	Method	Dataset					
		ASCADv1 (fixed)	ASCADv1 (random)	DPAv4 (Zaid version)	AES-HD	OTiAiT	OTP
	Random	112 ± 1	108 ± 4	11.5 ± 0.4	127 ± 1	1.20 ± 0.05	1.048 ± 0.008
First-order parametric methods	SNR	111.0 ± 0.2	123 ± 2	$\boxed{126 \pm 1}$	$\boxed{128.5 \pm 0.3}$	4.26 ± 0.07	1.33 ± 0.04
	SOSD	111.6 ± 0.2	125.6 ± 0.6	105.7 ± 0.9	$\boxed{128.3 \pm 0.3}$	3.94 ± 0.08	$\boxed{1.34 \pm 0.04}$
	CPA	118.2 ± 0.4	114 ± 2	111.5 ± 0.9	$\boxed{128.4 \pm 0.3}$	2.7 ± 0.2	1.32 ± 0.04
Neural net attribution	GradVis	124.9 ± 0.3	127 ± 1	121 ± 1	127 ± 1	1.8 ± 0.2	1.31 ± 0.05
	Saliency	124.8 ± 0.3	127 ± 1	$\underline{124 \pm 1}$	127 ± 1	2.8 ± 0.2	1.29 ± 0.05
	Input $*$ Grad	124.8 ± 0.3	127 ± 1	$\underline{124 \pm 1}$	127 ± 1	3.1 ± 0.2	1.29 ± 0.05
	LRP	124.8 ± 0.3	127 ± 1	$\underline{124 \pm 1}$	127 ± 1	3.1 ± 0.2	1.29 ± 0.05
	1-Occlusion	124.8 ± 0.3	127 ± 1	$\underline{124 \pm 1}$	127 ± 1	3.2 ± 0.2	1.29 ± 0.05
	5-Occlusion	125.0 ± 0.4	127.1 ± 0.6	$\underline{124.9 \pm 0.8}$	128 ± 1	$\boxed{\underline{4.4 \pm 0.1}}$	1.27 ± 0.03
	17-Occlusion	124.5 ± 0.2	127.1 ± 0.7	123.1 ± 0.7	128 ± 1	$\boxed{\underline{4.5 \pm 0.1}}$	1.23 ± 0.02
	65-Occlusion	122.4 ± 0.3	126 ± 2	119.5 ± 0.8	127 ± 2	4.1 ± 0.1	1.17 ± 0.02
	257-Occlusion	121.4 ± 0.6	122 ± 3	104.4 ± 0.6	127 ± 2	3.9 ± 0.1	$1/16 \pm 0.02$
	2^{nd}-order 1-Occlusion	124.8 ± 0.3	127 ± 1	$\underline{125 \pm 1}$	127 ± 1	3.6 ± 0.2	1.29 ± 0.05
	OccPOI	TODO	TODO	TODO	TODO	TODO	TODO
	GradVis (ZaidNet)	119 ± 3	n/a	113 ± 2	128.0 ± 0.5	n/a	n/a
	Saliency (ZaidNet)	119 ± 3	n/a	113 ± 2	128.0 ± 0.5	n/a	n/a
	Input $*$ Grad (ZaidNet)	119 ± 2	n/a	113 ± 2	128.0 ± 0.5	n/a	n/a
	1-Occlusion (ZaidNet)	119 ± 2	n/a	113 ± 2	128.0 ± 0.5	n/a	n/a
	5-Occlusion (ZaidNet)	121 ± 2	n/a	120 ± 1	$\boxed{\underline{128.3 \pm 0.4}}$	n/a	n/a
	17-Occlusion (ZaidNet)	121 ± 2	n/a	119 ± 2	$\boxed{\underline{128.3 \pm 0.3}}$	n/a	n/a
	65-Occlusion (ZaidNet)	120 ± 1	n/a	114 ± 4	$\boxed{\underline{128.3 \pm 0.4}}$	n/a	n/a
	257-Occlusion (ZaidNet)	122 ± 1	n/a	101 ± 2	$\boxed{\underline{128.4 \pm 0.4}}$	n/a	n/a
	2^{nd}-order 1-Occlusion (ZaidNet)	120 ± 2	n/a	113 ± 2	128.0 ± 0.5	n/a	n/a
	OccPOI (ZaidNet)	TODO	n/a	TODO	TODO	n/a	n/a
	GradVis (WoutersNet)	119.2 ± 0.9	n/a	112 ± 7	$\underline{128.1 \pm 0.6}$	n/a	n/a
	Saliency (WoutersNet)	119.3 ± 0.9	n/a	112 ± 7	$\underline{128.1 \pm 0.5}$	n/a	n/a
	Input $*$ Grad (WoutersNet)	119.3 ± 0.9	n/a	112 ± 7	$\underline{128.1 \pm 0.6}$	n/a	n/a
	1-Occlusion (WoutersNet)	119.3 ± 0.9	n/a	112 ± 7	$\underline{128.1 \pm 0.6}$	n/a	n/a
	5-Occlusion (WoutersNet)	120 ± 1	n/a	119 ± 3	$\boxed{\underline{128.3 \pm 0.4}}$	n/a	n/a
	17-Occlusion (WoutersNet)	121 ± 1	n/a	120 ± 2	$\boxed{\underline{128.4 \pm 0.3}}$	n/a	n/a
	65-Occlusion (WoutersNet)	121.4 ± 0.7	n/a	114 ± 3	$\boxed{\underline{128.4 \pm 0.3}}$	n/a	n/a
	257-Occlusion (WoutersNet)	122.0 ± 0.7	n/a	104 ± 4	$\boxed{\underline{128.4 \pm 0.3}}$	n/a	n/a
	2^{nd}-order 1-Occlusion (WoutersNet)	119.7 ± 0.9	n/a	112 ± 7	128.1 ± 0.6	n/a	n/a
	OccPOI (WoutersNet)	TODO	n/a	TODO	TODO	n/a	n/a
	ALL (ours)	$\boxed{\underline{125.5 \pm 0.4}}$	$\boxed{\underline{127.6 \pm 0.3}}$	$\underline{124.5 \pm 0.8}$	$\boxed{\underline{128.4 \pm 0.3}}$	4.3 ± 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 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 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 $\pm$ std. dev. over 5 random seeds. Observe that *our method is superior or comparable to all deep learning methods on 5 of the 6 datasets* and slightly worse on the remaining dataset. Additionally, deep learning methods perform better in comparison to parametric methods on the first-ordered datasets relative to their performance under the oSNR metric.