SFEM: Structural Feature Extraction Methodology for the Detection of Malicious Office Documents Using Machine Learning Methods

Aviad Cohen, Nir Nissim, Lior Rokack, Yuval Elovici

Motivation

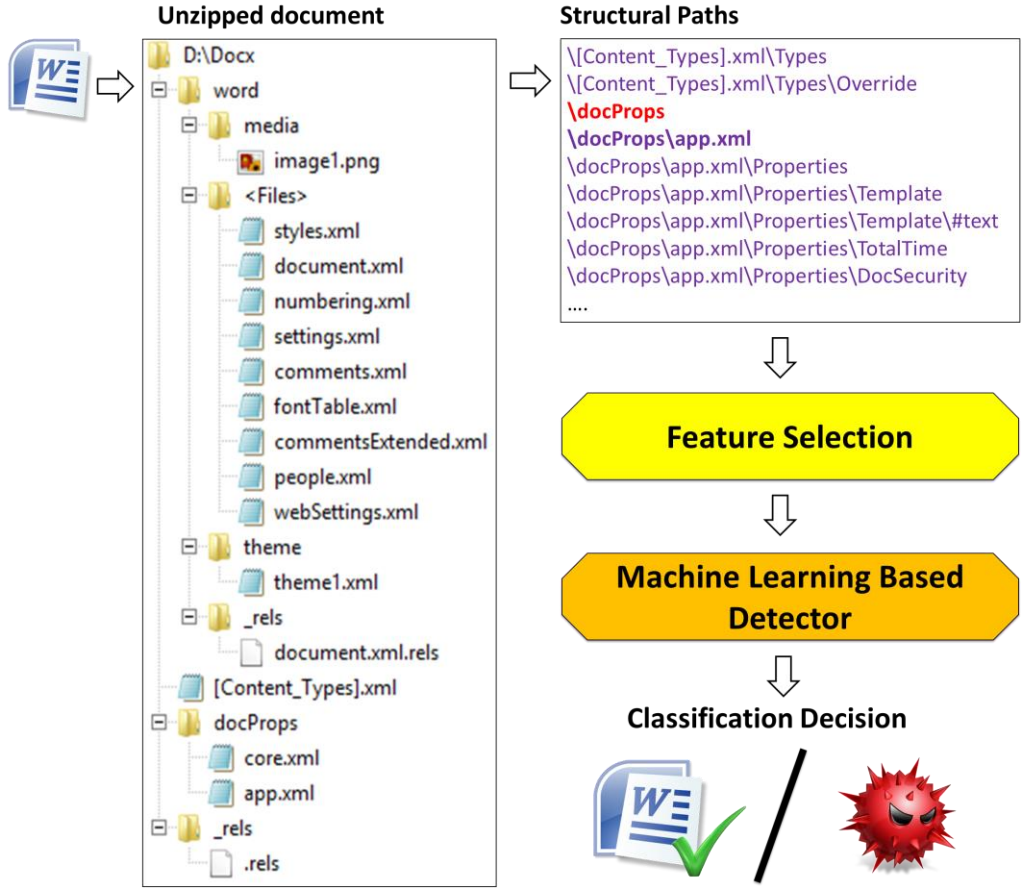
- Office documents are widely used by individuals and organizations.
- Office documents are not safe and can perform malicious actions.
- Attacker increasingly leverage Office documents in cyber attacks.
- Existing tools fail to detect new unknown malicious documents.

Goal

- Efficient detection of malicious XML-based Office documents (e.g., *.docx, *.xlsx, *.pptx, *.odt, *.ods)

Methodology

- SFEM – extracts discriminative features from Office documents based on their structure.
- Leveraging SFEM's features with Machine Learning for efficient detection of new unknown malicious documents.



Cyber Security Research Center

MalwareLab In the Cyber Security Research Center