	Method	Dataset					
		ASCADv1 (fixed)	ASCADv1 (random)	DPAv4 (Zaid version)	AES-HD	OTiAiT	OTP
	Random	111.6 ± 0.3	108 ± 5	13 ± 2	$\boxed{127 \pm 1}$	1.21 ± 0.04	1.05 ± 0.02
First-order parametric methods	SNR	117.2 ± 0.6	116.7 ± 0.7	11.4 ± 0.2	$\boxed{126 \pm 2}$	$\boxed{1.10 \pm 0.02}$	$\boxed{1.0125 \pm 0.0007}$
	SOSD	114.9 ± 0.5	105 ± 2	$\boxed{8.0 \pm 0.8}$	$\boxed{126 \pm 2}$	1.14 ± 0.03	$\boxed{1.027 \pm 0.002}$
	CPA	111.5 ± 0.4	114 ± 1	11.5 ± 0.3	$\boxed{126 \pm 2}$	1.49 ± 0.04	$\boxed{1.0125 \pm 0.0007}$
Neural net attribution	GradVis	$\boxed{\underline{107.0 \pm 0.5}}$	$\boxed{\underline{95 \pm 2}}$	12.1 ± 0.3	$\boxed{\underline{127 \pm 1}}$	1.4 ± 0.2	1.0142 ± 0.0008
	Saliency	$\boxed{\underline{107.1 \pm 0.5}}$	$\boxed{\underline{95 \pm 2}}$	11.8 ± 0.3	$\boxed{\underline{127 \pm 1}}$	1.39 ± 0.04	1.014 ± 0.001
	Input ∗ Grad	$\boxed{\underline{107.2 \pm 0.5}}$	$\boxed{\underline{95 \pm 2}}$	11.8 ± 0.4	$\boxed{\underline{127 \pm 1}}$	1.36 ± 0.04	1.0141 ± 0.0009
	LRP	$\boxed{\underline{107.2 \pm 0.5}}$	$\boxed{\underline{95 \pm 2}}$	11.8 ± 0.4	$\boxed{\underline{127 \pm 1}}$	1.36 ± 0.04	1.0141 ± 0.0009
	1-Occlusion	$\boxed{\underline{107.1 \pm 0.5}}$	$\boxed{\underline{95 \pm 2}}$	10.1 ± 0.2	$\boxed{\underline{127 \pm 1}}$	1.36 ± 0.04	1.0141 ± 0.0009
	5-Occlusion	$\boxed{\underline{107.4 \pm 0.4}}$	$\boxed{\underline{94 \pm 2}}$	9.6 ± 0.2	$\boxed{\underline{127 \pm 1}}$	1.43 ± 0.03	$\underline{1.013 \pm 0.002}$
	17-Occlusion	108.7 ± 0.4	$\boxed{\underline{96 \pm 2}}$	9.5 ± 0.2	$\boxed{\underline{127 \pm 1}}$	1.51 ± 0.02	1.021 ± 0.002
	65-Occlusion	111.6 ± 0.7	99 ± 2	10.1 ± 0.2	$\boxed{\underline{127 \pm 1}}$	1.60 ± 0.01	1.026 ± 0.007
	257-Occlusion	118.0 ± 0.8	104 ± 1	10.1 ± 0.2	$\boxed{\underline{127 \pm 1}}$	1.7 ± 0.2	$\underline{1.031 \pm 0.006}$
	2^{nd}-order 1-Occlusion	$\boxed{\underline{107.0 \pm 0.4}}$	$\boxed{\underline{95 \pm 2}}$	10.0 ± 0.2	$\boxed{\underline{127 \pm 1}}$	1.34 ± 0.04	1.0138 ± 0.0008
	OccPOI	TODO	TODO	TODO	TODO	TODO	TODO
	GradVis (ZaidNet)	108.8 ± 0.8	n/a	$\underline{9.3 \pm 0.2}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	Saliency (ZaidNet)	108.8 ± 0.8	n/a	$\underline{9.3 \pm 0.2}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	Input ∗ Grad (ZaidNet)	109.0 ± 0.6	n/a	$\underline{9.2 \pm 0.2}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	1-Occlusion (ZaidNet)	109.3 ± 0.6	n/a	$\underline{9.2 \pm 0.2}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	5-Occlusion (ZaidNet)	109 ± 1	n/a	10.2 ± 0.3	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	17-Occlusion (ZaidNet)	111 ± 1	n/a	9.9 ± 0.3	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	65-Occlusion (ZaidNet)	113 ± 1	n/a	9.9 ± 0.4	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	257-Occlusion (ZaidNet)	120 ± 2	n/a	11 ± 1	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	2^{nd}-order 1-Occlusion (ZaidNet)	108.8 ± 0.4	n/a	9.2 ± 0.2	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	OccPOI (ZaidNet)	TODO	n/a	TODO	TODO	n/a	n/a
	GradVis (WoutersNet)	109.9 ± 0.5	n/a	9.6 ± 0.3	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	Saliency (WoutersNet)	109.8 ± 0.5	n/a	9.6 ± 0.3	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	Input ∗ Grad (WoutersNet)	109.7 ± 0.5	n/a	$\underline{9.4 \pm 0.3}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	1-Occlusion (WoutersNet)	109.7 ± 0.4	n/a	$\underline{9.4 \pm 0.3}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	5-Occlusion (WoutersNet)	110.1 ± 0.7	n/a	10.4 ± 0.4	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	17-Occlusion (WoutersNet)	111.6 ± 0.5	n/a	10.0 ± 0.5	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	65-Occlusion (WoutersNet)	114.1 ± 0.7	n/a	10.0 ± 0.3	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	257-Occlusion (WoutersNet)	118 ± 1	n/a	11.0 ± 0.2	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	2^{nd}-order 1-Occlusion (WoutersNet)	109.2 ± 0.2	n/a	$\underline{9.4 \pm 0.3}$	$\boxed{\underline{126 \pm 2}}$	n/a	n/a
	OccPOI (WoutersNet)	TODO	n/a	TODO	TODO	n/a	n/a
	ALL (ours)	$\boxed{\underline{107.5 \pm 0.3}}$	101 ± 2	12.2 ± 0.4	$\boxed{\underline{126 \pm 2}}$	$\underline{1.23 \pm 0.03}$	1.0161 ± 0.0009

Table 1: Performance of leakage localization algorithms according to the Fwd-DNNO (forward DNN occlusion) test (smaller 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 most- to least-leaky as estimated by the leakiness assessment under test, and at each step compute its performance (quantified by rank, lower is better) on the test dataset. The Fwd-DNNO metric is given by the average value of these performance assessments (lower is better, because it indicates that claimed leaky features indeed had utility to the classifier). Of the two DNN occlusion metrics, this is more sensitive to *true/false positive* 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. This metric appears to have high variance and little discriminative power compared to the oSNR and Rev-DNNO metrics (as indicated by the large number of tied 'best' methods), and there is no clear best method according to Fwd-DNNO.