

Classification error (%) on the first 1000 test samples

model	mnist				fashion_mnist			
	original	attack	white-box	gray-box	original	attack	white-box	gray-box
UNENCRYPTED	A	1.49	CW I_2	100.00	8.30	CW I_2	100.00	
			CW I_0	100.00		CW I_0	100.00	
			CW I_∞	100.00		CW I_∞	100.00	
	B	2.10	FGSM	39.50	9.50	FGSM	77.20	
PERMUTATED	A	3.70	CW I_2	100.00	12.30	CW I_2	100.00	12.70
			CW I_0	100.00		CW I_0	100.00	12.50
			CW I_∞	100.00		CW I_∞	100.00	12.90
	B	4.20	FGSM	8.60	12.00	FGSM	29.80	
AES · ECB	A	18.40	CW I_2	future work		CW I_2	irrelevant	irrelevant
	B	19.30	FGSM	future work		FGSM	irrelevant	irrelevant
AES · CBC	A	67.60	CW I_2	irrelevant		CW I_2	irrelevant	irrelevant
	B	87.40	FGSM	irrelevant		FGSM	irrelevant	irrelevant
AES · CTR	A	3.70	CW I_2		17.40	CW I_2		17.20
	B	2.70	FGSM		16.70	FGSM		26.50

1

Classification error (%) on the first 1000 test samples

		mnist			fashion_mnist		
		original	attacked		original	attacked	
			white-box	gray-box		white-box	gray-box
NENCRYPTED	CW I_2	1.49	100.00		8.30	100.00	
	CW I_0		100.00			100.00	
	CW I_∞		100.00			100.00	
	FGSM	2.10	82.94		9.50	94.25	
ERMUTATE	CW I_2	3.70	100.00	4.50	12.30	100.00	12.70
	CW I_0		100.00	7.30		100.00	12.50
	CW I_∞		100.00	5.40		100.00	12.90
	FGSM	4.19	81.40	53.40	12.00	89.70	72.60
AES · ECB	CW I_2	18.40	future work	future work		irrelevant	irrelevant
	FGSM	19.30					
AES · CBC	CW I_2	67.60	irrelevant	irrelevant		irrelevant	irrelevant
	FGSM	87.40					
AES · CTR	CW I_2	3.70		4.20			
	FGSM	2.70		31.10			