Secure and Trustworthy Hardware and Machine Learning Systems for Internet of Things

by

SHAYAN TAHERI
M.S. Utah State University, 2015

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in the Department of Electrical and Computer Engineering in the School of Electrical Engineering and Computer Science
in the College of Engineering and Computer Science at the University of Central Florida
Orlando, Florida
Fall 2019
Major Professor: Jiann-Shiun Yuan

Overview

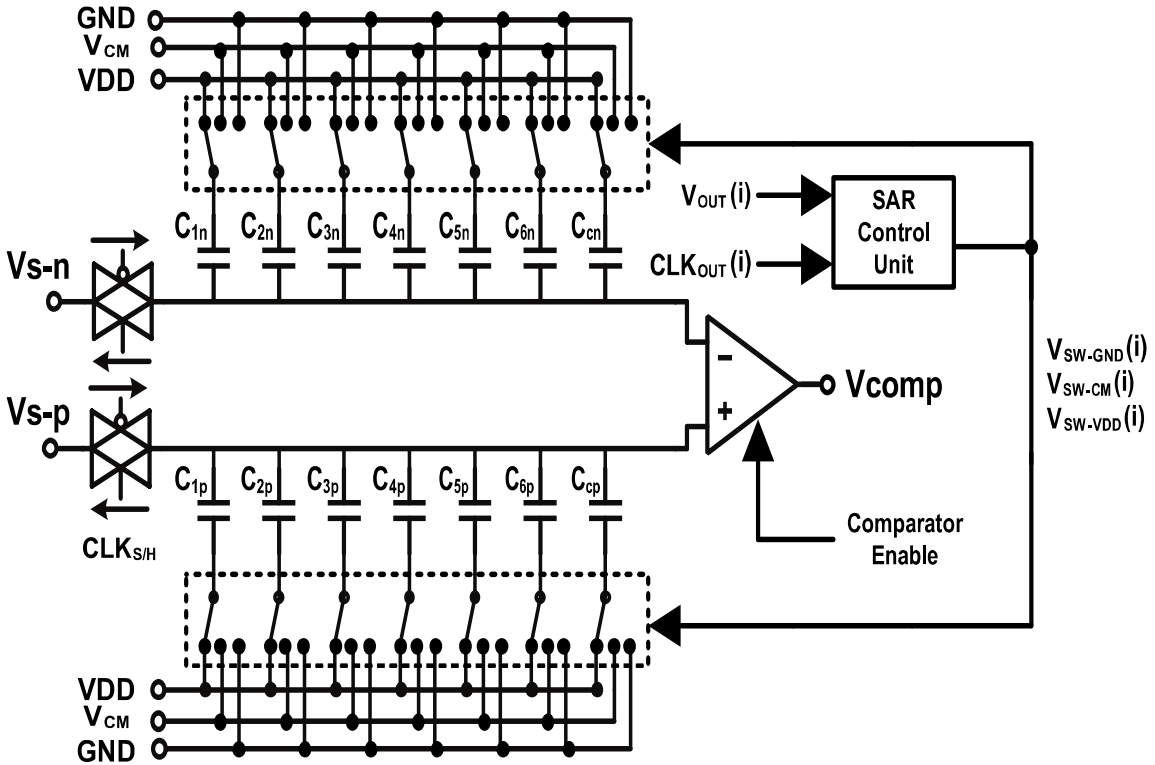
- Introduction: Cybersecurity – Hardware Security (HS) and Machine Learning Security (MLS)
- HS: Security of Analog to Digital Converter
- MLS: Leveraging Transfer Learning and Data Transformation in Botnet Detection
- MLS: Design of Privacy-Preserving and Secure Biometric Recognition System
- MLS: Detection of Adversarial Examples Using Adversarial Training and Noise Training
-
- Future Work -- MLS: Leveraging CycleGAN for making neural network robust
- Conclusion

Introduction

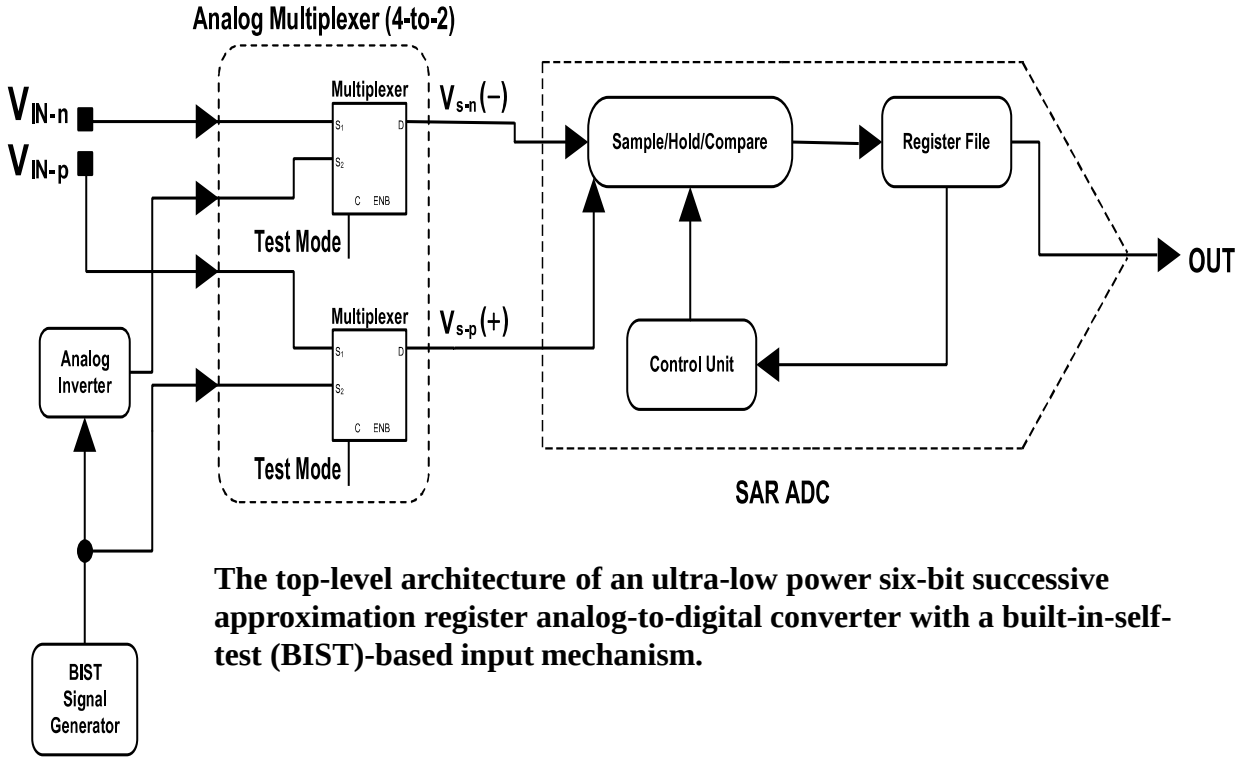
- **Hardware Security**
  - Hardware Trojan
- **Machine Learning Security**
  - Malware/Botnet Network Traffic Data
  - Adversarial Examples
  - Biometric Recognition System

HS: Security of Analog to Digital Converter

- ***This is the first attempt in the literature on the ADC security*! More than 98% of the work completed by me.**
- **Successive approximation register (SAR) analog to digital converter (ADC): Architecture**



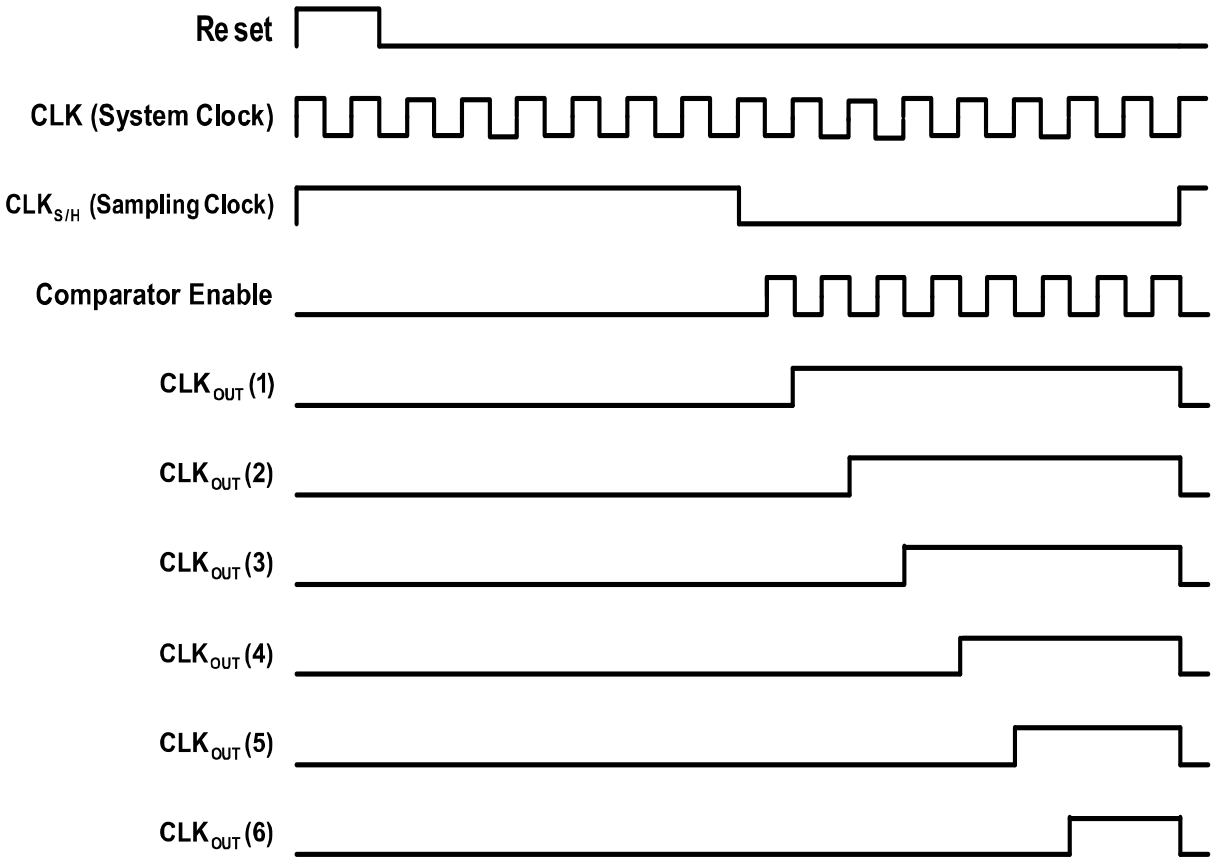
The sample/hold/compare block circuit.



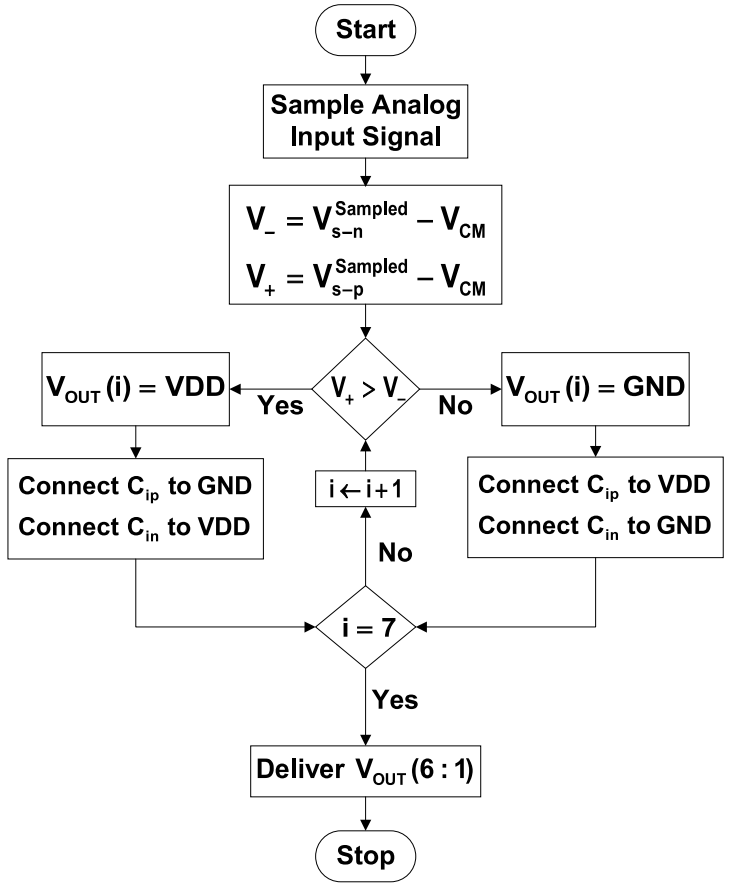
The top-level architecture of an ultra-low power six-bit successive approximation register analog-to-digital converter with a built-in-self-test (BIST)-based input mechanism.

HS: Security of Analog to Digital Converter

- **Successive approximation register (SAR) analog to digital converter (ADC): Operation**



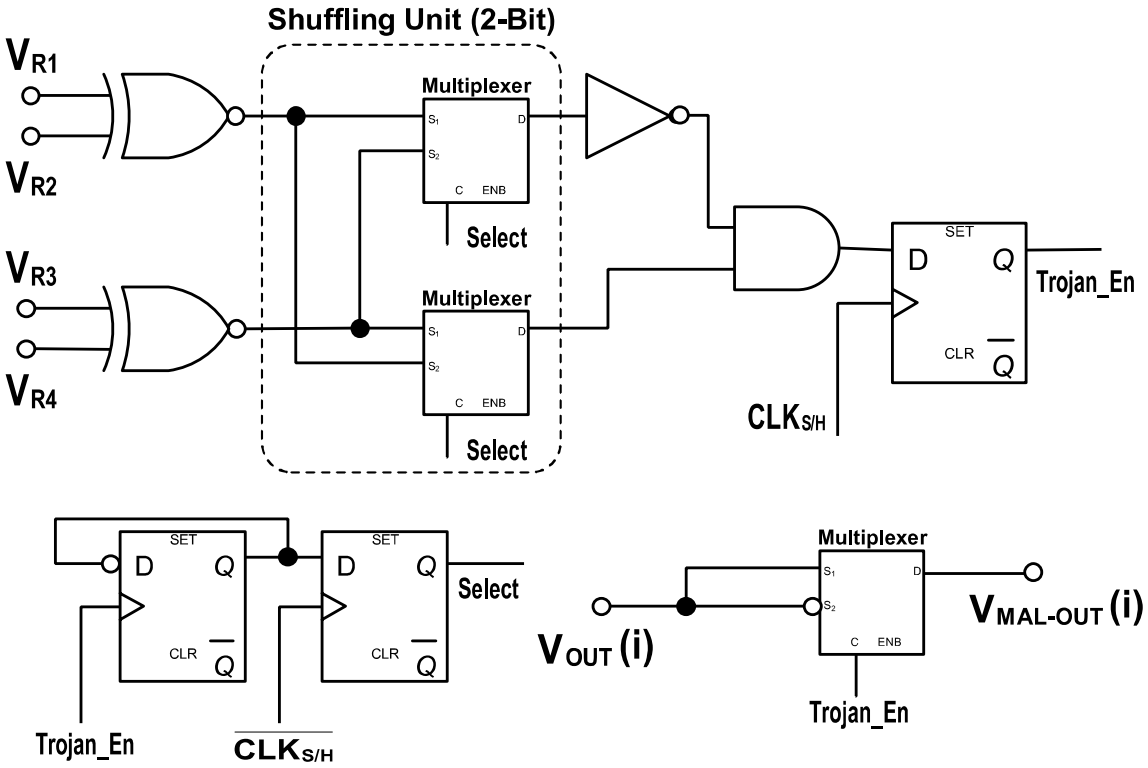
The timing diagram of the ADC operation.



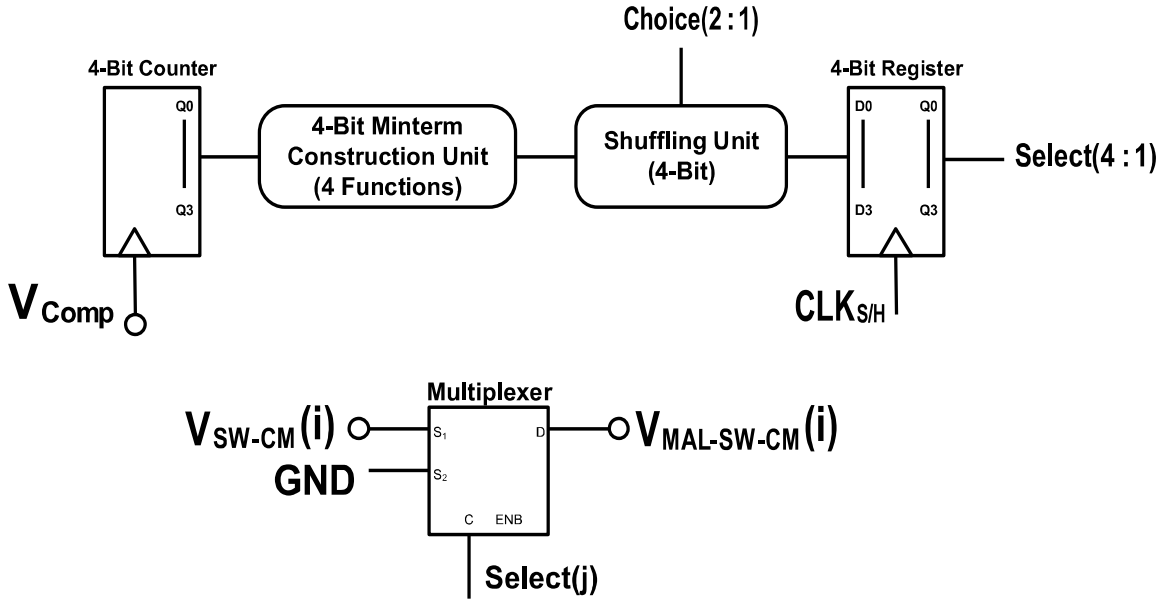
The operation flow of the original analog to digital converter (ADC) circuit.

HS: Security of Analog to Digital Converter

- **Successive approximation register (SAR) analog to digital converter (ADC): Threat Models**



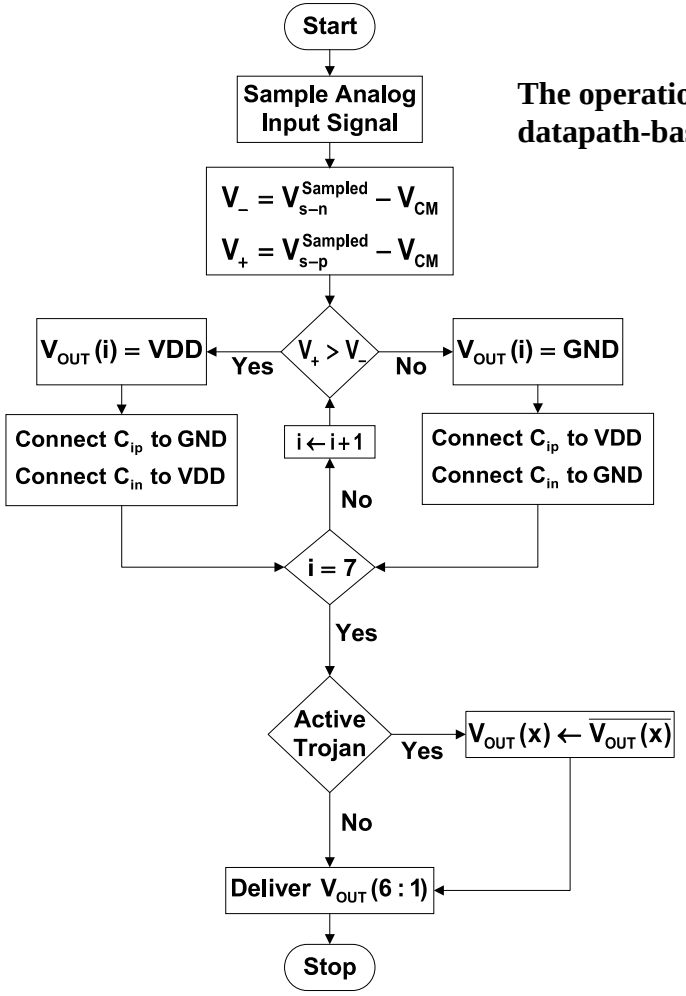
The Trojan circuit for the datapath-based threat.



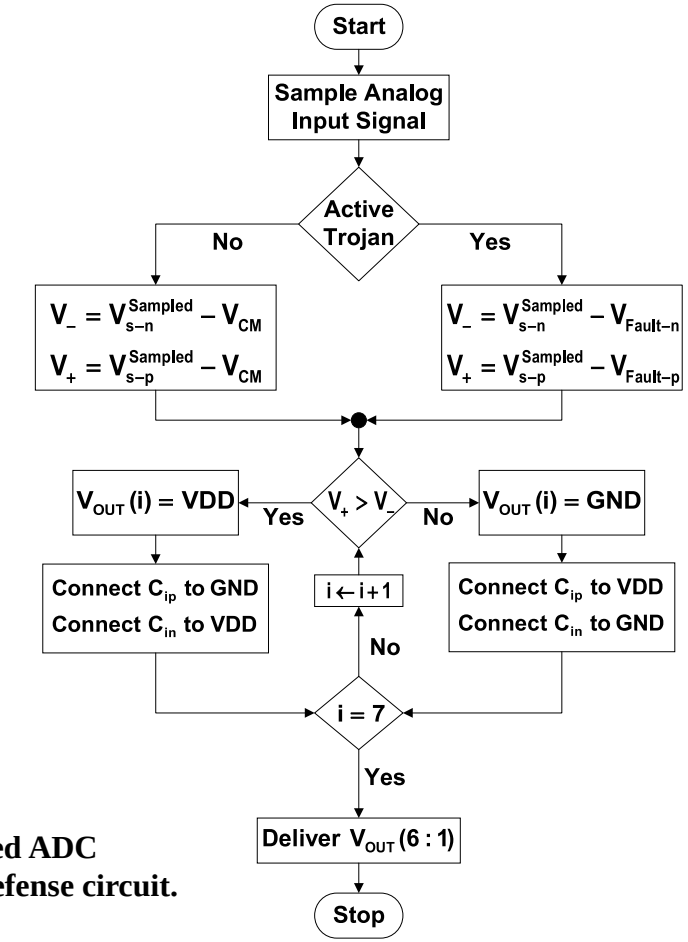
The Trojan circuit for the control-based threat.

HS: Security of Analog to Digital Converter

- **Successive approximation register (SAR) analog to digital converter (ADC): Operations of Threat Models**



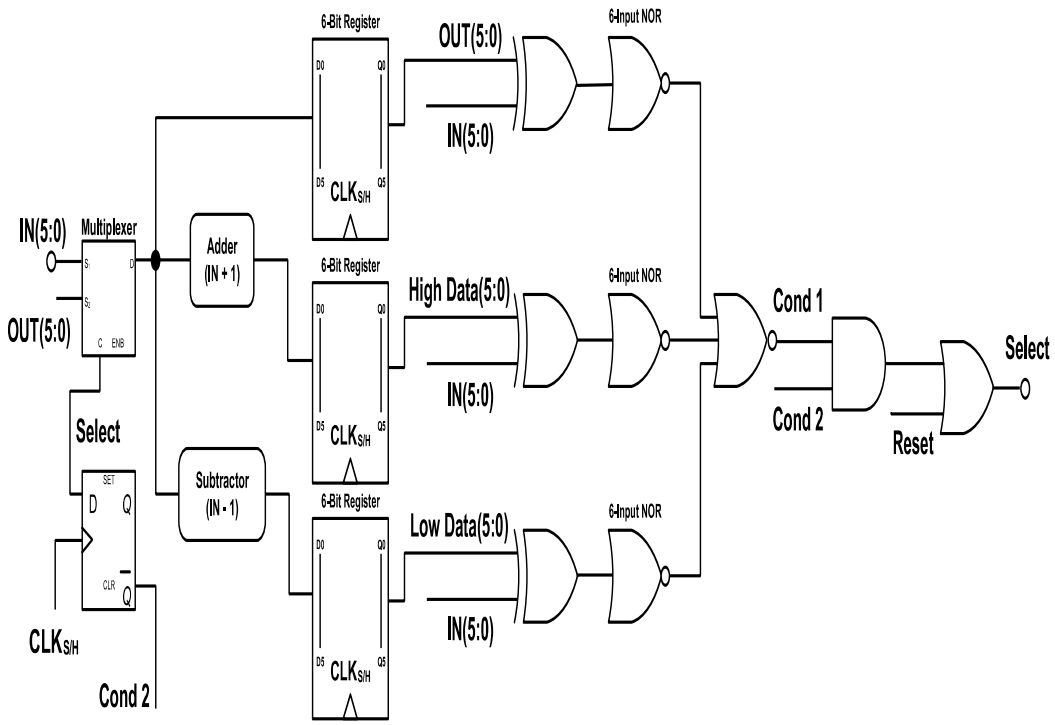
The operation flow of the infected ADC circuit by the datapath-based Hardware Trojan.



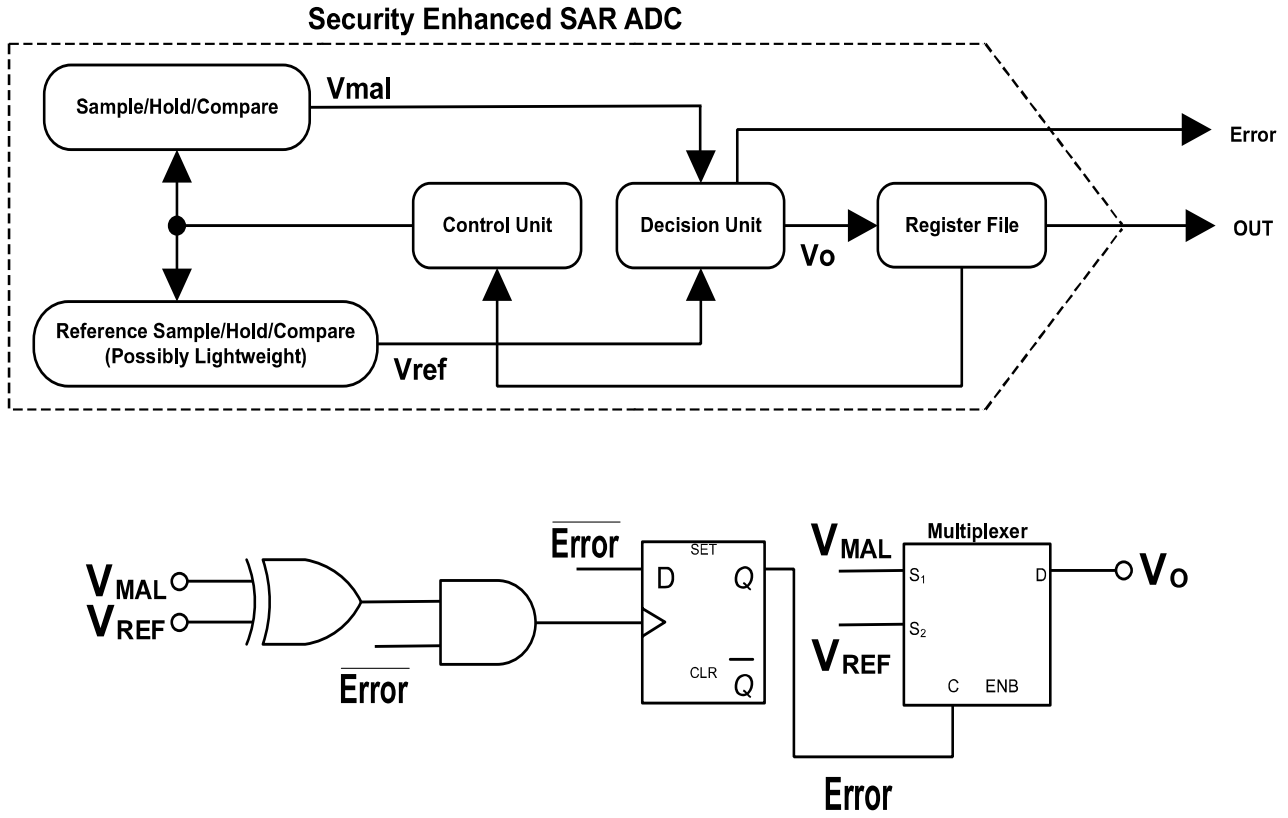
The operation flow of the secured ADC circuit by the datapath-based defense circuit.

HS: Security of Analog to Digital Converter

- **Successive approximation register (SAR) analog to digital converter (ADC): Defense Models**



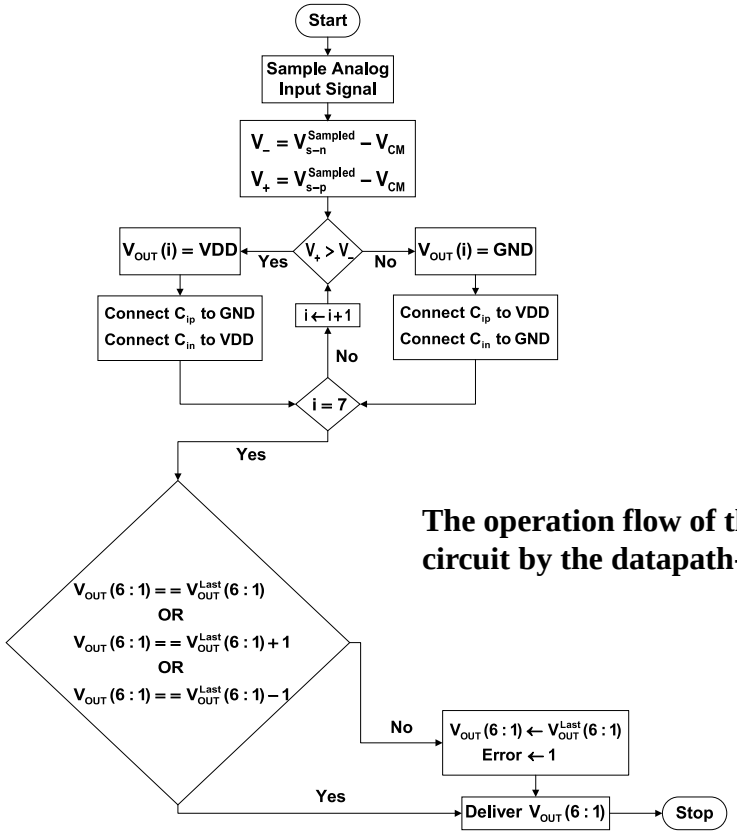
The defense circuit for the datapath-based countermeasure.



The defense circuit for the control-based countermeasure.

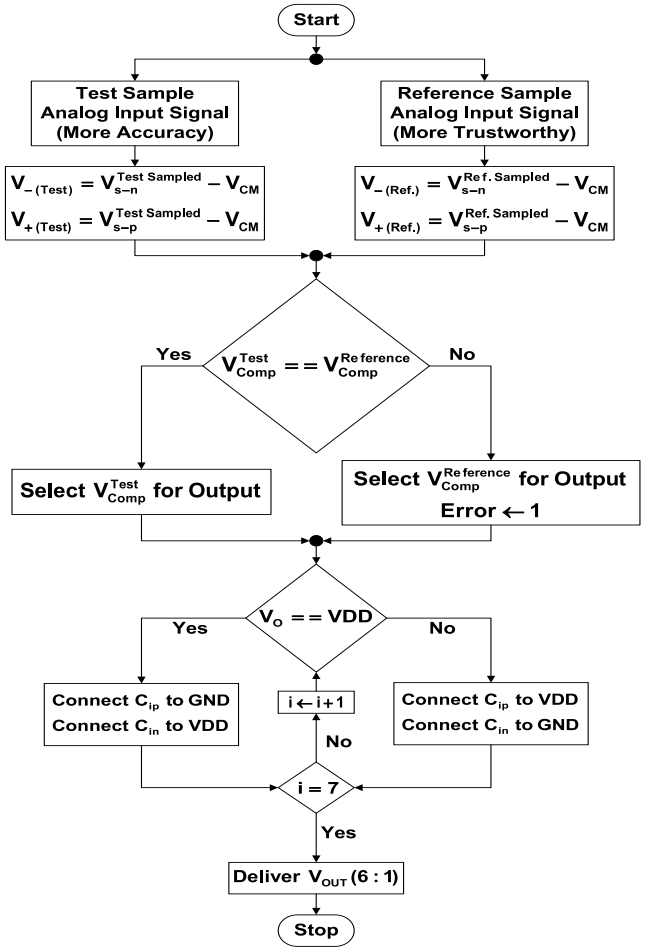
HS: Security of Analog to Digital Converter

- **Successive approximation register (SAR) analog to digital converter (ADC): Operations of Defense Models**



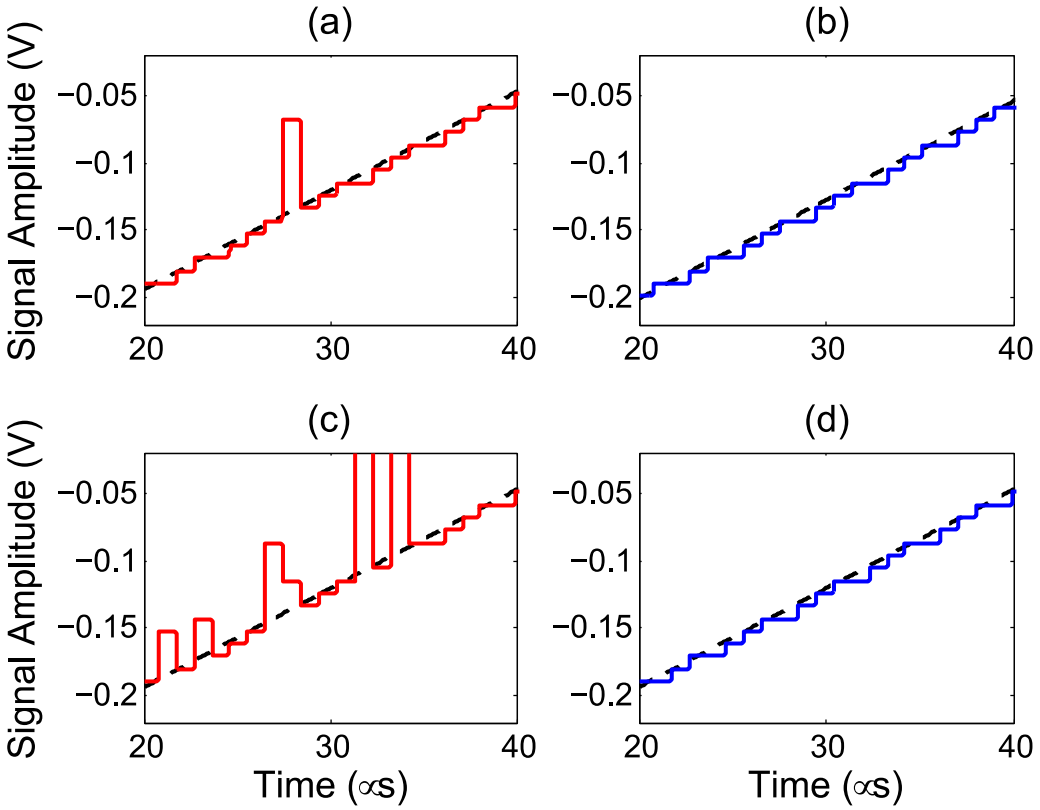
The operation flow of the secured ADC circuit by the datapath-based defense circuit.

The operation flow of the infected ADC circuit by the control-based Hardware Trojan.

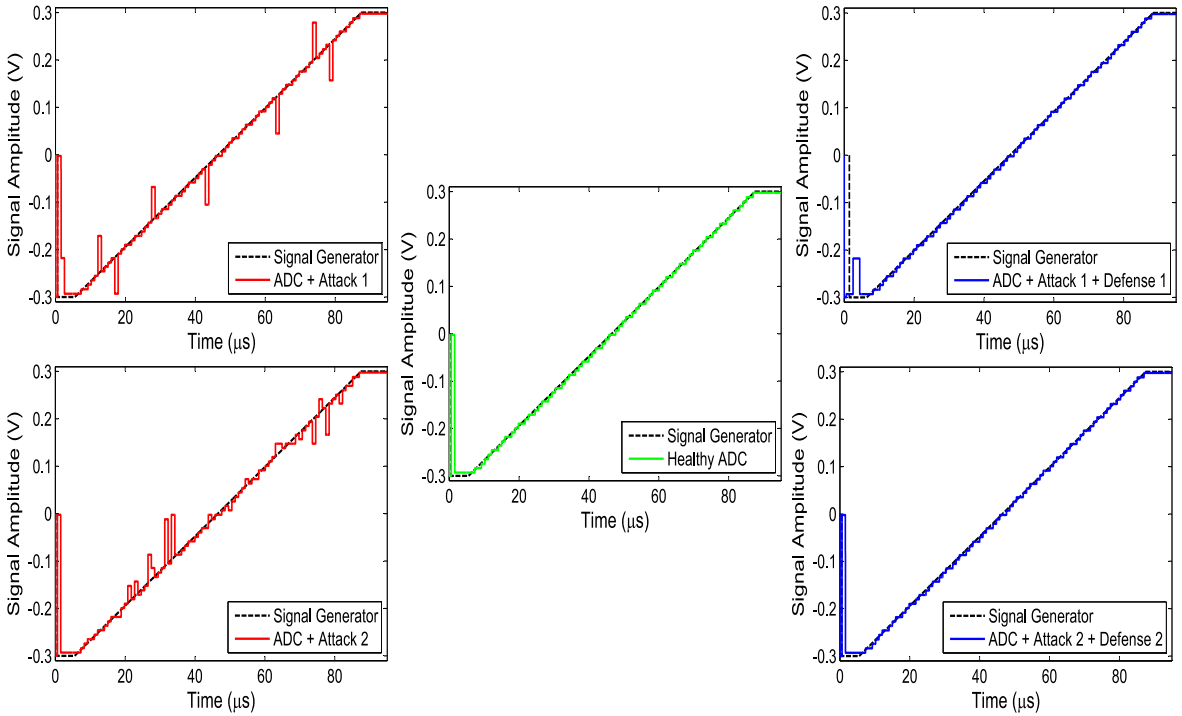


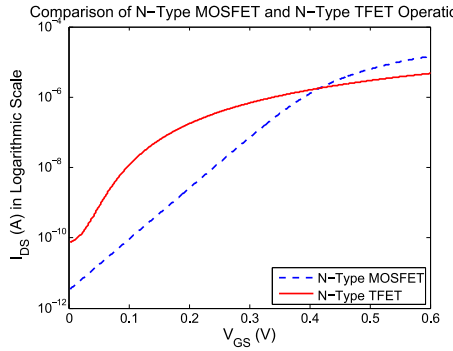
HS: Security of Analog to Digital Converter

- **Behavior of Normal and Infected SAR ADC using Tunnel Field-Effect Transistor (TFET)**



The ADC functionality evaluation in five different operating conditions.





HS: Security of Analog to Digital Converter

- **The performance evaluation of the normal and the infected SAR ADC:**

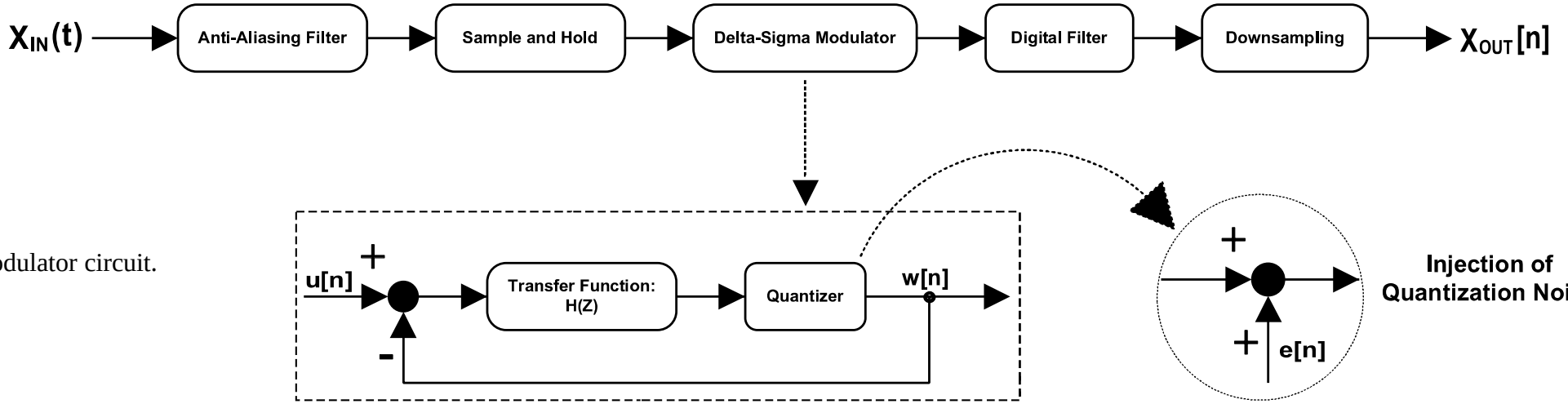
	Total Harmonic Distortion (%)	Effective Number of Bits	RMS Power Consumption for Signal Generator (nW)	RMS Power Consumption for Supply Voltage (μW)	Change in Total Cell Area (%)	Absolute Value of Mean of "Ref. Signal − Test Signal"	Standard Deviation of Adjacent Differences of "Ref. Signal − Test Signal"
Healthy ADC	3.57	4.49	22.21	142.8			
ADC + Attack 1	20.38	1.98	22.15	140.8	0.33	0.0011	1.4090×10^{-4}
ADC + Attack 1 + Defense 1	23.72	1.76	21.11	136.9	2.36	6.1810×10^{-5}	1.2310×10^{-4}
ADC + Attack 2	10.09	2.99	24.01	180.7	1.3	5.8099×10^{-5}	5.5301×10^{-5}
ADC + Attack 2 + Defense 2	3.57	4.49	42.6	312.8	100.23	2.5268×10^{-5}	5.3694×10^{-5}

UCF

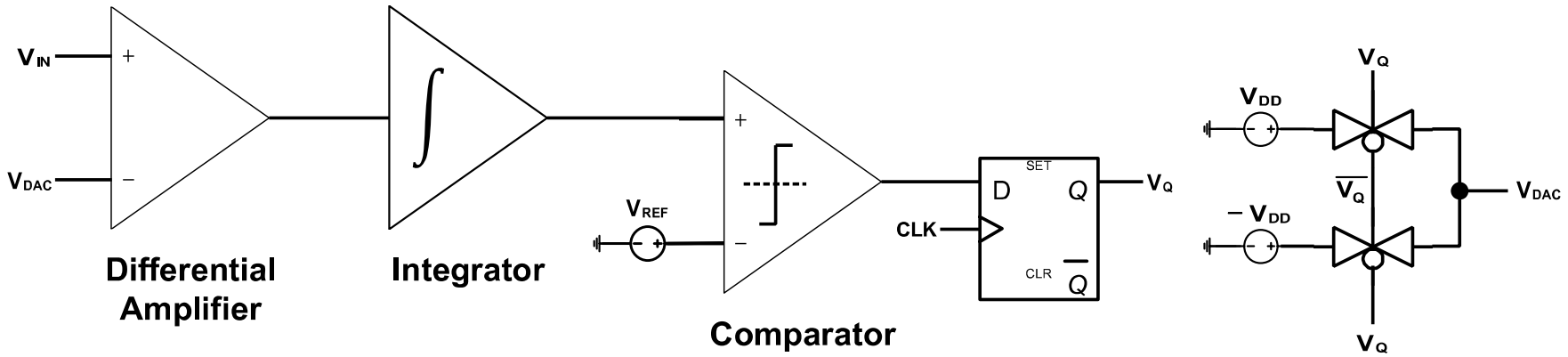
HS: Security of Analog to Digital Converter

Delta-Sigma analog to digital converter (ADC): Architecture and Operation

The delta-sigma analog to digital converter architecture along with a general model for the delta-sigma modulator.



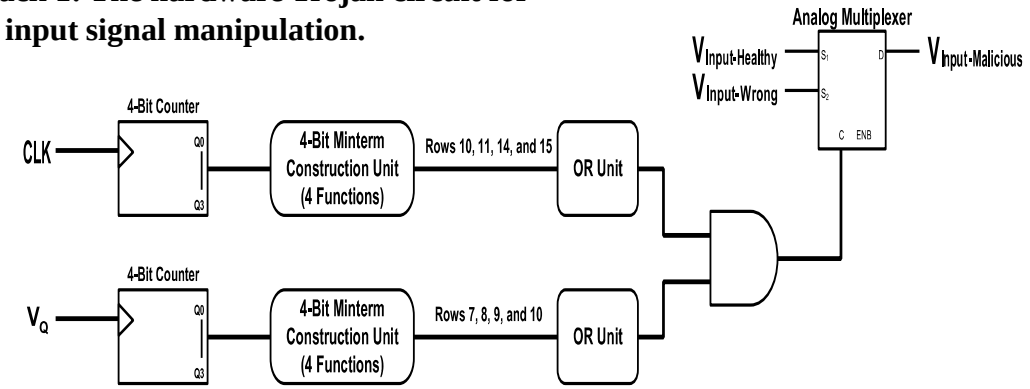
The delta-sigma modulator circuit.



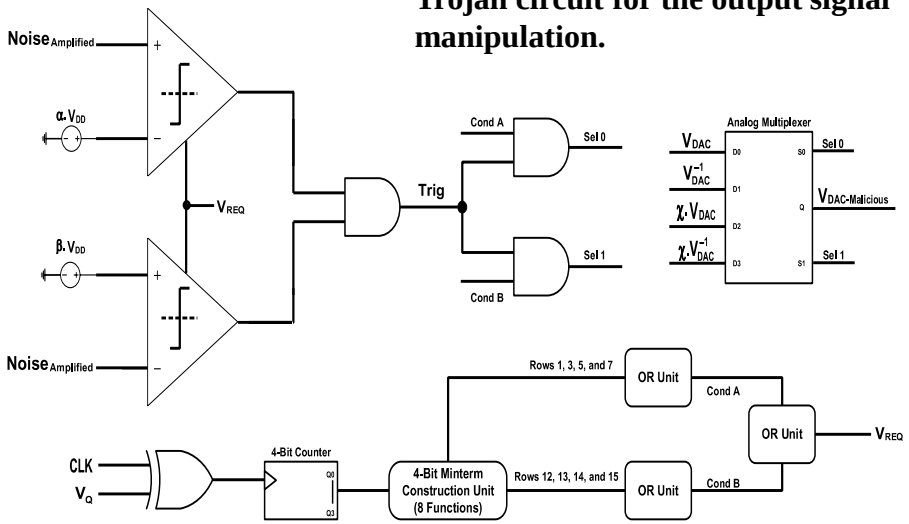
HS: Security of Analog to Digital Converter

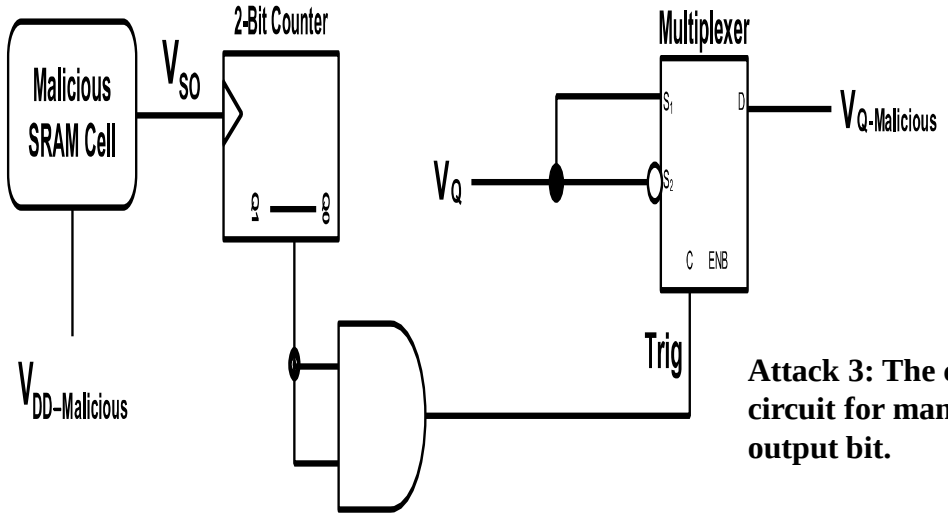
Delta-Sigma analog to digital converter: Threat Models

Attack 1: The hardware Trojan circuit for the input signal manipulation.

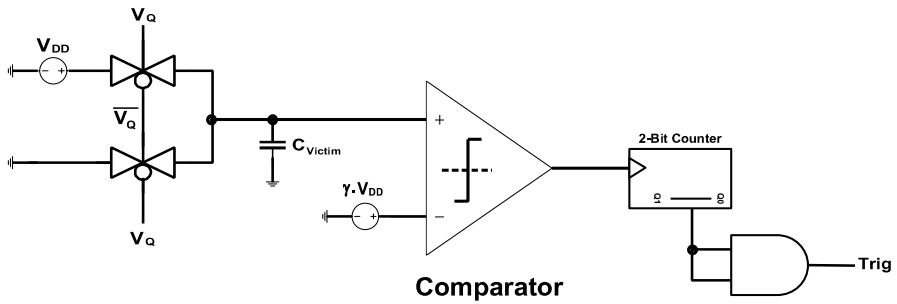


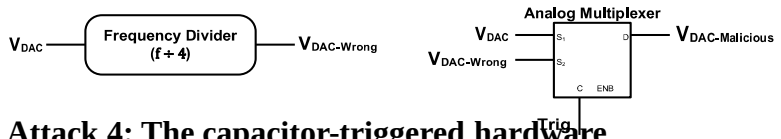
Attack 2: The noise-triggered hardware Trojan circuit for the output signal manipulation.





Attack 3: The controllable hardware Trojan circuit for manipulation of the modulator output bit.

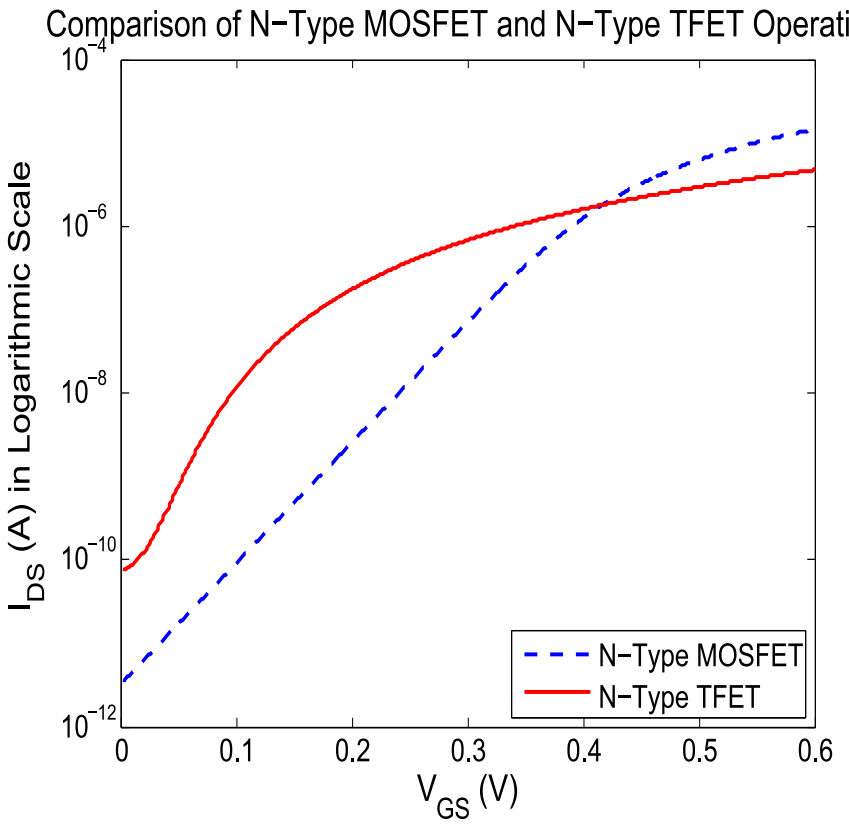


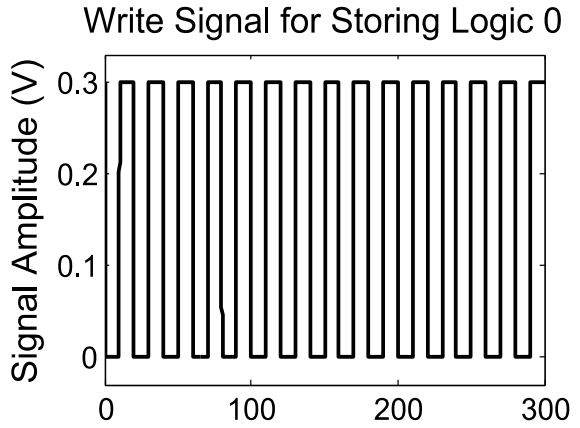


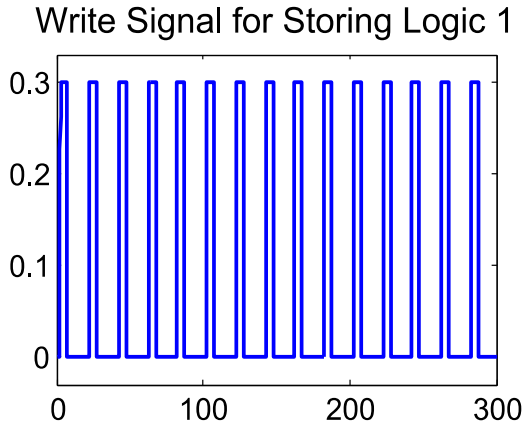
Attack 4: The capacitor-triggered hardware Trojan circuit for the output signal manipulation.

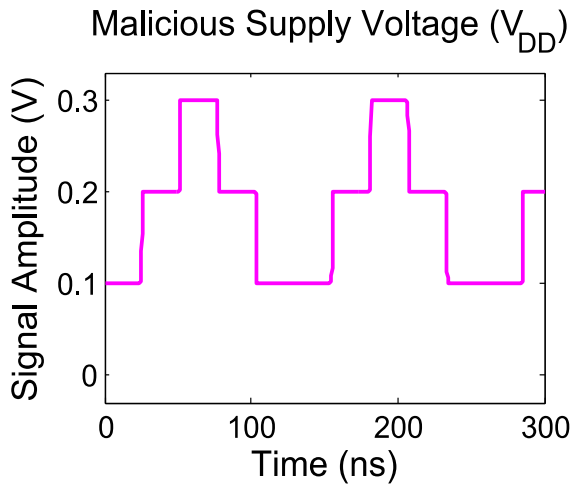
HS: Security of Analog to Digital Converter

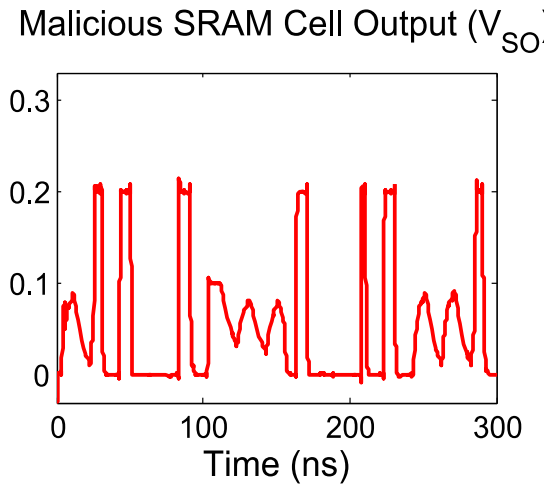
Delta-Sigma analog to digital converter: Malicious SRAM Cell Operation using TFET







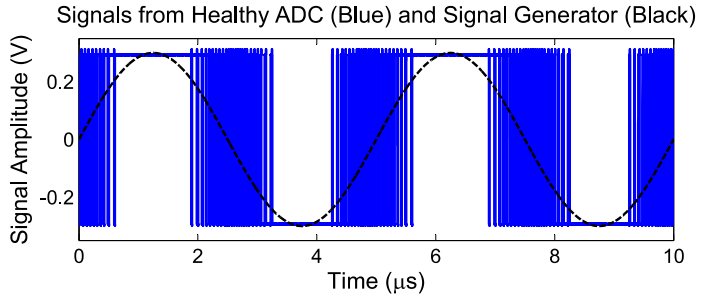


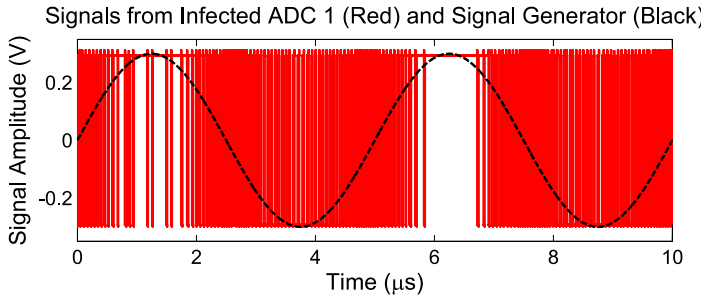


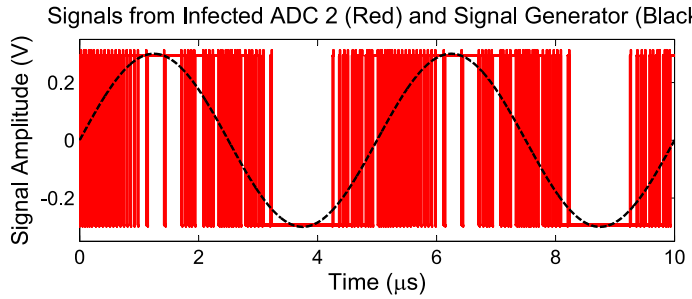
The circuit operation of malicious SRAM cell.

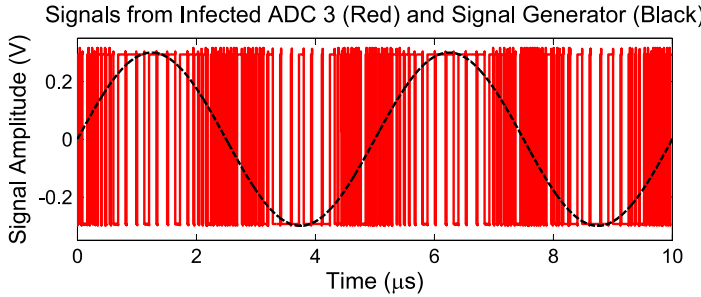
HS: Security of Analog to Digital Converter

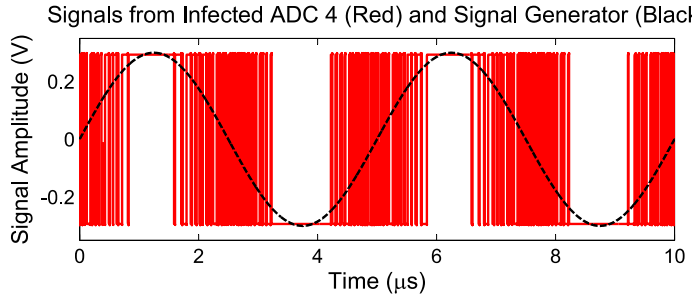
- **Behavior of Normal and Infected Delta-Sigma ADC using Tunnel Field-Effect Transistor (TFET)**

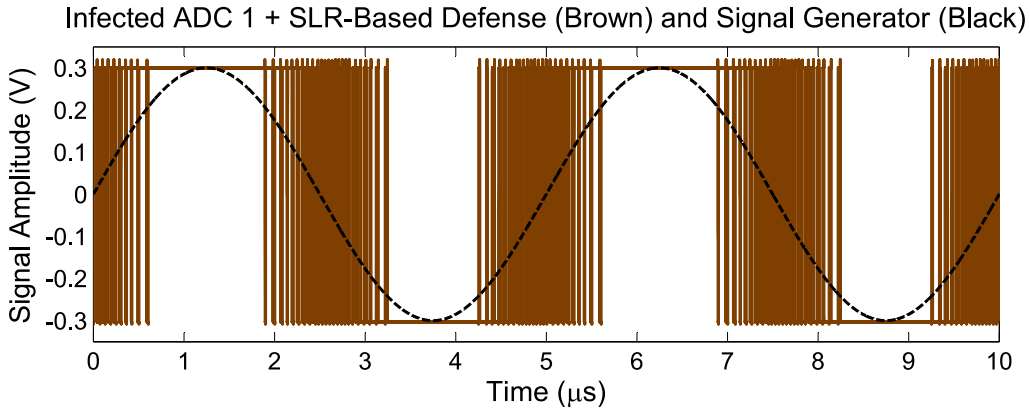


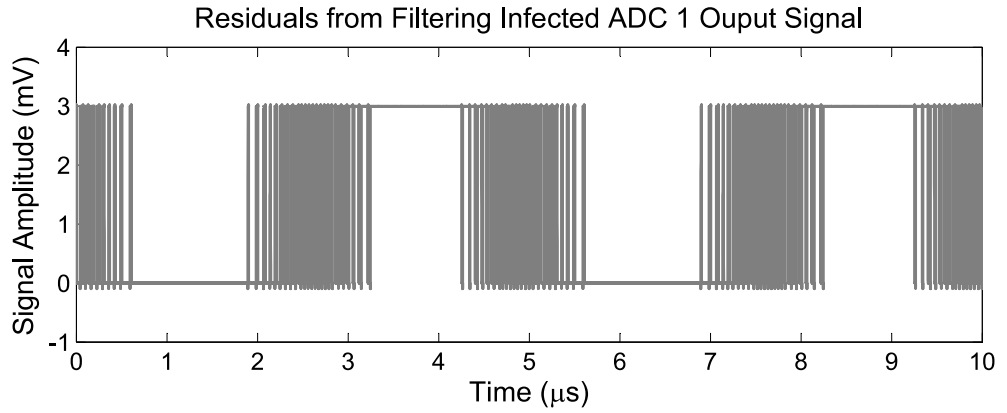












Analysis of Attacks: The delta-sigma modulator output signal in the multiple circuit structures.

HS: Security of Analog to Digital Converter

- **The performance evaluation of the normal and the infected delta-sigma ADC for Sine Waveform:**

Defense Analysis (1): The performance analysis of filters based on their corresponding modified R-squared values when a sinusoidal signal is applied to the modulator.

Type of Filter	Modified R-Squared Value for ADC Under Attack 1 (Sinusoidal Input)	Modified R-Squared Value for ADC Under Attack 2 (Sinusoidal Input)	Modified R-Squared Value for ADC Under Attack 3 (Sinusoidal Input)	Modified R-Squared Value for ADC Under Attack 4 (Sinusoidal Input)
Simple Linear Regression	1.0000	0.9987	0.9997	0.9998
Generalized Linear Regression	0.9999	0.9987	0.9996	0.9998
Moving Average	0.2293	0.0384	0.1563 (Reverse)	0.0260
Robust Regression (LAR Method)	1.0000	1.0000	1.0000	1.0000
Robust Regression (Bisquare Method)	1.0000	0.9987	0.9997	0.9998
Median Filtering	0.1511 (Reverse)	0.2193 (Reverse)	0.3339 (Reverse)	0.2494 (Reverse)
Savitzky-Golay Filtering	0.1308	0.0480	0.1579 (Reverse)	0.0078 (Reverse)
ARMAX Model	0.7945	0.5063	0.9282	0.4950
Region of Interest Filtering	0.1093	0.0266 (Reverse)	0.0479	0.0795 (Reverse)

UCF

HS: Security of Analog to Digital Converter

- **The performance evaluation of the normal and the infected delta-sigma ADC for Saw Waveform:**

Defense Analysis (2): The performance analysis of filters based on their corresponding modified R-squared values when a triangle signal is applied to the modulator.

Type of Filter	Modified R-Squared Value for ADC Under Attack 1 (Sawtooth Input)	Modified R-Squared Value for ADC Under Attack 2 (Sawtooth Input)	Modified R-Squared Value for ADC Under Attack 3 (Sawtooth Input)	Modified R-Squared Value for ADC Under Attack 4 (Sawtooth Input)
Simple Linear Regression	1.0000	0.9997	1.0000	1.0000
Generalized Linear Regression	1.0000	0.9997	1.0000	1.0000
Moving Average	0.1009 (Reverse)	0.4238 (Reverse)	0.4110 (Reverse)	0.4590 (Reverse)
Robust Regression (LAR Method)	1.0000	0.9997	1.0000	1.0000
Robust Regression (Bisquare Method)	1.0000	0.9997	1.0000	1.0000
Median Filtering	0.7509 (Reverse)	0.5139 (Reverse)	0.6886 (Reverse)	0.5341 (Reverse)
Savitzky-Golay Filtering	7.43×10^{-4} (Reverse)	0.3704 (Reverse)	0.4335 (Reverse)	0.2989 (Reverse)
ARMAX Model	0.5215 (Reverse)	0.3658	0.8223	0.0131 (Reverse)
Region of Interest Filtering	0.6192 (Reverse)	0.3918 (Reverse)	0.4910 (Reverse)	0.2780 (Reverse)

HS: Security of Analog to Digital Converter

- **The performance evaluation of the normal and the infected delta-sigma ADC for Triangle Waveform:**

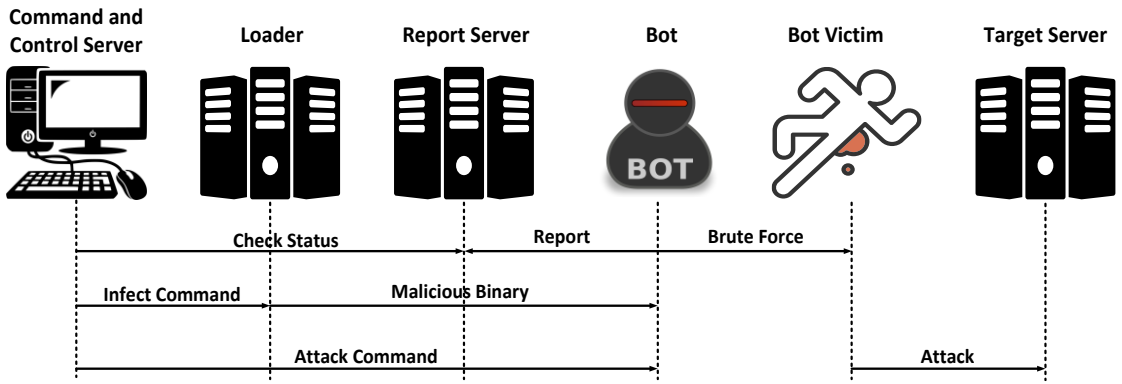
Defense Analysis (3): The performance analysis of filters based on their corresponding modified R-squared values when a sawtooth signal is applied to the modulator.

Type of Filter	Modified R-Squared Value for ADC Under Attack 1 (Triangle Input)	Modified R-Squared Value for ADC Under Attack 2 (Triangle Input)	Modified R-Squared Value for ADC Under Attack 3 (Triangle Input)	Modified R-Squared Value for ADC Under Attack 4 (Triangle Input)
Simple Linear Regression	0.9941	0.9882	0.9934	0.9933
Generalized Linear Regression	0.9941	0.9883	0.9934	0.9934
Moving Average	0.1787 (Reverse)	0.3153 (Reverse)	0.4146 (Reverse)	0.3108 (Reverse)
Robust Regression (LAR Method)	1.0000	1.0000	1.0000	1.0000
Robust Regression (Bisquare Method)	0.9197	0.9882	0.9934	0.9933
Median Filtering	0.3479 (Reverse)	0.5051 (Reverse)	0.6146 (Reverse)	0.5061 (Reverse)
Savitzky-Golay Filtering	0.1231 (Reverse)	0.3001 (Reverse)	0.4136 (Reverse)	0.3107 (Reverse)
ARMAX Model	0.5832	0.3310	0.9185	0.7649
Region of Interest Filtering	0.2253 (Reverse)	0.3755 (Reverse)	0.4137 (Reverse)	0.3276 (Reverse)

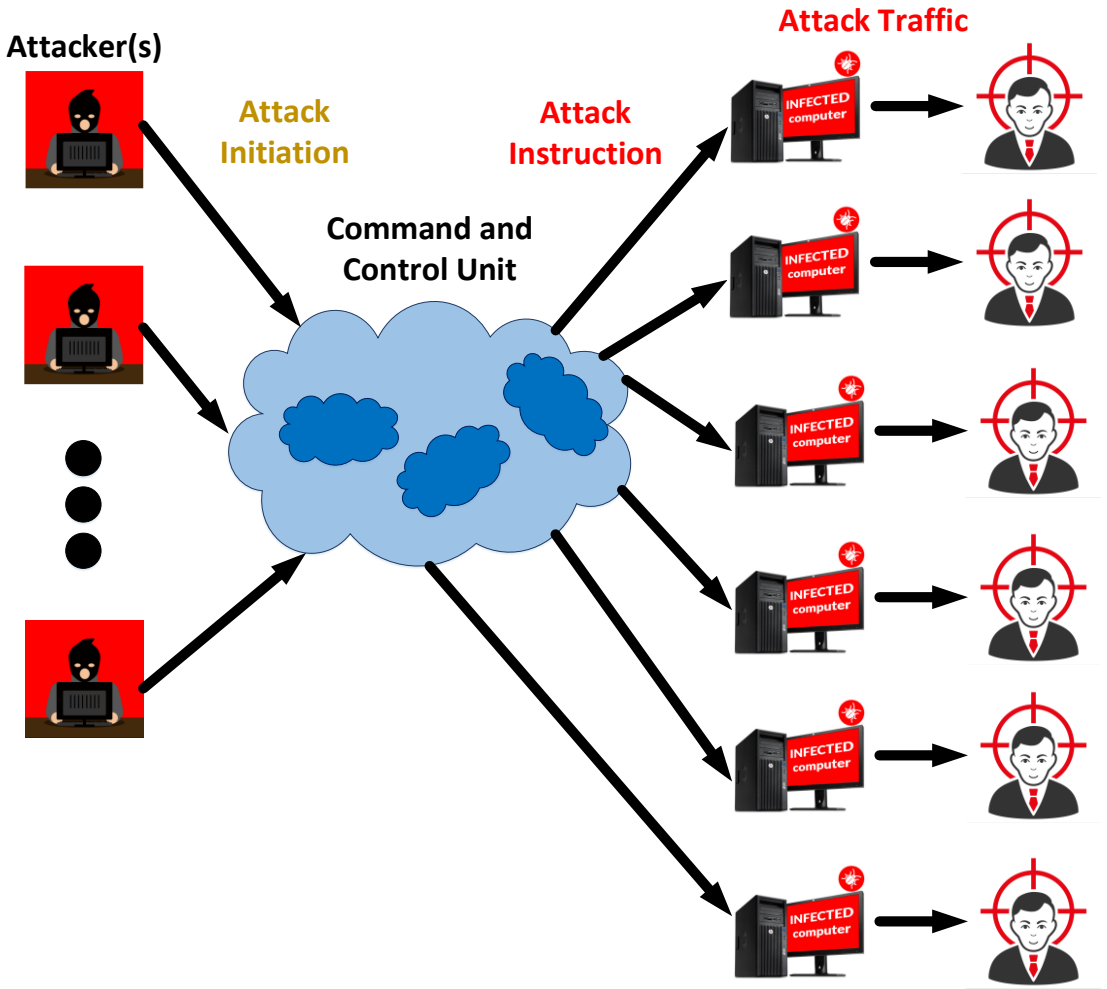
UCF

MLS: Leveraging Transfer Learning and Data Transformation in Botnet Detection

- More than 98% of the work completed by me!
- The operation flow of Botnet



The operations of the Mirai botnet and the network architecture for a typical botnet.



MLS: Leveraging Transfer Learning and Data Transformation in Botnet Detection

- **The Classification Measures**

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

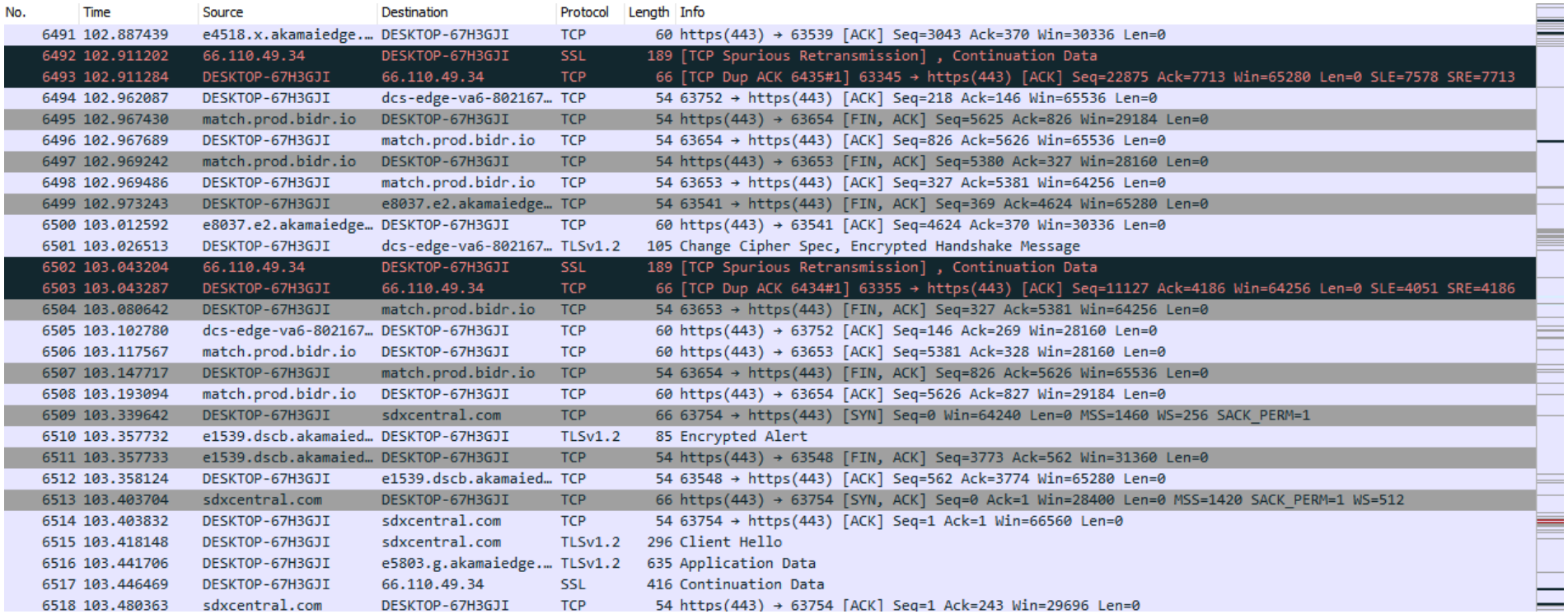
$$\text{F1 Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

$$\text{Kappa} = \frac{\text{Actual Accuracy} - \text{Random Accuracy}}{1 - \text{Random Accuracy}}$$

UCF

MLS: Leveraging Transfer Learning and Data Transformation in Botnet Detection

- **A snapshot of the captured traffic data from a local residential network.**



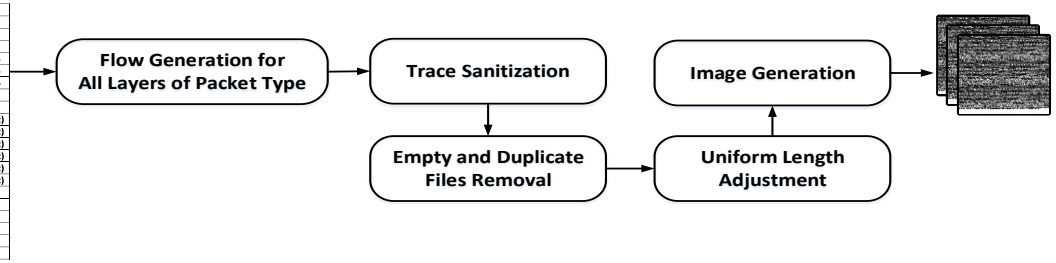
No.	Time	Source	Destination	Protocol	Length	Info
6491	102.887439	e4518.x.akamaiedge.…	DESKTOP-67H3GJI	TCP	60	https(443) → 63539 [ACK] Seq=3043 Ack=370 Win=30336 Len=0
6492	102.911202	66.110.49.34	DESKTOP-67H3GJI	SSL	189	[TCP Spurious Retransmission] , Continuation Data
6493	102.911284	DESKTOP-67H3GJI	66.110.49.34	TCP	66	[TCP Dup ACK 6435#1] 63345 → https(443) [ACK] Seq=22875 Ack=7713 Win=65280 Len=0 SLE=7578 SRE=7713
6494	102.962087	DESKTOP-67H3GJI	dcs-edge-va6-802167…	TCP	54	63752 → https(443) [ACK] Seq=218 Ack=146 Win=65536 Len=0
6495	102.967430	match.prod.bidr.io	DESKTOP-67H3GJI	TCP	54	https(443) → 63654 [FIN, ACK] Seq=5625 Ack=826 Win=29184 Len=0
6496	102.967689	DESKTOP-67H3GJI	match.prod.bidr.io	TCP	54	63654 → https(443) [ACK] Seq=826 Ack=5626 Win=65536 Len=0
6497	102.969242	match.prod.bidr.io	DESKTOP-67H3GJI	TCP	54	https(443) → 63653 [FIN, ACK] Seq=5380 Ack=327 Win=28160 Len=0
6498	102.969486	DESKTOP-67H3GJI	match.prod.bidr.io	TCP	54	63653 → https(443) [ACK] Seq=327 Ack=5381 Win=64256 Len=0
6499	102.973243	DESKTOP-67H3GJI	e8037.e2.akamaiedge…	TCP	54	63541 → https(443) [FIN, ACK] Seq=369 Ack=4624 Win=65280 Len=0
6500	103.012592	e8037.e2.akamaiedge…	DESKTOP-67H3GJI	TCP	60	https(443) → 63541 [ACK] Seq=4624 Ack=370 Win=30336 Len=0
6501	103.026513	DESKTOP-67H3GJI	dcs-edge-va6-802167…	TLSv1.2	105	Change Cipher Spec, Encrypted Handshake Message
6502	103.043204	66.110.49.34	DESKTOP-67H3GJI	SSL	189	[TCP Spurious Retransmission] , Continuation Data
6503	103.043287	DESKTOP-67H3GJI	66.110.49.34	TCP	66	[TCP Dup ACK 6434#1] 63355 → https(443) [ACK] Seq=11127 Ack=4186 Win=64256 Len=0 SLE=4051 SRE=4186
6504	103.080642	DESKTOP-67H3GJI	match.prod.bidr.io	TCP	54	63653 → https(443) [FIN, ACK] Seq=327 Ack=5381 Win=64256 Len=0
6505	103.102780	dcs-edge-va6-802167…	DESKTOP-67H3GJI	TCP	60	https(443) → 63752 [ACK] Seq=146 Ack=269 Win=28160 Len=0
6506	103.117567	match.prod.bidr.io	DESKTOP-67H3GJI	TCP	60	https(443) → 63653 [ACK] Seq=5381 Ack=328 Win=28160 Len=0
6507	103.147717	DESKTOP-67H3GJI	match.prod.bidr.io	TCP	54	63654 → https(443) [FIN, ACK] Seq=826 Ack=5626 Win=65536 Len=0
6508	103.193094	match.prod.bidr.io	DESKTOP-67H3GJI	TCP	60	https(443) → 63654 [ACK] Seq=5626 Ack=827 Win=29184 Len=0
6509	103.339642	DESKTOP-67H3GJI	sdxcentral.com	TCP	66	63754 → https(443) [SYN] Seq=0 Win=64240 Len=0 MSS=1460 WS=256 SACK_PERM=1
6510	103.357732	e1539.dscb.akamaied…	DESKTOP-67H3GJI	TLSv1.2	85	Encrypted Alert
6511	103.357733	e1539.dscb.akamaied…	DESKTOP-67H3GJI	TCP	54	https(443) → 63548 [FIN, ACK] Seq=3773 Ack=562 Win=31360 Len=0
6512	103.358124	DESKTOP-67H3GJI	e1539.dscb.akamaied…	TCP	54	63548 → https(443) [ACK] Seq=562 Ack=3774 Win=65280 Len=0
6513	103.403704	sdxcentral.com	DESKTOP-67H3GJI	TCP	66	https(443) → 63754 [SYN, ACK] Seq=0 Ack=1 Win=28400 Len=0 MSS=1420 SACK_PERM=1 WS=512
6514	103.403832	DESKTOP-67H3GJI	sdxcentral.com	TCP	54	63754 → https(443) [ACK] Seq=1 Ack=1 Win=66560 Len=0
6515	103.418148	DESKTOP-67H3GJI	sdxcentral.com	TLSv1.2	296	Client Hello
6516	103.441706	DESKTOP-67H3GJI	e5803.g.akamaiedge.…	TLSv1.2	635	Application Data
6517	103.446469	DESKTOP-67H3GJI	66.110.49.34	SSL	416	Continuation Data
6518	103.480363	sdxcentral.com	DESKTOP-67H3GJI	TCP	54	https(443) → 63754 [ACK] Seq=1 Ack=243 Win=29696 Len=0

A snapshot of the captured traffic data from a local residential network

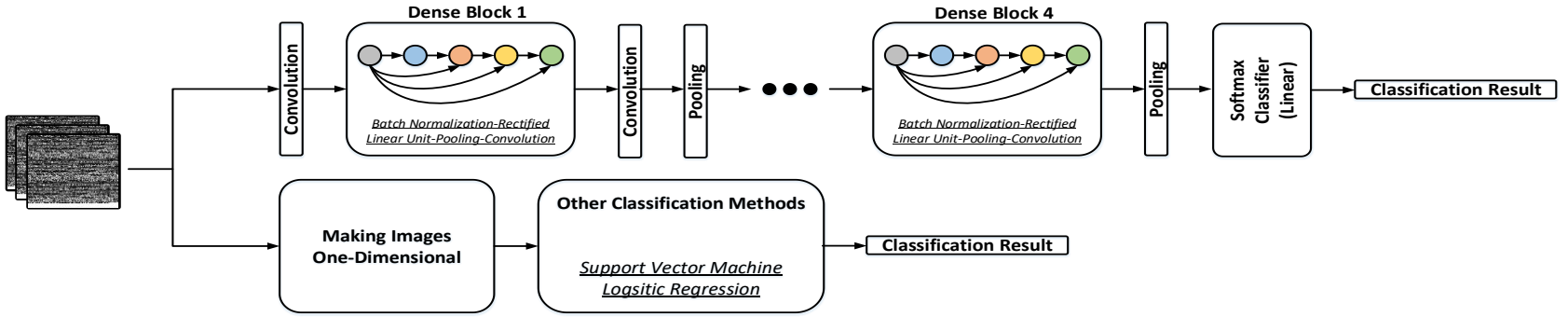
MLS: Leveraging Transfer Learning and Data Transformation in Botnet Detection

Input Data: Network Traffic Data

Number	Time	Source	Destination	Protocol	Length	Information
01	0.000000	Cisco_db:19:c3	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.1
02	8.982709	Cisco_db:19:c3	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.1
03	50.099564	Cisco_db:19:c3	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.1
04	50.369266	54:52:00:00:00:01	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.85
05	51.369054	54:52:00:00:00:01	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.85
06	52.369688	54:52:00:00:00:01	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.85
07	53.086840	Cisco_db:19:c3	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.1
08	59.086131	Cisco_db:19:c3	Broadcast	ARP	60	Who has 147.32.84.165? Tell 147.32.84.1
09	160.084662	PcsCompu_b5:b7:19	Broadcast	ARP	60	Gratuitous ARP for 147.32.84.165 (Request)
10	160.084668	PcsCompu_b5:b7:19	Broadcast	ARP	60	Gratuitous ARP for 147.32.84.165 (Request)
11	161.077511	PcsCompu_b5:b7:19	Broadcast	ARP	60	Gratuitous ARP for 147.32.84.165 (Request)
12	161.077519	PcsCompu_b5:b7:19	Broadcast	ARP	60	Gratuitous ARP for 147.32.84.165 (Request)
13	162.079007	PcsCompu_b5:b7:19	Broadcast	ARP	60	Gratuitous ARP for 147.32.84.165 (Request)
14	162.079013	PcsCompu_b5:b7:19	Broadcast	ARP	60	Gratuitous ARP for 147.32.84.165 (Request)
15	162.765253	147.32.84.165	147.32.84.255	NBNS	110	Registration NB SARUMAN<00>
16	163.510681	147.32.84.165	147.32.84.255	NBNS	110	Registration NB SARUMAN<00>
17	163.510692	147.32.84.165	147.32.84.255	NBNS	110	Registration NB SARUMAN<00>
18	163.926344	PcsCompu_b5:b7:19	Broadcast	ARP	60	Who has 147.32.84.1? Tell 147.32.84.165
19	163.926354	PcsCompu_b5:b7:19	Broadcast	ARP	60	Who has 147.32.84.1? Tell 147.32.84.165
20	163.929830	Cisco_db:19:c3	PcsCompu_b5:b7:19	ARP	60	147.32.84.1 is at 00:1e:49:db:19:c3



(a)

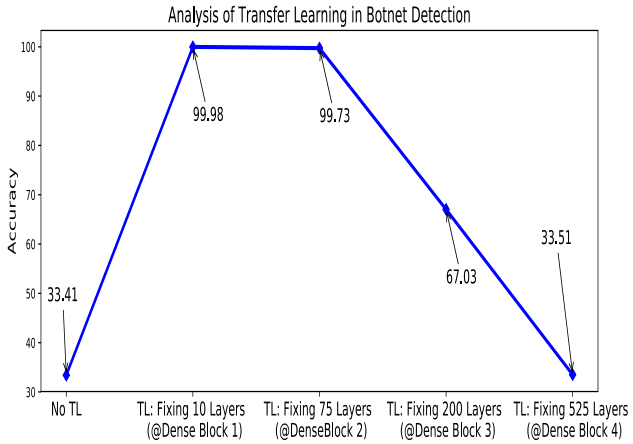


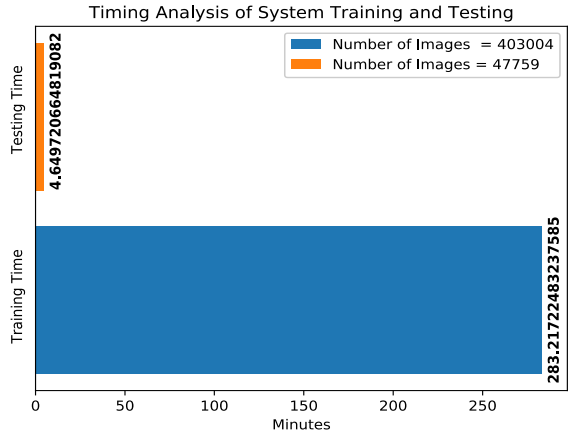
(b)

The flow of botnet detection using DenseNet, Support Vector Machine (SVM), and Logistic Regression: (a) transformation of network traffic data into image and (b) classification of transformed network traffic data into image using DenseNet (top) and SVM along with Logistic Regression (bottom).

MLS: Leveraging Transfer Learning and Data Transformation in Botnet Detection

- **Flow of botnet detection using DenseNet, Support Vector Machine**

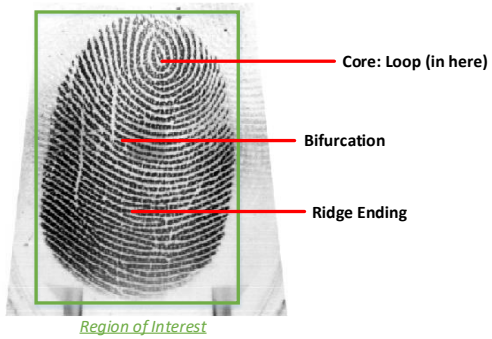


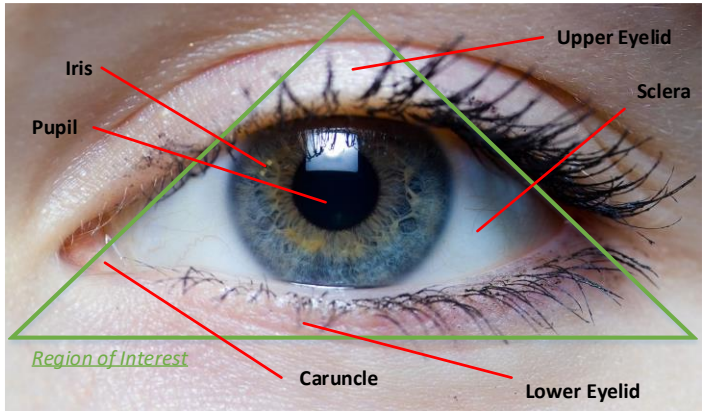


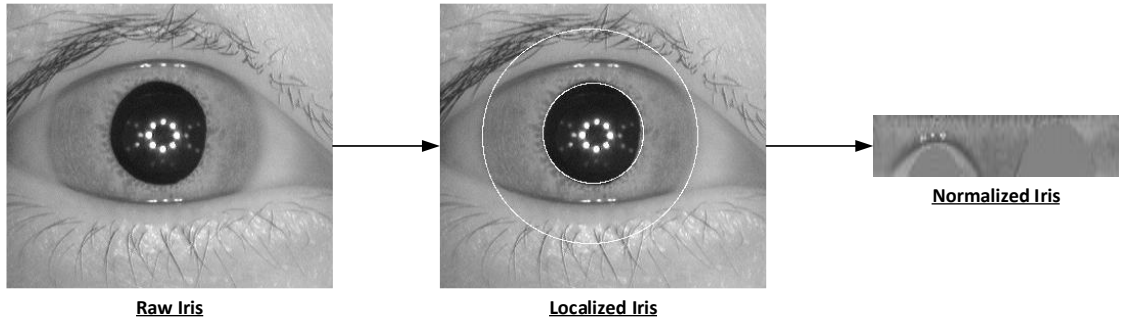
Score Merit	Accuracy (%)	Area under Curve (%)	Precision (%)	Recall (%)	F-1 Score (%)	Kappa (%)
CNN: DenseNet (Best Result) (for CTU-13 Dataset)	99.98	99.99	99.98	99.98	99.98	99.96
Support Vector Machine (SVM) (for CTU-13 Dataset)	83.15	75.80	80.73	98.07	88.56	57.72
Logistic Regression (for CTU-13 Dataset)	78.56	81.06	92.76	73.49	82.01	56.35
CNN: DenseNet (Best Result) (for Live Normal Network Traffic Data)	99.35	99.98	99.35	99.35	99.35	98.70
CNN: DenseNet (Best Result) (for ISOT HTTP Botnet Dataset [36], Only Botnet Traffic Data)	100	99.99	100	100	100	100
[6]	94.70	N/A	N/A	N/A	N/A	N/A
[7]: K-Nearest Neighbor (KNN) (for T1 training set)	90.20	N/A	89.50	91.00	90.30	N/A
[7]: C4.5 (for T1 training set)	90.10	N/A	89.10	91.20	90.10	N/A
[7]: Random Forest (RF) (for T1 training set)	90.80	N/A	90.70	91.00	90.80	N/A
[7]: Naïve Bayes (for T1 training set)	85.90	N/A	83.10	90.20	86.50	N/A
[7]: Clustering	98.39	N/A	86.45	84.47	85.45	N/A
[7]: Neural Network (NN)	89.38	N/A	92.50	85.70	88.97	N/A
[7]: Recurrent Neural Network (RNN)	83.09	N/A	95.41	69.53	80.44	N/A

MLS: Design of Privacy-Preserving and Secure Biometric Recognition System

- **Background**: Fingerprint and Iris Recognition



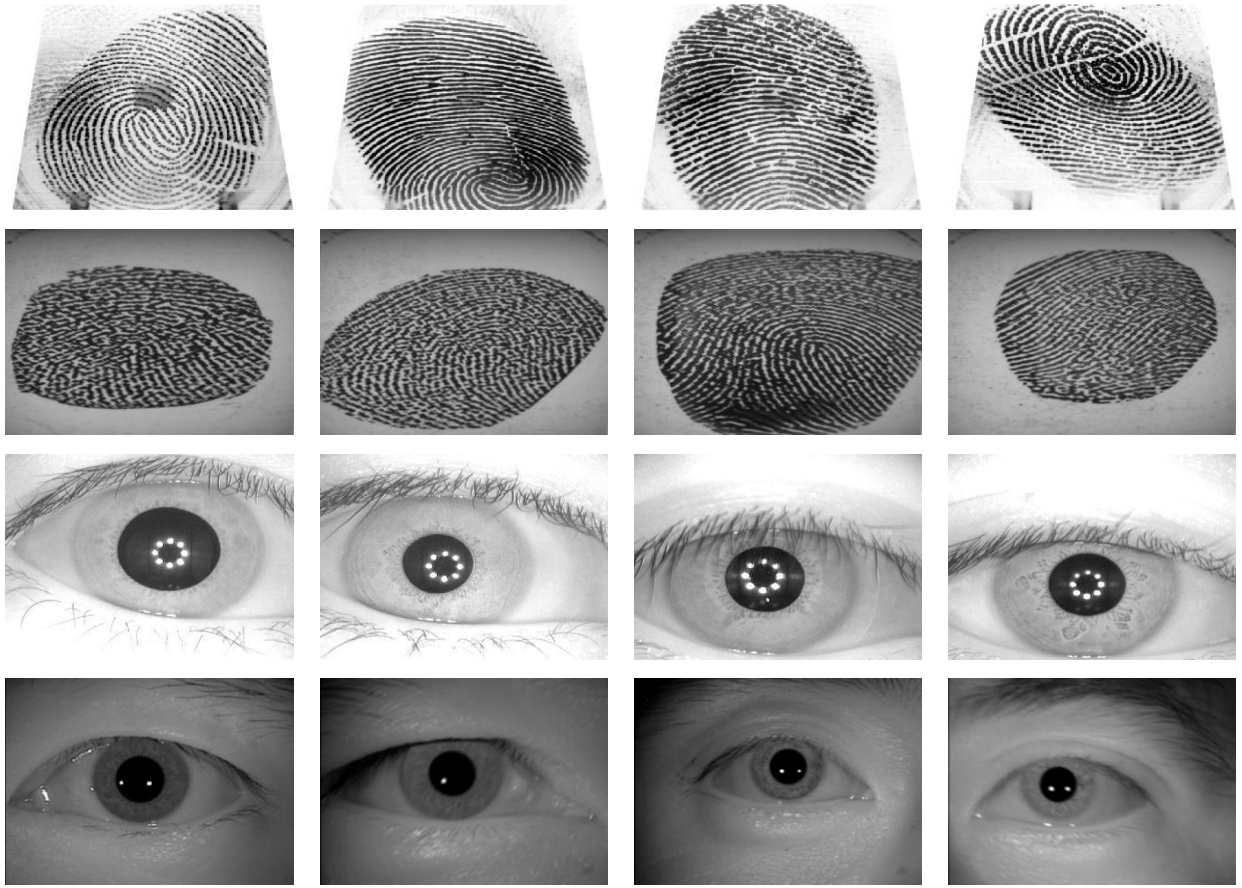




- ✓ **An authentic fingerprint.**
- ✓ **A human eye.**
- ✓ **The raw, localized, and normalized version of an iris.**

MLS: Design of Privacy-Preserving and Secure Biometric Recognition System

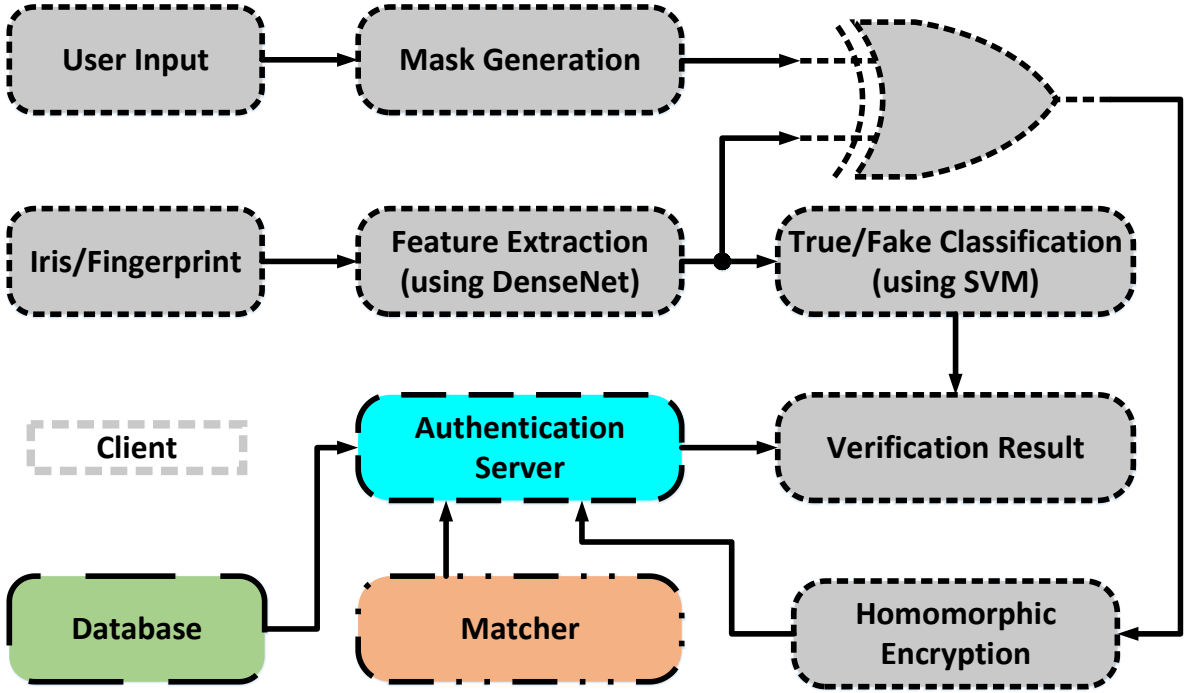
- **Background: The samples of Fingerprint and Iris.**



The samples of true fingerprint (first row), fake fingerprint (second row), true iris (third row), and fake iris (fourth row).

MLS: Design of Privacy-Preserving and Secure Biometric Recognition System

- **The proposed deep learning-based privacy preserving biometric recognition system.**

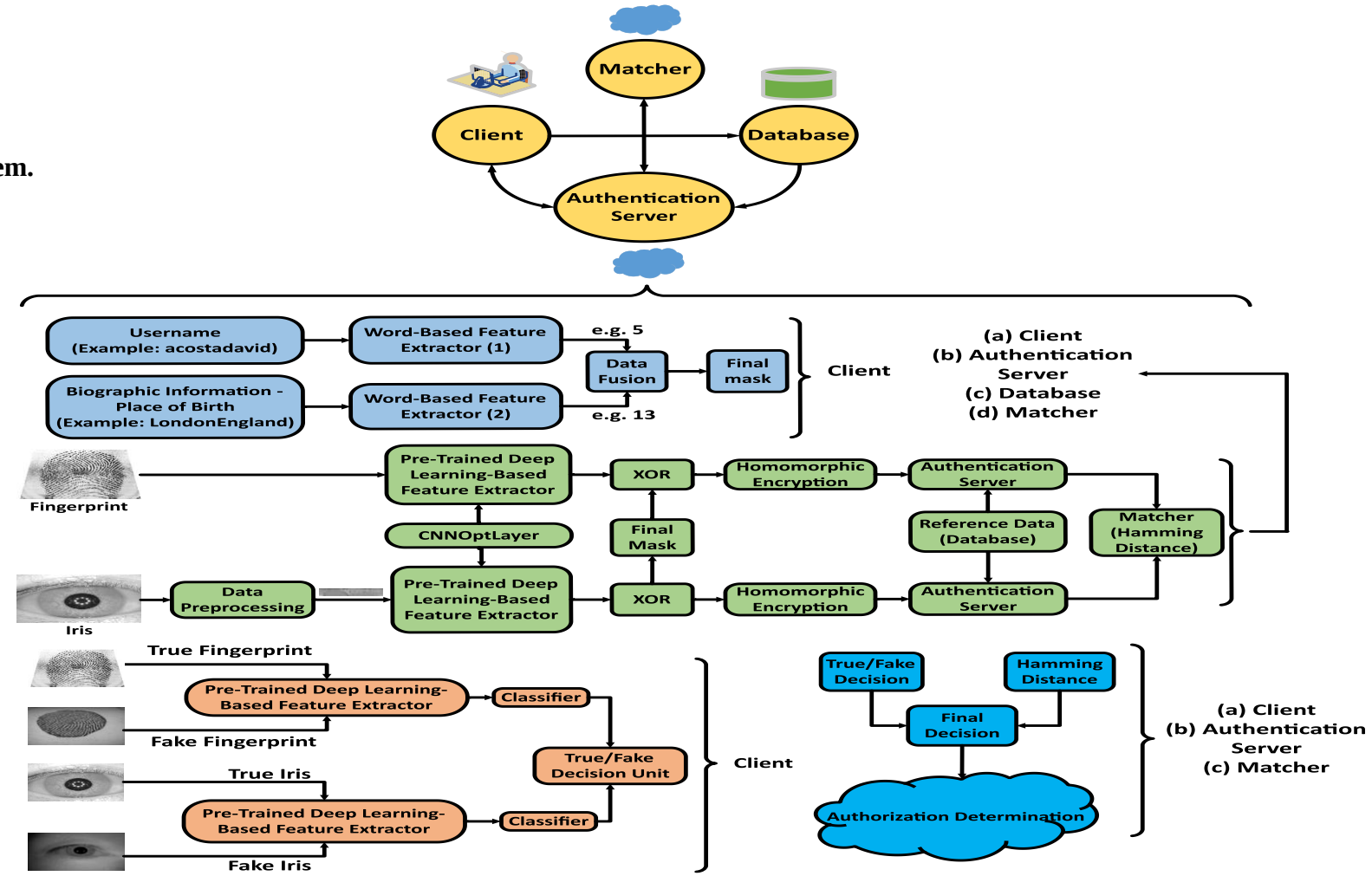


The proposed deep learning-based privacy preserving biometric recognition system.

MLS: Design of Privacy-Preserving and Secure Biometric Recognition System

- **More than 50% of the work completed by me!**
- The proposed deep learning-based privacy preserving biometric recognition system.

The details of the proposed system.



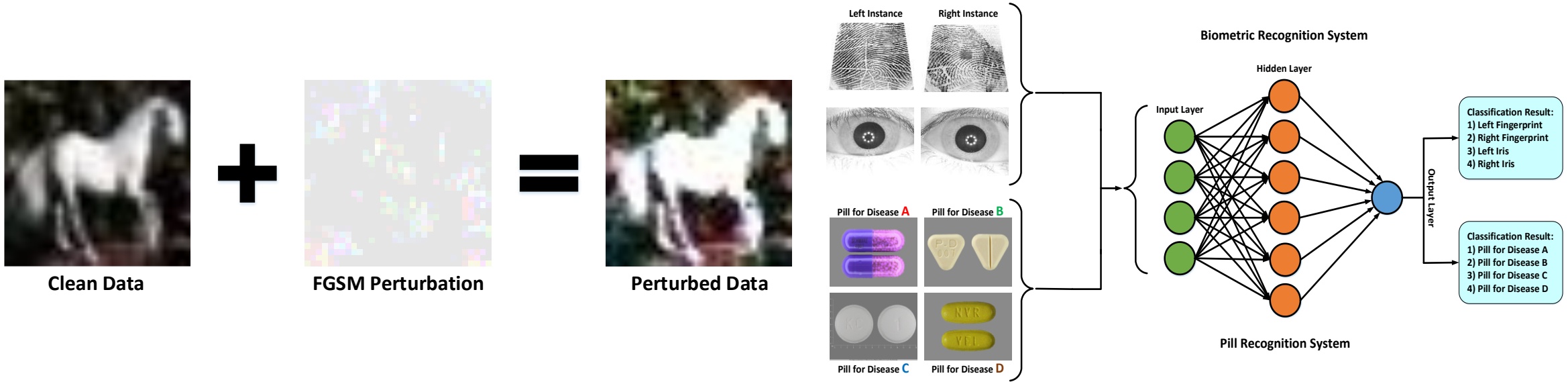
MLS: Design of Privacy-Preserving and Secure Biometric Recognition System

- **The simulation results of performance evaluation of biometric recognition system.**

Data Type	Masked	Layer	Threshold	TP	TN	FP	FN	F-Score
Fingerprint	No	12	10, 243	812	337,450	800	838	49.79
Fingerprint	Yes	12	10,243	812	338,246	4	838	65.86
Iris	No	21	6427	1414	338,165	85	236	89.81
Iris	Yes	21	6427	1414	338,248	2	236	92.24
Combined	No	12 + 21	17,108	1507	338,168	82	143	93.05
Combined	Yes	12 + 21	17,108	1507	338,250	0	143	95.47

MLS: Detection of Adversarial Examples Using Adversarial Training and Noise Training

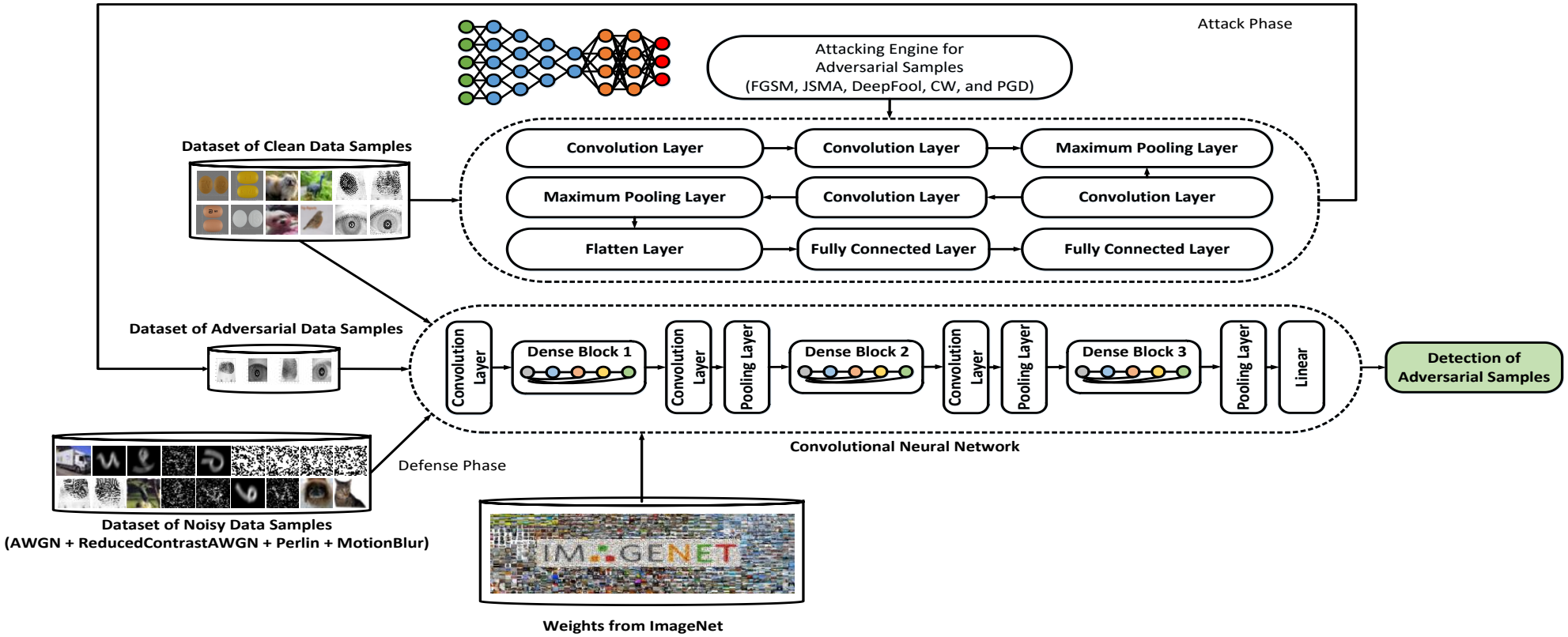
- More than 98% of the work completed by me!
- Applying Fast Gradient Sign Method (FGSM) on an image and pill/biometric data recognition structure:



The figure shows injecting perturbation from the FGSM attack into a sample image from the ten-class Canadian Institute For Advanced Research (CIFAR) dataset. A high amount of perturbation is chosen during the simulation for better visual presentation.

MLS: Detection of Adversarial Examples Using Adversarial Training and Noise Training

- **The system architecture for generation of adversarial examples by shallow neural network along with performing adversarial training and noise training on a deep neural network.**



The architecture of our system (Generator+Detector Network) for detecting adversarial examples.

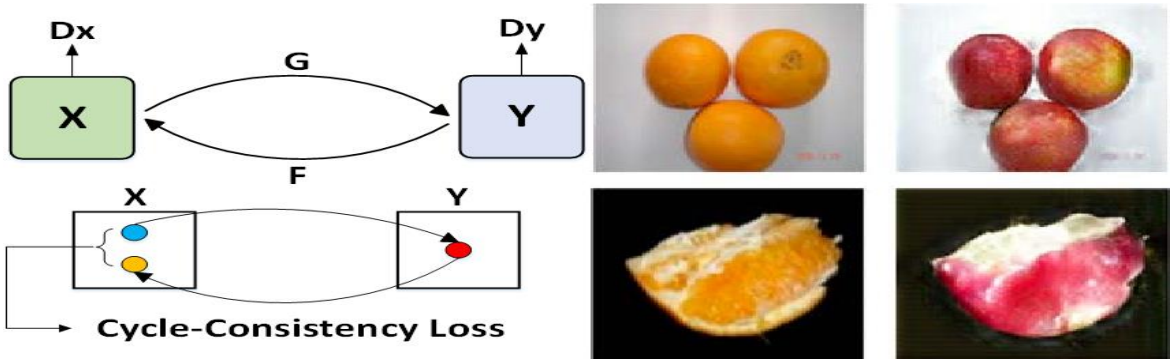
MLS: Detection of Adversarial Examples Using Adversarial Training and Noise Training

- **Performance evaluation of defense system for detection of adversarial examples:**

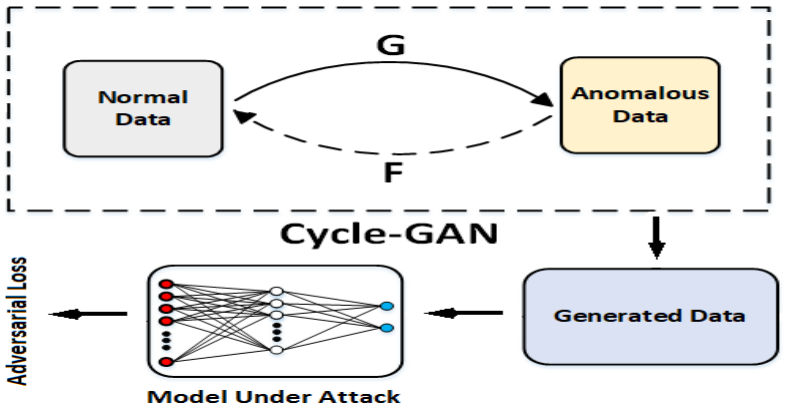
	Dataset	System Detection Accuracy on Clean Data	System Detection Accuracy on Attacked Data Without Defense	System Detection Accuracy on Attacked Data	
Ours – Biometric Dataset	**CASIA Dataset – Images of Iris and Fingerprint Data**	**90.58%**	**1.31%**	**Clean Data + Adversarial Data**	Clean Data + Adversarial Data + Noisy Data
				80.65%	93.4%
Ours – Pillbox Image Dataset	**Pillbox Dataset – Images of Pill**	**99.92%**	**34.55%**	**96.03%**	98.20%
[8]	**CIFAR**	**92%**	**10%**	**86%**	
[9]	**MNIST**	**N/A**	**19.39%**	**75.95%**	
[9]	**CIFAR**	**N/A**	**8.57%**	**71.38%**	
[10] – ResNet	**MNIST**	**88%**	**0% (Strongest Attack)**	**83% (Strongest Attack)**	
[10] – VGG	**MNIST**	**89%**	**36% (Strongest Attack)**	**85% (Strongest Attack)**	
[10] – ResNet	**CIFAR**	**85%**	**7% (Strongest Attack)**	**71% (Strongest Attack)**	
[10] -- VGG	**CIFAR**	**82%**	**37% (Strongest Attack)**	**80% (Strongest Attack)**	

Future Work -- MLS: Leveraging CycleGAN for Making Neural Network Robust!

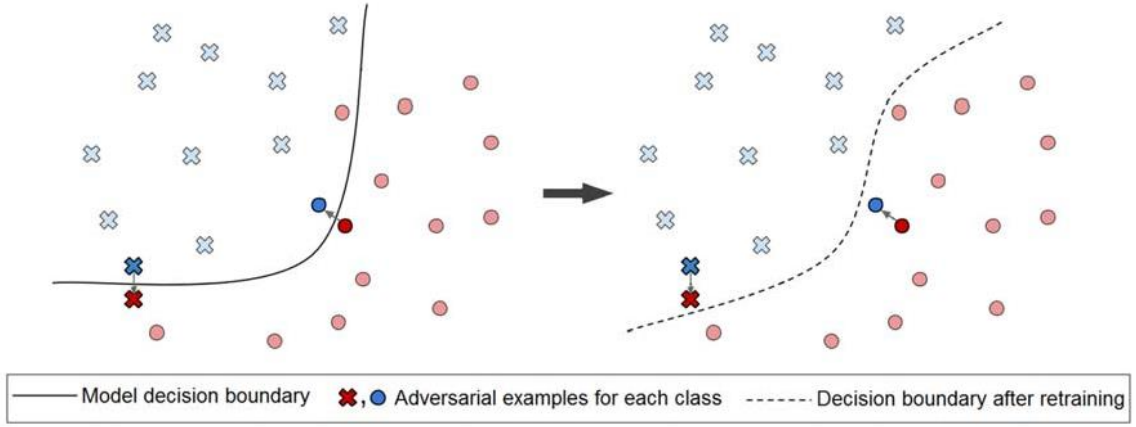
- **Using Cycle Generative Adversarial Network for generation of adversarial examples and retraining the neural network!**



The Cycle-GAN structure (left) and samples of the transformation (right).



The automated adversarial generation system.



The Visualization of the effect of retraining as defense mechanism.

References

1. Taheri, S., Lin, J. and Yuan, J.S., 2017. Security interrogation and defense for SAR analog to digital converter. *Electronics*, *6*(2), p.48.
2. Taheri, S. and Yuan, J.S., 2017. Mixed-Signal Hardware Security: Attacks and Countermeasures for ΔΣ ADC. *Electronics*, *6*(3), p.60.
3. Taheri, S., Salem, M. and Yuan, J.S., 2018. Leveraging Image Representation of Network Traffic Data and Transfer Learning in Botnet Detection. *Big Data and Cognitive Computing*, *2*(4), p.37.
4. Salem, M., Taheri, S. and Yuan, J.S., 2019. Utilizing transfer learning and homomorphic encryption in a privacy preserving and secure biometric recognition system. *Computers*, *8*(1), p.3.
5. Taheri, S., Salem, M. and Yuan, J.S., 2019. RazorNet: Adversarial Training and Noise Training on a Deep Neural Network Fooled by a Shallow Neural Network. *Big Data and Cognitive Computing*, *3*(3), p.43.
6. Chen, S.; Chen, Y.; Tzeng, W. Effective Botnet Detection Through Neural Networks on Convolutional Features. In Proceedings of the 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE), New York, NY, USA, 1–3 August 2018; pp. 372–378.
7. Hoang, X.D.; Nguyen, Q.C. Botnet Detection Based on Machine Learning Techniques Using DNS Query Data. Future Internet 2018, 10, 43.
8. Liu, X., Cheng, M., Zhang, H. and Hsieh, C.J., 2018. Towards robust neural networks via random self-ensemble. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 369-385).
9. Ranjan, R., Sankaranarayanan, S., Castillo, C.D. and Chellappa, R., 2017. Improving network robustness against adversarial attacks with compact convolution. *arXiv preprint arXiv:1712.00699*.
10. Song, Y., Kim, T., Nowozin, S., Ermon, S. and Kushman, N., 2017. Pixeldefend: Leveraging generative models to understand and defend against adversarial examples. *arXiv preprint arXiv:1710.10766*.

UCF