Detecting data exfiltration on computer networks: a machine learning approach

[1*]Samanwickrama WD, [2]Wickramasinghe IP and [1]Perera LDRD

[1]Faculty of Applied Sciences, Wayamba University of Sri Lanka, Sri Lanka
[2]Department of Mathematics, Prairie View A&M University, USA

*Corresponding author: dumindu@wyb.ac.lk

Data exfiltration is the unauthorized transfer of sensitive information from a target computer to a location where a threat actor controls. Due to the regular exchange of data back-and-forth from networked enterprises, differentiating data exfiltration from normal network traffic has become a daunting task. Existing traditional security controls are increasingly ineffective in detecting such attempts. Machine learning (ML) can be effectively utilized to bridge the gap. However, choosing a suitable ML algorithm for a given problem depends on various factors such as characteristics of data available and context of the problem. This work employs ML, in particular Local Outlier Factor algorithm, to monitor data exfiltration on computer networks. Experimental setup of research approach and very early stage results were provided. In this experiment, a network testbed was setup with a private VLAN which consists of 10 PCs installed with general purpose applications such as MS Office, and they has access to the Internet via a proxy server. Typical perimeter defenses are set as usual. No inbound connection is allowed for the internal network, and only HTTP (port 80) and HTTPS (port 443) traffic are allowed as outbound traffic. During the experiment, depending on their choice, benign users use internal PCs to mimic normal user activities such as word processing, Internet and Emails. Attacker resides externally (Internet) and targets a victim on the private network. Using a remote administration toolkit (RAT)1, attacker stealthy captured a screenshot of victim's desktop and stole a file from her hard disk. All data transferred via encrypted web traffic (HTTPS) to bypass the firewall and intrusion detection systems. During a 2.5 hours monitoring period, using tshark (terminal oriented version of Wireshark), about one million of packets captured from victim's PC to perform the analysis describing in this work. Experimental results are encouraging. Proposed method successfully isolates exfiltrated data flows (RAT traffic) from the rest of legitimate web traffic even traffic is encrypted and uses the same channel (HTTPS). However, in order to generalize these findings, an extensive validation is needed and left as the future work.

Keywords: Data exfiltration, Machine learning, Network security