| | Method | Dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | | ASCADv1 (fixed) | ASCADv1 (random) | DPAv4 (Zaid version) | AES-HD | OTiAiT | OTP |
| | Random | $112 \pm 1$ | $108 \pm 4$ | $11.5 \pm 0.4$ | $127 \pm 1$ | $1.20 \pm 0.05$ | $1.048 \pm 0.008$ |
| First-order parametric methods | SNR | $111.0 \pm 0.2$ | $123 \pm 2$ | $\boxed{126 \pm 1}$ | $\boxed{128.5 \pm 0.3}$ | $4.26 \pm 0.07$ | $1.33 \pm 0.04$ |
| | SOSD | $111.6 \pm 0.2$ | $125.6 \pm 0.6$ | $195.7 \pm 0.9$ | $\boxed{128.3 \pm 0.3}$ | $3.94 \pm 0.08$ | $\boxed{1.34 \pm 0.04}$ |
| | CPA | $118.2 \pm 0.4$ | $114 \pm 2$ | $111.5 \pm 0.9$ | $\boxed{128.4 \pm 0.3}$ | $2.7 \pm 0.2$ | $1.32 \pm 0.04$ |
| Neural net attribution | GradVis | $124.9 \pm 0.3$ | $127 \pm 1$ | $121 \pm 1$ | $127 \pm 1$ | $1.8 \pm 0.2$ | $1.31 \pm 0.05$ |
| | Saliency | $124.8 \pm 0.3$ | $127 \pm 1$ | $\underline{124 \pm 1}$ | $127 \pm 1$ | $2.8 \pm 0.2$ | $1.29 \pm 0.05$ |
| | Input $*$ Grad | $124.8 \pm 0.3$ | $127 \pm 1$ | $\underline{124 \pm 1}$ | $127 \pm 1$ | $3.1 \pm 0.2$ | $1.29 \pm 0.05$ |
| | LRP | $124.8 \pm 0.3$ | $127 \pm 1$ | $\underline{124 \pm 1}$ | $127 \pm 1$ | $3.1 \pm 0.2$ | $1.29 \pm 0.05$ |
| | 1-Occlusion | $124.8 \pm 0.3$ | $127 \pm 1$ | $\boxed{\underline{125 \pm 1}}$ | $127 \pm 1$ | $3.2 \pm 0.2$ | $1.29 \pm 0.05$ |
| | $m$-Occlusion[†] | $125.2 \pm 0.4$ | $127.1 \pm 0.6$ | $122.7 \pm 0.9$ | $127 \pm 2$ | $\boxed{\underline{4.3 \pm 0.1}}$ | $1.24 \pm 0.03$ |
| | $2^{\text{nd}}$-order 1-Occlusion | $124.8 \pm 0.3$ | $127 \pm 1$ | $\boxed{\underline{125.0 \pm 0.6}}$ | $127 \pm 2$ | $3.6 \pm 0.2$ | $1.29 \pm 0.05$ |
| | OccPOI | $110.8 \pm 0.5$ | $106 \pm 2$ | $23 \pm 20$ | $127 \pm 1$ | $1.20 \pm 0.05$ | $1.049 \pm 0.008$ |
| | OccPOI (Released result) | $111.1 \pm 0.3$ | n/a | n/a | $127 \pm 1$ | n/a | n/a |
| | OccPOI-Extended[*] | $112 \pm 1$ | $104 \pm 1$ | $11 \pm 1$ | $127 \pm 1$ | $1.9 \pm 0.2$ | $1.11 \pm 0.02$ |
| | GradVis (ZaidNet) | $119 \pm 3$ | n/a | $113 \pm 2$ | $128.0 \pm 0.5$ | n/a | n/a |
| | Saliency (ZaidNet) | $119 \pm 3$ | n/a | $113 \pm 2$ | $128.0 \pm 0.5$ | n/a | n/a |
| | Input $*$ Grad (ZaidNet) | $119 \pm 2$ | n/a | $113 \pm 2$ | $128.0 \pm 0.5$ | n/a | n/a |
| | 1-Occlusion (ZaidNet) | $119 \pm 2$ | n/a | $113 \pm 2$ | $128.0 \pm 0.5$ | n/a | n/a |
| | $2^{\text{nd}}$-order 1-Occlusion (ZaidNet) | $120 \pm 2$ | n/a | $113 \pm 1$ | $128.0 \pm 0.5$ | n/a | n/a |
| | OccPOI (ZaidNet) | $111 \pm 1$ | n/a | $20 \pm 9$ | $127 \pm 1$ | n/a | n/a |
| | OccPOI-Extended[*] (ZaidNet) | $112.8 \pm 0.8$ | n/a | $20 \pm 10$ | $127 \pm 1$ | n/a | n/a |
| | GradVis (WoutersNet) | $119.2 \pm 0.9$ | n/a | $112 \pm 7$ | $\underline{128.1 \pm 0.6}$ | n/a | n/a |
| | Saliency (WoutersNet) | $119.3 \pm 0.9$ | n/a | $112 \pm 7$ | $\underline{128.1 \pm 0.5}$ | n/a | n/a |
| | Input $*$ Grad (WoutersNet) | $119.3 \pm 0.9$ | n/a | $113 \pm 2$ | $\underline{128.1 \pm 0.6}$ | n/a | n/a |
| | 1-Occlusion (WoutersNet) | $119.3 \pm 0.9$ | n/a | $113 \pm 2$ | $\underline{128.1 \pm 0.6}$ | n/a | n/a |
| | $2^{\text{nd}}$-order 1-Occlusion (WoutersNet) | $119.7 \pm 0.9$ | n/a | $117.1 \pm 0.8$ | $128.1 \pm 0.6$ | n/a | n/a |
| | OccPOI (WoutersNet) | $112 \pm 1$ | n/a | $17 \pm 11$ | $127 \pm 2$ | n/a | n/a |
| | OccPOI-Extended[*] (WoutersNet) | $112 \pm 1$ | n/a | $16 \pm 2$ | $127 \pm 1$ | n/a | n/a |
| | ALL (ours) | $\boxed{\underline{125.5 \pm 0.4}}$ | $\boxed{\underline{127.6 \pm 0.3}}$ | $\underline{124.5 \pm 0.8}$ | $\boxed{\underline{128.4 \pm 0.3}}$ | $\boxed{\underline{4.3 \pm 0.1}}$ | $\boxed{\underline{1.38 \pm 0.04}}$ |

Table 1: Performance of leakage localization algorithms according to the Rev-DNNO (reverse DNN occlusion) test (**larger is better**). To compute this metric, we first train a supervised DNN classifier to map emission traces to the sensitive variable. We then occlude all its inputs and incrementally un-occlude them *from least- to most-leaky* (the opposite order as for the Fwd-DNNO test) as estimated by the method under test, and at each step compute its performance (quantified by rank, for which lower is better) on the test dataset. The Rev-DNNO metric is given by the average value of these performance assessments (higher is better, because it indicates that claimed nonleaky features indeed had little utility to the classifier). Of the two DNN occlusion metrics, this is more sensitive to *true/false negative* leakiness measurements because the performance of the classifier tends to jump and stay up as soon as it sees leaky measurements. Best result is $\boxed{\text{boxed}}$ and best deep learning result is $\underline{\text{underlined}}$. Results are reported as mean ± std. dev. over 5 random seeds. Observe that *our method is superior or comparable to all deep learning methods on all datasets*, and is *superior or comparable to all parametric methods* on every dataset except DPAv4. Additionally, deep learning methods perform better in comparison to parametric methods on the first-ordered datasets relative to their performance under the oSNR metric. [*]The algorithm as proposed in (Yap, 2025) is very time-consuming because it requires $O(T^2)$ non-parallelizable passes through the dataset. Due to time constraints, for ASCADv1 (random) and AES-HD we stop it early after 10× the runtime of ALL. [†]For each dataset, we use $m$ in `np.arange(1, 51, 2)` which maximizes the oSNR metric for $m$-occlusion. Specifically, we use $m = 3$ for ASCADv1 (fixed), $m = 7$ for ASCADv1 (random), $m = 43$ for DPAv4, $m = 33$ for AES-HD, $m = 3$ for OTiAiT, and $m = 7$ for OTP.