



NSS 2023:17th International Conference on Network and System Security

Transformer Based model to Detect BEC Phishing Attack

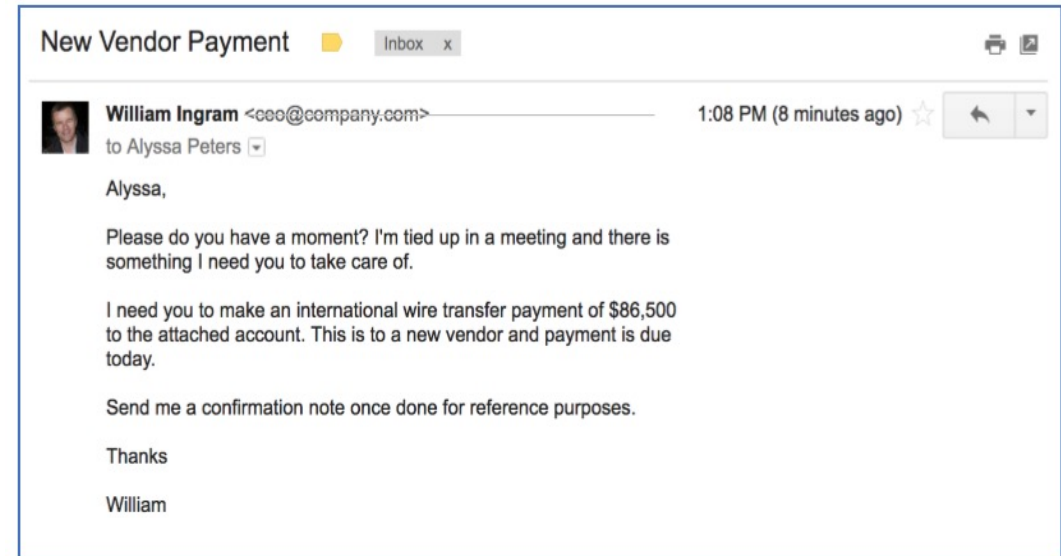
Amirah Almutairi, Dr. BooJoong Kang ,and Dr. Nawfal Fadhel

Outline:

- Introduction
- Related Work
- Proposed Method
- Experiment
- Discussion and Conclusion

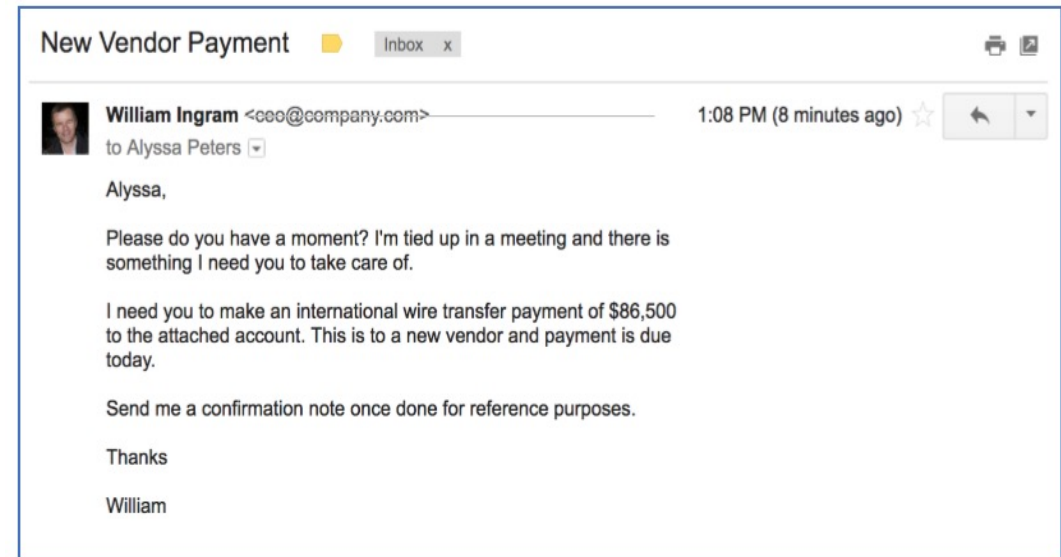
Business Email Compromise

- **What is BEC attack?**
 - BEC is Business Email Compromise is a specific type of phishing attack.



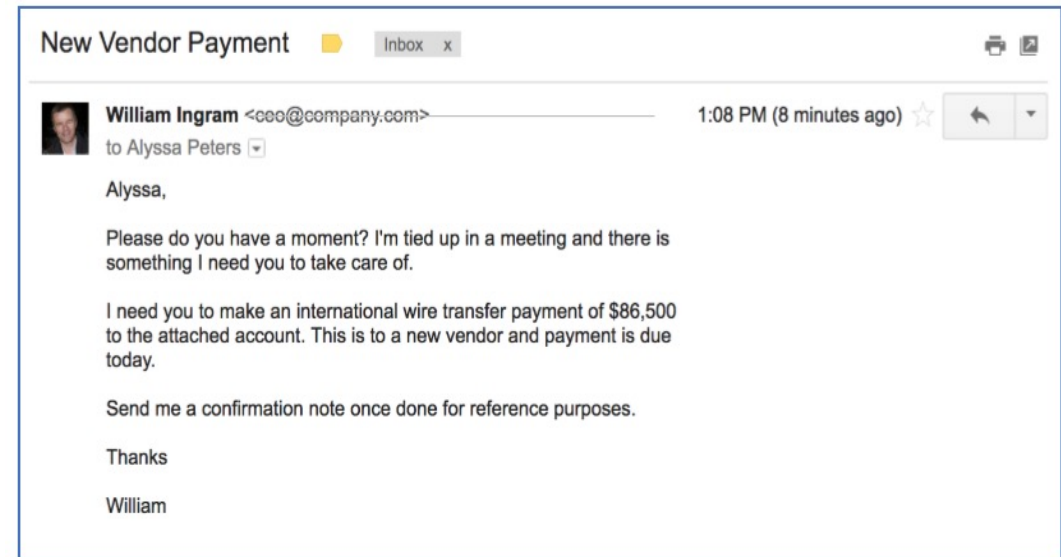
Business Email Compromise


- **What is BEC attack?**
 - BEC is Business Email Compromise is a specific type of phishing attack.
 - BEC emails are usually sent from trusted accounts that have been compromised by attackers, and the legitimate owners are often unaware of the attacks.



Business Email Compromise

- **What is BEC attack?**
 - BEC is Business Email Compromise is a specific type of phishing attack.
 - BEC emails are usually sent from trusted accounts that have been compromised by attackers, and the legitimate owners are often unaware of the attacks.
 - Around 60% of BEC incidents don't contain any malicious links, which makes it even harder to identify by current email security defense.






Related works:

- 
- Focused on metadata.

(1) Cidon, A., Gavish, L., Bleier, I., Korshun, N., Schweighauser, M., & Tsitkin, A. (2019). High precision detection of business email compromise. In *28th USENIX Security Symposium (USENIX Security 19)* (pp. 1291-1307).

(2) Vorobeva, A., Khisaeva, G., Zakoldaev, D., & Kotenko, I. (2021, October). Detection of Business Email Compromise Attacks with Writing Style Analysis. In *International Symposium on Mobile Internet Security* (pp. 248-262). Singapore: Springer Nature Singapore.



Related works:


- Focused on metadata.
- Using Feature selection.

(1) Cidon, A., Gavish, L., Bleier, I., Korshun, N., Schweighauser, M., & Tsitkin, A. (2019). High precision detection of business email compromise. In *28th USENIX Security Symposium (USENIX Security 19)* (pp. 1291-1307).

(2) Vorobeva, A., Khisaeva, G., Zakoldaev, D., & Kotenko, I. (2021, October). Detection of Business Email Compromise Attacks with Writing Style Analysis. In *International Symposium on Mobile Internet Security* (pp. 248-262). Singapore: Springer Nature Singapore.



- **Research Question:**

- How can transformer-based models be employed to detect BEC attacks in plain text emails, and what factors aid in their identification?
- 

Experiments Step 1:

- **Formulate the research hypothesis:**

Experiments Step 1:

- **Formulate the research hypothesis:**

In a scenario where there is no explicit indicator that email is from an attacker, for example: BEC attacks, transformer machine learning based models can be used to classify messages from trustworthy sources and attackers accurately.

Experiments Step 2:

Pre-process the Datasets.

- **Fraud Email:** 5,187 phishing emails and 6,742 Reliable emails.

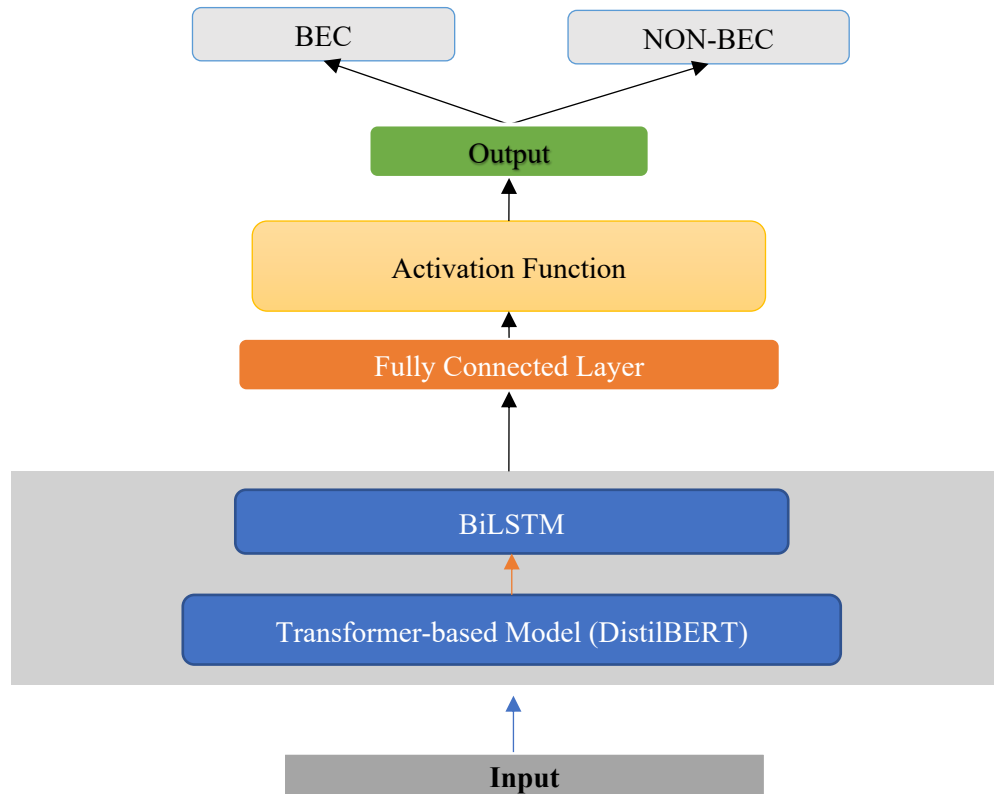
Experiments Step 2:

Pre-process the Datasets.

- **Fraud Email:** 5,187 phishing emails and 6,742 Reliable emails.
- **Trec 2007:** 50,199 phishing emails and 25,220 Reliable emails

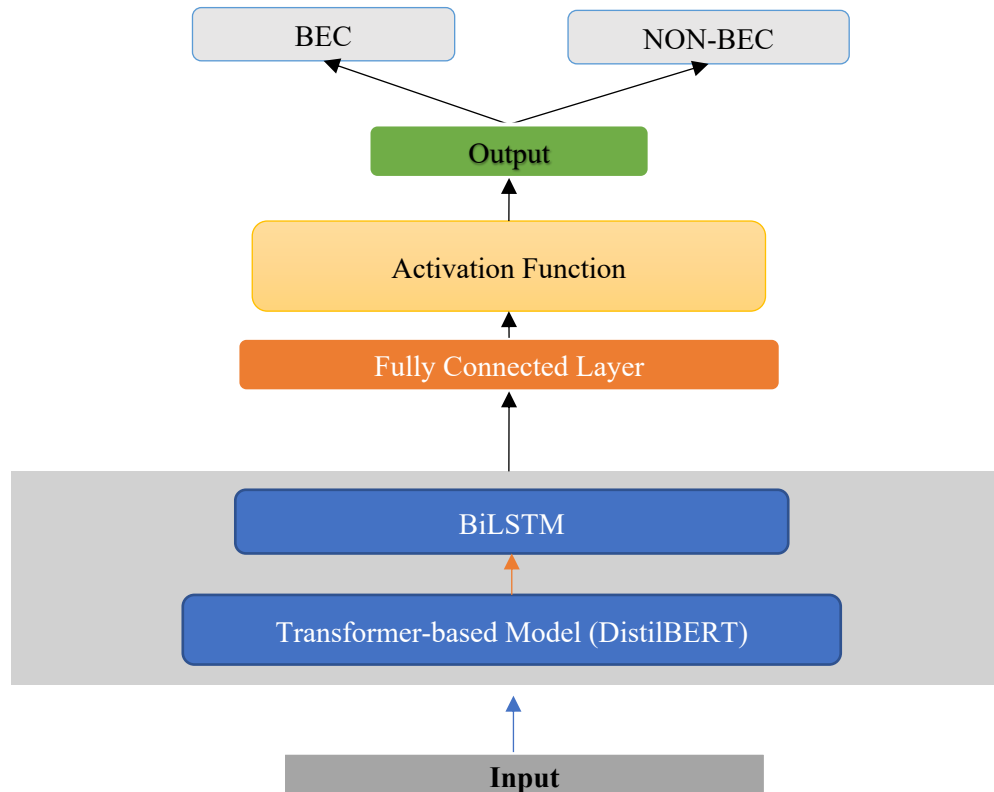
Experiments Step 3:

Design the model.



Experiments Step 3:

Design the model.



MAX length	150
batch size	16
learning rate	$2e^{-5}$
epochs	3
hidden_dim	50

TABLE 5.1: Hyperparameters

BERT

- BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.

BERT

- BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.
- BERT is a “large language model” that is trained on a **very large amount of text**.

BERT

- BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.
- BERT is a “large language model” that is trained on a **very large amount of text**.
- It is based on the Transformer architecture that we all share, and then fine-tune on any smaller dataset.



BERT

- BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.
- BERT is a “large language model” that is trained on a **very large amount of text**.
- It is based on the Transformer architecture that we all share, and then fine-tune on any smaller dataset.
- Unlike traditional language models that process text in a sequential manner, BERT utilizes a bidirectional approach.



BERT

- BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.
- BERT is a “large language model” that is trained on a **very large amount of text**.
- It is based on the Transformer architecture that we all share, and then fine-tune on any smaller dataset.
- Unlike traditional language models that process text in a sequential manner, BERT utilizes a bidirectional approach.
- BERT has the ability to capture the context of words in a sentence, allowing it to better understand the meaning behind the text.



BiLSTM

- BiLSTM stands for **B**idirectional **L**ong **S**hort-Term **M**emory

BiLSTM

- BiLSTM stands for **B**idirectional **L**ong **S**hort-**T**erm **M**emory
- It is an extension of the LSTM architecture that captures information from both past and future contexts in sequential data.

BiLSTM

- BiLSTM stands for **B**idirectional **L**ong **S**hort-Term **M**emory
- It is an extension of the LSTM architecture that captures information from both past and future contexts in sequential data.
- In a BiLSTM, the input sequence is fed into two separate LSTM layers: one processing the sequence in the forward direction and the other in the backward direction

Experiments Step 3:

- We implemented two baseline models that replicate the methods used in related studies to compare the results with our method.

Experiments Step 3:

- We implemented two baseline models that replicate the methods used in related studies to compare the results with our method.

Model name	Method	Features
LogReg	Logistic regression	TF-IDF counts
Xgboost	Distributed gradient boosting	TF-IDF counts
BERT	Long short-term memory network	DistilBERT embedding

Experiments Step 4:

Evaluate the proposed model using traditional evaluation metrics.

Model	Fraud mail			Trec07		
	P	R	F1	P	R	F1
LogReg	0.987	0.987	0.987	0.982	0.982	0.982
Xgboost	0.988	0.988	0.988	0.985	0.985	0.985
BERT	0.998	0.998	0.998	0.986	0.986	0.986

TABLE 4.1: Spam classification weighted average F1 results.
(The best F1 score for each dataset is highlighted).

Experiments Step 5:

Compare our results against current study:

Model name	Method	Dataset	Accuracy
KNN Xiao and Jiang (2020)	LSTM /TF-IDF	Fraud mail	0.94
(proposed)	BiLSTM / DistilBERT	Fraud mail	0.99

TABLE 5.4: Accuracy Comparison with some Existing Approach

Conclusion and Future Work:

- We proposed a novel approach for detecting Business Email Compromise (BEC) attacks using a Transformer Based model.

Conclusion and Future Work:

- We proposed a novel approach for detecting Business Email Compromise (BEC) attacks using a Transformer Based model.
- The effectiveness of Transformer models even when they only have access to written content, without additional information about sources, links or attachments.

Conclusion and Future Work:

- We proposed a novel approach for detecting Business Email Compromise (BEC) attacks using a Transformer Based model.
- The effectiveness of Transformer models even when they only have access to written content, without additional information about sources, links or attachments.
- Our study highlights the importance of considering language and content factors in the detection of phishing attacks.

Conclusion and Future Work:

- We proposed a novel approach for detecting Business Email Compromise (BEC) attacks using a Transformer Based model.
- The effectiveness of Transformer models even when they only have access to written content, without additional information about sources, links or attachments.
- Our study highlights the importance of considering language and content factors in the detection of phishing attacks.
- Future research will test our proposed model to other datasets and investigate the interpretability of the model's decisions.

Thank you!

Questions?

