

Principles of Analyzing Office Documents

Office Attacks



Macros



Features



Vulnerabilities

Document Analysis

Scripts

- VBA macros

Commands

Embedded files

Office Document Formats



Office Open XML Format

Zip archive containing XML

.docx, .docm

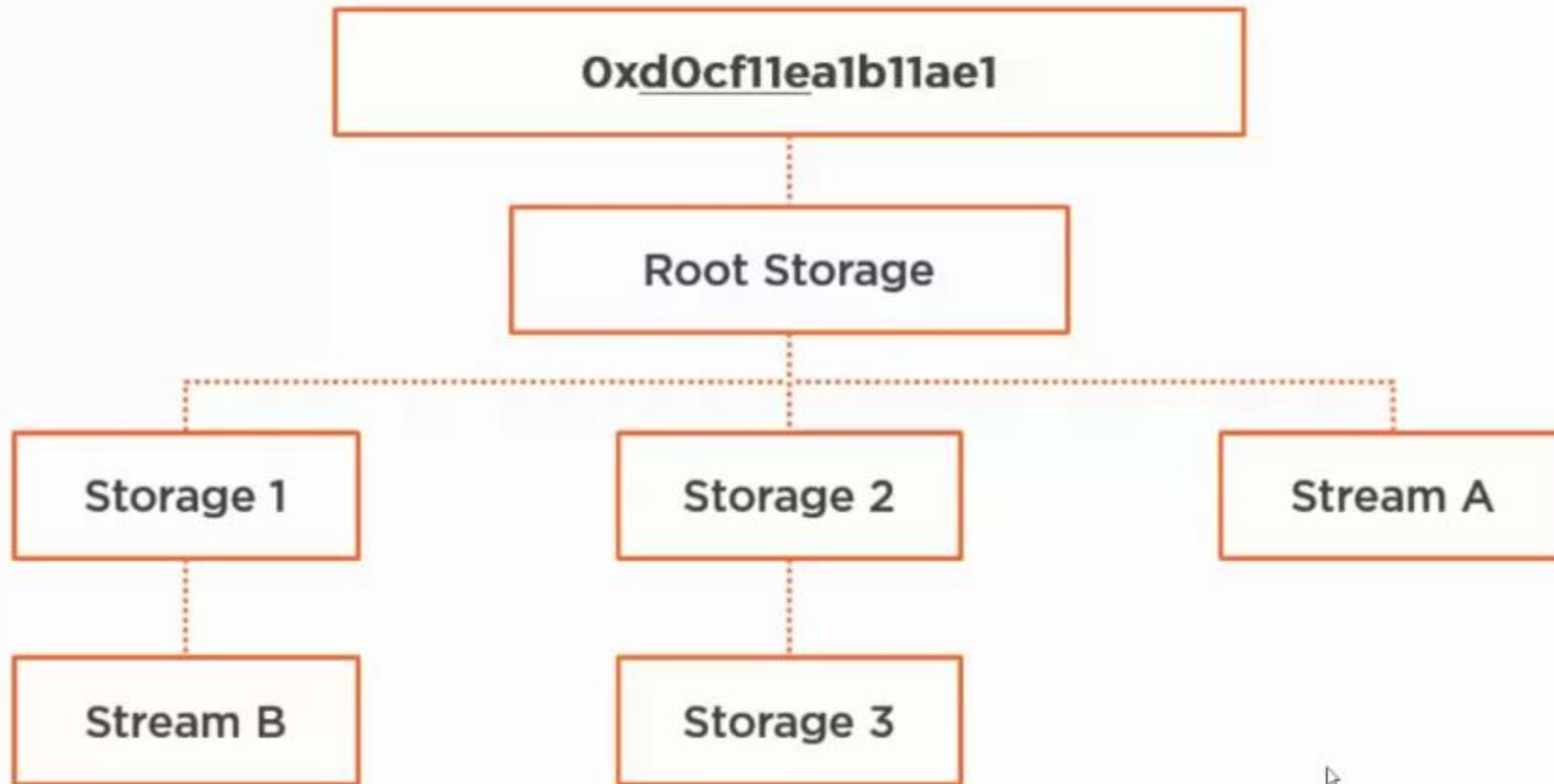


Structured Storage Format

Binary Format

.doc, .xls, .ppt

Structured Storage Format



oletools

olebrowse

- View and extract streams

oletimes

- Extract times

oleid

- Looks for malicious characteristics

olevba

- Extract VBA scripts

<https://www.decalage.info/python/oletools>

Office Open XML Format

document.docx

```
graph LR; docx[document.docx] -.-> CT[Content_Types.xml]; docx -.-> word_dir[word\]; docx -.-> doc_xml[word\document.xml]; docx -.-> doc_props_dir[docProps\]; docx -.-> core_xml[docProps\core.xml]; docx -.-> vba_dir[word\vbaProject.bin];
```

[Content_Types].xml

word\document.xml

docProps\core.xml

word\vbaProject.bin

Office Open XML Format

document.docm

[Content_Types].xml

word\document.xml

docProps\core.xml

word\vbaProject.bin

Tools

Metadata

exiftool

**Signature
Detection**

zipdump + yara

VBA

olevba

Workflow in Analyzing Office Documents

Determine the document type

- File identification tools

Search for malicious indicators

Extract and continue analysis

Thank you