By: Djoshua Moonen
Supervisor: Stjepan Picek &
Marina Krcek

Little or Large?

The effects of network size on AI explainability in Side-Channel Attacks

1. Knowledge Gap

Side-Channel Attacks

Side-Channel attacks are when a person or organization obtains information through the implementation of a computer system.

This work specifically looks at how encryption keys can be collected by observing power consumption during the encryption process

Research Question

In the literature, the benefits of reducing network complexity are mentioned. However, no evidence that backs this claim was found.

This lead us to ask: what are the effects of network size on AI explainability?

2. Experiment

Visualization

Heatmap:

- Input/output relation
- Low computational overhead

Dataset

ASCAD:

- Realistic
- Frequently used

Architectures

ASCAD:

- 5 blocks
- 1D convolution
- Average pooling
- Batch size 200
- Epochs 100

ZAID:

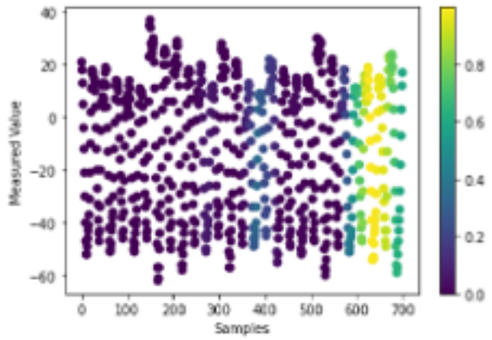
- 1 block
- 1D convolution
- Average Pooling
- Batch size 50
- Epochs 50

3. Results

Heatmap overlay

Area of interest:

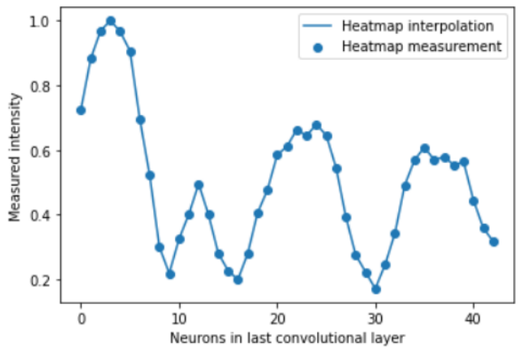
- ASCAD: horizontally
- ZAID: vertically



Heatmap as function

Patterns in misclassification

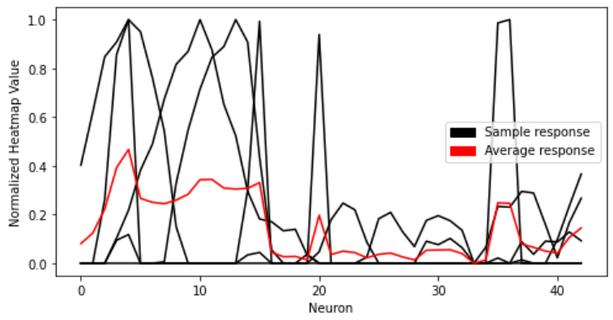
- ASCAD: spread & additional peaks
- ZAID: Too similar or uniform distribution



Class level pattern

Level of consistency:

- ASCAD: low
- ZAID: high



4. Discussion

Heatmap overlay

The difference in area of interest likely stems from the different amount of convolutions are performed.

Heatmap as function

It seems to be the case that neither of the two models perform better when a misclassification occurs.

Class level pattern

The model large in complexity was shown to be inconsistent, indicating lower levels of explainability.

5. Conclusion

Conclusions

A reduction in complexity leads to improved explainability because of:

- No notable changes in the misclassification patterns
- Improvement in classification explainability of the ZAID model

Future work

The following topics show promise for future research:

- The effect different datasets have.
- How different model behave
- Output from different visualization techniques

CSE3000 June 25th, 2020