Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

# A Survey of Data Exfiltration Prevention Techniques

**Peter S. Nyakomitta**
Faculty of Biological and Physical Sciences, Tom Mboya University College, Homa Bay, Kenya.
**Dr. Silvance O. Abeka**
School of Informatics and Innovative Systems (SIIS), Jaramogi Oginga Odinga University of Science and Technology, Bondo - Kenya.

**ABSTRACT**

**Data exfiltration is a serious cybercrime facing many organizations worldwide. Over the past few years, notable organizations such as the Google, Yahoo, the Pentagon, Iran nuclear facility and the United States military contractors and banks have fallen victims of data exfiltration. The current techniques for averting these threats revolve around firewalls, intrusion detection systems, intrusion prevention techniques, firewalls, anti-virus an anti-malware. However, despite heavy deployment of these devices, attackers still continue to wreck havoc on organizations and individuals, stealing their sensitive data. The aim of this paper was therefore to explore how the current techniques for data loss prevention fail. The results of this analysis revealed that these techniques either use whitelists, blacklists, signature-based scanning, behavioral analysis of programs which are not sufficient to counter attacks based on zero day vulnerabilities. Based on these shortcomings, a novel data exfiltration prevention algorithm is proposed towards the end of this paper. This algorithm is suggested to employ real-time traffic entropy coupled with heuristically computed functional correlations to detect data exfiltrations. The premises of this algorithm and its operations are discussed at the last section of this paper.**

Keywords: **Algorithm, anti-virus, anti-malware, Data exfiltration, IDS, IPS**

Date of Submission: Dec 10, 2020 Date of Acceptance: Dec 23, 2020

## I. INTRODUCTION

Data exfiltration is a form of illegal leakage of sensitive data from a particular organizational or individual system. According to Murtaza and Naveed (2016), this caliber of intrusion is hard to catch due to its careful planning and execution, often involving insider entities to leak facilitating information such as usernames and passwords. This is normally achieved through social engineering via email attachments or links which once opened or clicked, either directs the naïve user to a malicious website which installs the malware on the system, or directs runs and executes malicious activities. In most cases, the insider entity could be a person working in the target organization or a malicious hardware component bought from an unreliable third party.

A typical scenario is where an employee in an organization plugs in a universal serial bus (USB) stick infected with a malware to a machine connected to the company intranet. This malware rides on the autorun feature and executes itself as a background process (Neeshu and Shitanshu, 2016). Afterwards, this malware gains root privileges through the exploitation of zero-day vulnerabilities inherent in software already installed on the target machine such as Acrobat Reader and Internet explorer.

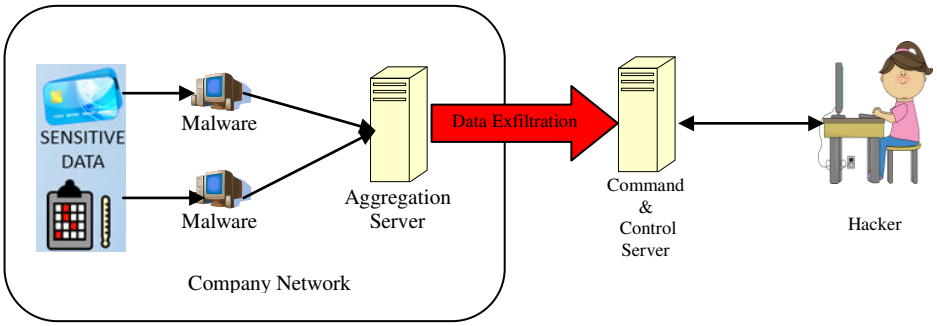


**Figure 1: Typical Data Exfiltration**

Figure 1 that follows gives a depiction of the data exfiltration process.

As this figure shows, the company network contains sensitive data such as credit card numbers which needs to be protected from the hackers. However, an hacker can employ social engineering tactics and send the malware in form of mail attachment which upon opening, installs itself on the victim network devices. The malware then establishes a communication with the hacker controlled command and control server for additional instructions. The hacker then locates the sensitive data within the

Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

network, aggregates it together and compresses it and finally transfers it outside the organization.

Due to its stealth nature, this malware cannot be detected by the anti-virus, anti-malware or host-based intrusion detection systems since it hides behind a legitimate process through code injection. A number of these malware have modules for communicating with their designers, who can then use the obtained root privileges to install extra payload on the target machine to enable them scan the victim's machine, search and discover the exact information being sought and relay it back to attacker server before destroying itself, hence doing ways with all traces of any attack.

This makes forensic analysis of the attack a complicated affair since these malware have not only the ability to change log files but also hiding themselves behind legitimate programs. Another scenario would be for the attackers to exfiltrate data using insider hardware such as network interface cards (NICs) bought from unreliable vendors (Hanu and Dharani, 2015). The vendors might be working with attackers and hence could implant malicious firmware on the NICs such that all packets travelling from the target organization are duplicated and sent to a remote machine controlled by the attacker.

## II. CHALLENGES OF CURRENT DATA EXFILTRATION PREVENTION TECHNIQUES

Due to their ability to completely bring down organizations by disclosing sensitive and classified information, data exfiltration has received much attention from both the academic field and the market field. A number of techniques and devices have been developed to try and at least detect early phases of data exfiltration. According to Barbara (2013), firewalls employ policy based approach to detect and prevent attacks and can therefore be utilized to deter data exfiltration at the reconnaissance phase during which port scans and internet protocol IP) sweeps are carried out to identify potential targets and vulnerabilities. Figure 2 demonstrates how a firewall works.

As demonstrated above, data exfiltration attacks are in most cases very difficult to detect due to their multifaceted approach, using people or hardware within organizations, malware that hides and destroys themselves after a given period of time, their ability to masquerade as legitimate programs to prevent detection, and the malware ability to communicate with attackers to receive additional payloads or instructions. In the following section, a number of techniques that are currently being employed to curb these attacks are presented.

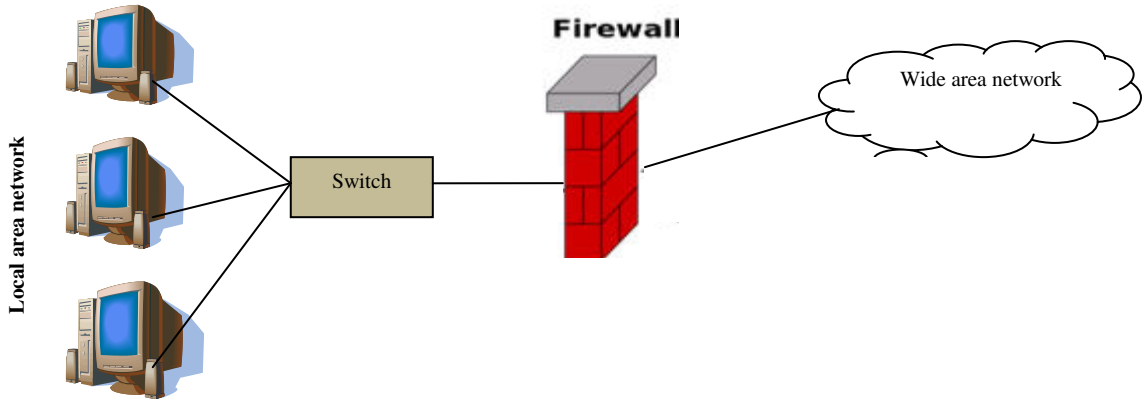


**Figure 2: Firewall Operations**

This figure illustrates that the firewall sits between enterprise local area network and the public wide area network such as the internet. It filters all the traffic coming into the local area network as well as that leaving the local area network. However, as Sara et al., (2016) points out, even with the presence of firewalls, the malware can encrypt and route messages and information through allowed routes such as the hyper-text transfer protocol's port 80 to the command and control servers. This is achieved through tunneling in which malicious payload is hidden in legitimate traffic.

The second device that can be utilized to curb data exfiltration is an anti-virus.

Randy (2015) explains that although these anti-viruses may or may not be able to contain malicious code, they can prevent a number of Trojans and other infections from spreading to other network devices by examining the files, folders, mail messages and web pages in the computers. When the scanner detects something that looks like a virus, it quarantines the suspect object and warns you about what it found. Figure 3 provides an illustration on the operation of an anti-virus.

Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

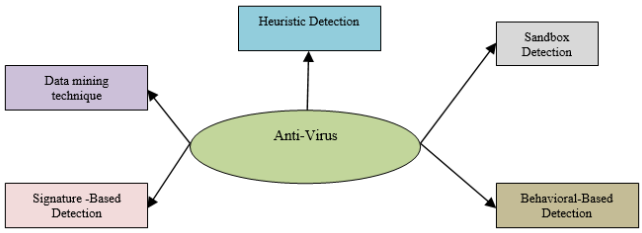


**Figure 3: Operation of Anti-virus**

In this figure, the anti-virus is shown to utilize five detection mechanisms for malicious software, namely the data mining, signature-based, heuristsic, behavioural and sandboxing techniques. Data mining techniques as employed in intrusion detection fall into one of two categories which are misuse detection and anomaly detection (Nadiammai and Hemalatha, 2013). In misuse detection, each instance in a data set is labeled as normal or intrusion and a learning algorithm is trained over the labeled data. These techniques have the ability of automatically retraining intrusion detection models on different input data that include new types of attacks, as long as they have been labeled appropriately. The strength of misuse detection is that it has a high degree of accuracy in identifying known attacks and their variations. Their main challenge is their inability to detect attacks whose instances have not yet been observed. On the other hand, anomaly detection builds models of normal behavior, and automatically detects any deviation from it, flagging the latter as suspect. The drawback of anomaly detection is its high rate of false alarms due to previously unseen yet legitimate system behaviors that may also be recognized as anomalies, and hence flagged as potential intrusions.

An antivirus using heuristic detection seeks to detect malware by examining the code in a malicious program and analyzing its structure. It might run a process that simulates the actual running of the code it is examining. This is meant to identify additional code logic that may help it determine if the suspected virus is really a threat.

A behavioral detection antivirus program works by looking for odd behavior in a program. It monitors the operating system, searching for suspicious events. As an illustration, if the antivirus program notices an attempt to change or modify a file or communicate over the web, it may take action and warn the user about the threat. It may also block the threat depending on its security settings. Sandboxing is normally employed to separate running programs, and executing untested or un-trusted programs or codes from unverified third parties, suppliers, users or websites, so as to mitigate any risk the malicious program may pose to the host machine or operating system.

On the flip side, Nikolaos (2015) point out that malware can hide in locations not scanned by anti-virus software. Consequently, during scans, the anti-virus may give erroneous reports regarding malware activities. In addition, some malware have the ability to stop their activities for some time so as to prevent their detection.
On their part, Intrusion Detection system (IDS), network intrusion detection systems (NIDS) and intrusion Prevention Systems (IPS) either use signature-based or anomaly based approaches. For the former case, all malicious pieces of software signatures are contained in an IDS or IPS database and the detection process involves comparing the signature of a given piece of software against the database signatures (Hanu and Dharani, 2015). As such, malicious software is identified when its signature matches one or more signatures in the database. Figure 4 is an illustration of the operation of an IDS.

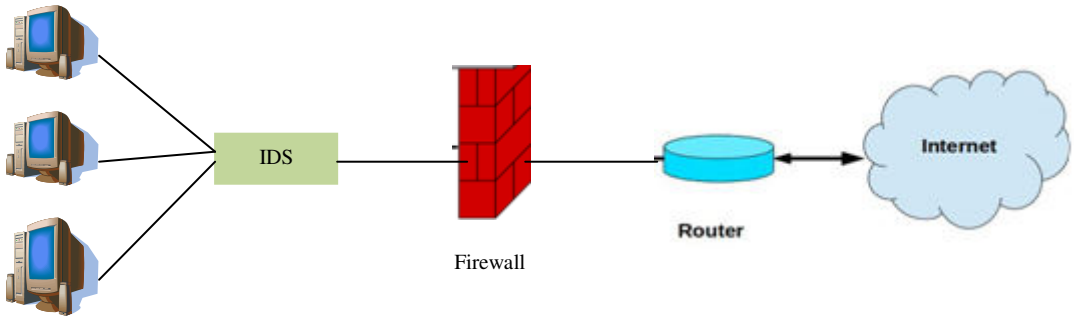


**Figure 4: Intrusion Detection System**

This figure points to the fact that an IDS sits between the intranet and the organizational firewall. It scans and detects any malicious activities within the network. However, as Barbara (2013) discusses, this approach is similar to the blacklist approach and can therefore not detect attacks for which a signature has not yet been created, such as zero-day exploits. Figure 5 provides a pictorial view of an IPS.
Just as was the case with IDS, the IPS sits next to the local area network to detect any anomalous activities. In addition to active detection, an IPS can actually carry out some preventive measures and therefore requires some

Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

form of manager to dictate the actions to be taken once malicious activities have

Just as was the case with IDS, the IPS sits next to the local area network to detect any anomalous activities. In addition to active detection, an IPS can actually carry out some preventive measure and therefore requires some form of manager to dictate the actions to be taken once malicious activities have been flagged down.

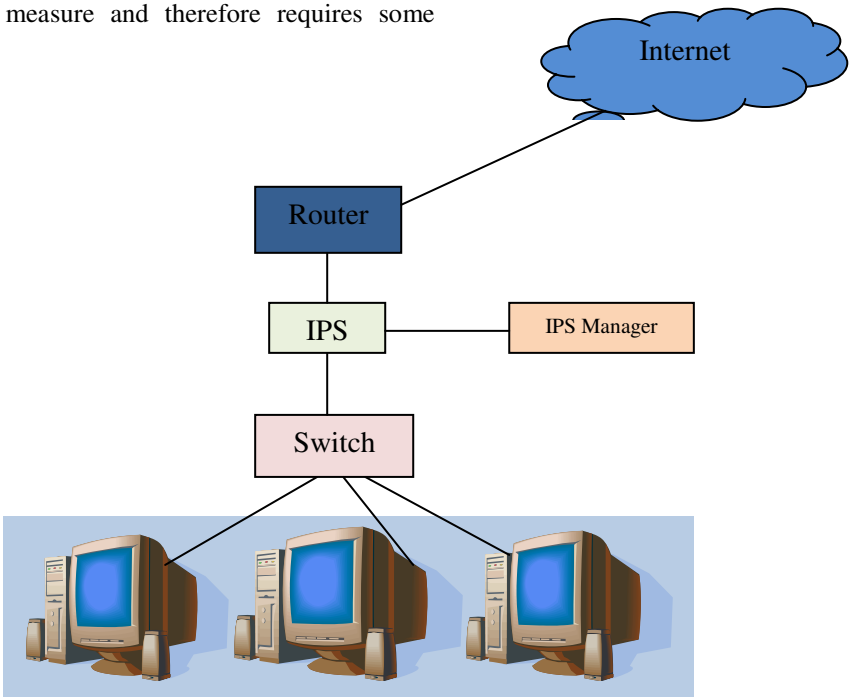


**Figure 5: Intrusion Prevention System**

The anomaly based approach relies on behavior analysis using machine learning to detect malicious piece of code in computer or network systems. Here, the behavior is classified as being acceptable or unacceptable and requires that a number of data sets be employed to train the learning agent o what constitute normal and abnormal behavious. As such, is a program's behavior deviates from what is considered normal behavior, it is flagged as being malicious.

However, machine-learning techniques are ideal in finding events similar to ones observed previously (Kateryna, 2017). Consequently, IDPS and IPS approaches demonstrate promising detection efficiency for specific training data sets, but are faced with grave operational limitations when used in operational environments. In addition, behavioral analysis may fail to detect known attacks that could easily be detected with signature-based IDS if these attacks do not differ significantly from what the system establishes to be normal behavior. Figure 6 provides an overview of the operation of the machine learning malware detection process.

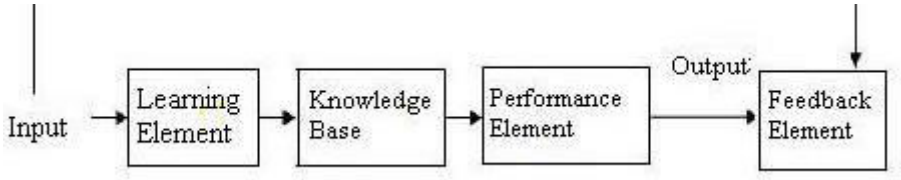


**Figure 6: Machine Learning For Malware Detection**

From this figure, it is clear that machine learning requires an input, which is the malicious code or program to be analyzed. It also has a database of all normal or abnormal behaviours and a learning element to detect the malware behavior. The performance element is crucial for providing feedback on the detection ability of the learning element.

Another technique that has gained popularity in as far as data loss exfiltration is concerned is the usage of anti-malware that make use of signature based scanning while a few of them use anomaly based approach. On the flip side, this anti-malware faces the same challenges as IDS and IPS. Ruchika (2013) points out that a grave concern is that attackers may test a great number of antivirus or anti-

Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

malware products and adapt their malware accordingly to escape detection.

It is clear that the current data exfiltration prevention devices face serious challenges in detecting the initial stages of the data exfiltration processes which normally involve the usage of malware to identify vulnerabilities that can be exploited to leak data from the system. If these initial stages of malware infiltration are not detected and mitigated, the then malware reports back to the attacker who can advance to gaining administrative privileges and therefore access to sensitive data. Once this is accomplished, the data is amalgamated, compresses, encrypted and tunneled through HTTP or HTTPS in the full presence of the firewall.

## III. STATE OF THE ART DATA EXFILTRATION MITIGATION APPROACHES

In the wake of numerous data exfiltration attacks such as those ones involving Google, Yahoo and the U.S military, a number of researchers have come up with other techniques, methods and procedures that they believe could help mitigate data exfiltrations. One of these techniques is full packet capture which can intercept network traffic and archive it for analysis. This is ideal for after-attack forensic analysis and offer opportunity for in-depth analysis. However, according to Puneet (2013), this approach is very expensive, requires usage of external tools for low level traffic inspection, are fairly incompatible with other systems such as IDS and IPS, and provide massive data that presents storage constraints especially for high speed networks.

Another recent development in the security sector is the use of security incident and event management systems that work by collecting network events from sources such as IDS, NIDS, antivirus and event logs from firewalls. Afterwards, statistical correlation is applied on the collected events to identify probable threats. However, as Gustav (2016) explain, this technique has limited efficiency in the detection of sophisticated attacks, and has limited time window during which the event correlations and hence incidents spread over a larger time period will never be correlated. As such, a carefully executed attack in form of a series of seemingly unrelated episodes can never be detected. In addition, the event correlation is carried out centrally and hence limited by the availability of resources.

Due to the complex and stealth nature of data exfiltration attacks, researchers are also exploring on mechanism that can help them the intruder attack vectors. Deception techniques such as honeypots, honeynets, honeywords, honeyusers and honeyfiles have been developed for this purpose. Honeypots and honeynets are deployed in internet facing and intranets to detect botnets and other attacks over wireless networks. On their part, honeyfiles have been employed to detect unauthorizes access to resources while honeywords and honeyusers have been utilized to track down compromised credentials. On the other hand, as Nikolaos (2015) explains, all these deception techniques fail to take into consideration the complicated maneuvers employed by experienced and skilled attackers. For instance, honeyfiles consist of codes that execute when the document is opened and conveys the report to the monitoring unit. An experienced attacker will decide not to open the file but exfiltrate and open it from an offline machine where the monitoring unit cannot be accessed.

## IV. PROPOSED ALGORITHM FOR DATA EXFILTRATION PREVENTION

A review of the current data exfiltration mechanisms have revealed that all of them fall short of the expectations in one way or the other. As such, this paper proposes a new data exfiltration algorithm for effective and efficient data loss prevention. This algorithm is based on information entropy, heuristics and functional correlations to detect and perform some preventive data exfiltration measures.

Information entropy will be employed to segregate plaintext and encrypted traffic. The plaintext information utilizes the twenty six letters of the alphabet and hence will have lower entropy compared with encrypted information which can make use of the alphabet, numeric and special characters. Consequently, comparing the same set of data, high entropy will be observed if these data items are encrypted and lower entropy will be recorded when this data items are just in plaintext.

Heuristics scanning will be employed to observe the behavior of the traffic that will be passing across the network. This behavior will be of two types: normal behavior and anomalous behavior. In case any anomalous behavior is observed, the algorithm will be designed in such a way that it will first alert and then take some preventive actions such as terminating the connection between the hacker controlled command and control centers and the victim's computer systems. On the other hand, the traffic with expected behavior will be allowed to traverse the network.

The last component of the proposed algorithm is the functional correlations of the network traffic, based on four decision trees: expected encrypted received encrypted; expected encrypted received plaintext; expected plaintext received plaintext; and expected plaintext received encrypted as shown in Figure 7. This algorithm will intercept traffic that is being passed across the network, and determine the port that is being used for communication purpose. This will inform the action to be taken based on the expected and the actual format of traffic received.

As this figure shows, the proposed algorithm will intercept traffic, and check the port in which the data packets are being sent over. Since data exfiltration is likely to happen over hypertext transfer protocol (HTTP) or secure hypertext transfer protocol (HTTPS), the format of the data being transferred over these ports will be of great interest. In normal circumstances, the information being transferred over HTTP is in plaintext and that transferred over HTTPS is in cipher-text.

Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

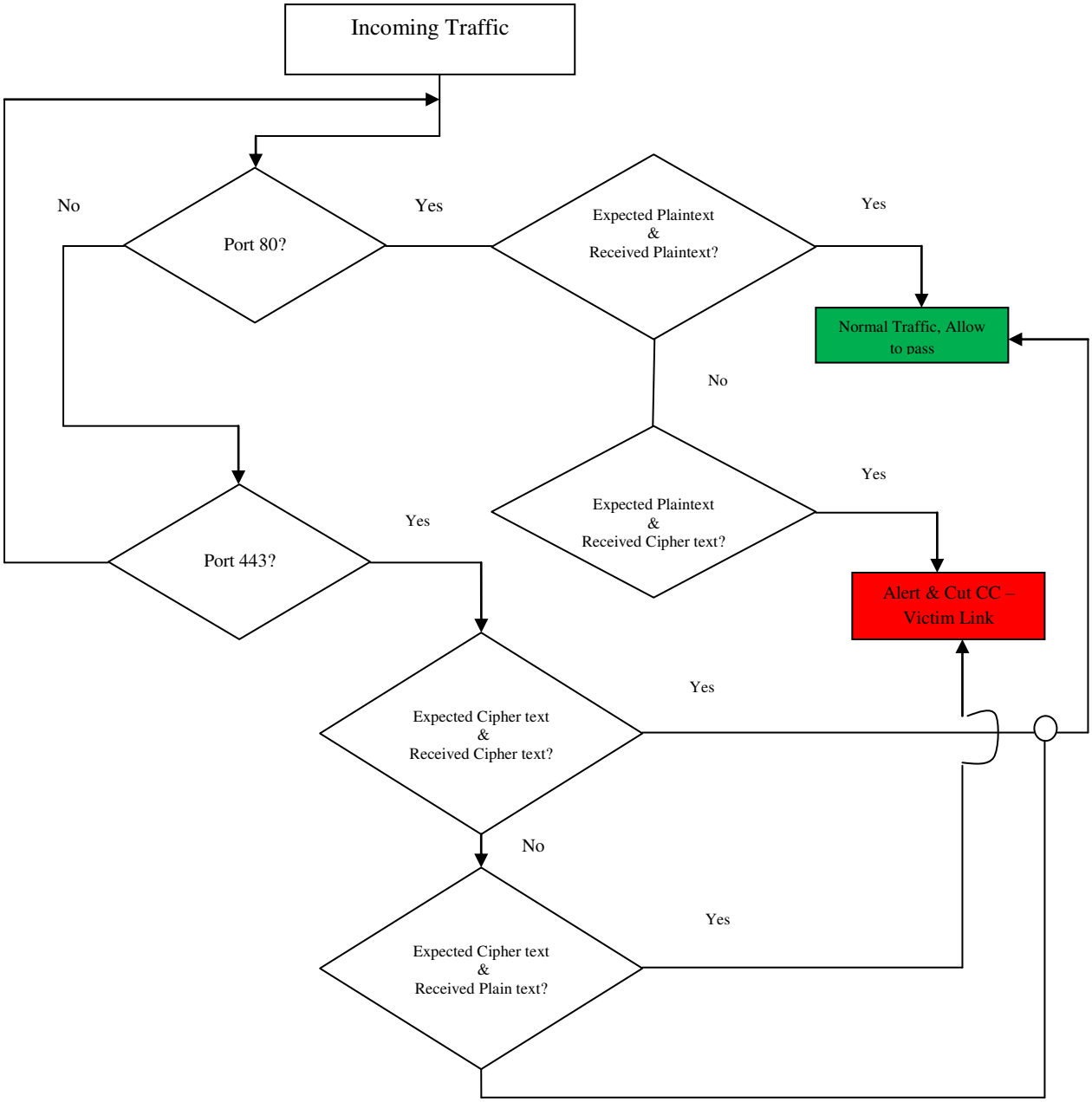


**Figure 7: Proposed Data Exfiltration Prevention Algorithm Data Flow Diagram**

Therefore, the algorithm will check to determine if this rule is being adhered to. In case of any deviation from this rule, an alert will be triggered and preventive measures such as the disconnection of the victim machine from the hacker controlled command and control (CC) will be carried out.

In so doing, the proposed algorithm will incorporate the security rules in its operation, akin to the firewall. It will also detect anomalies, just like an intrusion detection system and carry out some preventive measures in the same way that an intrusion prevention system does. As such, on top of the inclusion of the current malware detection mechanisms, the developed algorithm will incorporate some machine learning and functional correlations to supplement the signature-based approaches employed in the current IDS, IPS and anti-virus.

## V. CONCLUSIONS AND RECOMMENDATIONS

In this paper, a survey of the current and state of the art data exfiltration prevention methods and techniques has been carried out. From this analysis, it has been noted that all of them fall short of expectations in one way or the

Volume: 12 Issue: 03 Pages: 4585-4591(2020) ISSN: 0975-0290

other, making it possible for attackers to carry out unauthorized data transfers from their victim machines in the full presence of IDS, firewalls, IPS and anti-virus programs. Towards the end of this paper, an algorithm that could potentially address some of the challenges noted in the current data exfiltration prevention methods has been provided. One of the pillars of this algorithm is information entropy computations that would help discriminate cipher text from plaintext. Secondly, the algorithm carries out some machine learning to segregate normal traffic behavior and anomalous traffic behavior. The last component of this algorithm is the functional correlation that will be bound to the communication ports, where four branches of the decision tree will emanate from and help in the termination of malicious activities. Owing to its strengths, this algorithm is therefore recommended for implementation in organizations that handle crucial and critical data for their customers and in military data centers.

**REFERENCES**

[1]. Murtaza A., & Naveed G. (2016). Critical Analysis on Advanced Persistent Threats. International Journal of Computer Applications. Volume 14, Issue No.13 (pp. 46-50).

[2]. Neeshu S., Shitanshu J. (2016). A Study and Review on Advanced Persistent Threats. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 4, Issue 11. (pp. 19925-19932).

[3]. Hanu P., & Dharani J. (2015). Advanced Persistent Threat Detection System. International Journal of Science and Research. Volume 4, Issue 4. (pp. 1990- 1993).

[4]. Barbara H. (2013), Advanced Persistent Threats: Detection, Protection and Prevention. SOPHOS. (pp. 1-10).

[5]. Sara M., Arunesh S., Milind T.,and Pratyusa M. (2016).Data Exfiltration Detection and Prevention: Virtually Distributed POMDPs for Practically Safer Networks. Springer International Publishing.

[6]. Randy D. (2015). Data Loss Prevention. The SANS Institute. (pp. 1-30).

[7]. Nadiammai G., and Hemalatha M. (2013).Effective approach toward Intrusion Detection System using data mining techniques. Egyptian Informatics Journal. Vol. 13. (pp. 37-50).

[8]. Nikolaos V. (2015). Detecting Advanced Persistent Threats through Deception Techniques. Phd Thesis, Information Security and Critical Infrastructure Protection (INFOSEC) Laboratory. Department of Informatics. (pp. 1-174).

[9]. Frank K. (2015). Detection of APT Malware through External and Internal Network Traffic Correlation. Masters Thesis, University of Twente. (pp. 1-82).

[10].Kateryna C. (2017). Machine Learning Methods For Malware Detection And Classification. University of Applied Sciences. (pp.1-93).

[11].Ruchika M. (2013). Schemes for Surviving Advanced Persistent Threats. PhD Thesis, Department of Computer Science and Engineering, University at Buffalo, State University of New York. (pp. 1-158).

[12].Gustav L. (2016). Bypassing modern sandbox Technologies. Masters Thesis, Department of Electrical and Information Technology, Faculty of Engineering, LTH, Lund University. (pp. 1-94)

[13].Puneet S. (2013). A Multilayer Framework To Catch Data Exfiltration. Masters Thesis, Department of Computer Science and Electrical Engineering. (pp. 1-18).