UNDERSTANDING A CNN ATTACK: AS CRUCIAL AS SUCCEEDING IT

CONTACT:
Loïc Masure (1, 2, 3)
Cécile Dumas (1, 2)
Emmanuel Prouff (3)

Today: CNNs are used to extract sensitive information in Side Channel Analysis. How do they extract relevant information from the trace? This well known problem in Machine Learning implies a lack of expertise to design further counter measures. New idea: use CNNs to localize sensitive information and PoIs.

Vanilla Backprop (VBP)
- Computes the loss function and its gradient via backprop through the layers to the input trace.
- The loss function is more likely to be impacted by small perturbations at relevant PoIs. Therefore the gradient should be higher at those points.
- Problem: the loss function for complex CNNs is prone to be less smooth, leading to irrelevant peaks unless training with lots of data.

Guided Backprop (GBP)
- Same principle as VBP except for ReLU layers.
- Activation in a ReLU layer:
  \[ f_i^{t+1} = ReLU(f_i^t) = \max(f_i^t, 0) \]
- Regular backprop:
  \[ R_i^t = (f_i^t > 0).R_i^{t+1} \]
- Guided Backprop:
  \[ R_i^t = (f_i^t > 0).(R_i^{t+1} > 0).R_i^{t+1} \]
- Gives better and sparser relevance maps.

Layerwise Relevance Prop (LRP)
- Each type of layer has its own rule of relevance propagation.
- More details at heatmapping.org
- Performance highly depends on an explainable architecture.

Results
- We introduced means to interpret the success of a CNN attack.
- It can help the developer to understand the vulnerability and therefore help for a better design against such threats.
- Sensitivity methods (VBP, GBP) perform a better characterization than SNRs against masking.
- Those methods are also robust against desynchronization, provided the CNN is trained on desynchronized data (see animated demo).
- Visualizing the characterization helps evaluating whether a CNN overfits, and therefore leads the choice for an optimal architecture.

Affiliations
1. CEA, LETI, MINATEC Campus, F-38054 Grenoble, France, name.surname@cea.fr
2. Univ. Grenoble Alpes, F-38000, Grenoble, France
3. Sorbonne Universités, UPMC Univ Paris 06, POLSYS, UMR 7606, LIP6, F-75005, Paris, France