Project No.	**Project Title**	
2022-01-124	Detection of malicious LSP files from the CAD domain using ML methods	
Academic Advisor		**Co-Advisor**
Dr. Nir Nissim		Trivikram Muralidharan, Tomer Panker
Team Members		
Alex Yevsikov		
yevsikov@post.bgu.ac.il		

Abstract

Computer Aided Design (CAD) files are used for creating digital designs for various engineering structures – from the smallest chips in the high-tech industry, to large scale building and bridges in the civil engineering space. CAD is a growing market with many competing companies and varying software packages for creating and editing CAD files. Overall there are over 100 different CAD file types. File sharing is an integral part of CAD, this together with the high value of the intellectual property being traded, makes the CAD design files appealing for cyber-criminals. Considering the high number of documented vulnerabilities in CAD files, we have found that most of the exploits and malicious payloads are deployed either through LSP or FAS files. These files hold script in the AutoLisp language that is native to AutoCAD. Nowadays, organizations block the entrance of executable files either via email or direct download so attackers utilize non-executable files (NEF) to launch their attacks**.** While solution for detecting many malicious NEFs were proposed (PDFs, DOCX, JPEGs etc.), the CAD frontier remained exposed and can be exploited for cyber-attacks. Antivirus software are widely used detection mechanisms but are only effective against known malicious CAD files or their relatively similar variants. In this study we first provide an extensive background on the CAD file ecosystem, identify the gaps in defending against CAD exploits and propose a new static analysis-based detection mechanism for malicious LSP files from the CAD domain that is based on machine learning algorithms. Particularly, we propose two novel feature extraction methodologies (Keywords and Structural) for LSP files by which we extract a discriminative set of informative features. These features encompass common functions and commands that are referenced in the files and associated with the file's potential functionality. We have collected 1,034 malicious LSP files and 2,427 benign ones for our experiments. We have compared the detection capabilities of our detectors that were based on six different ML algorithms trained on data set that was based on our two proposed feature extraction methodologies and took into consideration 4 different feature representations. We have also compared our proposed detectors which outperformed state of the art methods such as MinHash and deep learning based CNNs. Among our detectors, there were several configurations that surpassed an AUC score of 0.999. The best performance in the task of malicious LSP file detection was obtained by an ANN trained over 100 keyword features in binary representation with a TPR of 0.994 and an FPR of 0.005.

Keywords: AutoLisp, Computer Aided Design, Malware Detection, Machine Learning, Feature Extraction