

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
Spring 2020

Major Professor: Jiann-Shiun Yuan



Overview

- Introduction: Cybersecurity – Hardware Security (HS) and Artificial Intelligence Security (AIS)
- HS: Security of Analog to Digital Converter
- AIS: Leveraging Transfer Learning and Data Transformation in Botnet Detection
- AIS: Design of Privacy-Preserving and Secure Biometric Recognition System
- AIS: Detection of Adversarial Examples Using Adversarial Training and Noise Training
- Conclusion

Introduction

- **Hardware Security**
- **Analog to Digital Converter**
- **Artificial Intelligence Security**

Introduction – Hardware Security

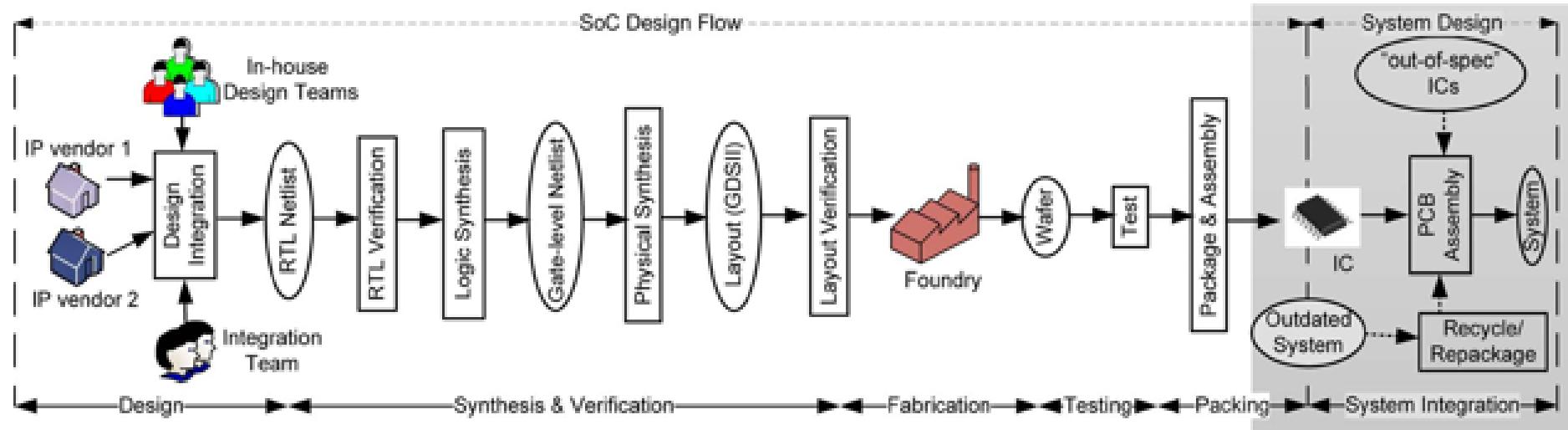
- Hardware is not secure and protected anymore as oppose to the traditional view that saw it entirely without malicious flaws.
- The concept of hardware security was formally introduced after emergence of hardware Trojans and the following proposing countermeasures for them.
- In Hardware Security, we deal with securing and protecting data stored in the hardware devices as well as making the design and the manufacturing process trusted.
- Hardware security and trust plays critical roles in computing processes due to their applications in many infrastructures.
- Application of semiconductor devices in military, financial, economic, and other critical infrastructure has raised significant concerns regarding hardware security.
- Making the hardware secure is associated with cost, power consumption, performance, and reliability concerns.

Introduction – Hardware Security

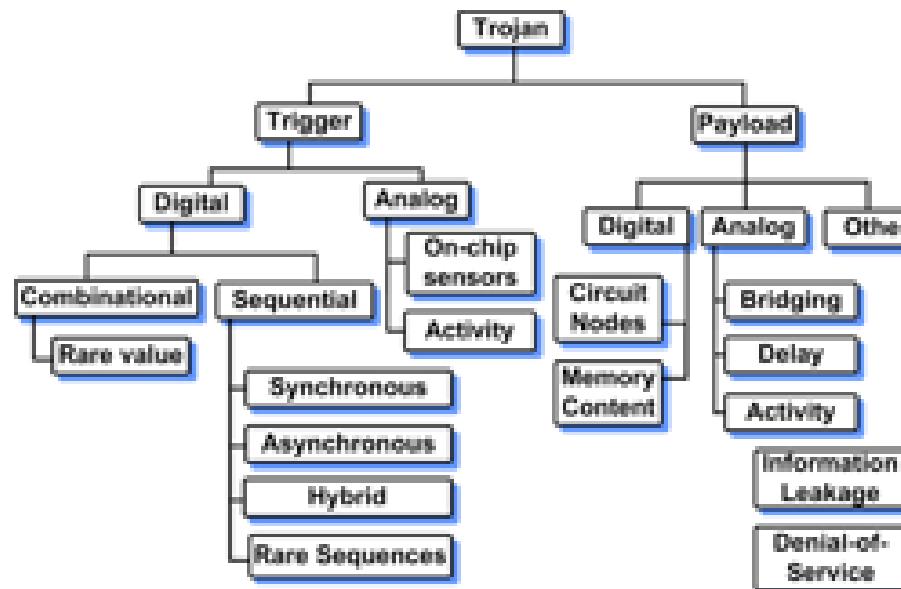
- Nowadays, the intellectual property (IP) cores may contain malicious logic and/or design flaws that can be exploited by attackers after the IP cores got integrated into the chip platforms.
- A hardware Trojan is defined as a malicious and intentional modifications of a circuit design that results in undesired behavior when the circuit is deployed.
- Hardware Trojans are designed to be stealthy by intelligent adversaries, which is a major difference from manufacturing defects that have been extensively researched for decades.
- Hardware Trojans change the functionality and degrade the system operation and performance.
- HT designs are categorized in a variety of forms according to different features.

Introduction – Hardware Security

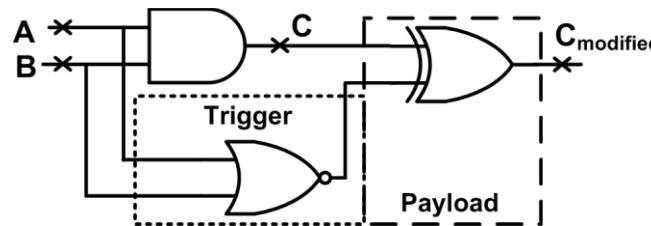
Semiconductor supply chain: IC design flow.



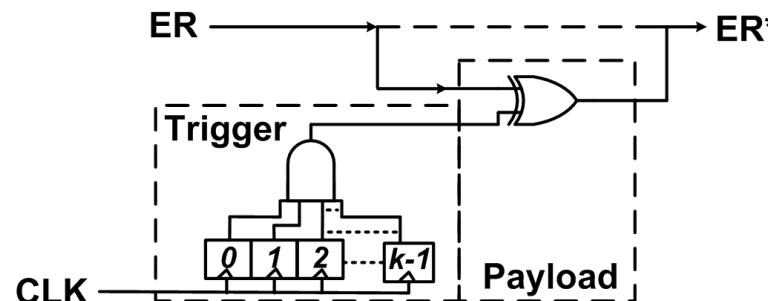
Trojan taxonomy based on trigger and payload mechanisms.



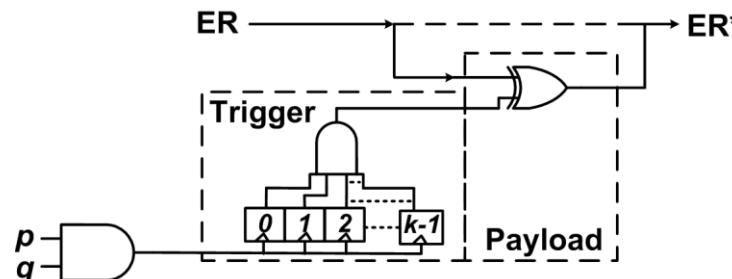
Introduction – Hardware Security



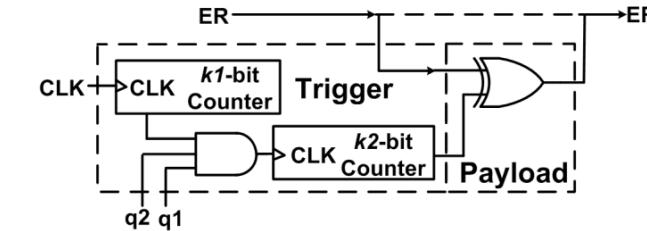
(a) Combinatorially triggered Trojan



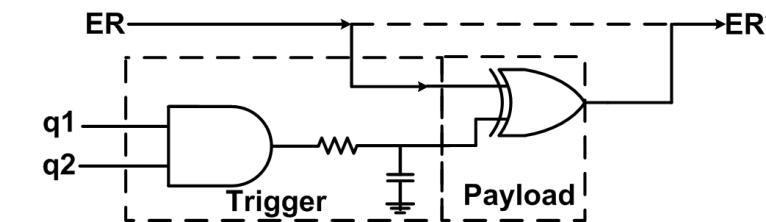
(b) Synchronous counter ("time-bomb") Trojan



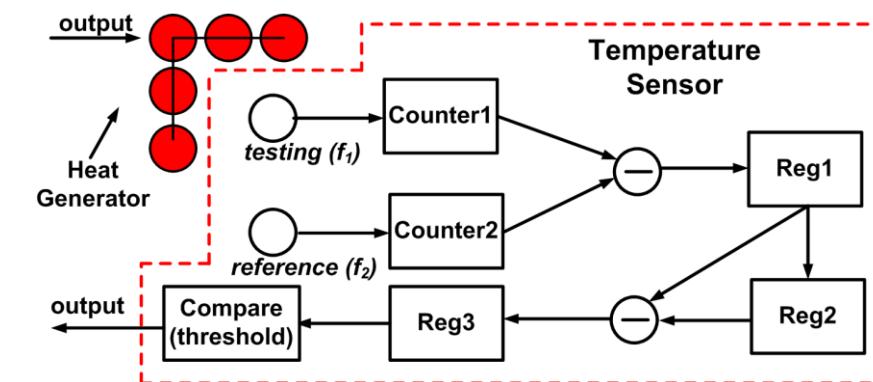
(c) Asynchronous counter Trojan



(d) Hybrid counter Trojan



(e) Analog Trojan triggered based on logic value



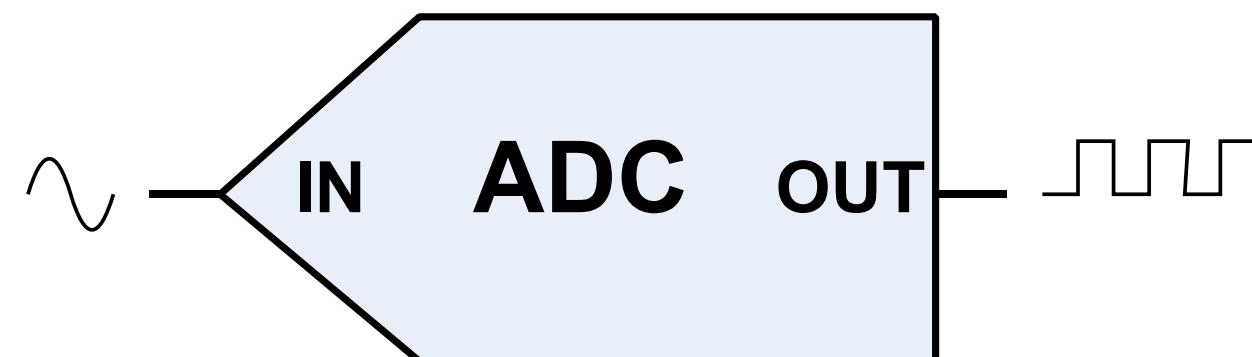
(f) Analog Trojan triggered based on circuit activity

Different Trojans and their trigger mechanisms.

Introduction – Analog to Digital Converter (ADC)

- Data converters are a fundamental building block for many circuits.
- They easily process a digital quantity in compare to equivalently processing information in the analog domain.
- Applications in functions such as digitizing voice, image and wireless telecommunications signals.
- They may have either linear or non-linear scalability.
- ADCs are continuous or non-continuous in nature.
- There is conversion of analog signal to discrete values.
- Types of ADC

- ✓ Flash ADC
- ✓ Delta-Sigma ADC
- ✓ Dual Slope (integrating) ADC
- ✓ Successive Approximation ADC



Introduction – Artificial Intelligence Security

- Artificial Intelligence (AI) has a broad variety of applications some of which we already know and encounter in our everyday life.
- AI has a wide range of applications, such as facial recognition, speech recognition, and robotics, but its application scope goes far beyond the three aspects of image, voice, and behavior.
- Nowadays, the AI systems are employed in the IoT world and they get engaged in hardware and software platforms.
- Integrated AI systems involve four components: Perception, Learning, Decisions, and Actions.
- These systems operate in complex environments that require each component to interact and be interdependent.
- Among these prevailing research fields, the application of AI related to Cybersecurity has become a ubiquitous research topic.
- Deep Learning plays an important role in cyberspace defense, and AI systems may also be attacked or deceived, resulting in incorrect classification or prediction results.
- Security issues of AI software and hardware systems need attention and finding possible solutions for them is challenging.

Introduction – Artificial Intelligence Security

- Possible security issues can be stated as: bad training mechanism, a bug in the system, the training data is not a representative of the given environment, and attacks in adversarial environments.
- Attacks in adversarial environments are intended to undermine the integrity and usability of various AI applications, and mislead neural networks by employing adversarial samples, causing classifiers to derive wrong classification.
- Possible applications of AI in the Cybersecurity domain can be named as: malware monitoring, intrusion detection, bypassing possible threats, and intelligence analysis.
- Finding threat models in researching the security of AI system is critical.
- Threat models identify measurable properties that define trustworthiness so a defender can incorporate robustness, privacy, and fairness into decision-making algorithms.
- Given a specific threat model, the system will have to reason about adversarial interference and define requisite conditions to achieve these trustworthiness properties.
- With fully and correct utilization of AI, many of the emerging problems for Cybersecurity can be overcome.

HS: Security of Analog to Digital Converter

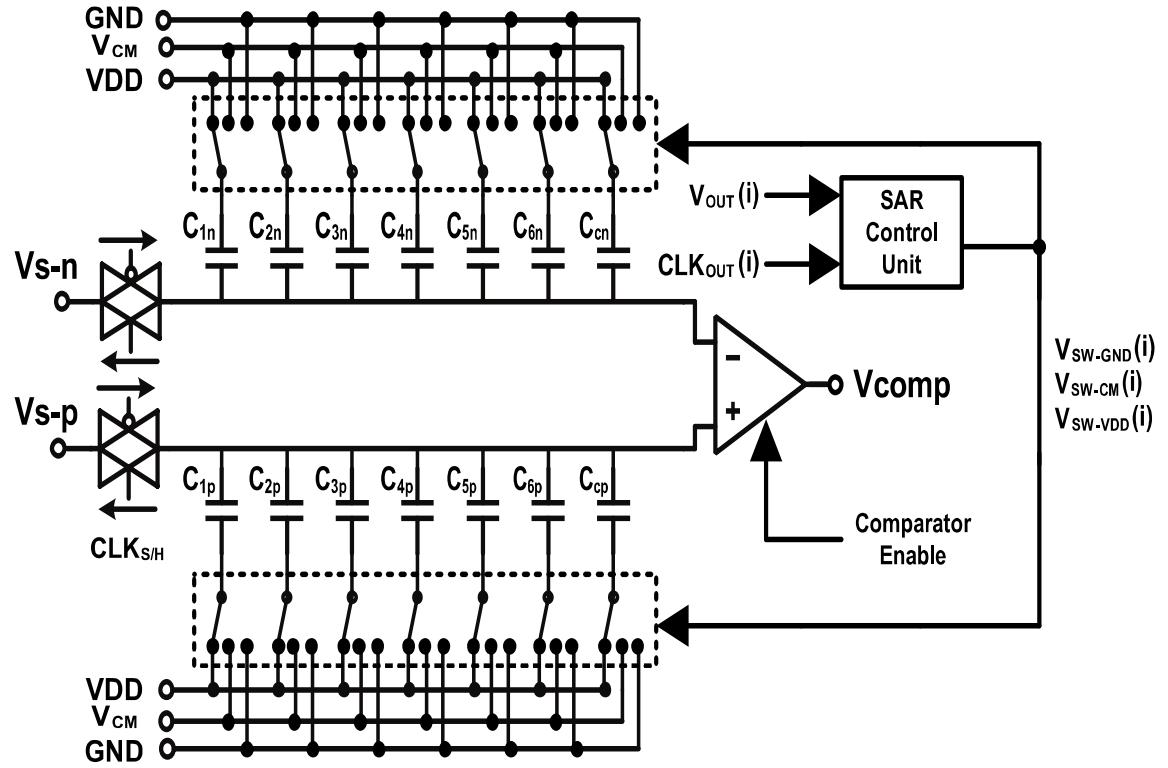
- Successive approximation register (SAR) analog to digital converter (ADC): Architecture
- SAR ADC: Operation
- SAR ADC: Threat Models
- SAR ADC: Operations of Threat Models
- SAR ADC: Defense Models
- SAR ADC: Operations of Defense Models
- Behavior of Normal and Infected SAR ADC using Tunnel Field-Effect Transistor (TFET)
- The performance evaluation of the normal and the infected SAR ADC

HS: Security of Analog to Digital Converter

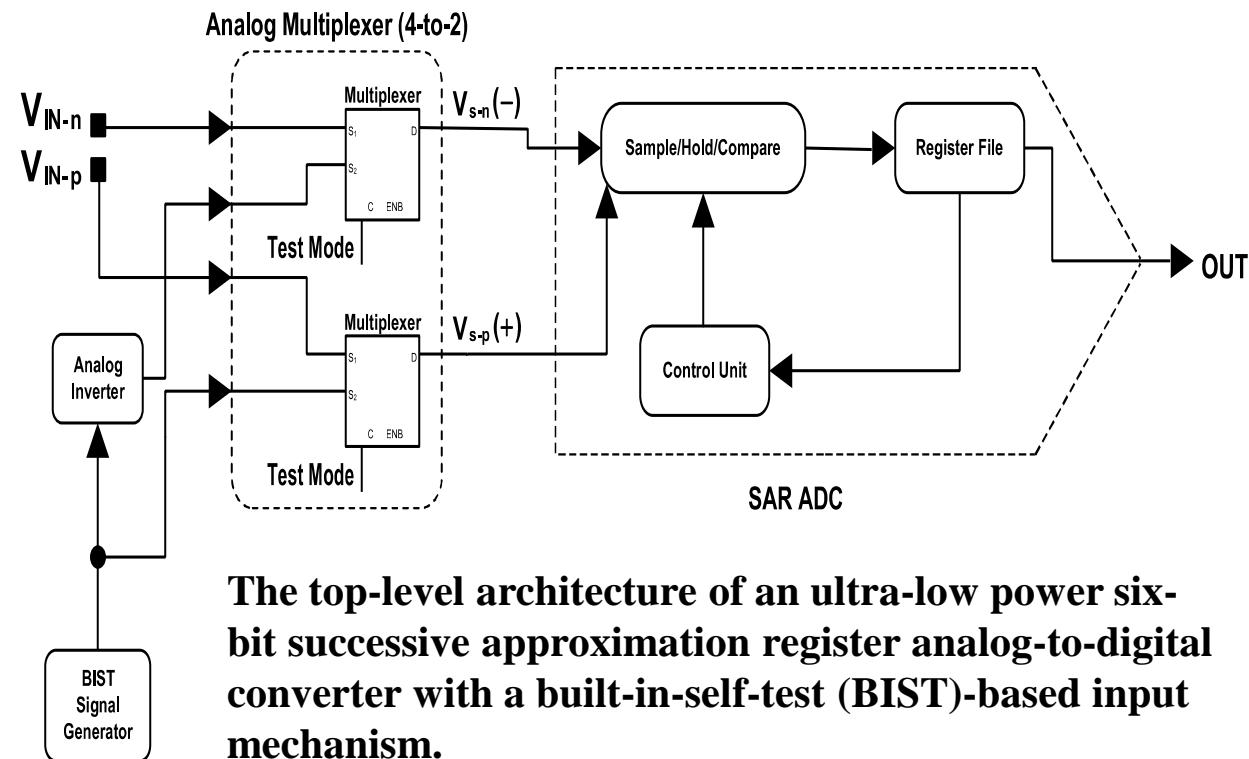
- Delta-Sigma analog to digital converter: Architecture and Operation
- Delta-Sigma ADC: Threat Models
- Delta-Sigma ADC: Malicious SRAM Cell Operation using TFET
- Behavior of Normal and Infected Delta-Sigma ADC using Tunnel Field-Effect Transistor (TFET)
- The performance evaluation of the normal and the infected delta-sigma ADC for Sine Waveform
- The performance evaluation of the normal and the infected delta-sigma ADC for Saw Waveform
- The performance evaluation of the normal and the infected delta-sigma ADC for Triangle Waveform

HS: Security of Analog to Digital Converter - Successive approximation register (SAR) analog to digital converter (ADC): Architecture

- 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



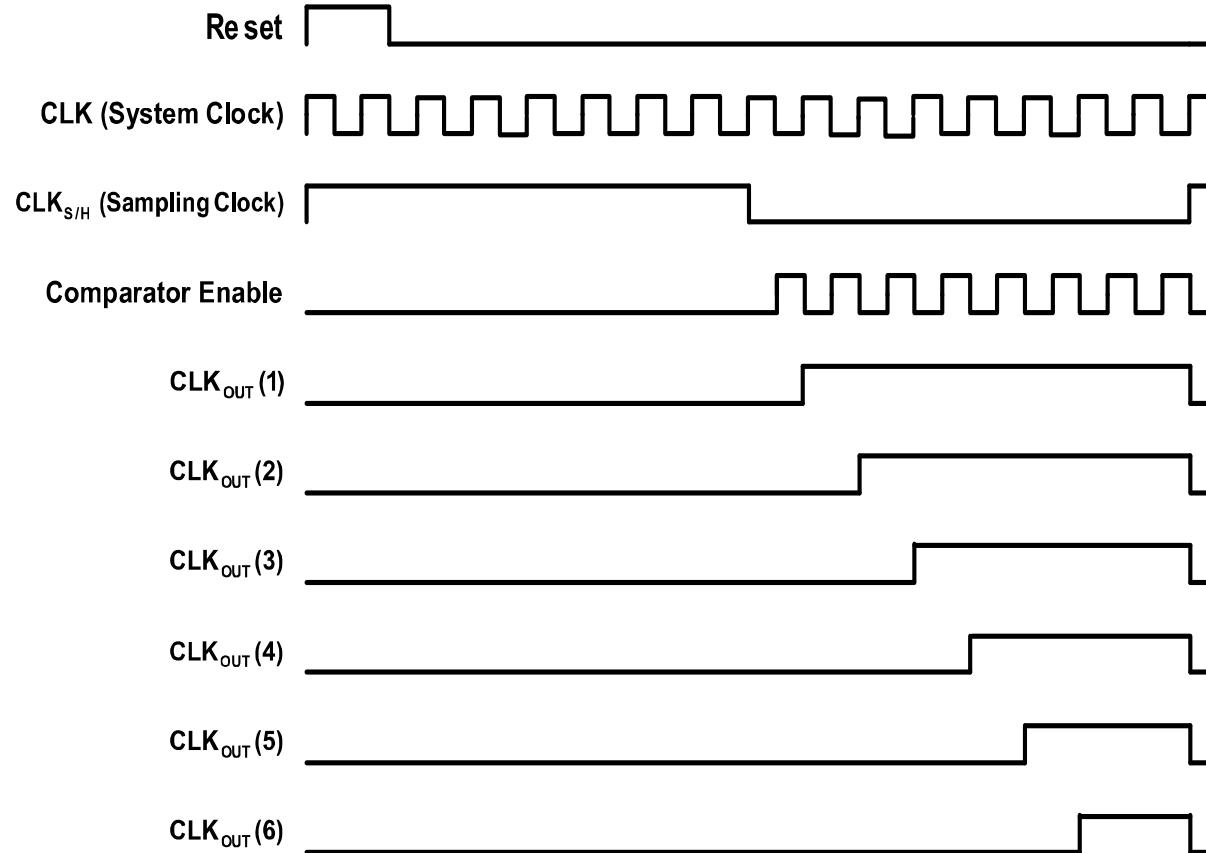
The sample/hold/compare block circuit.



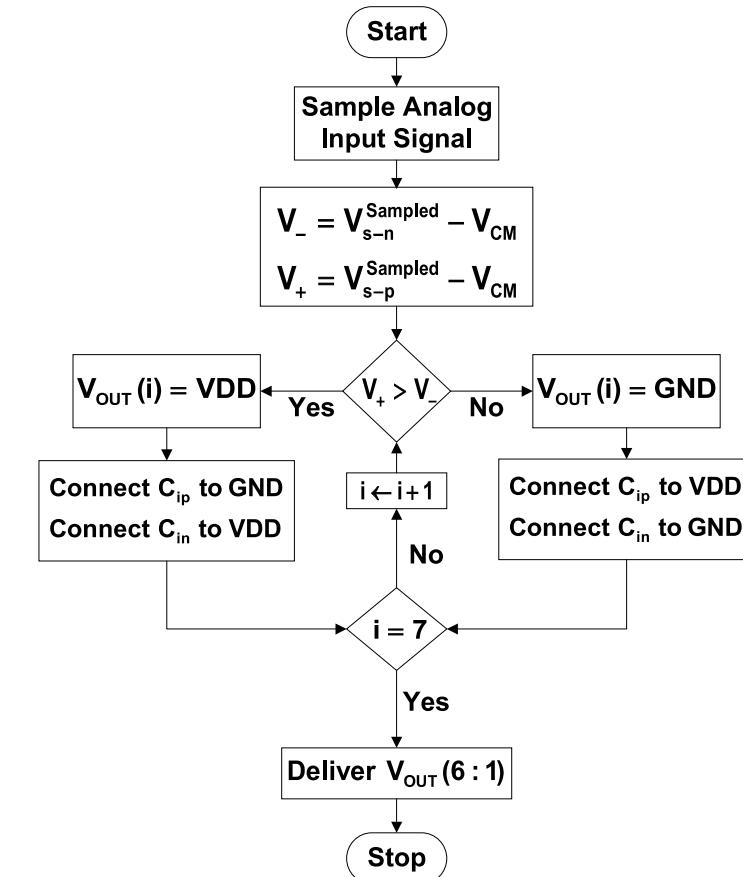
The top-level architecture of an ultra-low power six-bit successive approximation register analog-to-digital converter with a built-in-self-test (BIST)-based input mechanism.

HS: Security of Analog to Digital Converter - SAR ADC: Operation

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



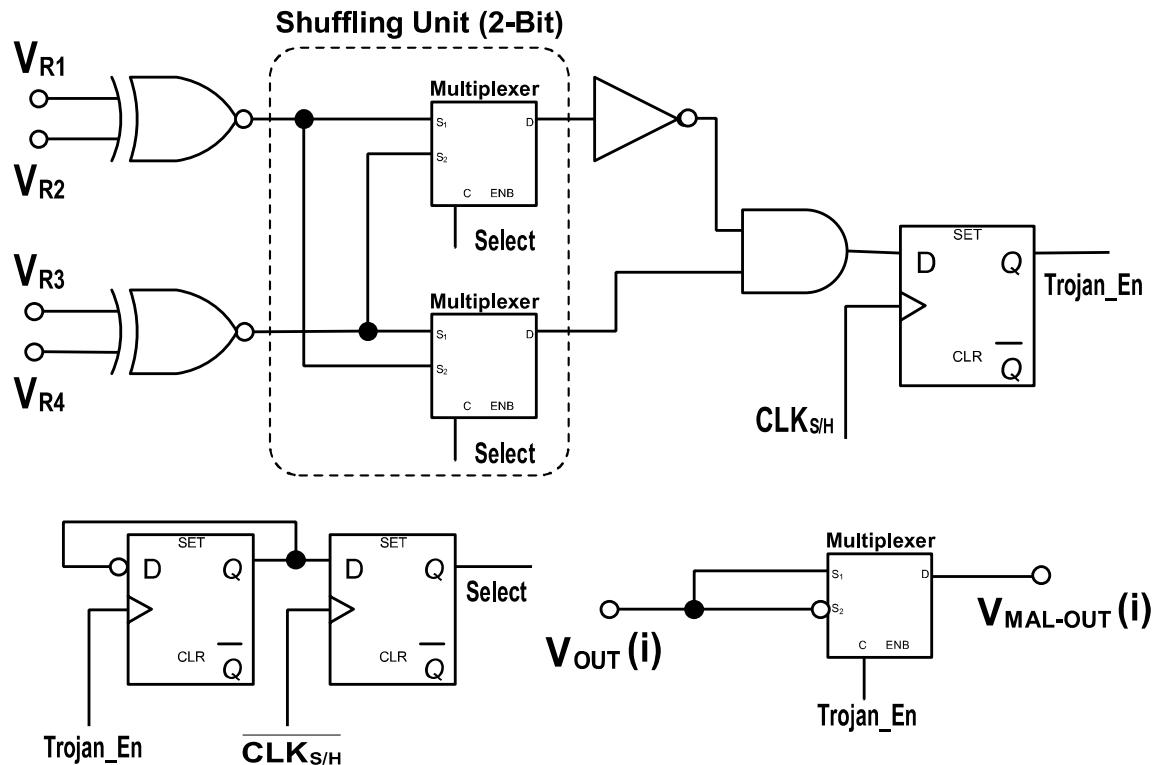
The timing diagram of the ADC operation.



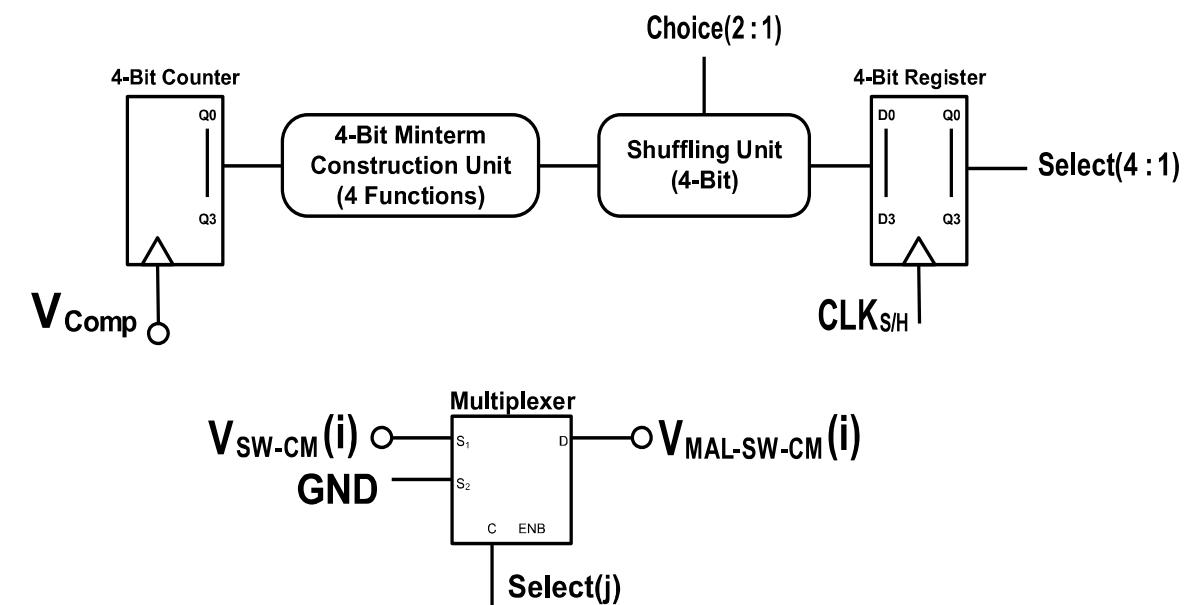
The operation flow of the original analog to digital converter (ADC) circuit.

HS: Security of Analog to Digital Converter - SAR ADC: Threat Models

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

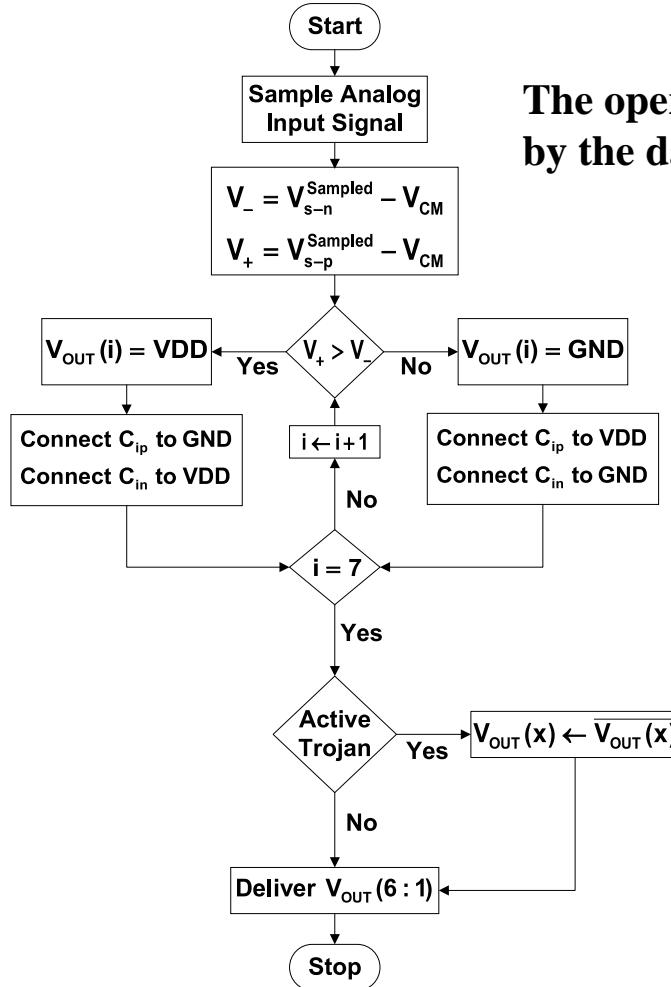


The Trojan circuit for the datapath-based threat.

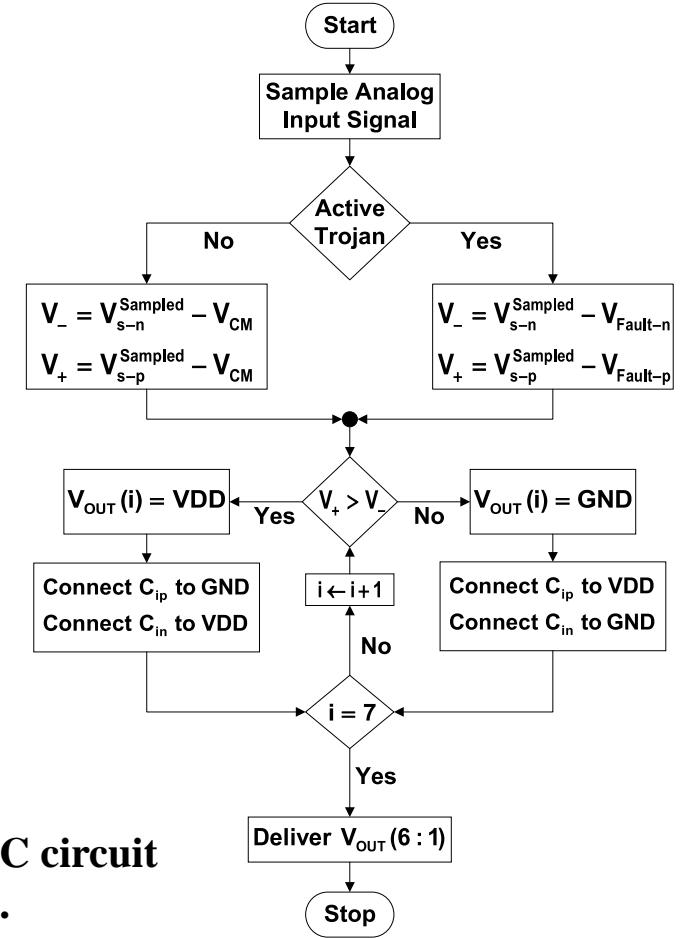


The Trojan circuit for the control-based threat.

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



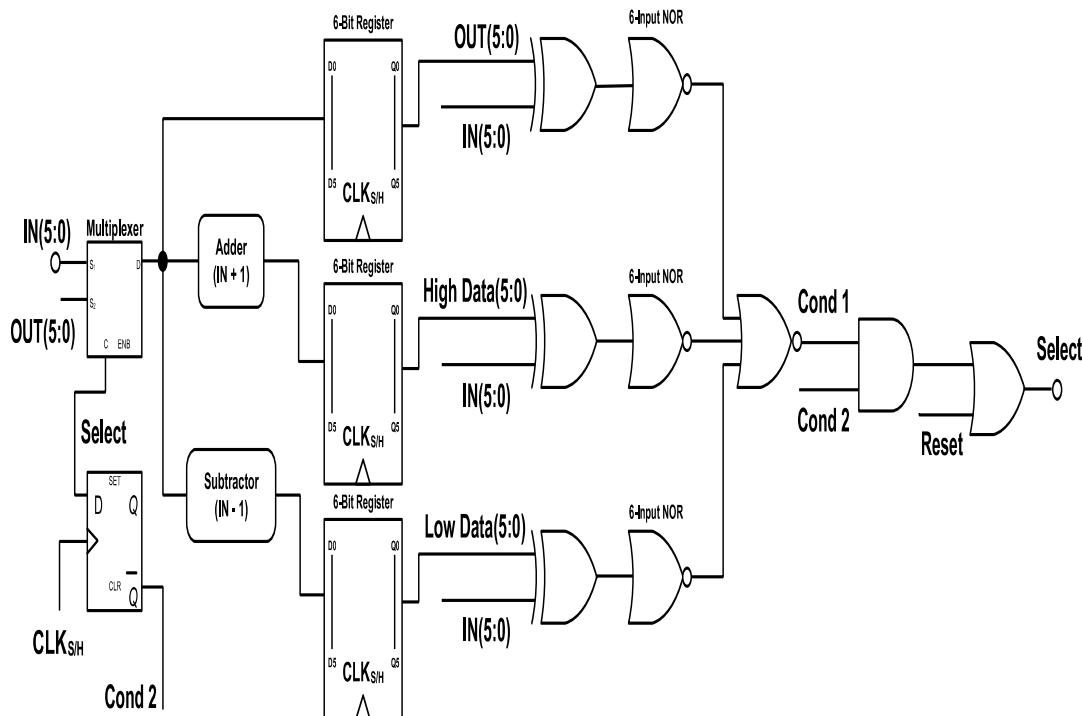
The operation flow of the infected ADC circuit by the datapath-based Hardware Trojan.



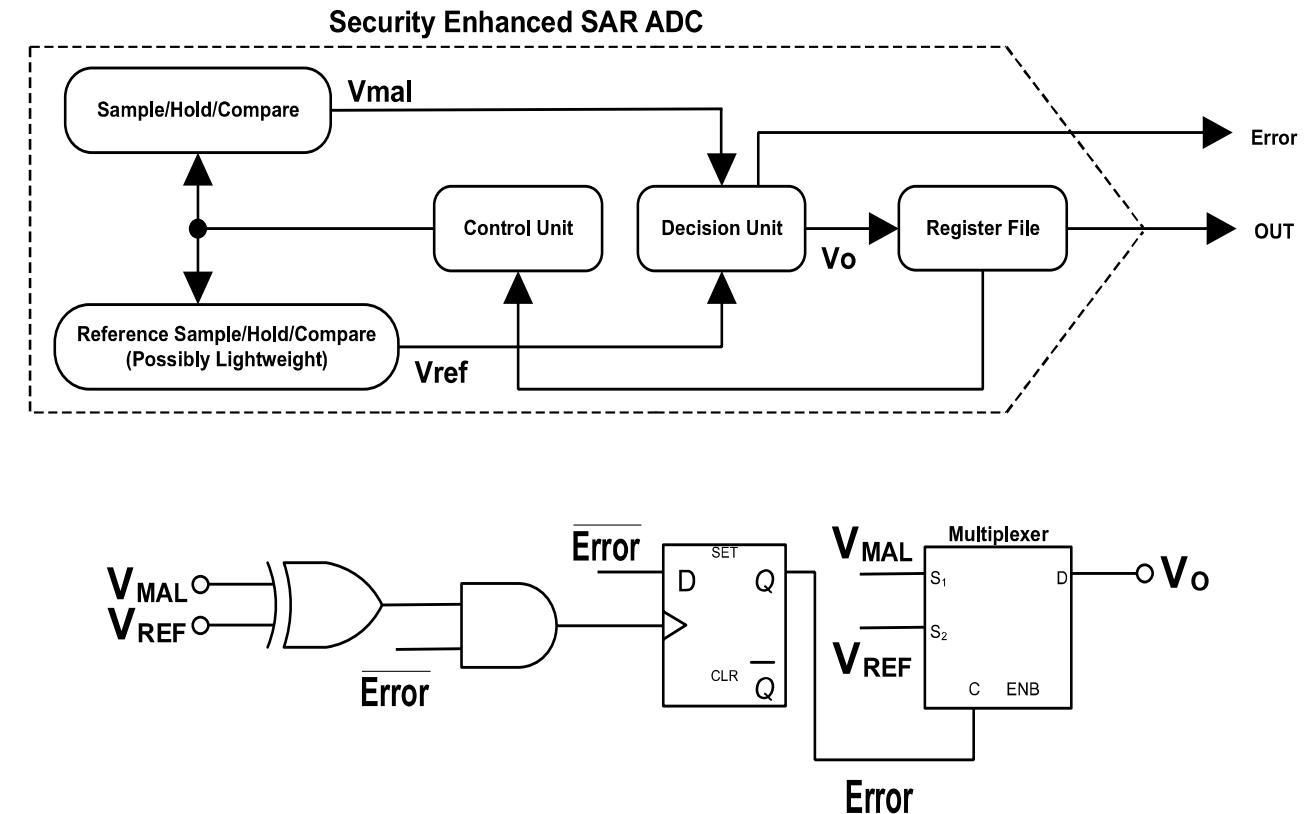
The operation flow of the secured ADC circuit by the datapath-based defense circuit.

HS: Security of Analog to Digital Converter - SAR ADC: Defense Models

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

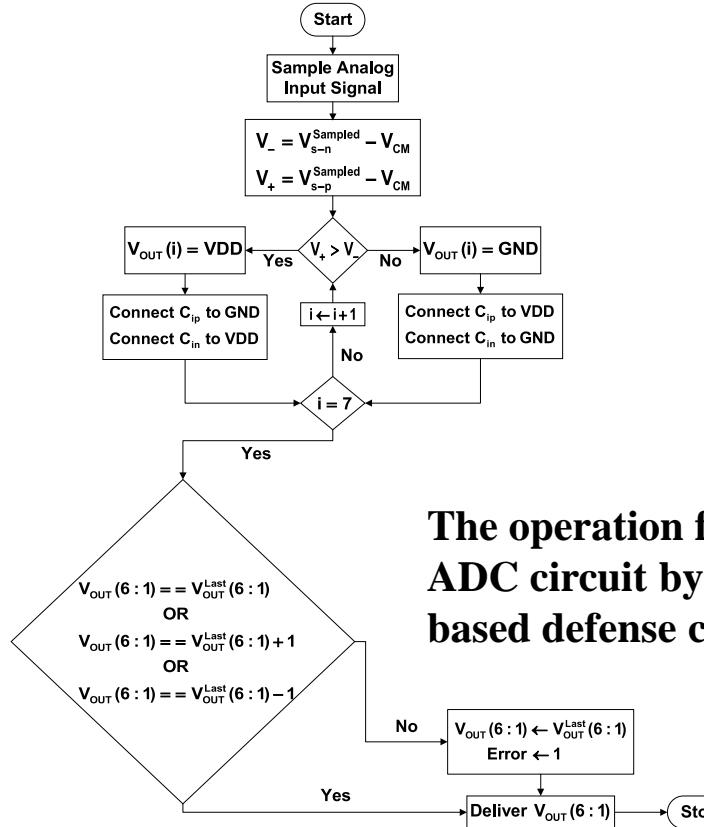


The defense circuit for the datapath-based countermeasure.

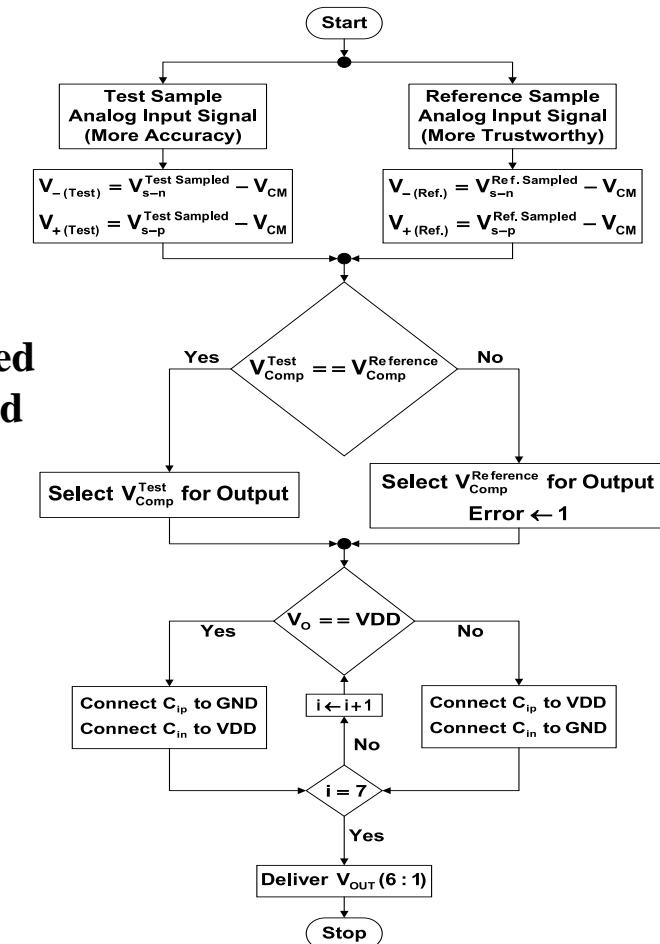


The defense circuit for the control-based countermeasure.

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

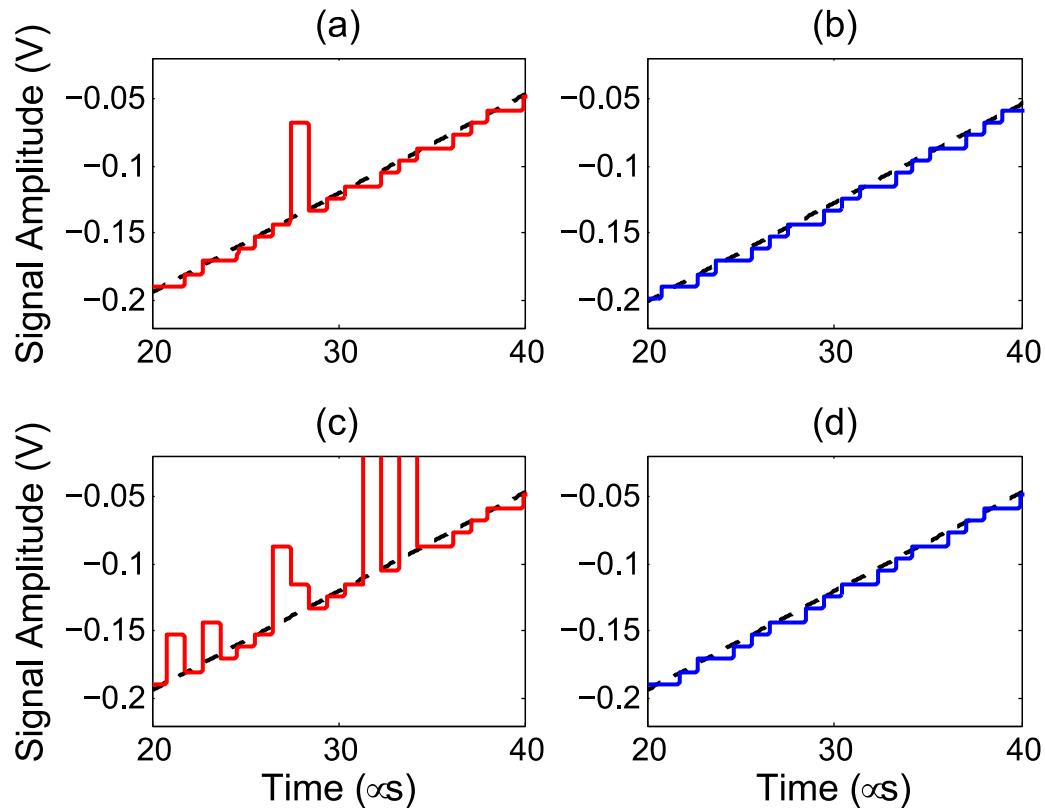


The operation flow of the infected ADC circuit by the control-based Hardware Trojan.

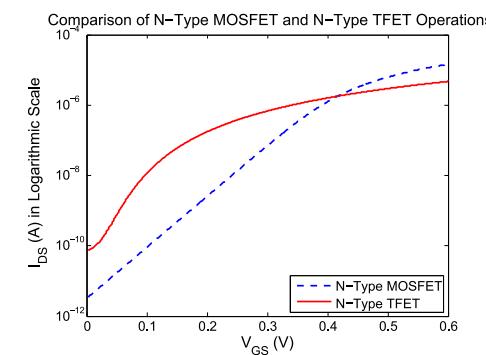
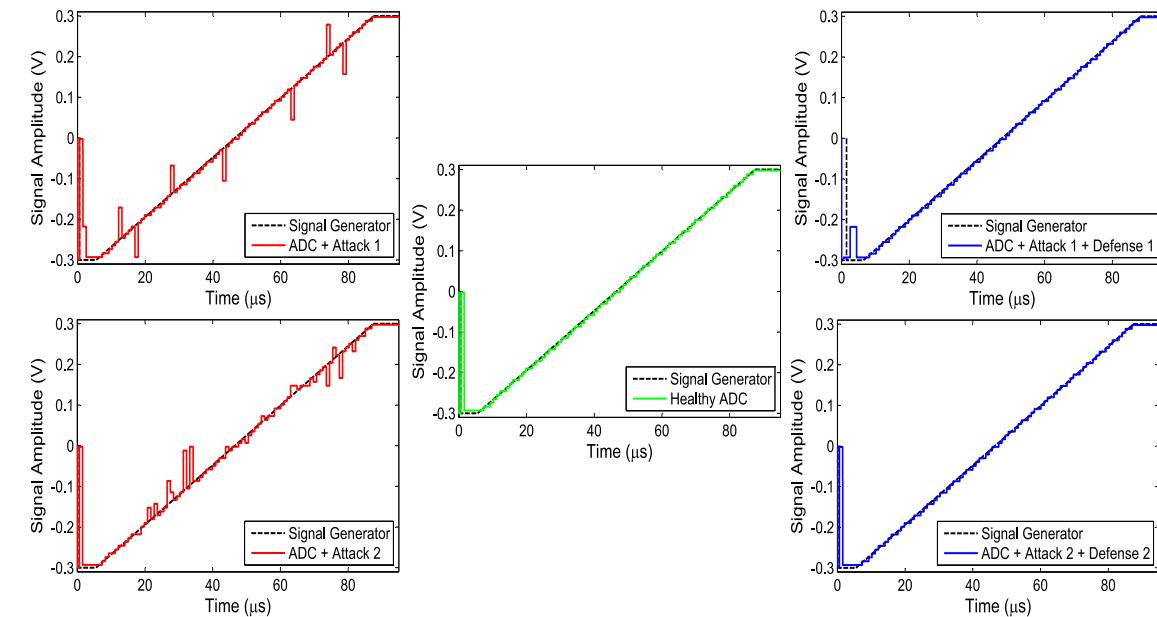


HS: Security of Analog to Digital Converter - Behavior of Normal and Infected SAR ADC using Tunnel Field-Effect Transistor (TFET)

- 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

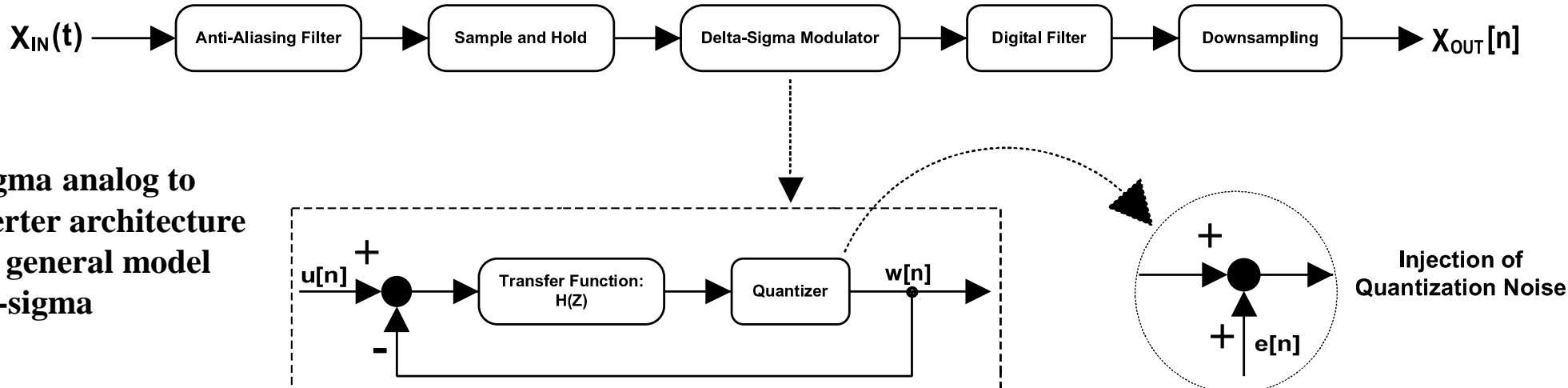
20

- 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 (μ W)	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	3.34	3.74	43.26	277.7	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}

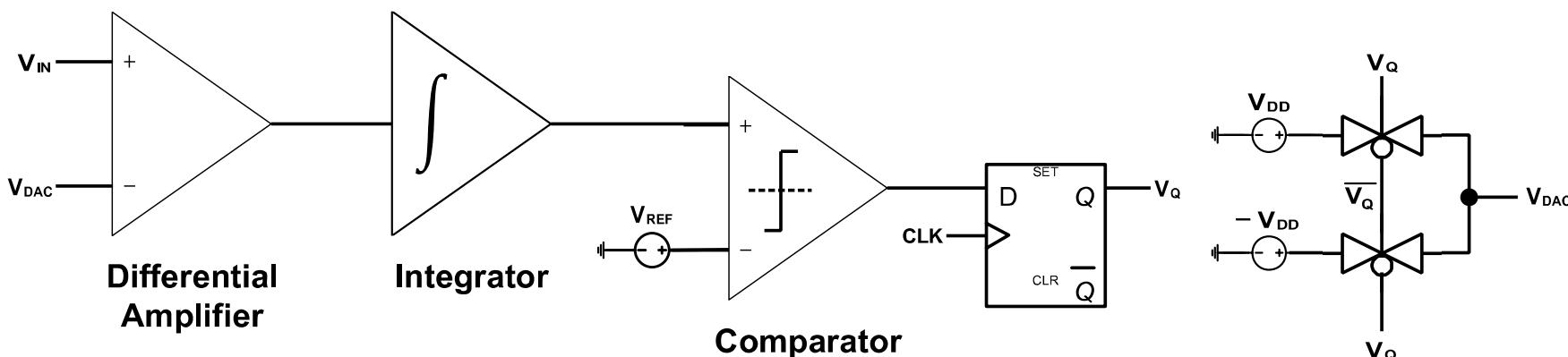
HS: Security of Analog to Digital Converter - Delta-Sigma analog to digital converter: Architecture and Operation

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

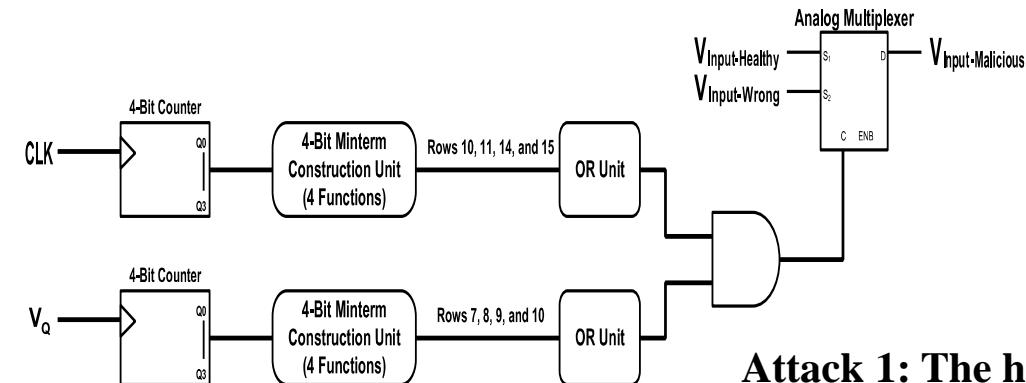


The delta-sigma analog to digital converter architecture along with a general model for the delta-sigma modulator.

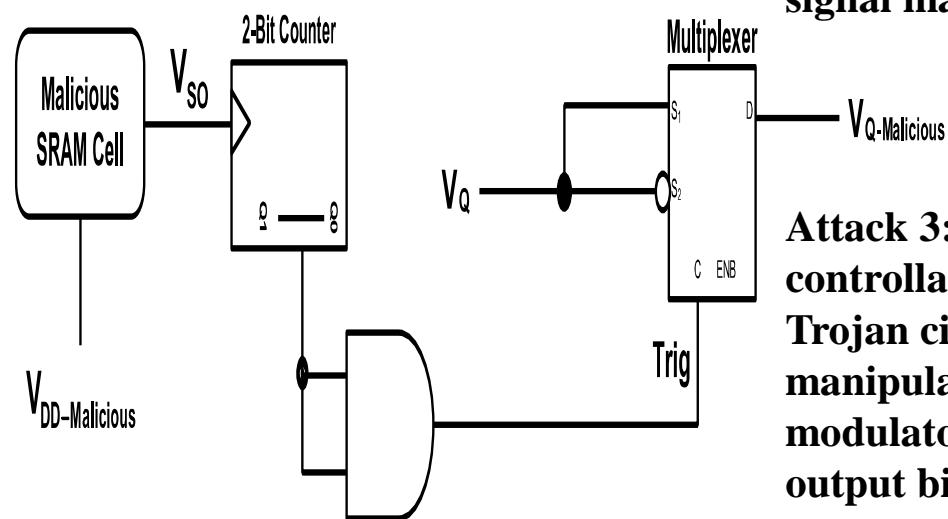
The delta-sigma modulator circuit.



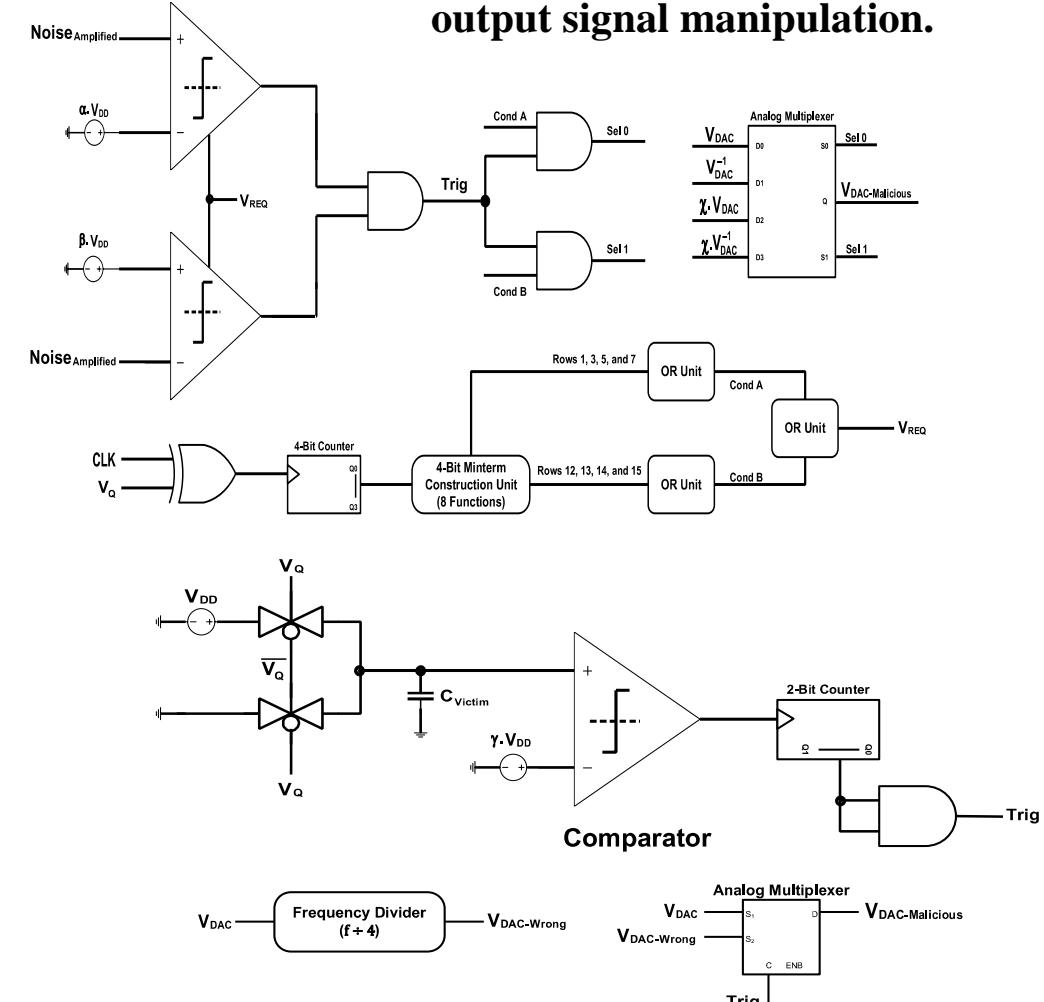
- Delta-Sigma analog to digital converter: Threat Models



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



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

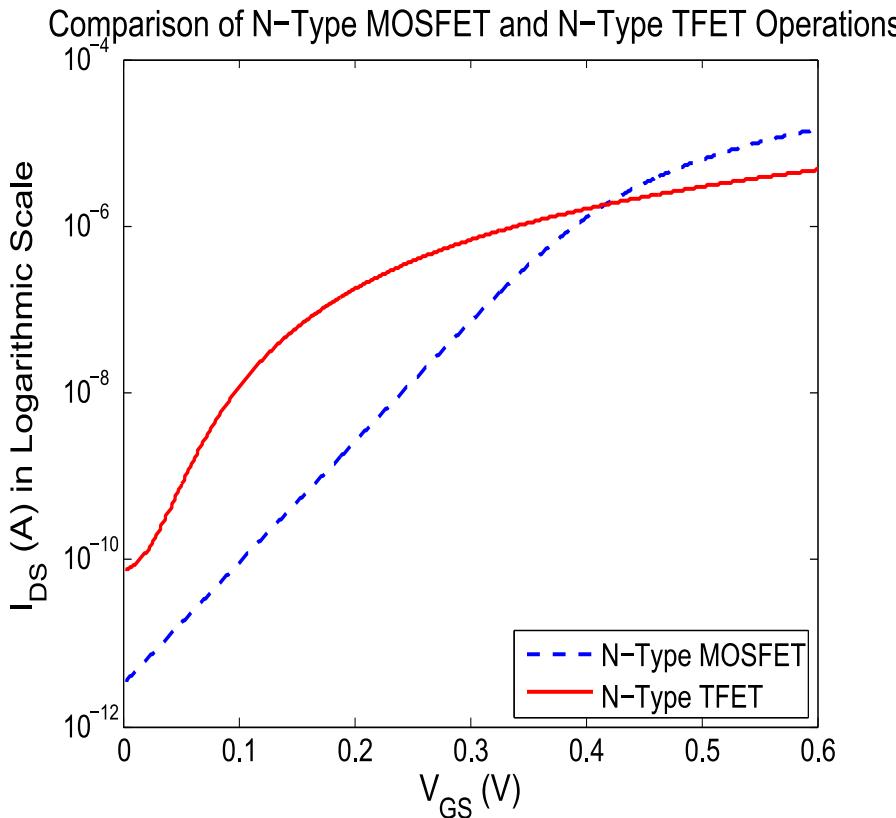


Attack 4: The capacitor-triggered hardware Trojan circuit for the output signal manipulation.

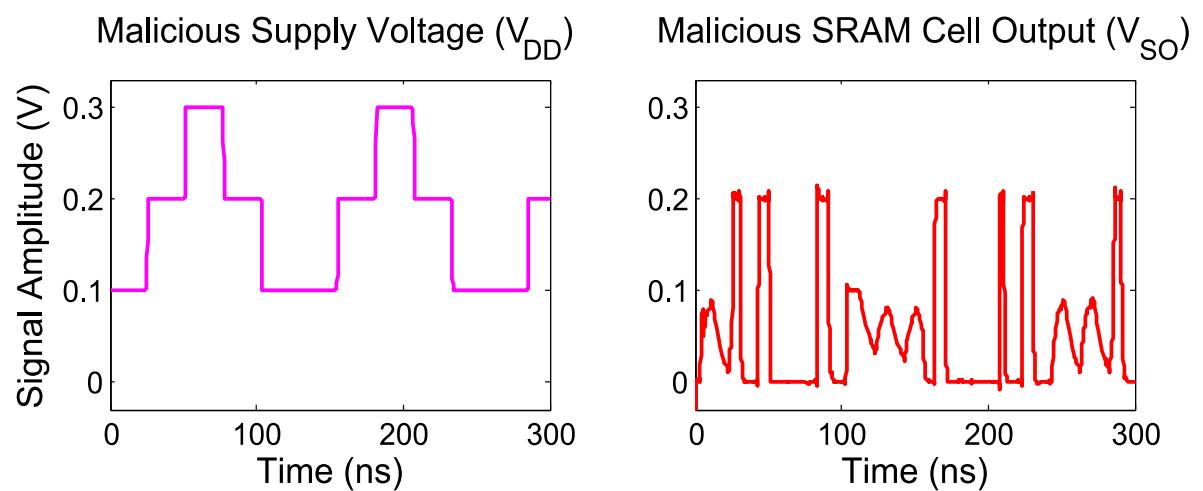
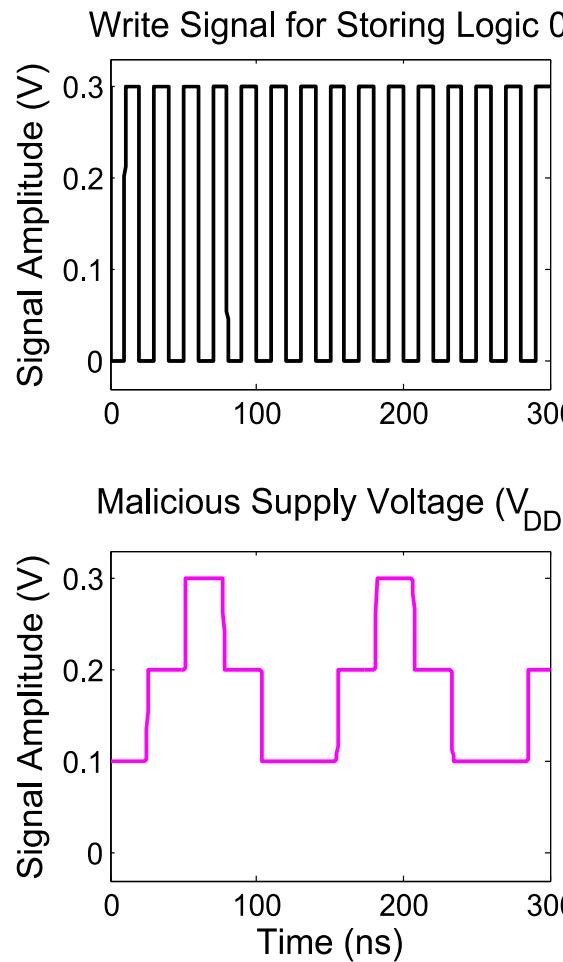
HS: Security of Analog to Digital Converter - Delta-Sigma ADC: Malicious SRAM Cell Operation using TFET

23

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

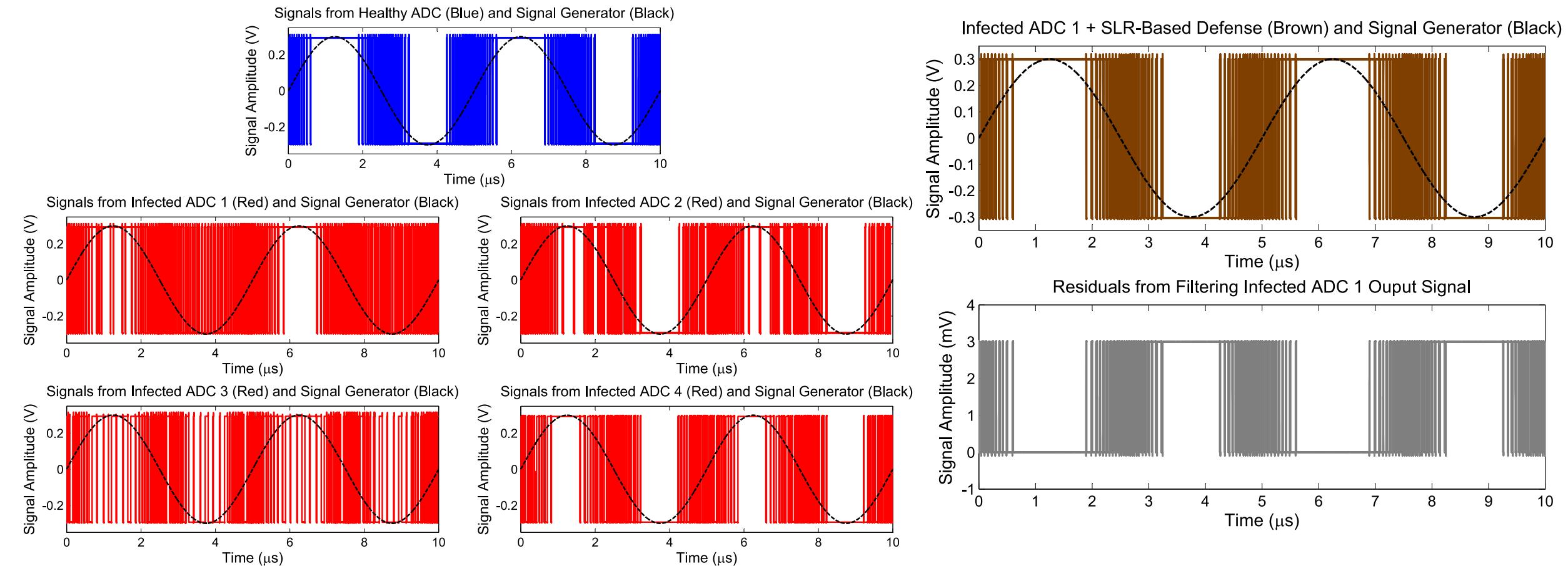


The circuit operation of malicious SRAM cell.



HS: Security of Analog to Digital Converter - Behavior of Normal and Infected Delta-Sigma ADC using Tunnel Field-Effect Transistor (TFET)

- 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.

- 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)

- 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

27

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

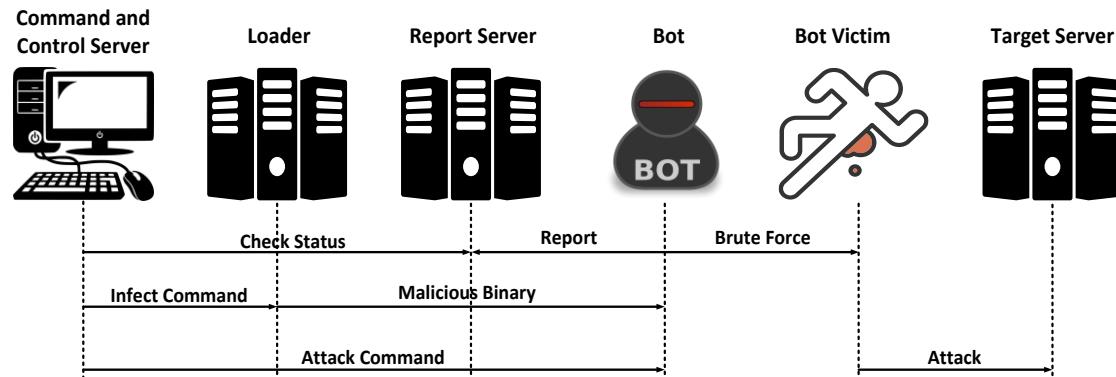
Type of Filter	Modified R-Squared Value for ADC Under Attack 1 (Triangle Input)	Modified R-Squared Value for ADC Under Attack 2 (Triangle Input)	Modified R-Squared Value for ADC Under Attack 3 (Triangle Input)	Modified R-Squared Value for ADC Under Attack 4 (Triangle Input)
Simple Linear Regression	0.9941	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)

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.

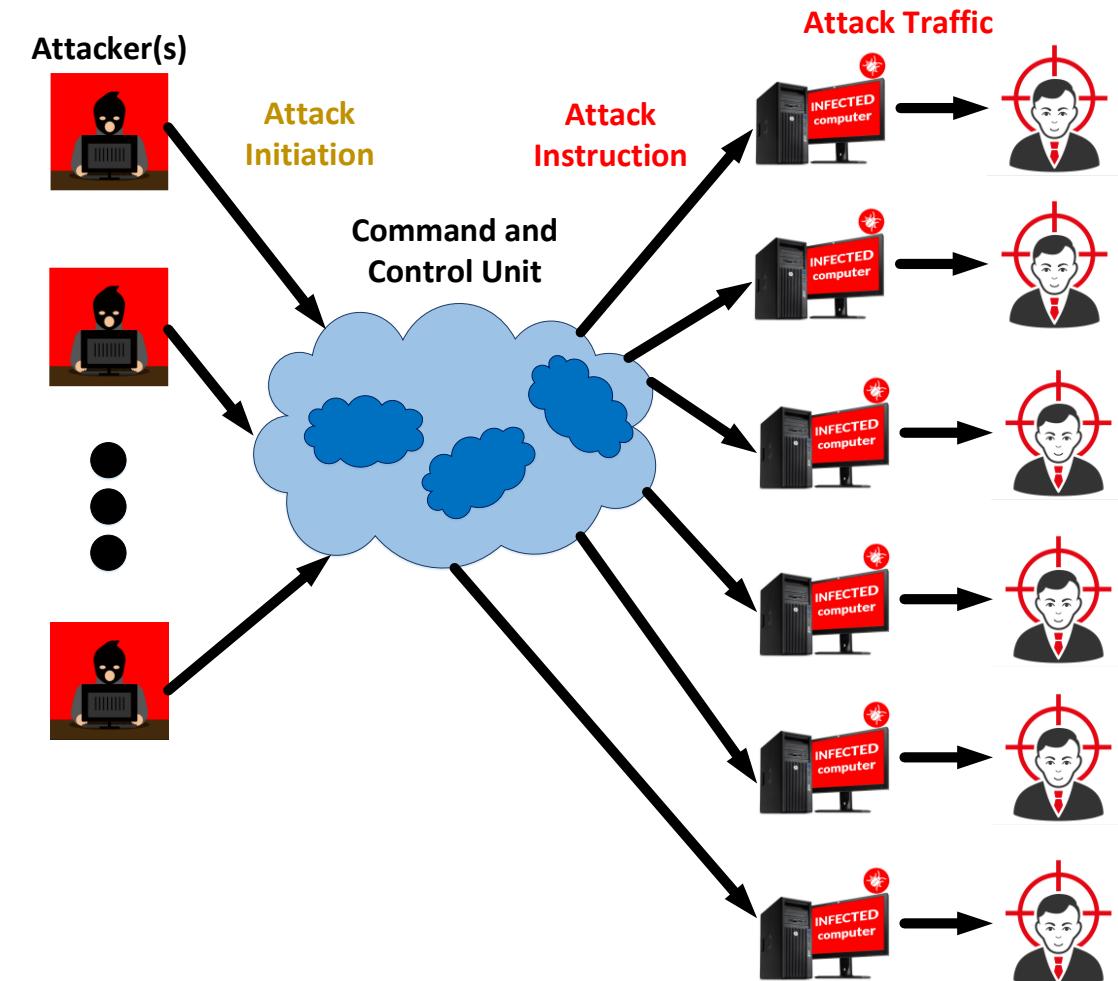
- Overview and Introductory Material
- The Classification Measures
- Network Traffic Data Acquisition
- System Architecture and Operation
- System Performance and Results

AIS: Leveraging Transfer Learning and Data Transformation in Botnet Detection - Overview and Introductory Material

- Overview and Introductory Material
- 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.



- The Classification Measures

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

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

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

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

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

AIS: Leveraging Transfer Learning and Data Transformation in Botnet Detection – Network Traffic Data Acquisition

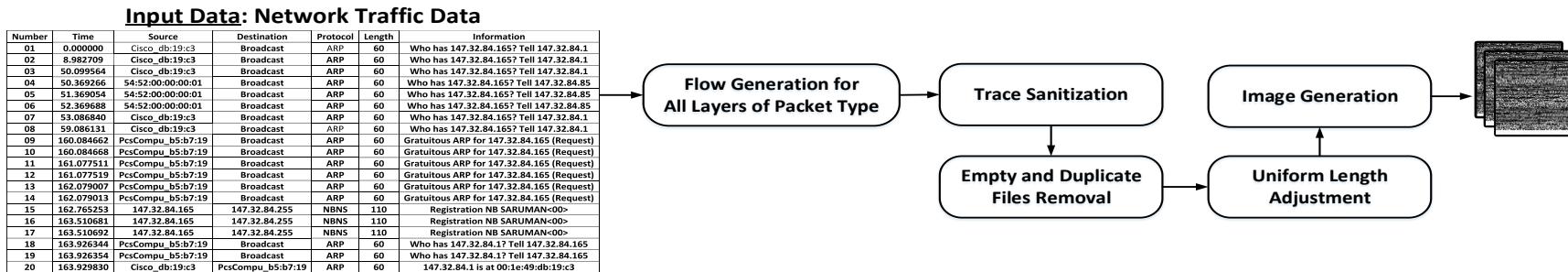
- Network Traffic Data Acquisition
- A snapshot of the captured traffic data from a local residential network.

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.dsrb.akamaied...	DESKTOP-67H3GJI	TLSv1.2	85	Encrypted Alert
6511	103.357733	e1539.dsrb.akamaied...	DESKTOP-67H3GJI	TCP	54	https(443) → 63548 [FIN, ACK] Seq=3773 Ack=562 Win=31360 Len=0
6512	103.358124	DESKTOP-67H3GJI	e1539.dsrb.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

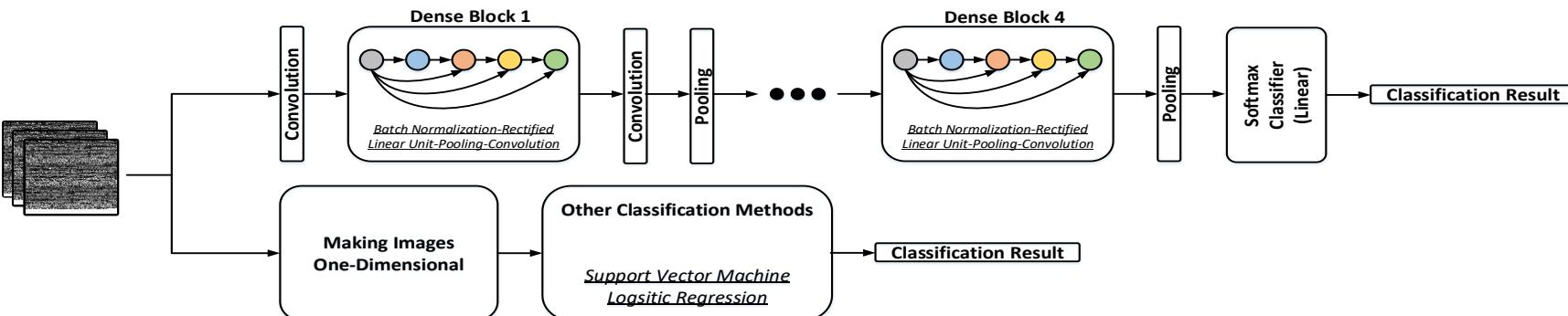
A snapshot of the captured traffic data from a local residential network.

AIS: Leveraging Transfer Learning and Data Transformation in Botnet Detection – System Architecture and Operation

- System Architecture and Operation



(a)

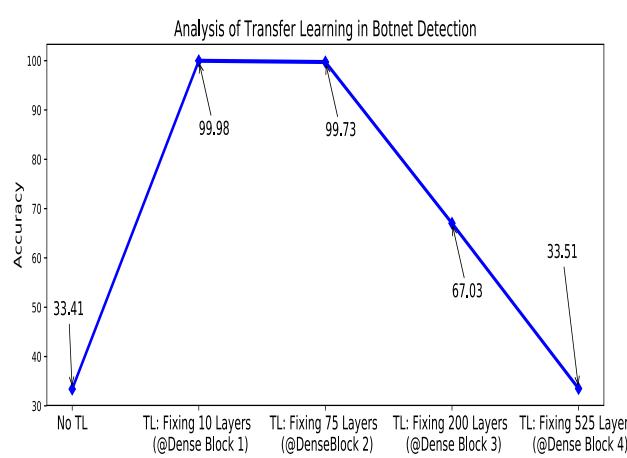
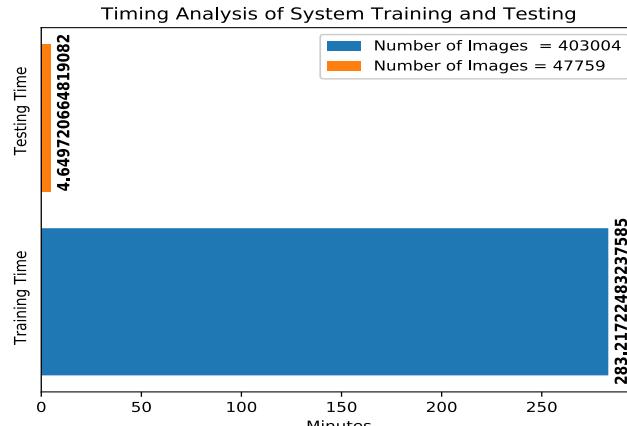


(b)

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).

AIS: Leveraging Transfer Learning and Data Transformation in Botnet Detection – System Perf. and Results

■ System Performance and Results

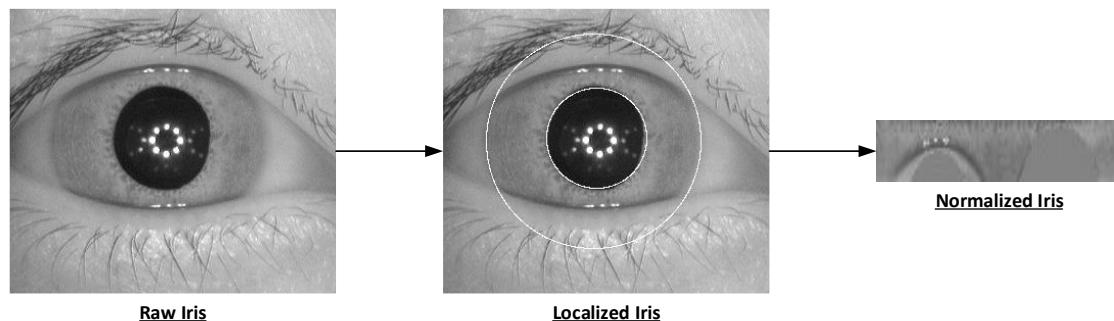
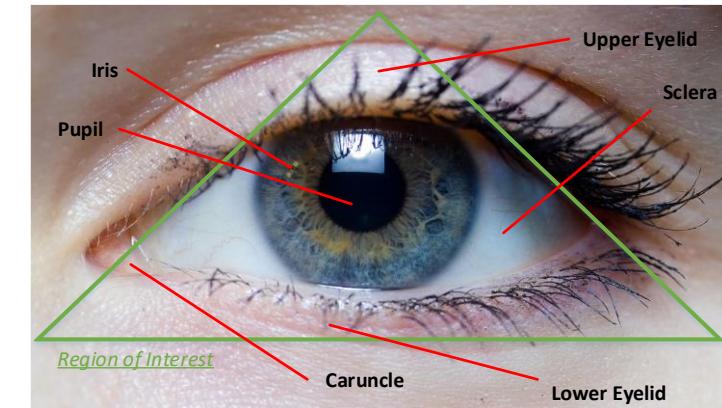
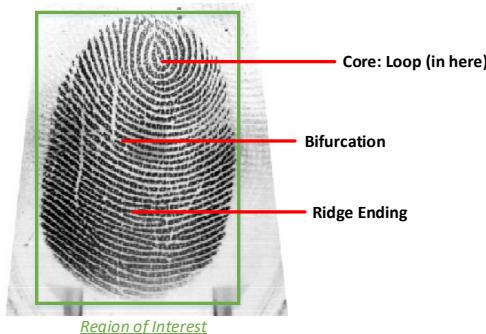


Score Merit	Accuracy (%)	Area under Curve (%)	Precision (%)	Recall (%)	F-1 Score (%)	Kappa (%)
CNN: DenseNet (Best Result) (for CTU-13 Dataset)	99.98	99.99	99.98	99.98	99.98	99.96
Support Vector Machine (SVM) (for CTU-13 Dataset)	83.15	75.80	80.73	98.07	88.56	57.72
Logistic Regression (for CTU-13 Dataset)	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 [6], 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

- Overview and Introductory Material
- System Architecture and Operation
- System Performance and Results

AIS: Design of Privacy-Preserving and Secure Biometric Recognition System – Overview and Introductory Material

- Overview and Introductory Material
- Background: Fingerprint and Iris Recognition

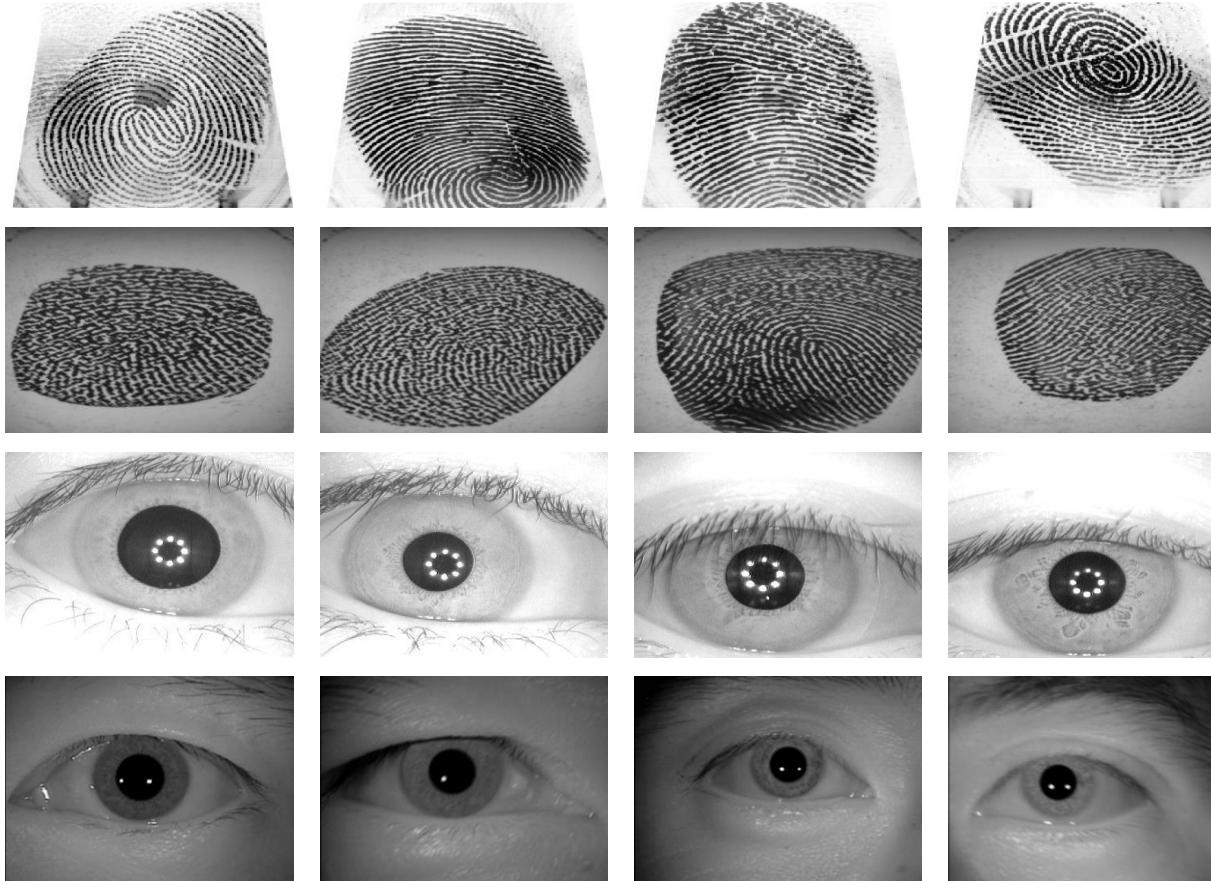


- ✓ An authentic fingerprint.
- ✓ A human eye.
- ✓ The raw, localized, and normalized version of an iris.

AIS: Design of Privacy-Preserving and Secure Biometric Recognition System – Overview and Introductory Material

36

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

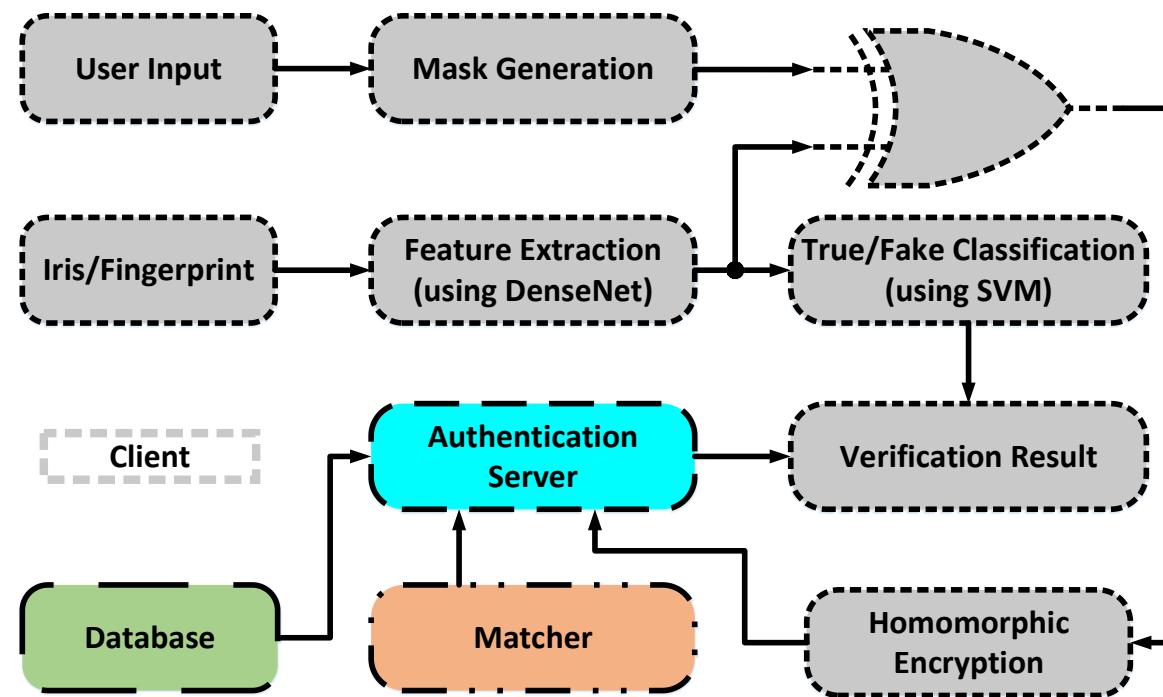


The samples of true fingerprint (first row), fake fingerprint (second row), true iris (third row), and fake iris (fourth row).

AIS: Design of Privacy-Preserving and Secure Biometric Recognition System – System Architecture and Operation

37

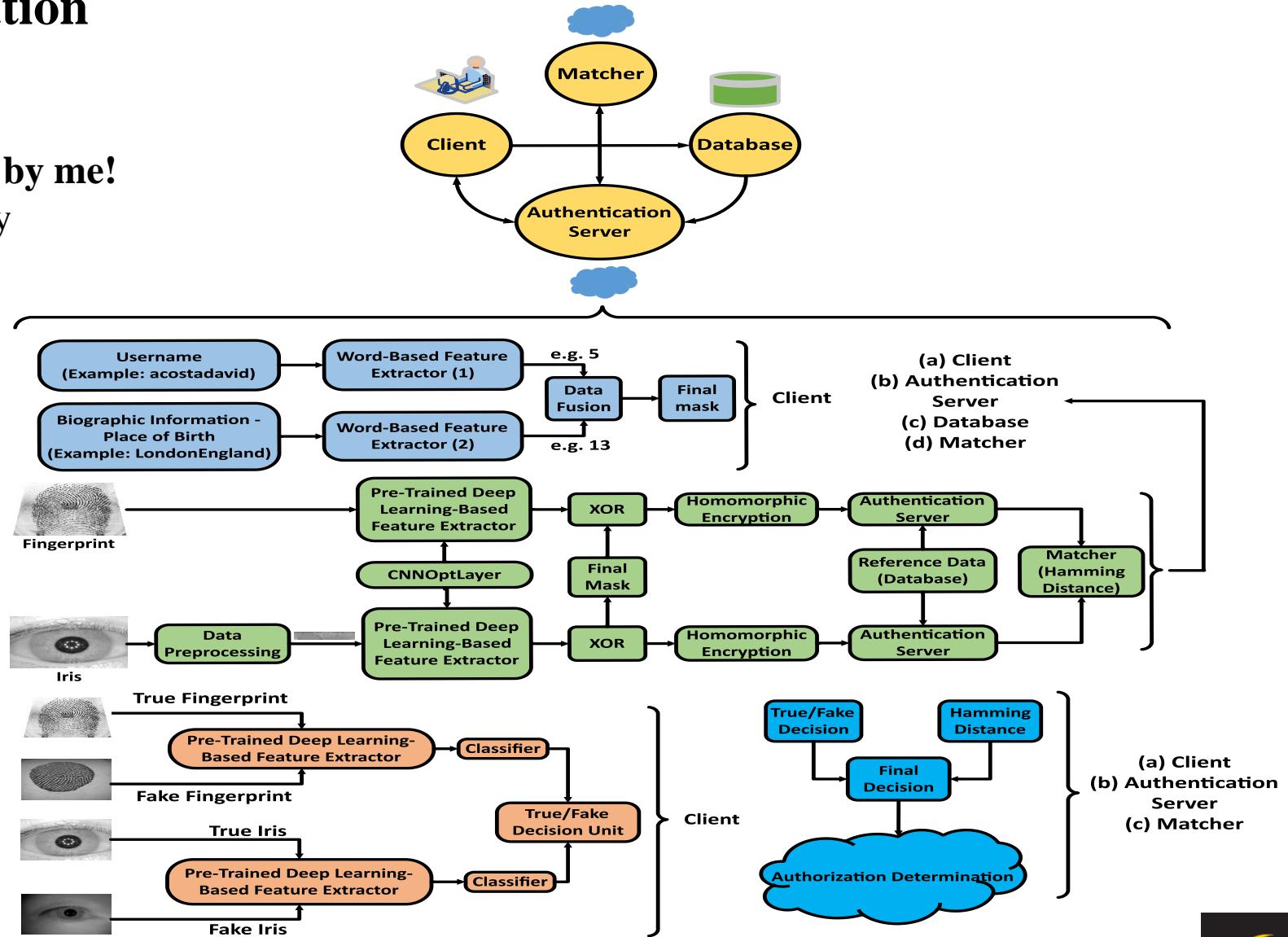
- System Architecture and Operation
- The proposed deep learning-based privacy preserving biometric recognition system.



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

AIS: Design of Privacy-Preserving and Secure Biometric Recognition System – System Architecture and Operation

- 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.

AIS: Design of Privacy-Preserving and Secure Biometric Recognition System – System Performance and Results

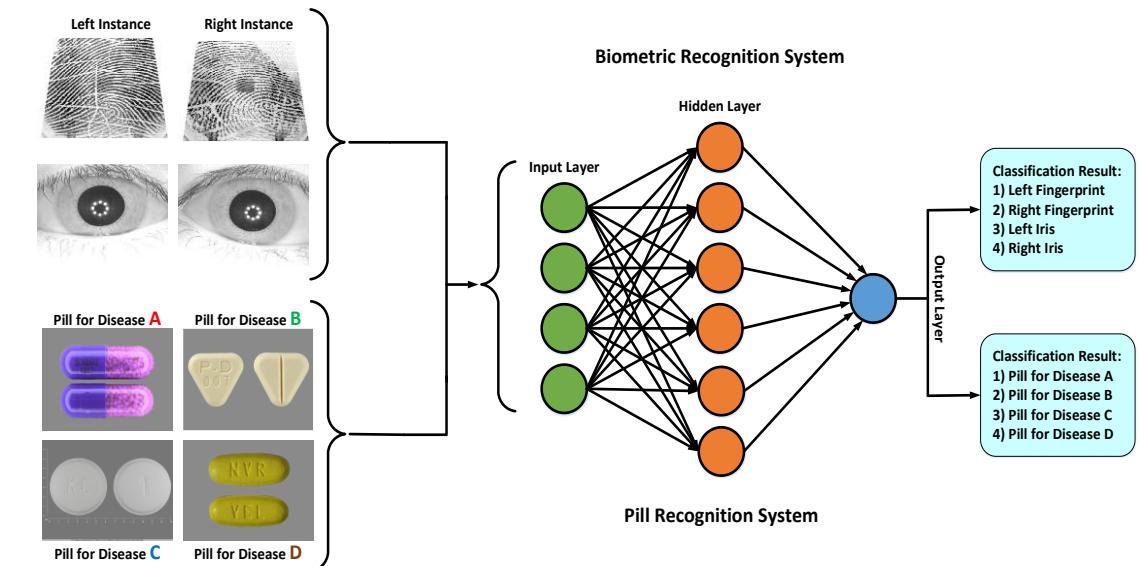
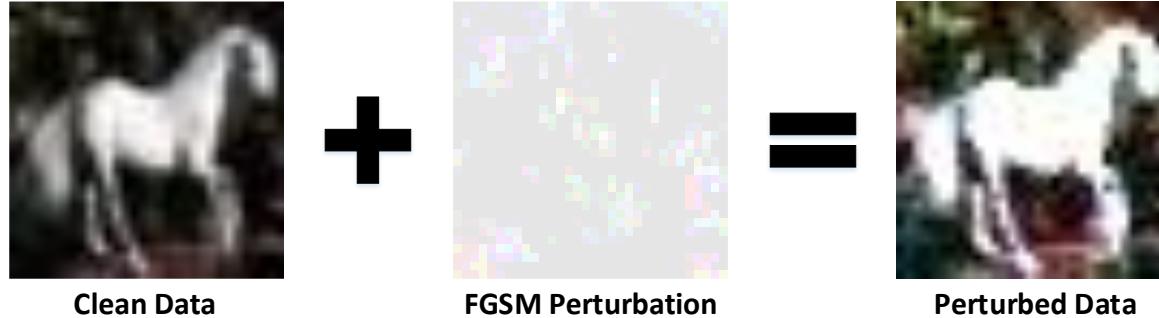
- **System Performance and Results**
- **The simulation results of performance evaluation of biometric recognition system.**

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

- Overview and Introductory Material
- System Architecture and Operation
- System Performance and Results

AIS: Detection of Adversarial Examples Using Adversarial Training and Noise Training - Overview and Introductory Material

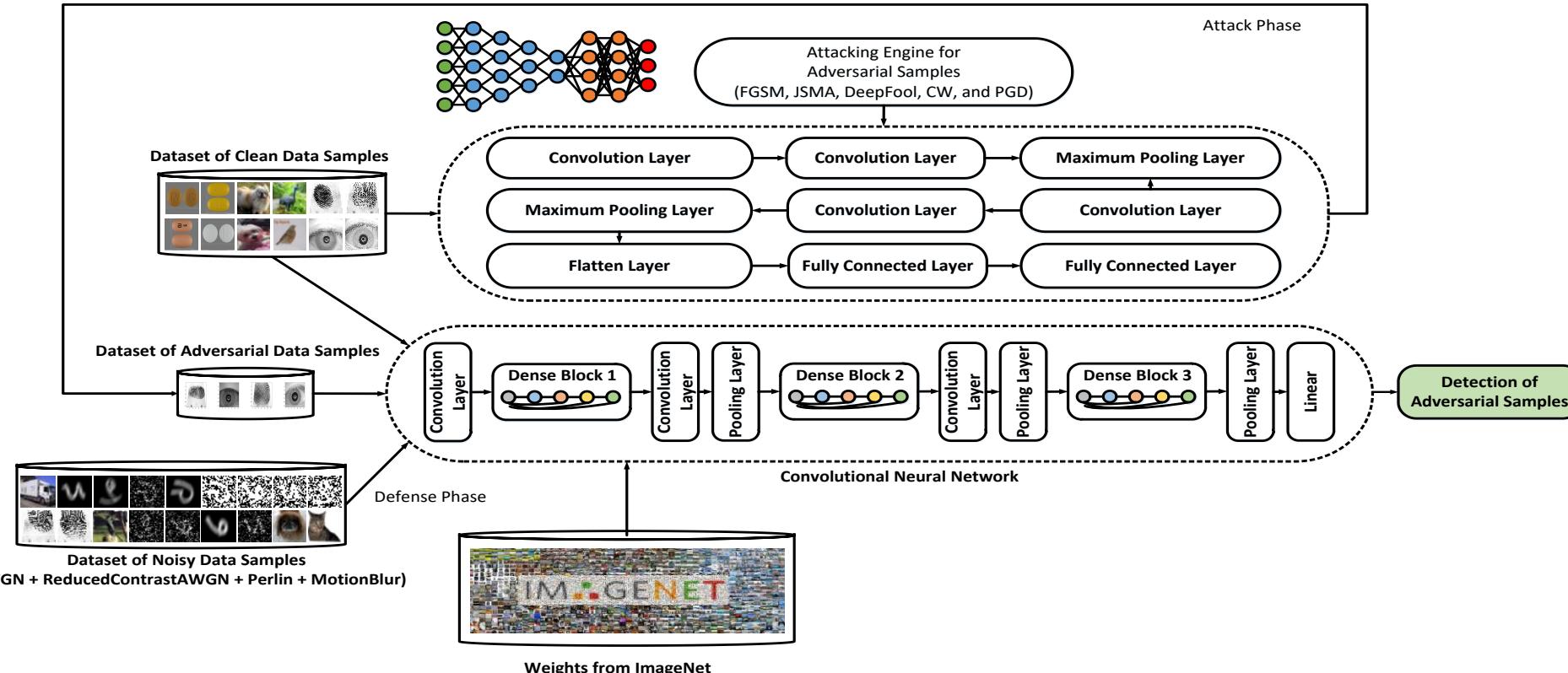
- Overview and Introductory Material
- More than 98% of the work completed by me!
- Applying Fast Gradient Sign Method (FGSM) on an image and pill/biometric data recognition structure:



- ✓ 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 - System Architecture and Operation

- System Architecture and Operation
- 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.

AIS: Detection of Adversarial Examples Using Adversarial Training and Noise Training - System Performance and Results

- System Performance and Results
- 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	
				Clean Data + Adversarial Data	Clean Data + Adversarial Data + Noisy 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)	

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 $\Delta\Sigma$ 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*.
11. Chakraborty, R.S., Narasimhan, S. and Bhunia, S., 2009, November. Hardware Trojan: Threats and emerging solutions. In 2009 IEEE International high level design validation and test workshop (pp. 166-171). IEEE.
12. Rostami, M., Koushanfar, F. and Karri, R., 2014. A primer on hardware security: Models, methods, and metrics. *Proceedings of the IEEE*, 102(8), pp.1283-1295.