POSSIBLE MEASURES
Malware generally affects the system folders by taking the root access. The operating system like Windows is highly sensible to the malware as it can sense the executable file which is in extension of .exe [38]. Malware basically come in the form of .exe file which, after entering into the system, directly attacks the system folders which causes the severe vulnerability to the system. Not only the Computer operating system gets affected by the malware attack, even the malware can attack in the mobile system also. During the attack, there are different symptoms that can be seen while operating the system [37-39]. While opening the system, it gets slower and takes a long time to open an application and unwanted crashes of the system. This is because the attacks of the malware sometimes destroy or corrupt the system files and more specifically the Dynamic Link Library or .dll file and the system file or .sys file. They are two main backbone system extensions which are destroyed by the malware attack. Sometimes, the malware encrypts the target file and it cannot be opened as all the content is hidden and is controlled remotely which is the effect of the rootkit and spyware [37,39]. Nowadays, the malware comes in the application of Artificial Intelligence and so the source code cannot be decoded to understand what the type of the malware is and what will be the target object that it is

IJSTR©2019
www.ijstr.org
pointing to. So, there are different means to obstruct to make the system secure. The general use is Installation of Anti Malware, Activating Firewall, Activation of network monitor to watch traffic, upgrading the system patches and software along with the Antivirus or Antimalware [38-40]. To prevent these type of attacks, different types of algorithm works with the specific cryptography like RSA, AED, DES, Blowfish etc. Recently, with the implementation of Neural Network technology, it has been easier to detect the malware [31-33]. Malware is also dangerous for commuting the business. Most of the recent businesses are in the form of online transactions and so all the business exchanges are made online. While this transaction happens, the bank details and the confidential documents are also shared [32,37]. This may cause a huge risk to the organization. This is because while those transactions are made, the hackers may snatch the data and make a theft of that. The other possibility is that, they will watch the data that is being transferred and use those credentials to make transaction of their own [28,33]. As a result, the client and the business persons both will be misguided as the amount can be credit or debit card may be robbed without any notice. This is called the Credit or Debit card fraud and the customers come to know about it after the theft happens and the money and information is stolen. The hackers also apply the strategy in the false transaction of the customers for which the product will be purchased with the name of the customer and from those Credit or Debit card with the bank details [38]. There are different source of hijacking the user account such as the pointing towards the ATM, grabbing the access of the system device, catching the details by rootkit, hijacking the online shopping site, where the bank details is attached, spoofing the mobile application which is allowed by the user for the device access and so on [39]. One of the good and intelligent measure of that kind of problem is the data hiding and changing of the password of the network and the system device regularly. Even the network password is more sensitive that the device password as the hackers watch on this. Very recently, with the implementation of Cyber Fog, the data hiding techniques have got much improvement [36-39]. The architecture of the cyber fog is shown below:

In this kind of system, the data is truncated and stored in the cloud of devices with no specific serial id. The part data is stored in different devices with ciphered format so that, if the hacker will steal off one piece of data, firstly, it cannot be identified and secondly, it cannot be treated as the data block. So, totally, the hackers are misguided [38-40]. But this is the aspect of the data hiding in cloud computing or may be the physical cloud. When the data is being transferred, then this concept will not work and the only thing that is applied is the cryptography. Apart from different existing approaches, in this paper, the new cryptographic approach will be made to secure the data during storage as well as while the transaction [41].

5. PROPOSED ALGORITHM

In this paper, the anomaly-based malware detection will be proposed in a new way so that the system will have the ability to prohibit the malware attack by any means. As the malware is the malicious code, so, in this section, the proposal is made to make another anti malware algorithm to fight against the malware [42]. In the conventional programs, the malicious code differs from each other depending upon the types. So, detection of those different kind of malware require different algorithms as in the commercial sense, the anti-malware is made such that it can detect the specific category of malware along with other basic malwares [39-40]. If so, it will be hard to fight against different kind of malware as in each day, new algorithm for the malware generation is being compiled and so the development of the malware is made advanced every day. So, no single malware detector may be paid or in free version can dynamically detect the newly generated malware for the definition problem. So, the new concept will have to be introduced so that it will be dynamic enough to understand the definition that is declared inside the malicious code of that malware. This can be done well using Anomaly detection and in a new way with new implementation [28,36,39]. Anomaly detection is basically used in data mining to detect the malfunctioning data in a big and structured data section. This is exclusively used for detecting the rare items, some events and observations which raise the suspicious activity and tend to hamper the system badly [40]. Anomaly detection is very helpful to detect the network intrusion which is the objective of the malware.

**Fig-2: Cyber Fog**
In this technique, the data is observed at the network gateway. The data represents the packet that is being transacted. After observing the data, those will be recorded for future purpose so that the analysis can be done in future also. After this analysis, the features will be extracted. Here the feature is meant for the inheritable features of the suspected packets [41]. As the suspected packets are already been analyzed, the signature, ports of data transaction, their IP and MAC address etc. all will be taken as the features [43]. The most catchable feature through which the identification can be done is the hexadecimal stamp of the packet which is provided by the Wireshark. Those suspected packets will then be sent to the anomaly detector where the entire packet will be analyzed for detection and identification of the packet by comparing those with the non-suspected items and packets [44]. When the identification is confirmed for the malware, the prevention algorithm is drawn by applying the Rule-Based Detector which will apply rule for methods to be used for prevention of such malware. The level of the protection will be determined by detecting the sensitivity parameters and finally all the decisions will be combined together for the association that can predict the suspicious packet and to prevent them by raising the alert for the suspicious attacks.

The proposed algorithm is shown below

![Proposed Algorithm Diagram](image-url)
transmission protocols [39-41]. The filter is applied so that the suspicious packets within each protocol can be identified. That will make it easier for protection rather by checking out each packet separately. After that the table is generated which includes the details of the suspicious activities. This table includes the attributes like the packet or frame number, host IP, destination IP, host MAC address and the destination MAC address. These all are required for checking the IP vulnerability and helps the system to make a blockage of that particular IP. After the suspicious table is done, the packets are analysed and features are being extracted. In this case, the hexadecimal format of the packet data is extracted [44]. As the malware is the software code, it is basically coded in assembly language which can be detected by binary or hexadecimal format. Along with the extraction, the special characters that are included with the data also will be extracted to denote the density of vulnerability of that packet. Generally, by the rule of malware detection, the more special characters are included in a packet; the level of suspicion will raise more. Lastly, the signature of the packet is extracted to know the identity of the packet [43]. After the feature extraction is done, the protection will be executed. In the protection section, the ANN or the Artificial Neural Network is applied as discussed. ANN will detect the traceroute of the host to destination of the packet routing and thereby rejects the IP addresses which holds the malicious data [45]. The analysis of the algorithm is shown in the below section.

6. RESULTS AND ANALYSIS

The basic structure of the anomaly detection is shown in the above section. Based on the anomaly detection and the algorithm defined, the analysis is done. This structure has some significant block through which the malware definition can be identified dynamically. The blocks altogether determine the alert for the malware attack. So, first a network is taken as the reference and observe the packet routing through it. Each packet is traced and captured in the Network Packet Capture file so that it can be used as in future use and analysis. This packet capture file is then analyzed using the network monitoring tools. Here the Wireshark is used for analysis the packet tracer file. When the packet trace and analysis will be done and the malware activity is detected, the extracted report is analyzed using the malware analyzer. For analysis of Malware activity, the IDA Tool is used through which the report data is analyzed and the final report for the malware is prepared. After all the activities are done, the algorithm will be applied to protect the system from such reported threat and thereby secure the network from the malware attack and injection.

The analysis results are shown below step by step:

6.1. Observation of the Packet Data:
In the gateway section of the network several data transactions may occur and the data may come from different section of IP and MAC address. But all those IP are not free from suspect and may pirate the database. The data needs to be monitored so that the understanding can be made whether the data belonging to a particular IP is reliable or not. This can be done using different tools. One of the widely used tool is the Wireshark through which the network monitoring can be done. An example view of the network data monitor is shown below:

![Fig-5: Network capture by Wireshark](image)

This network capture includes two distinct section, one section shows the data activity from the network and another shows the Protocol hierarchy which has the synopsis of the entire section with the basic hierarchy of protocols. The protocol hierarchy of the captured data is shown below:

![Fig-6: Protocol Hierarchy](image)

Now in Fig-4, there are different data which are accessed through the network and some of those are retained with suspicious activity. Such suspicious activities are shown by the Black color. Such suspicious IP with their details is shown below:
So, they are the suspicious issues that are detected by the tool and identified as the malicious content. Though only the detection is made by obtaining the data from the entire section of the data transaction made through the network.

6.2. Feature Extraction:

Now the data that is analyzed over the network have some distinct features like the ASCII form of the transacted data, the character space determines the hexadecimal representation of data. The malware data is basically interpreted in hexadecimal format to make easier access to the system. The special character space determines the hexadecimal representation of distinct features like the ASCII form of the data. The malware data is basically interpreted in hexadecimal character space the vulnerability of the malware that is caught.

With the observation of such data, the following result is obtained which signifies the data containing in the packet is malicious.

The obtained data is shown below:

<table>
<thead>
<tr>
<th>Star frame</th>
<th>Mac address of host</th>
<th>Mac address of destination</th>
<th>IP address of host</th>
<th>IP address of destination</th>
<th>Host Port</th>
<th>Destination Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>93</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.01</td>
<td>33</td>
<td>08</td>
</tr>
<tr>
<td>126</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.101</td>
<td>192.168.97.4</td>
<td>22</td>
<td>33086</td>
</tr>
<tr>
<td>113</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.4</td>
<td>17</td>
<td>77734</td>
</tr>
<tr>
<td>114</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.4</td>
<td>17</td>
<td>77734</td>
</tr>
<tr>
<td>119</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.4</td>
<td>17</td>
<td>77734</td>
</tr>
<tr>
<td>120</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.4</td>
<td>17</td>
<td>77734</td>
</tr>
<tr>
<td>124</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.4</td>
<td>17</td>
<td>77734</td>
</tr>
<tr>
<td>125</td>
<td>02:60:ac:9a:bc:02</td>
<td>02:60:ac:9a:bc:03</td>
<td>192.168.97.42</td>
<td>192.168.97.4</td>
<td>17</td>
<td>77734</td>
</tr>
</tbody>
</table>

In this character section, the concentration of special characters is higher which is determined by Wireshark as a highly suspicious activity. The least suspicious activity or a clean file is determined by all dots.

The obtained data is shown below:

The ASCII space of the signature is as follows:
`<C@..../>..`...

After detecting the signature and the features, the frame information is being extracted to show the suspicious data representation. This will show the entire details of the data. Basically, the analysis is made for one frame with host IP 192.168.97.42 which has the destination of 192.168.97.4 with the backend protocol of TCP. The frame information is shown below:

Frame 11365: 54 bytes on wire (432 bits), 54 bytes captured (432 bits)

| Encapsulation type: Ethernet (1) |
| Arrival Time: Sep 30, 2015 21:43:32.661438000 India Standard Time |
| Epoch Time: 1443629612.661438000 seconds |
| Time since reference or first frame: 194.782965000 seconds |
| Time delta from previous displayed frame: 116.155788000 seconds |
| Time delta from previous captured frame: 0.000057000 seconds |
| Frame Number: 11365 |
| Frame Length: 54 bytes (432 bits) |
| Capture Length: 54 bytes (432 bits) |

The Spirent Test Signature is the important part of the analysis as it informs about the validation process of the frames. The validation is the process through which it can be determined that the packet is well validated by the network or not. If this is not validated, the packet obviously is suspicious. The Spirent test result is shown below:

Frame 11365: 54 bytes on wire (432 bits), 54 bytes captured (432 bits)
0100 .... = Version: 4
.... 0101 = Header Length: 20 bytes (5)
Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
   Total Length: 40
   Identification: 0x0000 (0)
   Flags: 0x4000, Don't fragment
   Time to live: 64
   Protocol: TCP (6)
   Header checksum: 0xf750 [validation disabled]
   [Header checksum status: Unverified]
   Source: 192.168.97.42
   Destination: 192.168.97.4
Transmission Control Protocol, Src Port: 17007, Dst Port: 57734, Seq: 1, Ack: 1, Len: 0
Spirent Test Center Signature
   Raw Data: 426fe18600000000afde527c5014000046000000
   IV: 0x42
   StreamID: 1757758186
   ChassisSlotPort: 26821
   0.... = StreamType: Soft
   Stream Index: 17130
   Sequence Number (SM): 229886231799485
   Timestamp: 251.166575329 seconds
   .... ..1. = Pseudo-Random Binary Sequence: True
   .... ...1 = TSLR: EndOfFrame
   Unknown: 46000000

After the successful execution of the feature extraction, the hierarchy of the malicious code for the same frame will be done using the IDA tools and it will test the main service routine of the malware. The protocol hierarchy is shown below:

![Fig-7: Protocol Hierarchy](image)

From the analytical figure, the main service section is derived at the second layer and above it is the offset address from where it starts the execution.

6.3. Malware Protection

The last section of the algorithm deals with malware protection section. After all the identification is done with the above section finally the protection algorithm is deployed. Basically, the protection algorithm uses the Artificial Neural Network. Sometimes it happens that the root host may not be infected but the file received by the root host is identified as the malware. This is because of routing of the packet. Once the data starts routing, it crosses different nodes or hops through it. While doing this, may be the one of those nodes are affected and the program of the malware code is such that the nodes becomes hidden. Then the source nodes are hard to be identified. Artificial Neural Network is the option and technology which can detect the visible as well as the hidden nodes also. The general diagram for Artificial Neural Network is as follows:

![Fig-8: Artificial Neural Network](image)

The input is actually the source node and the output are the destination node as detected by the Wireshark. The intermediate node may be visible or invisible so that the destination cannot catch the IP of the hopping node or nodes. If those are not properly identified, the protection mechanism cannot be drawn properly as the antivirus or the antimalware always first catch the IP from where the infected file is being transacted and after that it will make analysis over it and with the percentage of severity, the proper algorithm will be applied like here in this paper is Artificial Neural Network. To design and apply the model, some steps are to be followed. As the routing and capturing of the packet is the random fashion, so, the unsupervised learning seems to be efficient. But for training of the model, first it will be provided with sufficient data so that it can recognize the type of malware. So, for this instance, the learning will be supervised first and the Bayesian network will help in that purpose. If the Wireshark table is followed, the data can be obtained for each and every instance. Some of the packets are shown below:

<table>
<thead>
<tr>
<th>Pckt No</th>
<th>Time</th>
<th>Source IP</th>
<th>Destination IP</th>
<th>Protocol</th>
<th>Length</th>
<th>Info</th>
</tr>
</thead>
</table>

![Table 2: Packet Table](image)
If the comparison is made with the previous data, it can be seen that the Packet No 11365, Packet No 11922 & Packet No 11460 are infected. Now looking into the Info column, only for that instances, the Info part differs from other which includes the following attributes:

\[
P(c | x) = P(x_1 | c) \times P(x_2 | c) \times \ldots \times P(x_n | c) \times P(c)
\]

Where,

- \( P(c | x) \) = posterior probability of class (c, target) given predictor (x, attributes).
- \( P(c) \) = prior probability of class.
- \( P(x|c) \) = likelihood which is the probability of predictor given class.
- \( P(x) \) = prior probability of predictor.

So, using this technique, the probable word from the tabular data can be identified and the algorithm can get trained. After training is over with sufficient data, the testing part will be done. While testing, for the first phase, the data is provided by taking it from Wireshark directly or by tracing new packet with new experiment. After the internal check is complete, the model can be deployed. While deploying the model, the runtime data or the real time data is provided in tabular format so that by the analysis with the Artificial Neural Network, with help of Naïve Bayes technique, the data can be well identified as malware. When the packet is identified as infected by malware, the corresponding ports are rejected. Artificial Neural Network works in the backbone of the neural search. While searching, it seeks for the next node of the data as per the working procedure. While, the search is carried out by the Artificial Neural Network, the hidden nodes are discovered internally which is helpful to make the algorithm fulfilment. Those infected ports are being rejected by the system. Once the port is rejected that is when the port is blacklisted, the data or packet from those ports will remain blocked and no data transaction can be made through those ports. This is smoothest way to make the system secure by using the artificial intelligence. To make the system operation fruitful, the entire system is designed in coding section.

### 6.4. Coding and Analysis

For the design of the system of malware protection, Matlab is used for coding. In Matlab programming language, the library is very rich and the required framework in already in-built to make the execution successful. In the coding, the Artificial Neural Network is used as the library through which the analysis is made. The packages like Artificial Neural Network and the Naïve Bayes are the prebuilt packages in Matlab for most of the cases and so it will be easier to compute the outcome. The required libraries for the software execution are as follows: In this coding, the Neural Node will be used to detect the node array all over the network. ANN is the rich library under Matlab (prebuilt or by using Toolbox) through which the network mining is possible. Network mining is the technique through which the port identification can be done whether it may be public or hidden. First of all, the tabular data is taken as the input to the model as visualized as follows:
Next the data visualization is done with respect to the available protocol and their count of data packet to serve the protocol with higher use. The visualization is shown below:

**Fig-9: Protocol vs Count**

So, it can be seen that HTTP carries the highest traffic compared to others. Now it will be determined that among those protocols, which one carries the highest vulnerability. While doing this, the Gaussian Naïve Bayes will be in use to detect the text that is depicted in the Info section whether the data contains the string mentioned earlier.

**Fig-10: Count of Severity**

It can be seen that TCP Protocol is carrying the highest malware traffic compared to the others. It means the packet that are get routed through the TCP Protocol have faced the highest possibilities to be infected by the malware. So, the model should look into the particular protocol which carries the high probability of having malware infected packet. The other protocols also be observed but that protocol will be taken into consideration in the most. The probability of malware infection of each protocol is shown below:

**Fig-11: Protocol Probability of having malicious packets**

So, it is clear that the TCP protocol for this experiment is carrying the highest number of malicious traffic. So, when the model will analyses the network packet, the highest and the most sensitive filter is to be applied upon the TCP. Now the ANN with work like the interaction of each and every port can find and will check for the severity of having the high vulnerable code. The following function will work in this purpose:
Clear; close all; clc;
addpath('lib')
testData = load('data.mat');
nnOptions = {};
modelNN = learnNN(testData.X, testData.y, nnOptions);
plotConfMat(modelNN.confusion_valid);

%% Predicting on Malware
rI = randi(size(testData.X, 1));
p = predictNN(testData.X(rI,:), modelNN);

This interacting function will check each ad every IP for the vulnerability. Ipshellcheck() is the function to send each and every frame to this for the checking. After check for each IP address, ANN will work for the decision making whether the current port is significant to infection or not. If the current port is infected, the port is totally blocked by the model. The following piece of code is the class for the designing of Artificial Neural Network:
input_layer_size  = size(X, 2);
layers = [input_layer_size, hiddenLayers, nrOfLabels];
modelNN.activationFn = activationFn;
modelNN.lambda = lambda;
modelNN.maxIter = maxIter;
modelNN.layers = layers;
modelNN.doNormalize = doNormalize;

X(isnan(X))=0;
Y = Y - min(Y) + 1;

m = size(X, 1);
validation_set_size = round(m*validPercent/100); % e.g. 10% for the validation
rand_indices = randperm(m);
X = X(rand_indices, :);
Y = Y(rand_indices);
X_valid = X(1:validation_set_size, :);
Y_valid = Y(1:validation_set_size);
X(1:validation_set_size) = [];
Y(1:validation_set_size) = [];
modelNN.X = X;
modelNN.Y = Y;
modelNN.X_valid = X_valid;
modelNN.Y_valid = Y_valid;

In this code the ports continue to invoke the system that is the ANN class and get determined for the infection. The self.crt_par is the parameter for holding the value of the current node. This class is so designed that it can detect the hidden layer and the nodes in the network also. So, when the hidden layers are detected, they will be examined for the vulnerability. The function of the class middle_hidden_layer() performs the respective operation as described earlier. So, these pieces of code block together and reject the malicious ports and there by reduce the possibility of the malware attack to the system.

At the ultimate, the model will be trained so that it can automatically identify the network vulnerability. The training of the model is done using the following code portion.
function modelNN = learnNN(X, Y, varargin)
lambda = 500;
nrOfLabels = numel(unique(Y));
hiddenLayers = round(sqrt(nrOfLabels*size(X, 2)));
activationFn = 'sigm';
validPercent = 20;
doNormalize = 1;
nnMu = 0;
nnSigma = 1;
if nargin>2
    nnOptions = varargin(2);
    if ~mod(numel(nnOptions), 2)
        for ii=1:2:numel(nnOptions)
            if isnumeric(nnOptions(ii+1))
                nOfn = numel(nnOptions(ii+1));
                if nOfn>1
                    eval(sprintf('%s=repeat(%d, nOfn, 1) ;',
                                nnOptions(ii), nnOptions(ii+1)));
                else
eval(sprintf('%s=%d ;',
                                nnOptions(ii),
nOptions(ii+1)));
            else
                eval(sprintf('%s=%s ;',
                                nnOptions(ii),
nOptions(ii+1)));
            end
        end
    else
        error('Number of Options should be Even.')
    end
end
initial_nn_params = randInitializeWeights(layers);
options = optimset('MaxIter', maxIter);
cf = @(p) nnCostFunction(p, layers, X_norm, Y, lambda,
activationFn);
startT = cputime;
[theta, modelNN.costArray] = fmincg(cf, initial_nn_params, options);
elapsed = cputime - startT;
fprintf('Required CPU Time: %f
', elapsed);
modelNN.Theta = unpackTheta( theta, layers);
if doNormalize
    X_valid_norm = (X_valid - repmat(nnMu,size(X_valid,1),1))... /repmat(nnSigma,size(X_valid,1),1);
else
    X_valid_norm = X_valid;
end
X_valid_norm(isnan(X_valid_norm))=0;

for la=1:nrOfLabels
    confusion_valid(la, lp) = numel(find(p_valid==lp&Y_valid==la));
end
confusion_train(la, lp) = numel(find(p_train==lp&Y==la));
end
end
modelNN.confusion_valid = confusion_valid;
modelNN.confusion_train = confusion_train;
modelNN.nnMu = nnMu;
modelNN.nnSigma = nnSigma;
modelNN.trainingTimestamp = datestr(now,'yy_mm_dd_HH_MM_SS');

if exist('savePath', 'var')
    % Saving the computed parameters
    save(savePath, 'modelNN');
end
end

Next to that, the predictive mode is shown which predicts whether a port is being transacted by malicious element. The predictive model is shown below:

if nargin==2
    modelNN = varargin{1};
    Theta = modelNN.Theta;
    activationFn = modelNN.activationFn;
    X = (X - repmat(modelNN.nnMu,size(X,1),1))...
        ./repmat(modelNN.nnSigma,size(X,1),1);
    X(isnan(X)) = 0;
else
    if nargin==3
        Theta = varargin{1};
        activationFn = varargin{2};
    else
        error('predictNN: Invalid Number of Arguments');
    end
end

if strcmp(activationFn, 'tanh')
    aF = @tanh;
else
    aF = @sigmoid;
end
m = size(X, 1);
end
h = aF([ones(m, 1) h] * Theta(ii));
end

[-, p] = max(h, [], 2);
end

So, using this model and the matlab code with the application of the Artificial Intelligence, the probability of malware attacks can be reduced as the model will check the network in double fold, first by checking the features in it and second by checking the vulnerable ports from where the data is being routed.

7. CONCLUSION

From the very beginning of the model, the issue was the malware attacks and the probable protection for it. First the network is traced. The network tracing means that the activities will be observed and hence recorded. Meanwhile the system will store the result in a tabular file because as per the plan, the protection model is done using Matlab and so it will be easier to read the excel file that is the structured rather than by using log file or unstructured data. Now, those into the record, the analysis is done. While doing the analysis, first the packet captured file is send under analysis. There the feature analysis, protocol analysis and all will be done and this is the kind of Feature engineering. This is necessary because the features actually say the attribute of the capture file as well as the packet data. It is observed that the packet is transferred from one host and goes to another destination with a via media of the current network. While analysis and the feature extraction are done, it is observed that some of the packets are highly vulnerable to the system as it includes the malicious code. Generally, they can be detected using network monitoring tool and is signifies for the identical issue that may harm the present system. This malicious code can be identified by the information or Info section of the analysis table in Wireshark. The packets with the particular attributes will be treated as the malware. By the ANN application, they can be blocked for good. But in this approach, not only the infected file will be blocked, but also the IP from where the packet is coming, also will go to the blacklist. So, it is one of the efficient methods through which the node can be blocked until it produces a good packet. But there may be the problem where some of the nodes may be hidden by the hackers and cannot be easily identified unless the port is read. The use of ANN removes the barrier and it enables the model to search for the hidden node. The hidden node may be the root host node or may be the intermediate node. This will be found out the ANN model. After the searching for the malicious node is over and the node is identified, the node will be immediately blocked and will be put under observation. If the node is releasing any white packet, the node will be unblocked then and again if it will produce a black packet of data, again the node will be blocked. This is the technology that is used in this paper. The data is analyzed and produce the analytical result of the packet which contains the malicious code and others. The visualization is shown to produce the outcome of the analysis and the model deployment. The entire algorithm with the working procedure is discussed in the respective section and that will be the measure for the issue created by the malicious code into the system.

REFERENCES


