DitDetector: Bimodal Learning based on Deceptive Image and Text for Macro Malware Detection

Jia YAN, Ming WAN, Xiangkun JIA, Lingyun YING*, Purui SU* and Zhanyi WANG

TCA/SKLCS, Institute of Software, Chinese Academy of Sciences

School of Cyber Security, University of Chinese Academy of Sciences

School of Computer Science and Technology, University of Chinese Academy of Sciences

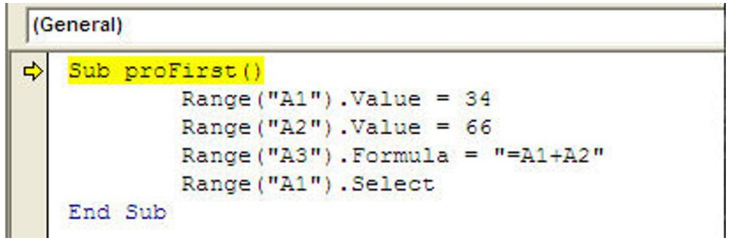
QI-ANXIN Technology Group Inc.

ISCAS
中国科学院软件研究所
Institute of Software Chinese Academy of Sciences

奇安信
QI-ANXIN

Macro Malware

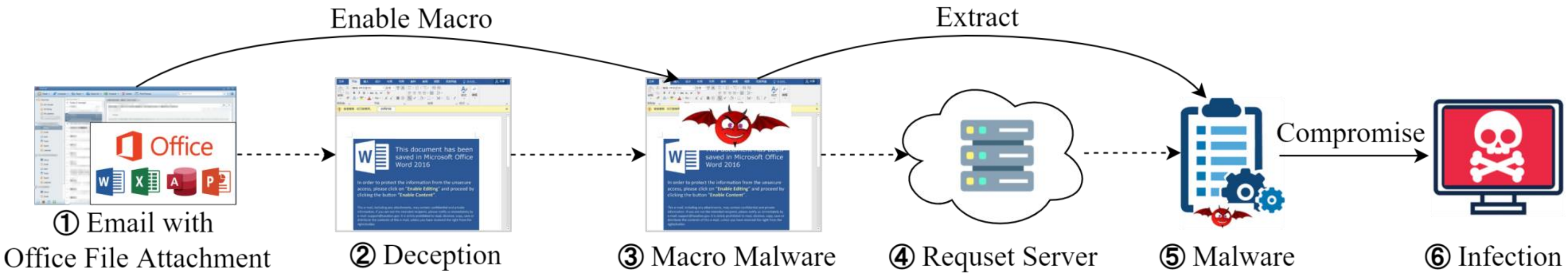
- Macro
  - Embedded in Microsoft Office Suite
  - Functionality, automate repetitive tasks



(General)

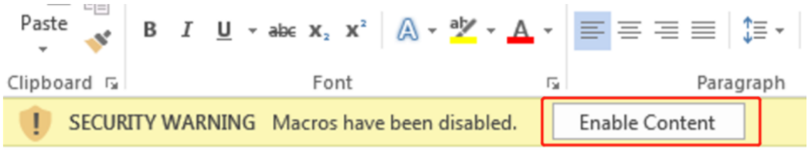
```
Sub proFirst()
        Range("A1").Value = 34
        Range("A2").Value = 66
        Range("A3").Formula = "=A1+A2"
        Range("A1").Select
End Sub
```

- Malware
  - Easily write and obfuse
  - Download attack payload or run itself



Macro Malware

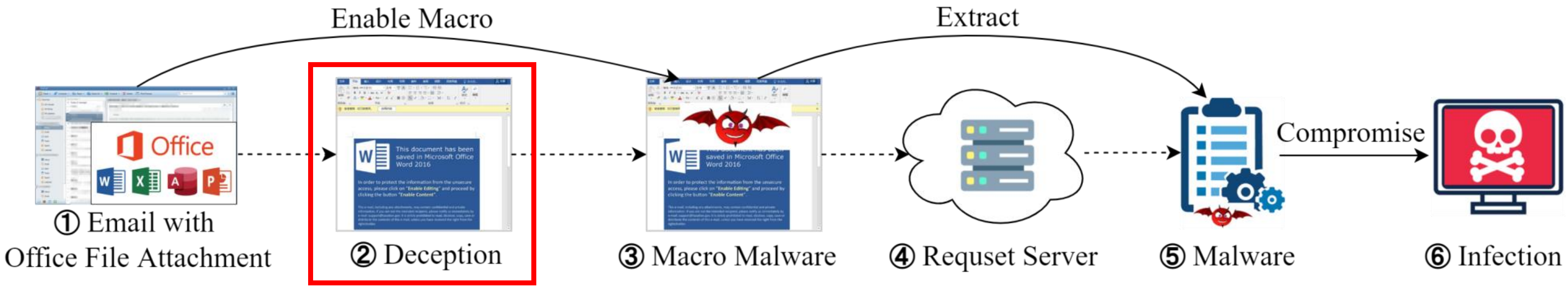
- Defense
  - Default disable-disable strategy



SECURITY WARNING Macros have been disabled.

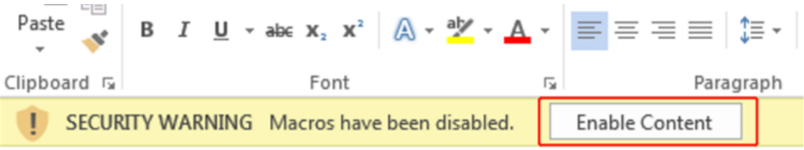
- Detection
  - Spam filtering
  - Document analysis
  - Traffic analysis
  - Runtime detection

- Deceptive Information



Macro Malware

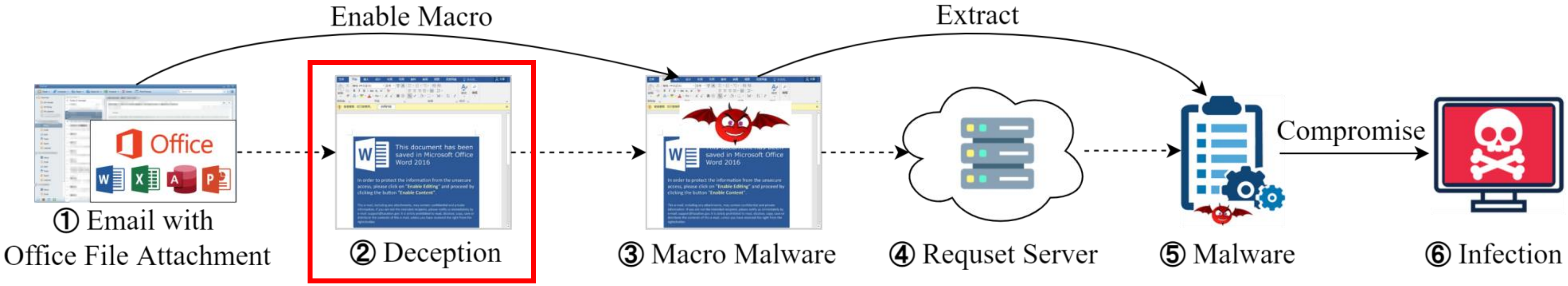
- Defense
  - Default disable-disable strategy



SECURITY WARNING Macros have been disabled. Enable Content

- Deceptive Information

- Detection
  - Spam filtering
  - Document analysis
  - Traffic analysis
  - Runtime detection



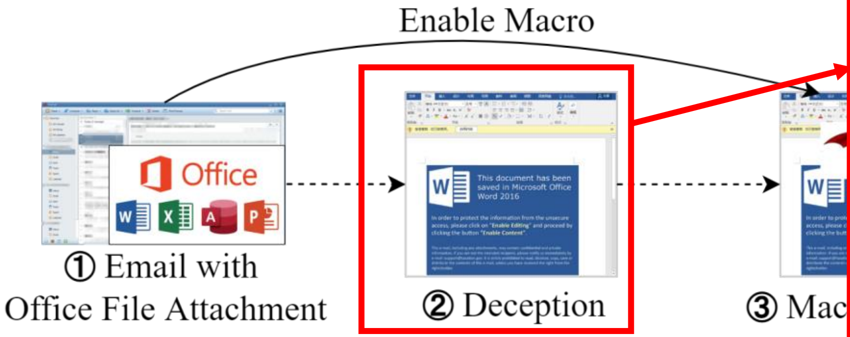
Macro Malware

- Document analysis
  - Signature
    - Specific strings
  - File metadata
    - Structural feature extraction
  - Macro code
    - Plain text
    - Machine Learning Algorithm
  - Deceptive information
    - Keyword hits
    - Visual element matching

- Challenges
  - Variants evasion
  - Different types of file formats
    - Compound File Binary (CFB)
    - Office Open XML (OOXML)
  - Multiple types of macros
    - VBA macro
    - Excel 4.0 macro
  - Advance malware
    - Remote template injection

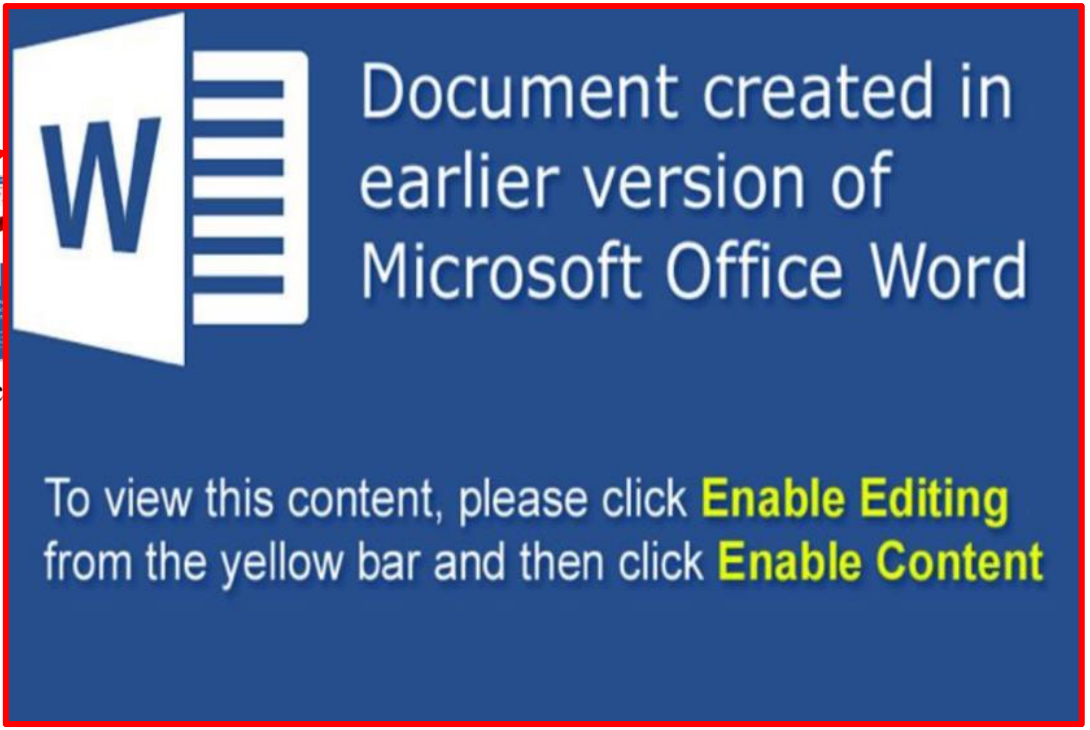
Macro Malware

- Document analysis
  - Signature



  - Machine Learning Algorithm
- Deceptive information
  - Keyword hits
  - Visual element matching

- Challenges


Document created in earlier version of Microsoft Office Word

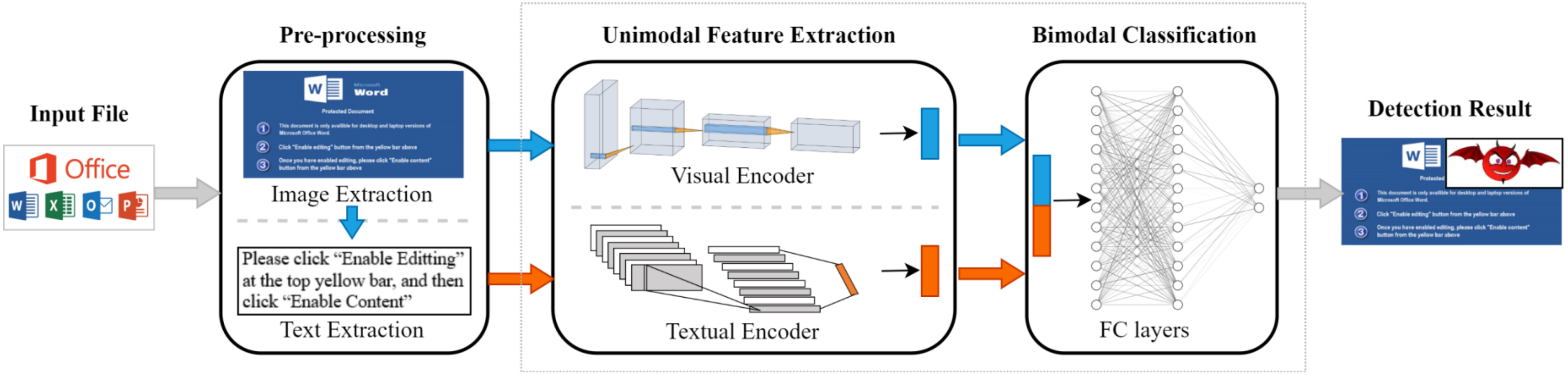
To view this content, please click **Enable Editing** from the yellow bar and then click **Enable Content**

DitDetector

A bimodal learning-based detector, to detect deceptive information

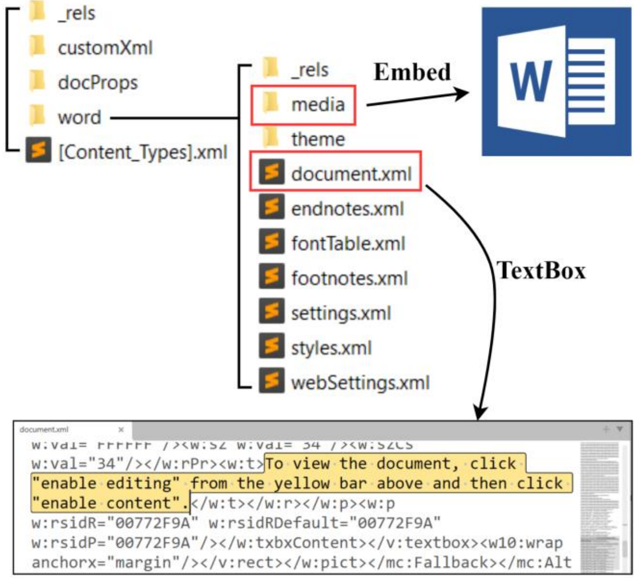
- End to end
- Complementarity
  - Image
  - Text

- Pre-processing
- Unimodal Feature Extraction
  - Visual Encoder
  - Textual Encoder
- Bimodal Classification



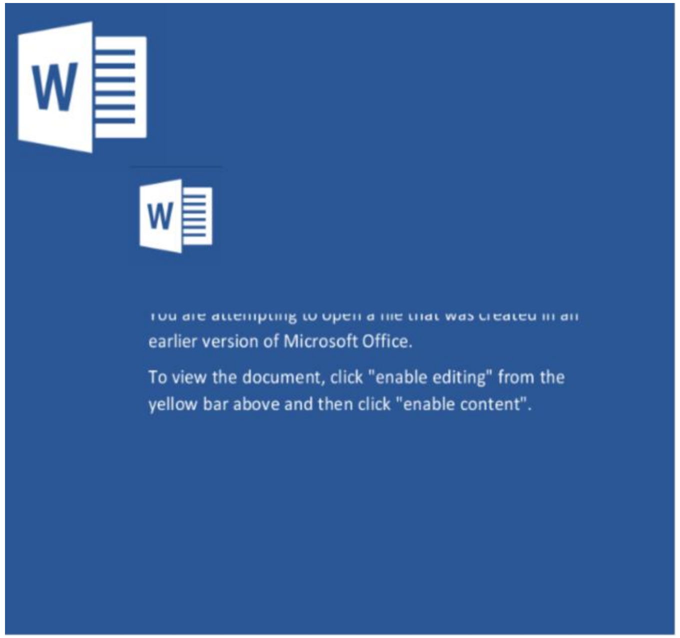
Pre-Processing

- Decompress document
  - Multiple independent individuals
  - File formats limitation
    - OOXML ✓
    - CFB ✗



```
w:val="FFFFFF"/><w:sz w:val="34"/><w:szCs
w:val="34"/></w:rPr><w:t>To view the document, click
"enable editing" from the yellow bar above and then click
"enable content".</w:t></w:r></w:p><w:p
w:rsidR="00772F9A" w:rsidRDefault="00772F9A"
w:rsidP="00772F9A"/></w:txbxContent></v:textbox><w10:wrap
anchorx="margin"/></v:rect></w:pict></mc:Fallback></mc:Alt
```

- Preview image extraction
  - Oracle open-source tool
  - Get the layout and complete information

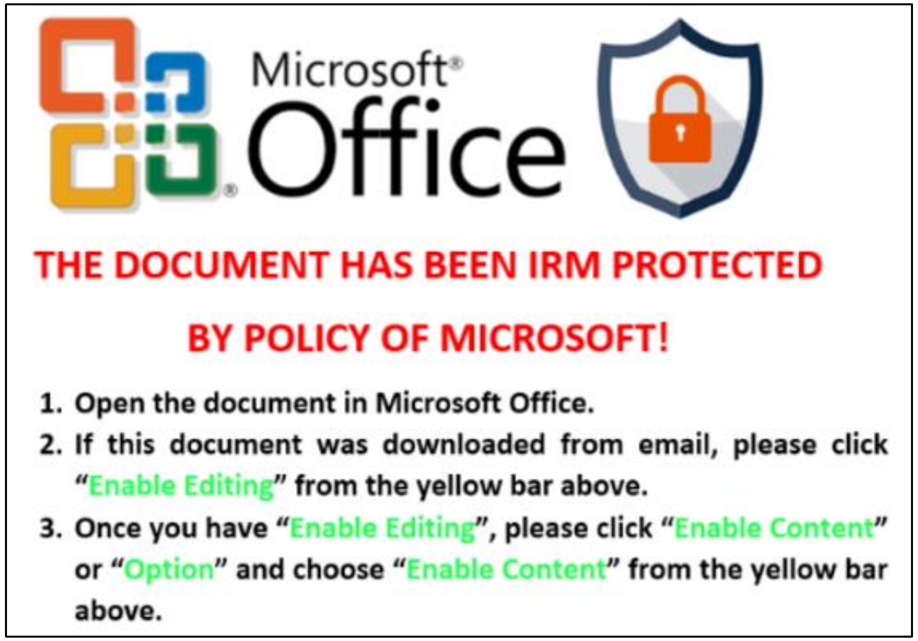


You are attempting to open a file that was created in an earlier version of Microsoft Office.

To view the document, click "enable editing" from the yellow bar above and then click "enable content".

Pre-Processing

- Text Extraction
  - Challenge
    - File fromat limitation
    - Text embeds in image
  - Solution
    - Optimical Character Recognition
    - e.g., Tesseract


Microsoft®
Office

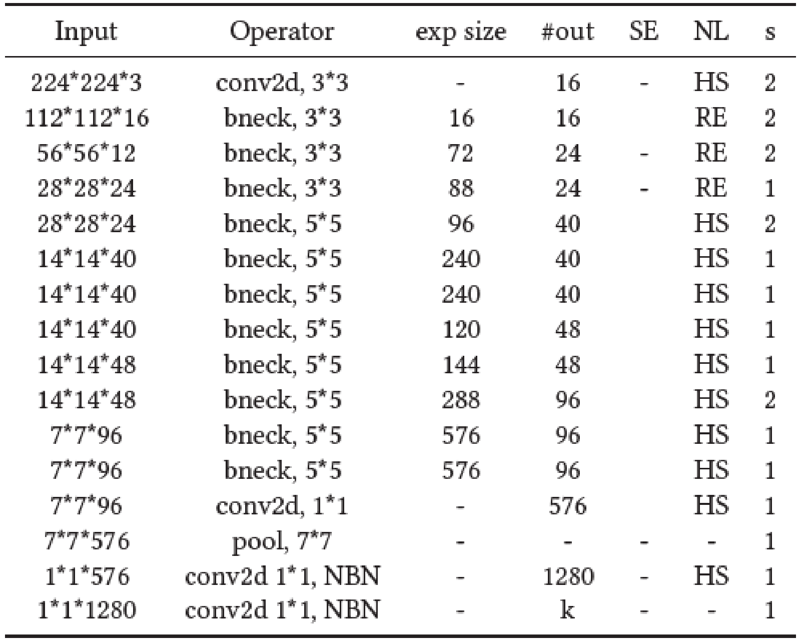
THE DOCUMENT HAS BEEN IRM PROTECTED

BY POLICY OF MICROSOFT!

1. Open the document in Microsoft Office.
2. If this document was downloaded from email, please click "Enable Editing" from the yellow bar above.
3. Once you have "Enable Editing", please click "Enable Content" or "Option" and choose "Enable Content" from the yellow bar above.

Detection

- Visual encoder
  - MobileNetV3-Small[1]



Input	Operator	exp size	#out	SE	NL	s
224*224*3	conv2d, 3*3	-	16	-	HS	2
112*112*16	bneck, 3*3	16	16		RE	2
56*56*12	bneck, 3*3	72	24	-	RE	2
28*28*24	bneck, 3*3	88	24	-	RE	1
28*28*24	bneck, 5*5	96	40		HS	2
14*14*40	bneck, 5*5	240	40		HS	1
14*14*40	bneck, 5*5	240	40		HS	1
14*14*40	bneck, 5*5	120	48		HS	1
14*14*48	bneck, 5*5	144	48		HS	1
14*14*48	bneck, 5*5	288	96		HS	2
7*7*96	bneck, 5*5	576	96		HS	1
7*7*96	bneck, 5*5	576	96		HS	1
7*7*96	conv2d, 1*1	-	576		HS	1
7*7*576	pool, 7*7	-	-	-	-	1
1*1*576	conv2d 1*1, NBN	-	1280	-	HS	1
1*1*1280	conv2d 1*1, NBN	-	k	-	-	1

- Text encoder
  - TextCNN[2]

$$\mathrm{z}'_{ks} = \mathrm{Conv1D}(\mathrm{x}_t, ks), \quad ks = 3, 4, 5$$

$$\mathrm{z}_{ks} = \mathrm{MaxPooling}(\mathrm{z}'_{ks}), \quad ks = 3, 4, 5$$

$$\mathrm{z}_t = z_3 \oplus z_4 \oplus z_5$$

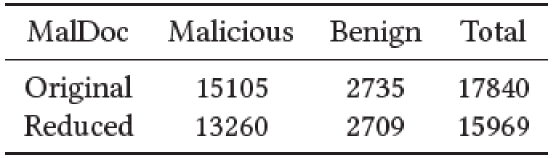
- Bimodal Classification
  - Joint representation
    - Visual representation
    - Textual representation
  - 2 fully connetecd layers

[1] Howard, Andrew, et al. "Searching for mobilenetv3." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

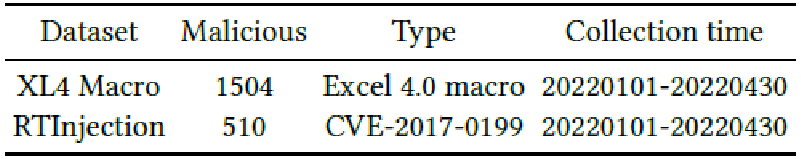
[2] Zhang, Ye, and Byron Wallace. "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." arXiv (2015).

Evaluation

- Setup
  - Three dataset
    - MalDoc[3]
    - XL4 Macro
    - RTInjection



MalDoc	Malicious	Benign	Total
Original	15105	2735	17840
Reduced	13260	2709	15969



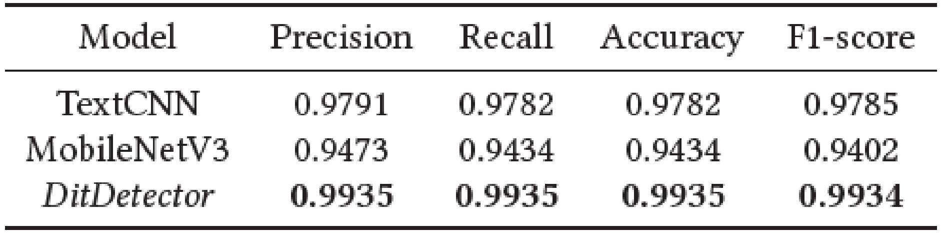
Dataset	Malicious	Type	Collection time
XL4 Macro	1504	Excel 4.0 macro	20220101-20220430
RTInjection	510	CVE-2017-0199	20220101-20220430

- Experiments
  - Ablation Study
    - Textual detection
    - Visual detection
    - Bimodal detection - DitDetector
  - Compare with methods
    - 4 state-of-the-art(SOTA) ML models
      - RFC, MLP, SVC and XGBoost
    - 5 AntiVirus engines
      - Kaspersky, Symantec, Microsoft, McAfee and Sophos

[3] Vasilios Koutsokostas, et al. Malicious MS Office documents dataset. https://doi.org/10.5281/zenodo.4559436

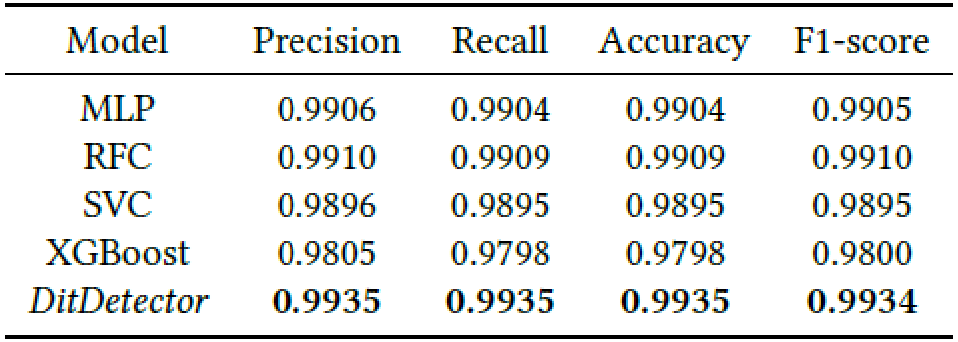
Evaluation on MalDoc

- DitDetector outperforms better on ablation study

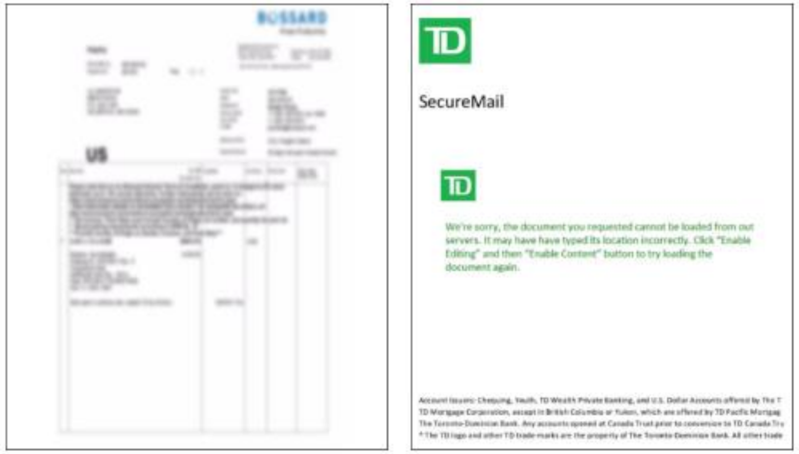


Model	Precision	Recall	Accuracy	F1-score
TextCNN	0.9791	0.9782	0.9782	0.9785
MobileNetV3	0.9473	0.9434	0.9434	0.9402
DitDetector	**0.9935**	**0.9935**	**0.9935**	**0.9934**

- DitDetector does not lose the performance



Model	Precision	Recall	Accuracy	F1-score
MLP	0.9906	0.9904	0.9904	0.9905
RFC	0.9910	0.9909	0.9909	0.9910
SVC	0.9896	0.9895	0.9895	0.9895
XGBoost	0.9805	0.9798	0.9798	0.9800
DitDetector	**0.9935**	**0.9935**	**0.9935**	**0.9934**



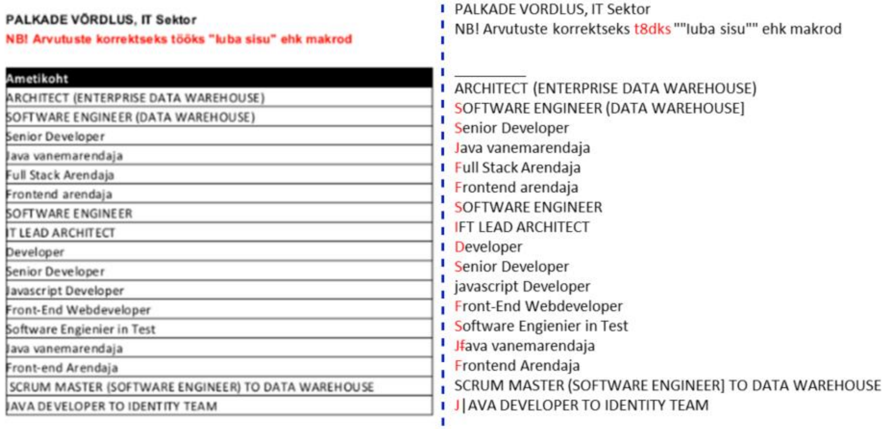
SecureMail

We're sorry, the document you requested cannot be loaded from out servers. It may have have typed its location incorrectly. Click "Enable Editing" and then "Enable Content" button to try loading the document again.

Account issuers: Chequing, Youth, TD Wealth Private Banking, and U.S. Dollar Accounts offered by The T
TD Mortgage Corporation, except in British Columbia or Yukon, which are offered by TD Pacific Mortgag
The Toronto-Dominion Bank. Any accounts opened at Canada Trust prior to conversion to TD Canada Tru
* The TD logo and other TD trade-marks are the property of The Toronto-Dominion Bank. All other trade

(a) Visual information

(b) Textual information



PALKADE VÕRDLUS, IT Sektor

NB! Arvutuste korrektseks tööks "luba sisu" ehk makrod

Ametikoht
ARCHITECT (ENTERPRISE DATA WAREHOUSE)
SOFTWARE ENGINEER (DATA WAREHOUSE)
Senior Developer
Java vanemarendaja
Full Stack Arendaja
Frontend arendaja
SOFTWARE ENGINEER
IT LEAD ARCHITECT
Developer
Senior Developer
Javascript Developer
Front-End Webdeveloper
Software Engienier in Test
Java vanemarendaja
Front-end Arendaja
SCRUM MASTER (SOFTWARE ENGINEER) TO DATA WAREHOUSE
JAVA DEVELOPER TO IDENTITY TEAM

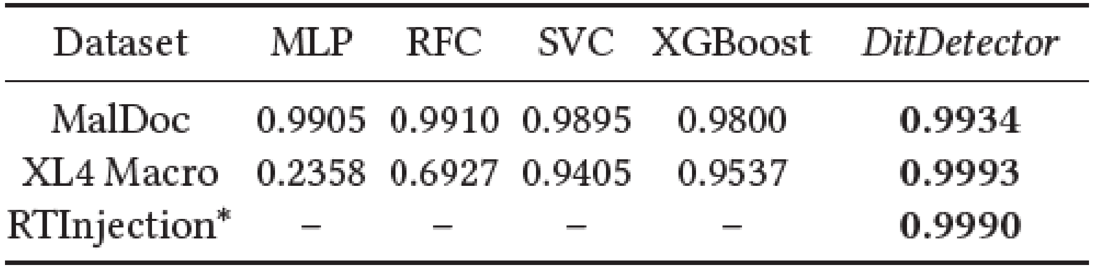
PALKADE VORDLUS, IT Sektor
NB! Arvutuste korrektseks t8dks ""luba sisu"" ehk makrod

____
ARCHITECT (ENTERPRISE DATA WAREHOUSE)
SOFTWARE ENGINEER (DATA WAREHOUSE]
Senior Developer
Java vanemarendaja
Full Stack Arendaja
Frontend arendaja
SOFTWARE ENGINEER
IFT LEAD ARCHITECT
Developer
Senior Developer
javascript Developer
Front-End Webdeveloper
Software Engienier in Test
Jfava vanemarendaja
Frontend Arendaja
SCRUM MASTER (SOFTWARE ENGINEER] TO DATA WAREHOUSE
J|AVA DEVELOPER TO IDENTITY TEAM

Estonian deceptive information

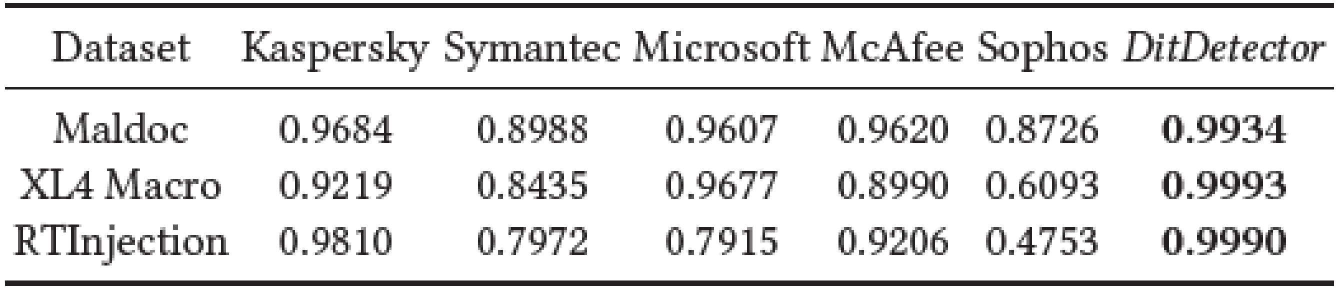
Evaluation on Real-world Dataset

- DitDetector outperforms robust



Dataset	MLP	RFC	SVC	XGBoost	*DitDetector*
MalDoc	0.9905	0.9910	0.9895	0.9800	**0.9934**
XL4 Macro	0.2358	0.6927	0.9405	0.9537	**0.9993**
RTInjection*	–	–	–	–	**0.9990**

- DitDetector performs better than AV engines



Dataset	Kaspersky	Symantec	Microsoft	McAfee	Sophos	*DitDetector*
Maldoc	0.9684	0.8988	0.9607	0.9620	0.8726	**0.9934**
XL4 Macro	0.9219	0.8435	0.9677	0.8990	0.6093	**0.9993**
RTInjection	0.9810	0.7972	0.7915	0.9206	0.4753	**0.9990**

Evasion and Countermeasures

- Macro code detection evasion
  - Good at VBA macro
  - Unstable on XL4 macro
  - Fail to detect remote template injection
    - Document at first attack stage has no macro code
- Unimodal deceptive information detection evasion
  - Replace the regular form of misleading with blurred images
  - Change the wording to persuade users
- DitDetector solves the above problems
  - Exploiting the complementary properties of visual and textual encoders
    - textual encoder focuses on malicioussemantics and can counter text evasion
    - visual encoder learns unusual imageelements for robustness

Discussion

- DitDetector faces real-world challenges
  - ML-based adversarial attack

  - Perform a more comprehensive defense solution at different stages of the attack chain
- Some limitations should be improved
  - Sensitive to the language family
    - Now works well in Indo-European languages

    - Improve the compatibility of the encoder for more languages

Conclusion

- We design and implement DitDetector
  - Detect malicious office documents related to macro malware
  - Counter adversarial samples, e.g., non-macro documents
- We evaluate DitDetector on three datasets
  - Outperform four compared SOTA machine learning methods and five AV engines
- Open Source
  - https://gitee.com/yjasper/dit-detector

Thank you!
Q&A

Jia YAN

yanjia2016@iscas.ac.cn