Resilience and Security in Cyber-Physical Systems: Self-Driving Cars and Smart Devices

Karthik Pattabiraman

Guanpeng (Justin) Li, Maryam Raiyat, Amita Kamath, Mohammad Rafiuzzaman, Kumseok Jung, Julien Gascon-Samson

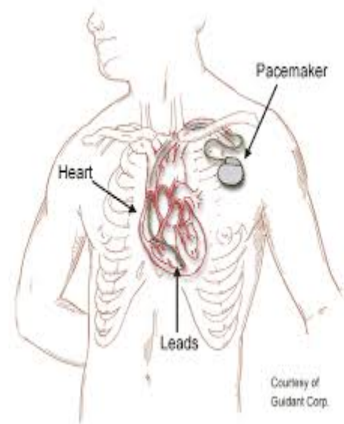
with

Siva Hari, Michael Sullivan, Tim Tsai, Joel Emer, Steve Keckler

My Research

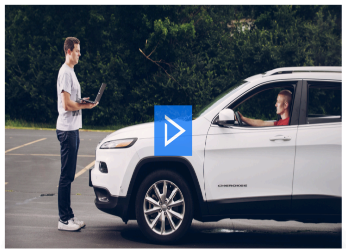
- **Building error resilient and secure software systems**
- **Three main areas:**
  – Software resilience techniques [SC'17][DSN'17][SC'16][DSN'16][DSN'15][DSN'14][DSN'13][DSN'12]
  – Web applications' reliability [ASE'17][ICSE'16][ICSE'15][ICSE'14A][ICSE'14B][ASE'14][ASE'15]
  – CPS Security [FSE'17][ACSAC'16][EDCC'15][HASE'14]
- **This talk**
  – CPS Security and Resilience

Cyber-Physical Systems (CPS)



Cyber-Physical Systems (CPS)

HACKERS REMOTELY KILL A JEEP ON THE HIGHWAY—WITH ME IN IT



Smart meters can be hacked to cut power bills

By Mark Ward
Technology correspondent, BBC News

16 October 2014 | Technology

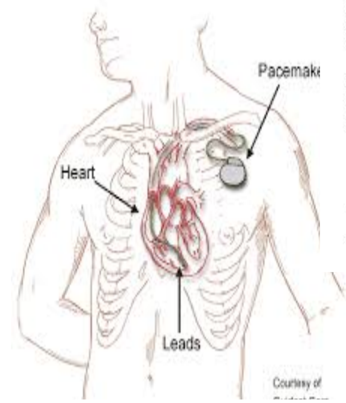
Share

Smart meters could help people do a better job of managing power use

Smart meters widely used in Spain can be hacked to under-report energy use, security researchers have found.

Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses

Daniel Halperin[†] University of Washington; Thomas S. Heydt-Benjamin[†] University of Massachusetts Amherst; Benjamin Ransford[†] University of Massachusetts Amherst; Shane S. Clark University of Massachusetts Amherst; Benessa Defend University of Massachusetts Amherst; Will Morgan University of Massachusetts Amherst; Kevin Fu, PhD* University of Massachusetts Amherst; Tadayoshi Kohno, PhD* University of Washington; William H. Maisel, MD, MPH* BIDMC and Harvard Medical School



Nest Thermostat Glitch Leaves Users in the Cold

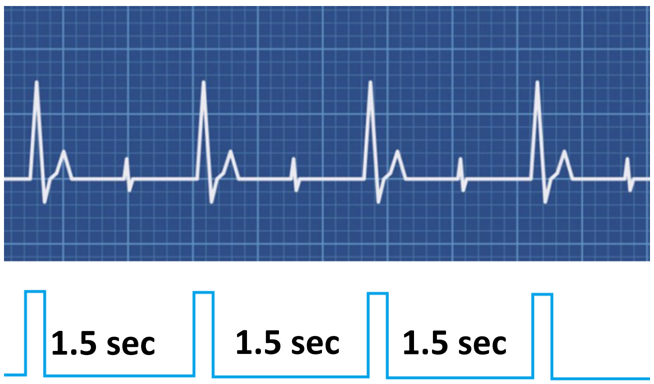
Disruptions
By NICK BILTON JAN. 13, 2016

The Nest Learning Thermostat is dead to me, literally. Last week, my once-beloved "smart" thermostat suffered from a mysterious software bug that drained its battery and sent our home into a chill in the middle of the night.

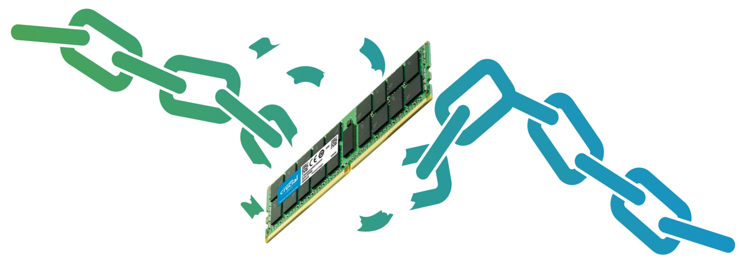
Although I had set the thermostat to 70 degrees overnight, my wife and I were woken by a crying baby at 4 a.m. The thermometer in his room read 64 degrees, and the Nest

CPS Challenges

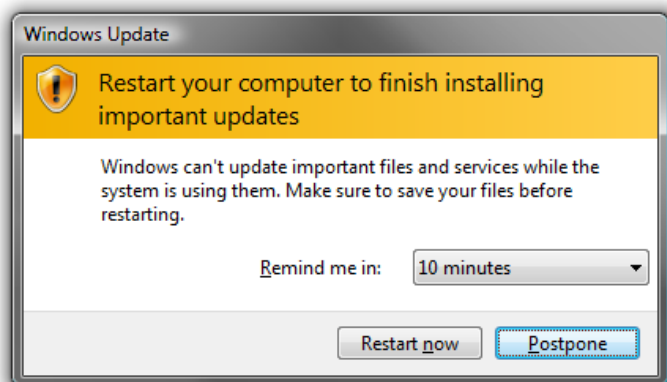
Real-time constraints



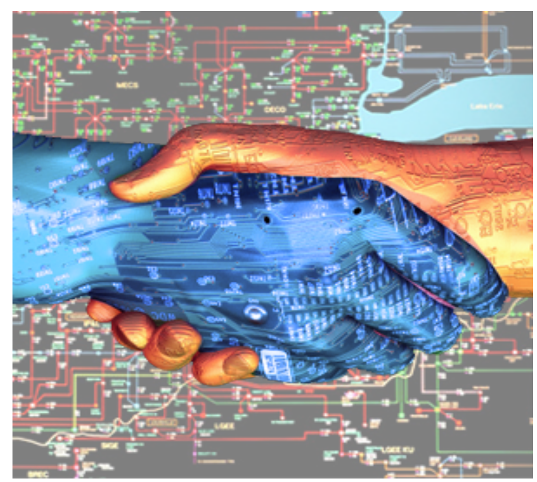
Resource constraints



Hard to Upgrade



No human-in-the-loop



This Talk

- Motivation
- Resilience of Deep Neural Networks in Self-Driving Cars from Soft Errors [SC'17 – to appear]
- Intrusion Detection Systems for Smart Embedded Devices using Dynamic Invariants [FSE'17]
- Ongoing work and conclusion

DNNs in Self-Driving Cars

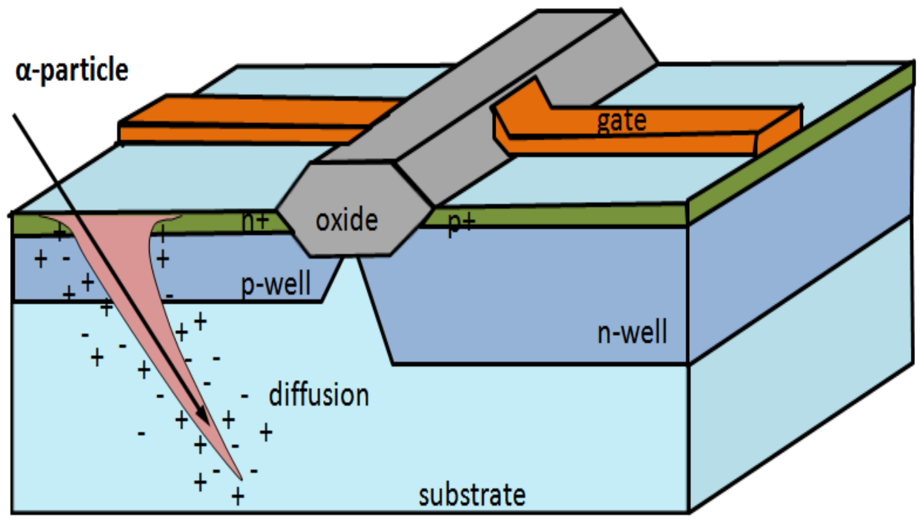
DNN applications are widely deployed in safety critical applications

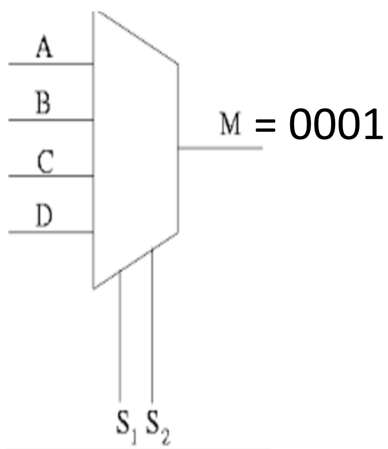
- autonomous-driving cars – specialized accelerators for real-time processing

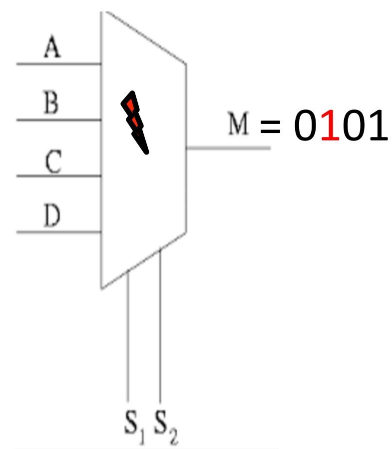
Silent Data Corruptions (SDCs)

- Results in wrong prediction of DNN application
- Safety standard requires SoC FIT<10 overall (ISO 26262)

Soft Errors

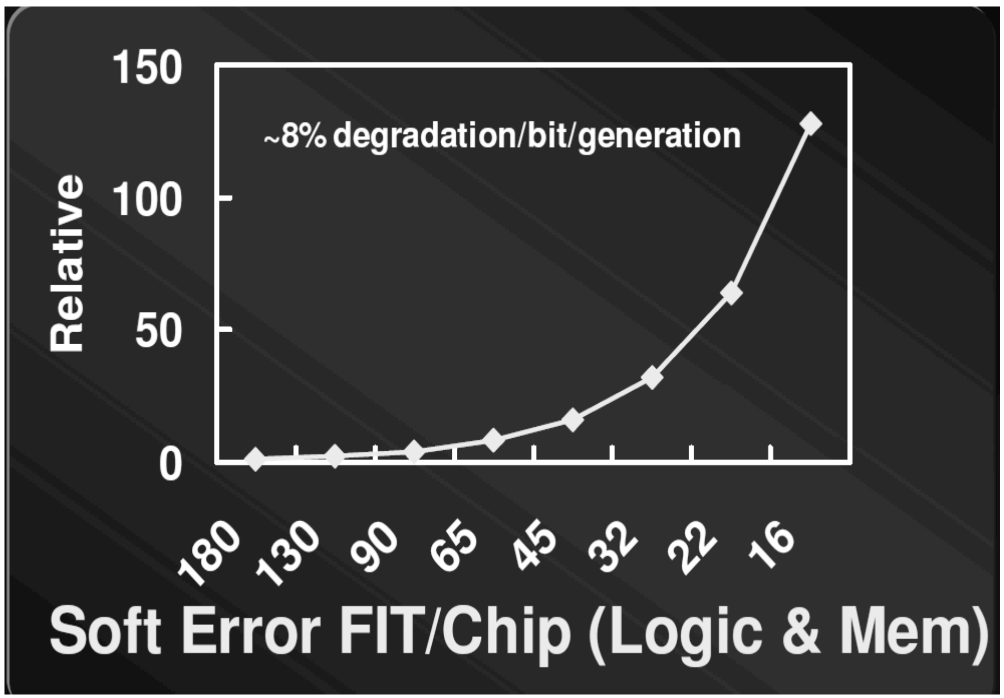

α-particle
gate
oxide
n+
p+
p-well
n-well
diffusion
substrate


A
B
C
D
M = 0001
S_1 S_2


A
B
C
D
M = 0101
S_1 S_2

Soft Error Problem

- Soft errors are increasing in computer systems



Source: Shekar Borkar (Intel) - Stanford talk

Current Solutions

Traditional Solutions

DMR for all latches in execution units

ECC/Parity on all storage elements

Incurs high overhead

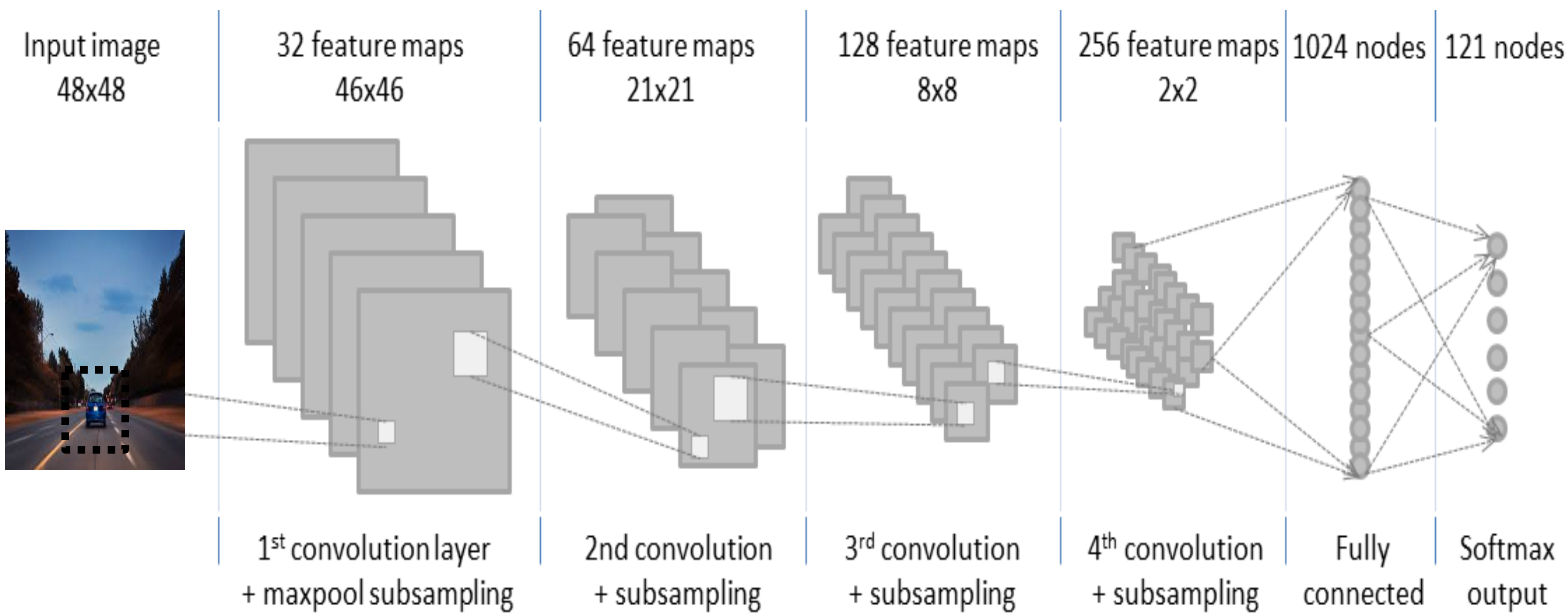
Recent Work

Generic micro-architectural solutions

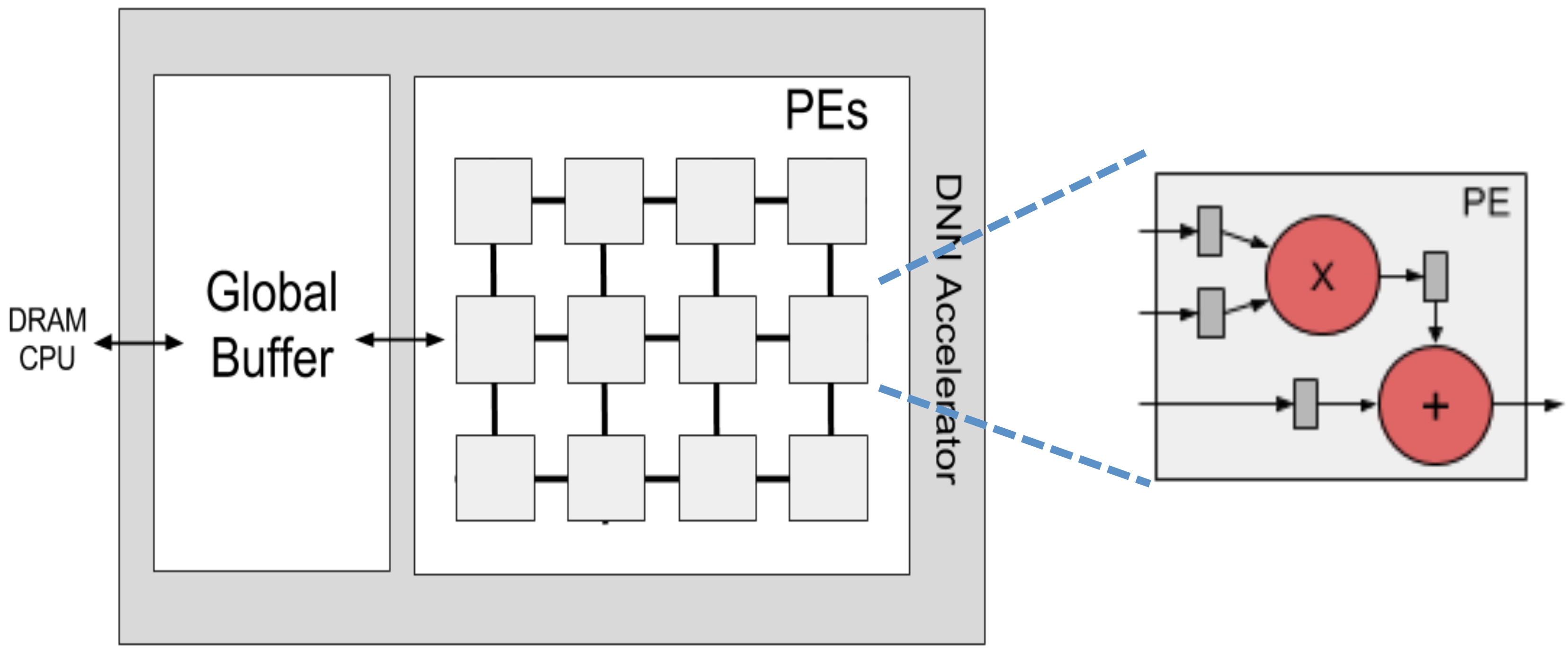
DNN-algorithm agnostic

Nonoptimal for DNN systems

Deep learning Neural Network (DNN)



DNN Accelerator Architecture (e.g., Eyeriss – MIT)



Goal

Understand error propagation in DNN accelerators through fault injection

- Quantification
- Characterization

Based on the insights, mitigate failures:

- Efficient way to detect errors
- Hardware: Selective duplication
- Software: Symptom-based detection

Fault Injection: Parameters

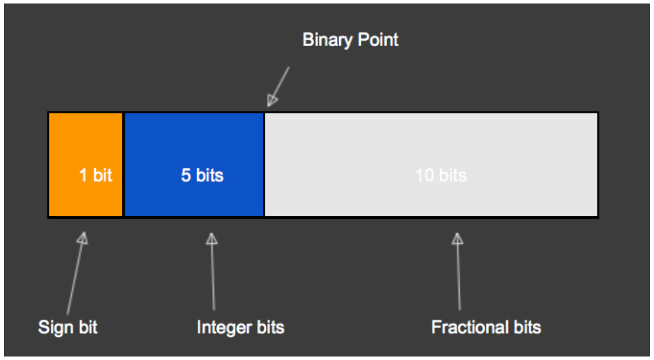
DNNs

Network	Dataset	No. of Output Candidates	Topology
ConvNet	CIFAR-10	10	3 CONV + 2 FC
AlexNet	ImageNet	1,000	5 CONV(with LRN) + 3 FC
CaffeNet	ImageNet	1,000	5 CONV(with LRN) + 3 FC
NiN	ImageNet	1,000	12 CONV

Data Types

Fixed Point (FxP): 16-bit and 32-bit

Float Point (FP): Full- and half-precision



Fault Injection Study: Setup

Fault Injection

- 3,000 random faults per each latch in each layer

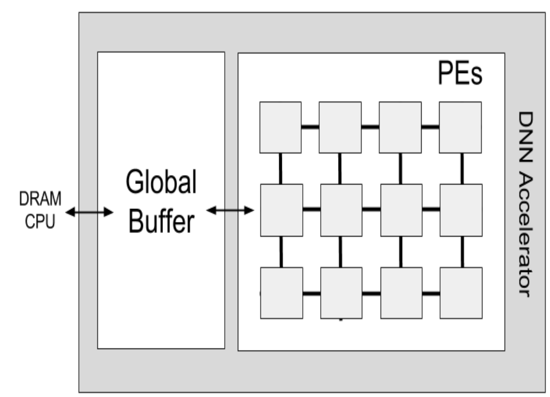
Simulator

- DNN simulation in Tiny-CNN in C
- Fault injections at C line code

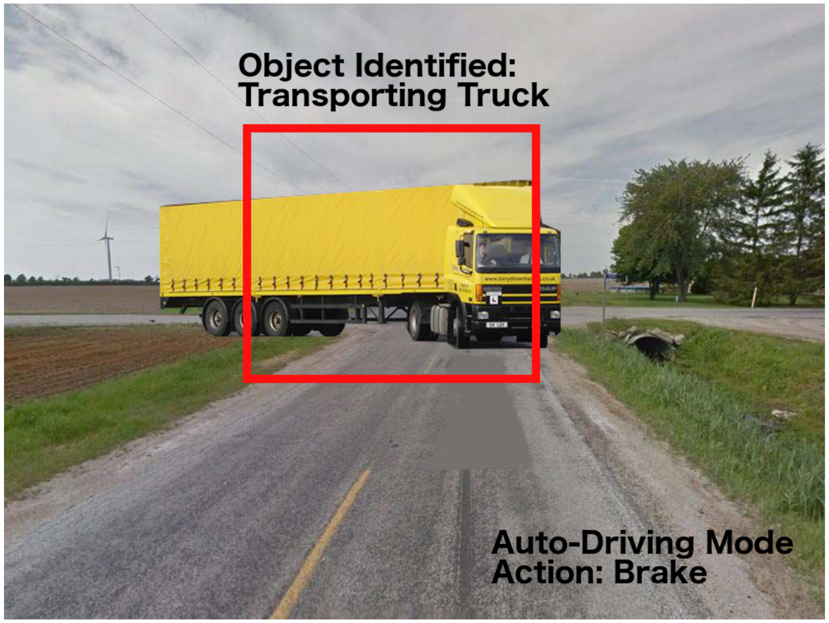
Fault Model

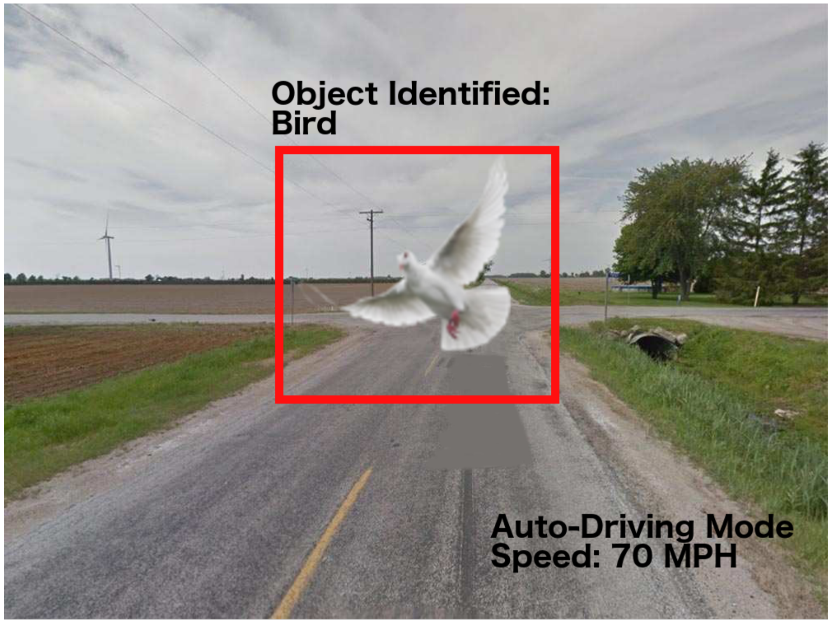
- Transient single bit-flip
- Execution Units: Latches
- Storage: buffer SRAM, scratch pad, REG

```
...
foreach layer:
  ...
  foreach weight:
    ...
    foreach input:
      ...
        R_L2.2 = inject_fault(R_L2.2)
        R_L3 = R_L2.2 + R_L5
      ...
...
```



Silent Data Corruption (SDC) Consequences





A single bit-flip error → misclassification of image by the DNN

Research Questions (RQs)

- RQ1: What are SDC rates in different DNNs using different data types?
- RQ2: Which bits are sensitive to SDCs in different data types?
- RQ3: How do errors affect values that result in SDCs?
- RQ4: How does an error propagate layer by layer?

SDC Types

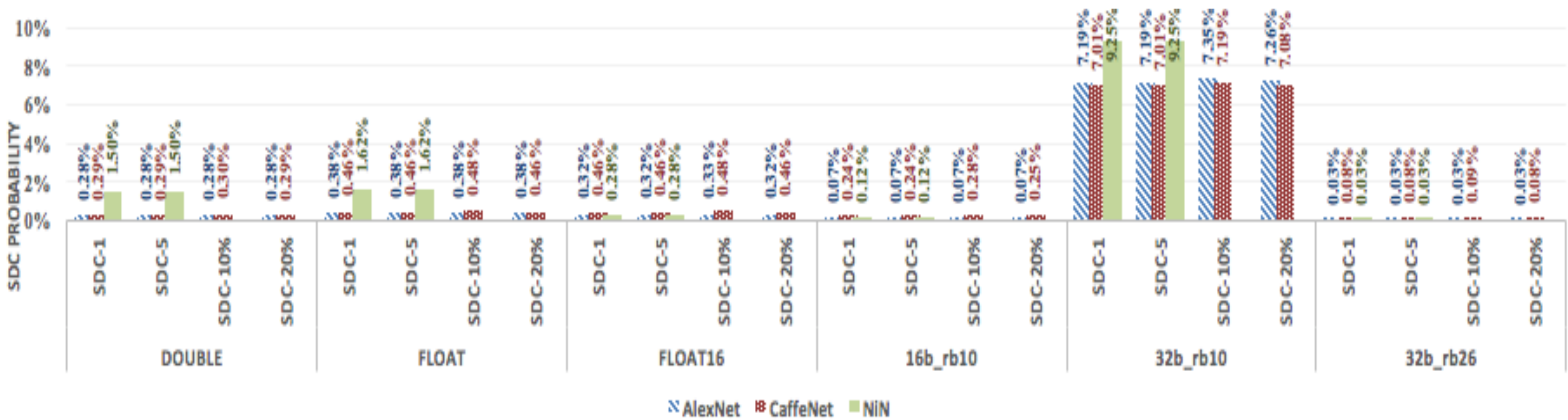
SDC1:
Mismatch between winners from faulty and fault-free execution.

SDC5:
Winner is not in top 5 predictions in the faulty execution.

SDC10%:
The confidence of the winner drops more than 10%.

SDC20%:
The confidence of the winner drops more than 20%.

RQ1: SDC in DNNs



1.All SDCs defined have similar SDC probabilities

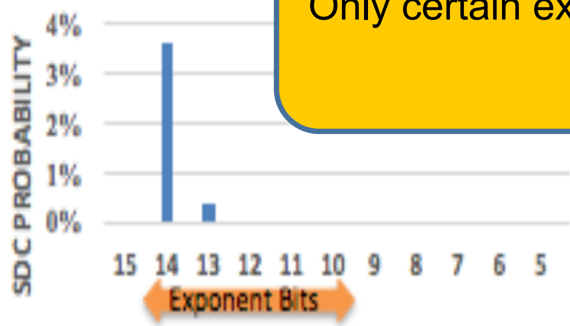
2.SDC probabilities are different in different DNNs

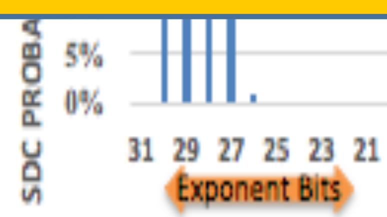
3.SDC probabilities vary a lot using different data types

RQ2: Bit Sensitivity

FP data types:

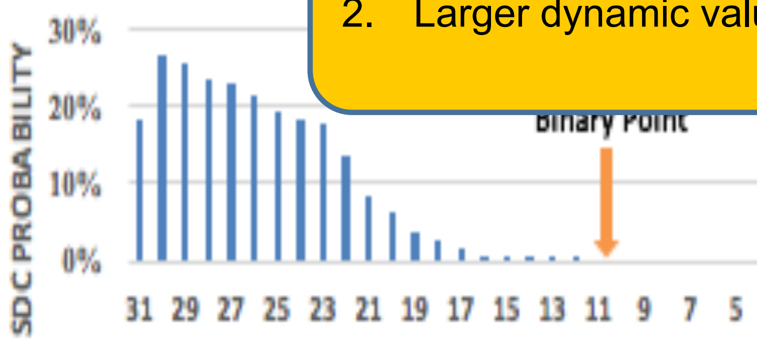
Only certain exponent bits are vulnerable to SDCs

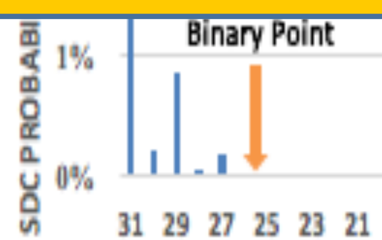




FxP data types:

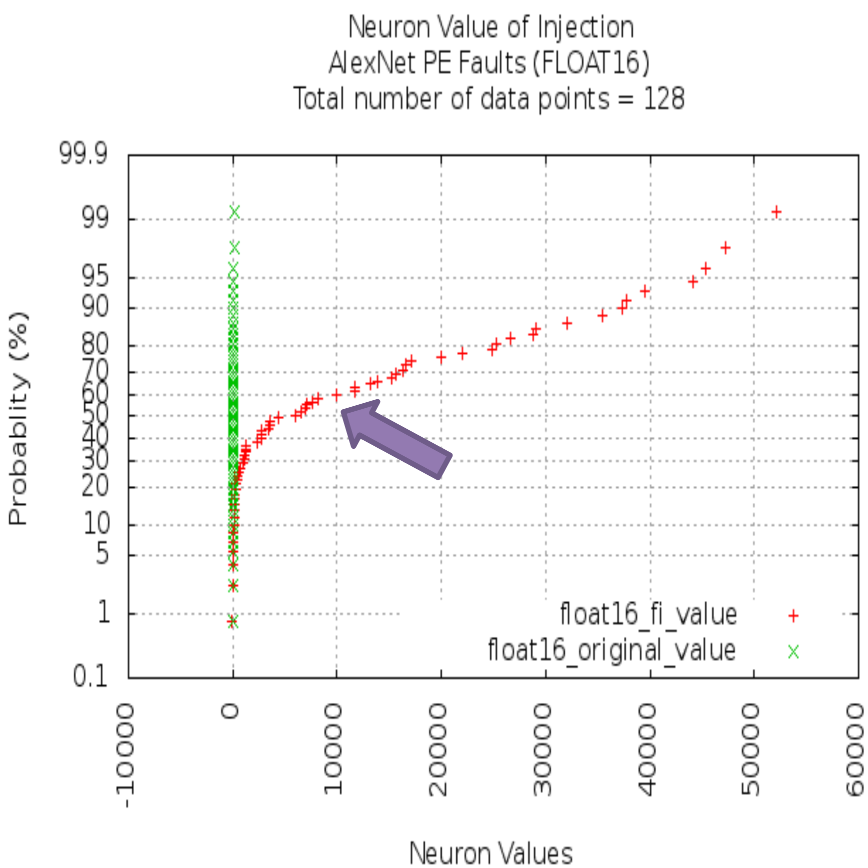
1. High-order bits are vulnerable
2. Larger dynamic value range allows more vulnerable bits



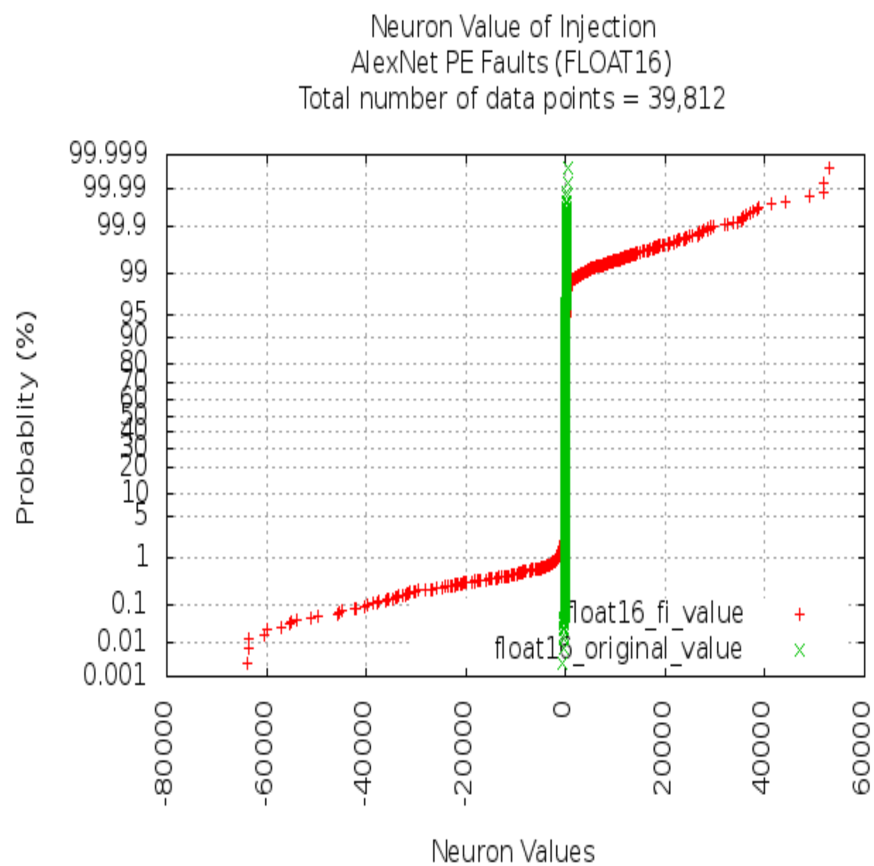


RQ3: Value Changes

AlexNet, PE Errors, Float16



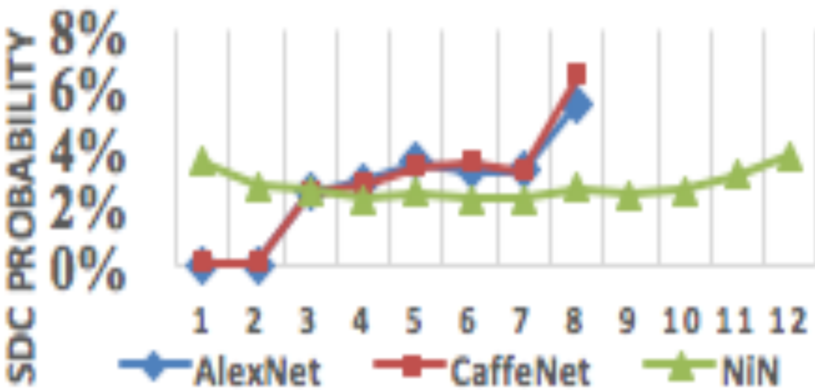
SDC

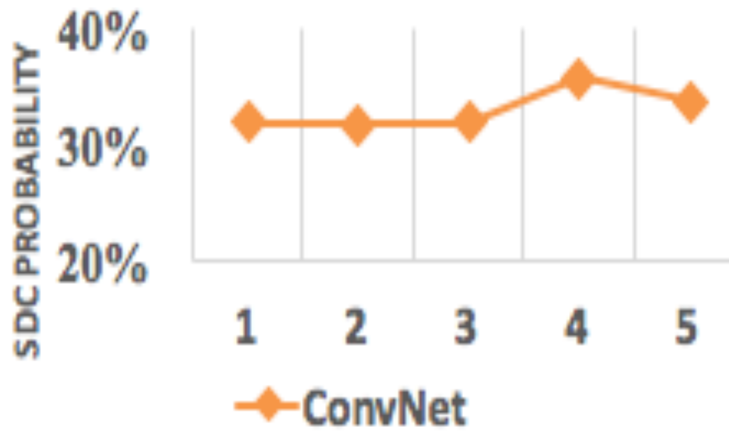


Benign

If a neuron value is changed to be a large value under a fault, it likely causes SDC

RQ4: SDC in Different Layers

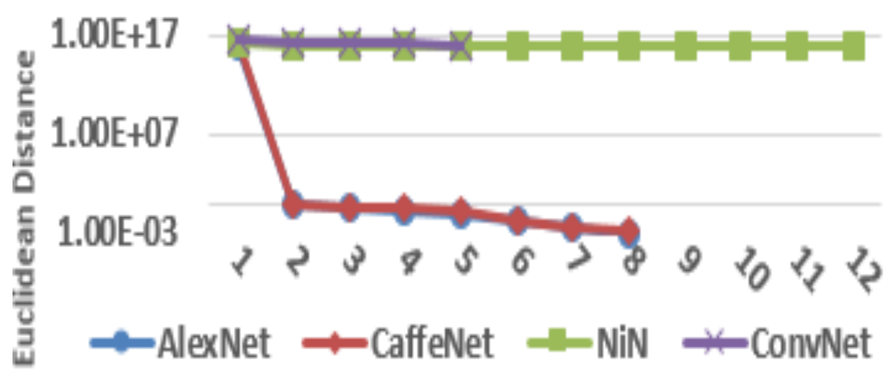




1.Layers 1&2 have lower SDC probabilities in AlexNet and CaffeNet

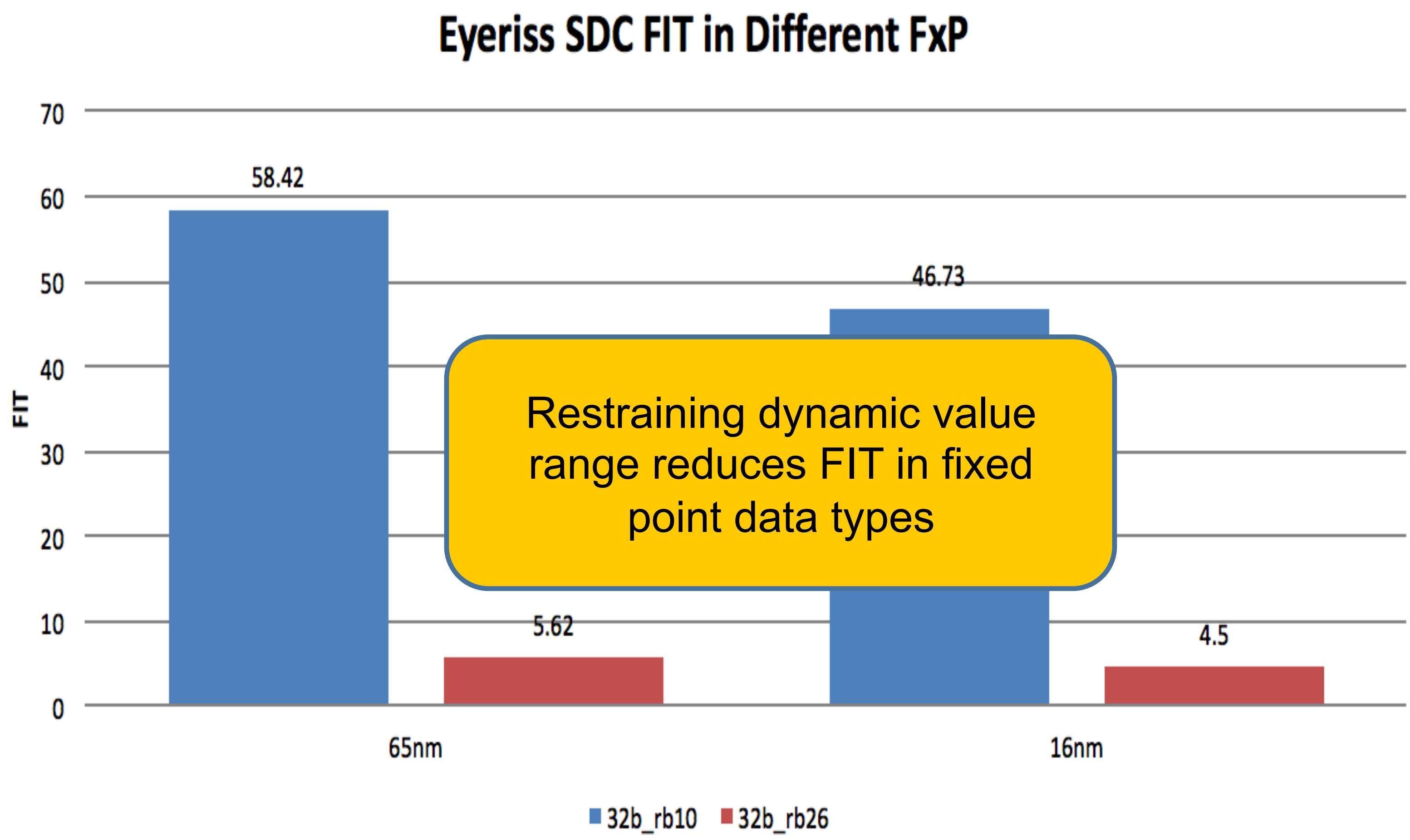
2.SDC probability increases as layer numbers increase

RQ4: Euclidean Distance of Values



1. Euclidean distance decreases by layers
2. Local Response Normalization (LRN) in Layer 1&2 re-normalizes values back towards normal range in AlexNet and CaffeNet

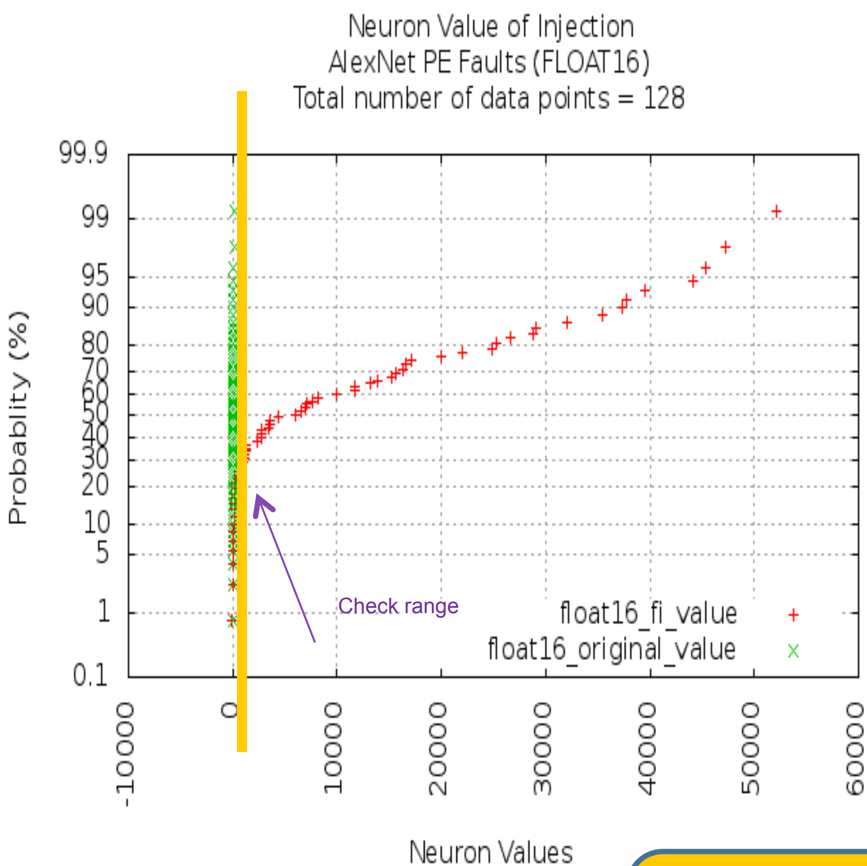
Mitgation: Data Type Choice



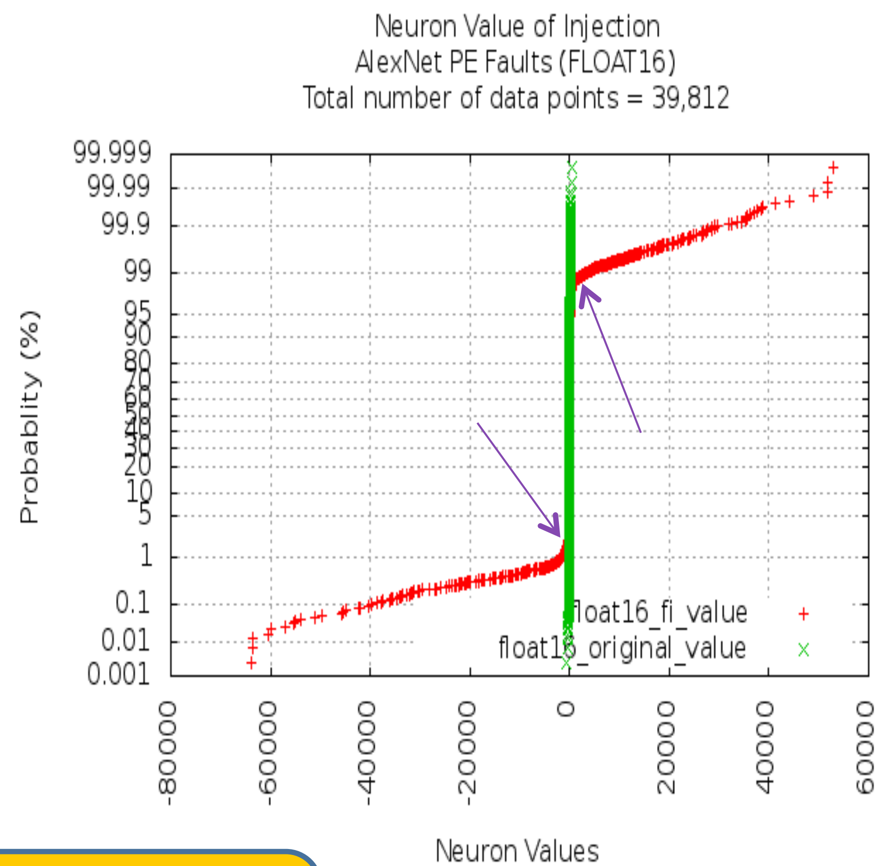
*Scaling factor = 2 by each tech. generation
All raw FIT rates are projected based on the FIT at 28nm [Neale, IEEE TNS]

Mitigation: Symptom-Based Error Detector (Software)

AlexNet, PE Faults, Float16



SDC



Benign

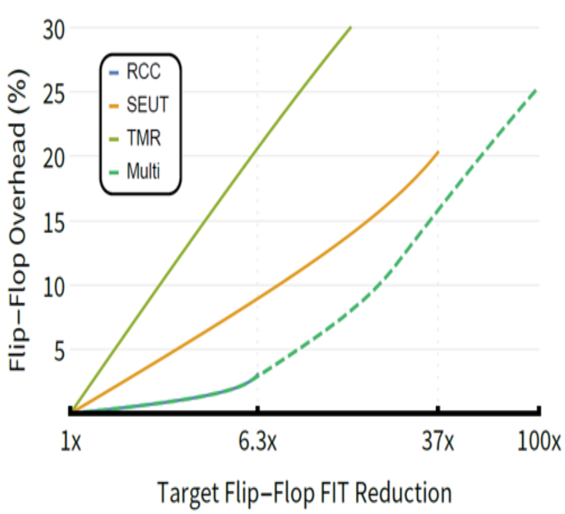
Recall: 92.5%
Precision: 90.21%
On selected data types

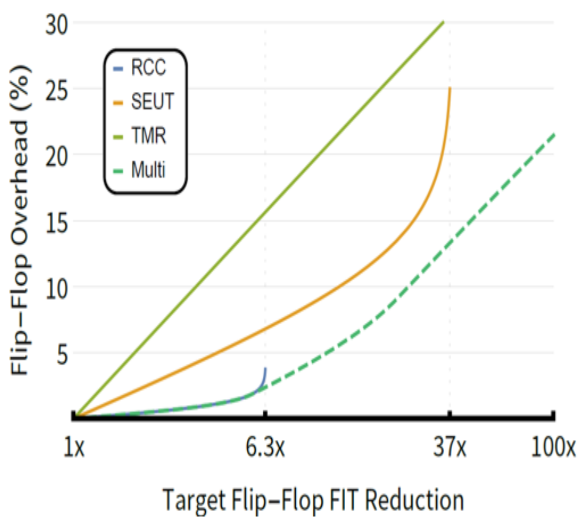
Mitigation: Selective Latch Hardening (Hardware)

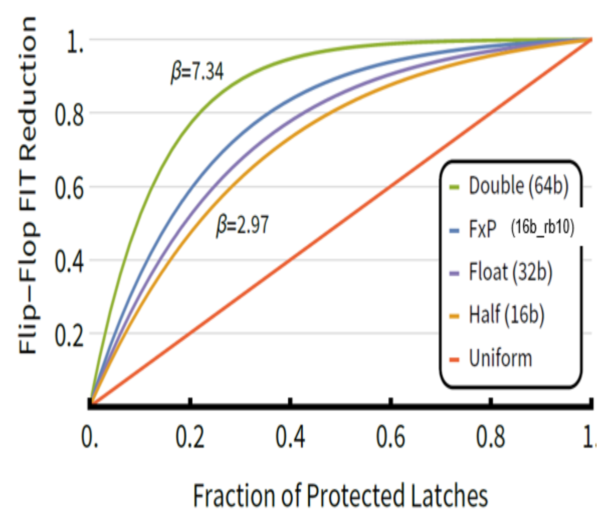
Latch hardening design choices:

Latch Type	Area Overhead	FIT Rate Reduction
Baseline	1x	1x
Strike Suppression (RCC)	1.15x	6.3x
Redundant Node (SEUT)	2x	37x
Triplicated (TMR)	3.5x	1,000,000x

~20% overhead provides 100x reduction in FIT







Summary of DNNs

1. Characterized error propagation in DNN accelerators based on data types, layers, value types and DNN topologies
2. Mitigation Methods:
    - restraining value range of data type
    - value range checker
    - selective latch hardening

This Talk

- Motivation
- Resilience of Deep Neural Networks in Self-Driving Cars from Soft Errors [SC'17 – to appear]
- Intrusion Detection Systems for Smart Embedded Devices using Dynamic Invariants [FSE'17]
- Ongoing work and conclusion

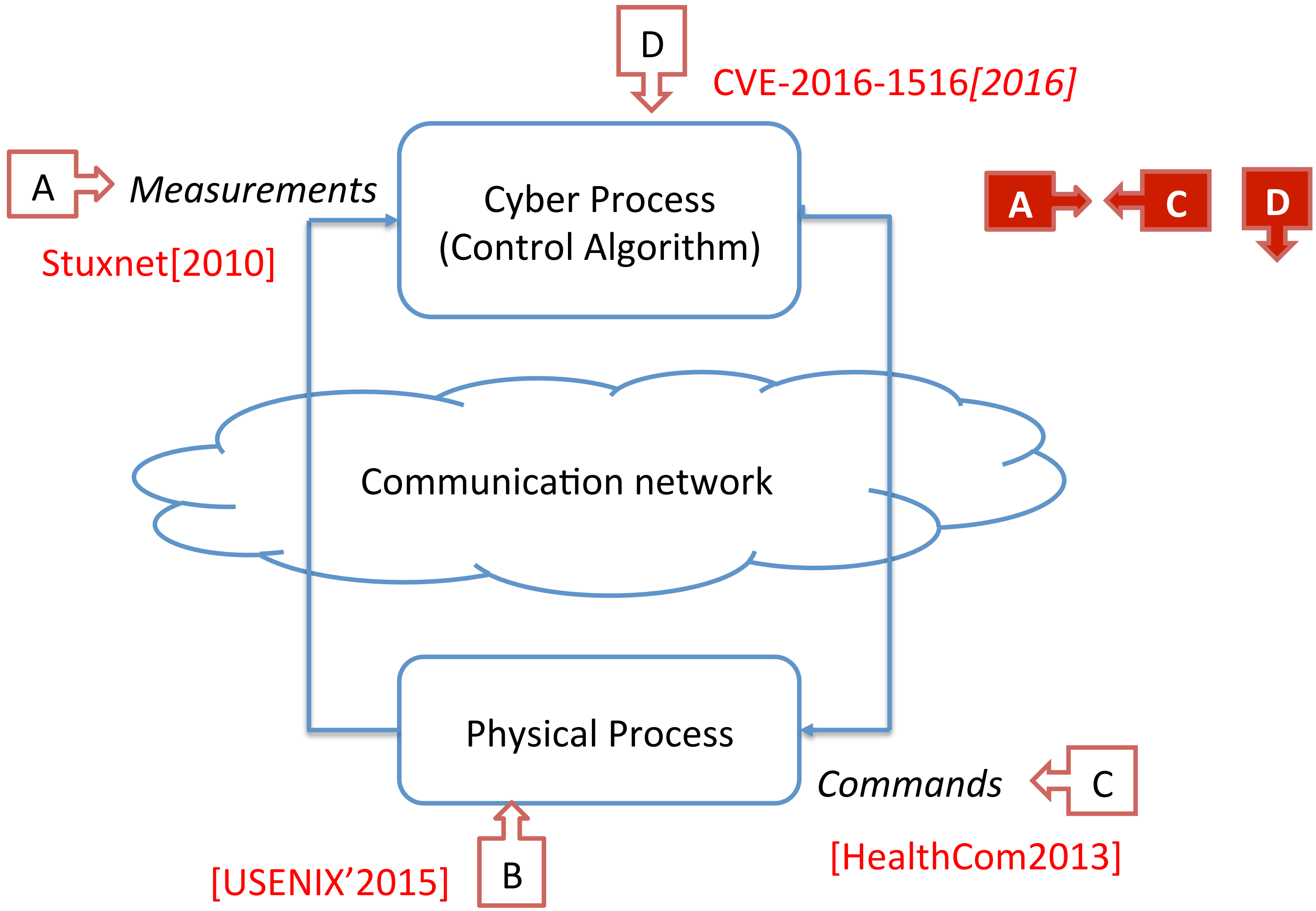
Motivation

- **Goal: Provide low-cost security for CPS**
  – Satisfying resource and real-time constraints
  – No human intervention needed
  – Is able to detect zero day attacks

Insight: Leverage properties of CPS for intrusion detection

- Simplicity and timing predictability
- Learn invariants based on dynamic execution
- Monitor invariants at runtime for violations

Threat Model



Intrusion Detection Systems (IDS)

Signature-based IDSs [CSUR2014]

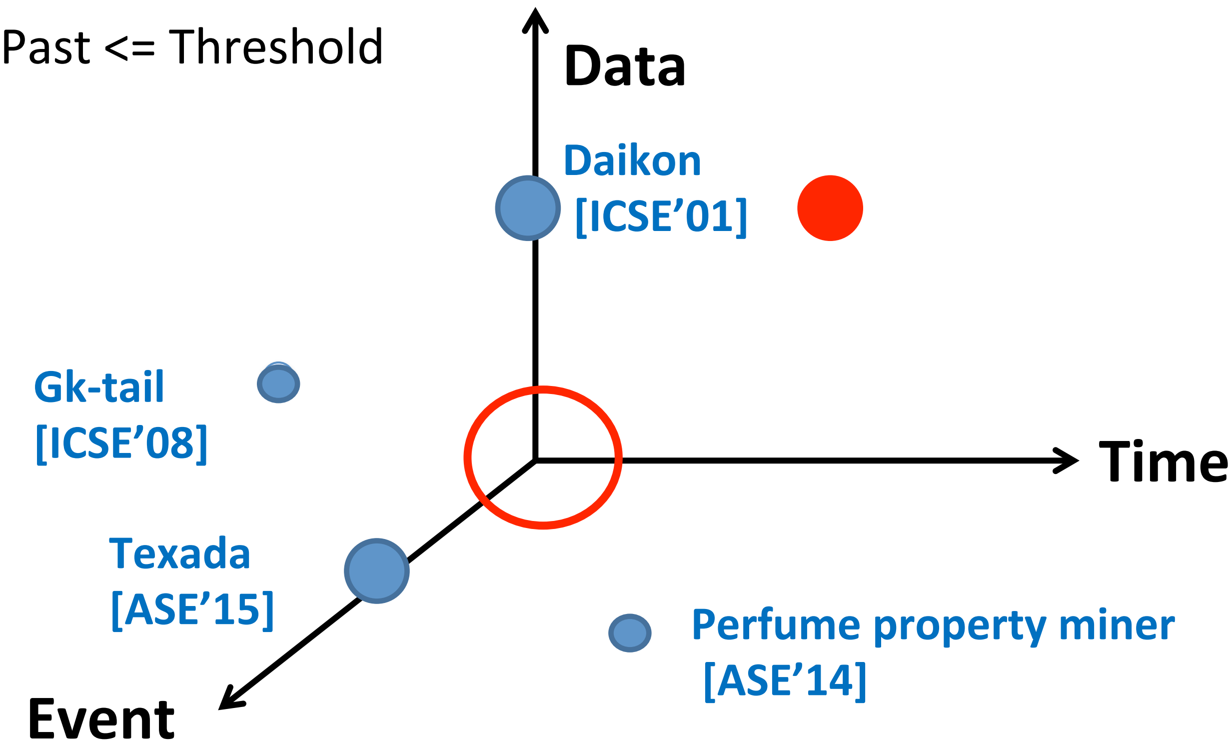
Anomaly-based IDSs [Computers&Security2009]

Specification-based IDSs [SmartGridCom2010]

- Static analysis
- Dynamic analysis

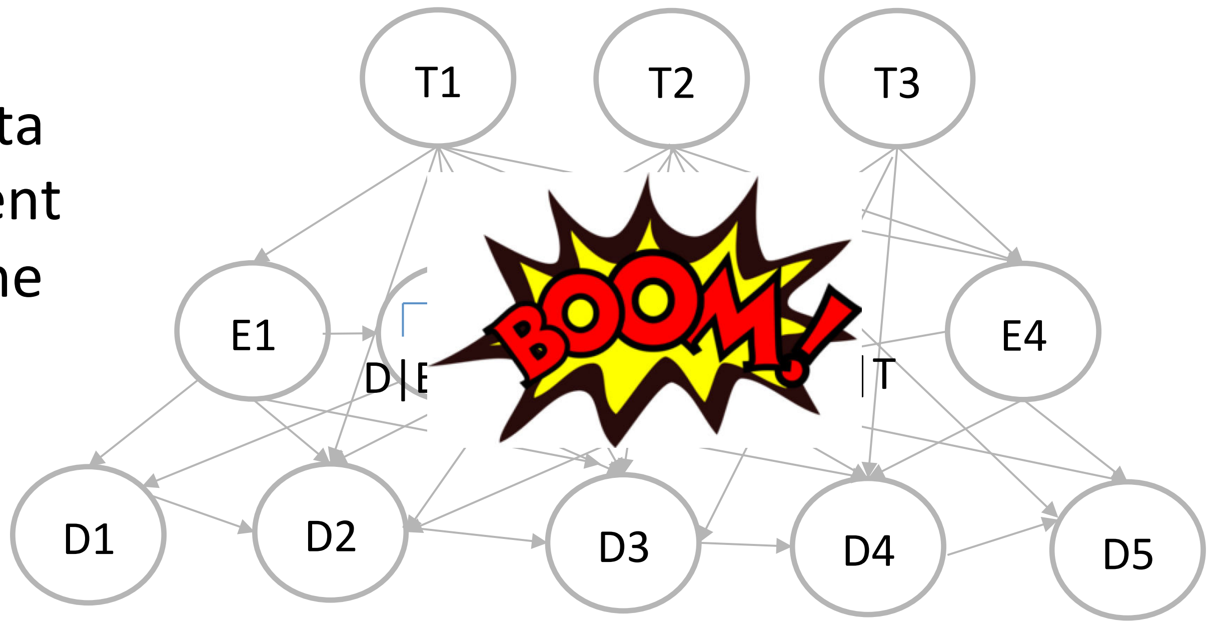
Dynamic Analysis Techniques

- Invariant Examples
  - Energy usage >=0
  - Current − Past <= Threshold



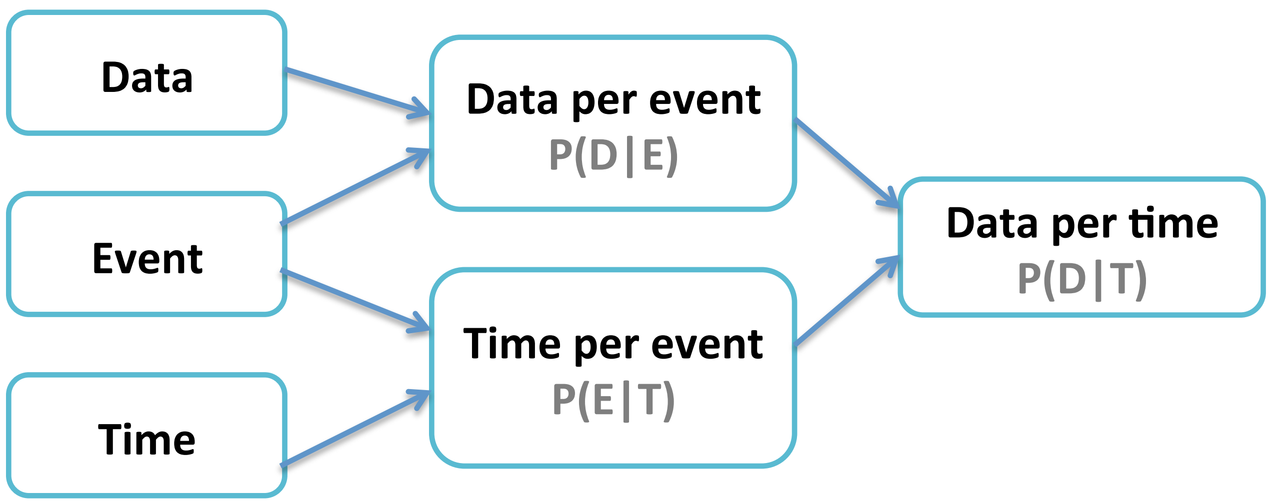
Main Idea

D: Data
E: Event
T: Time

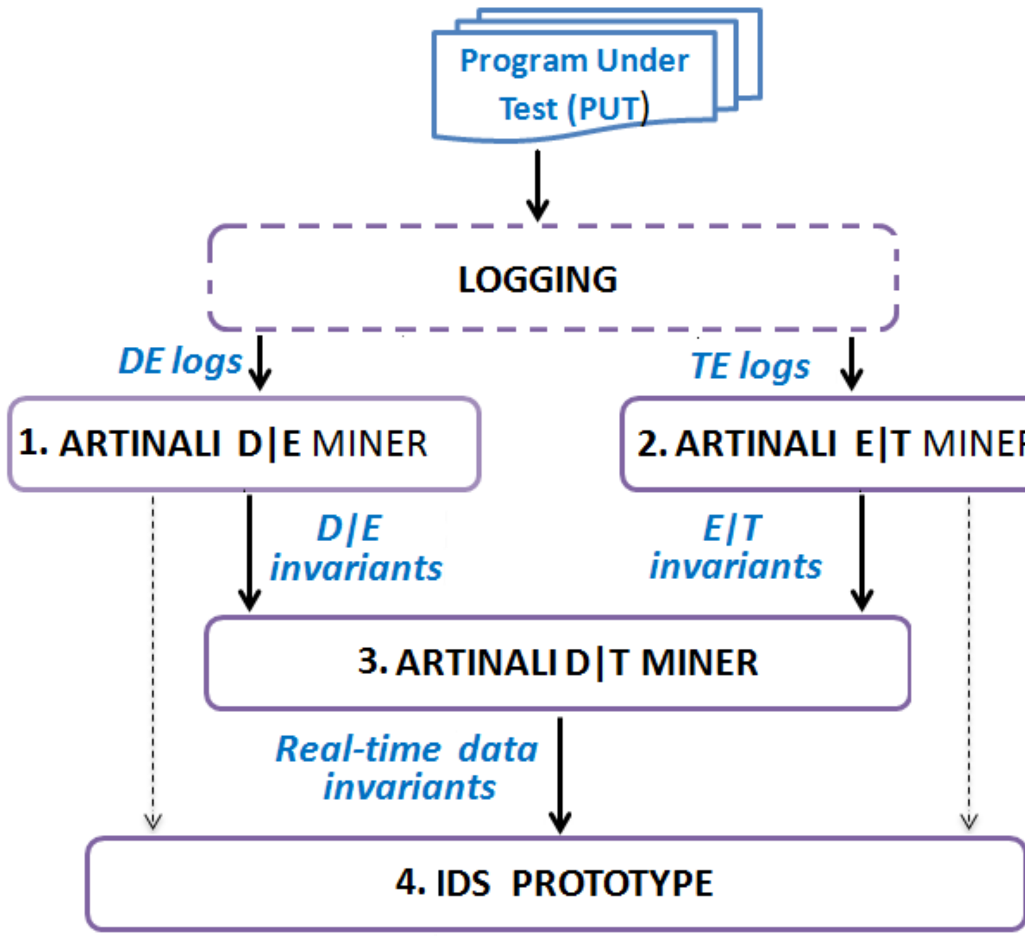


Methodology

- **ARTINALI**: **A R**eal **T**ime-specific **I**nvariant i**N**ference **AL**gor**I**thm
  - 3 dimensions and 6 classes of invariants

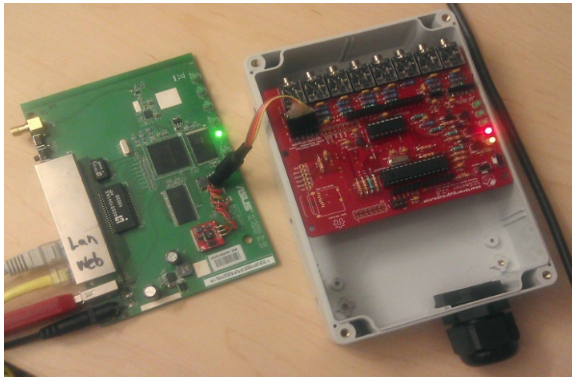


ARTINALI Implementation

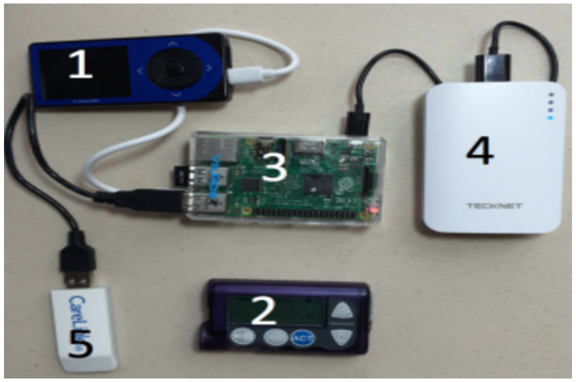


CPS Platforms for Evaluation

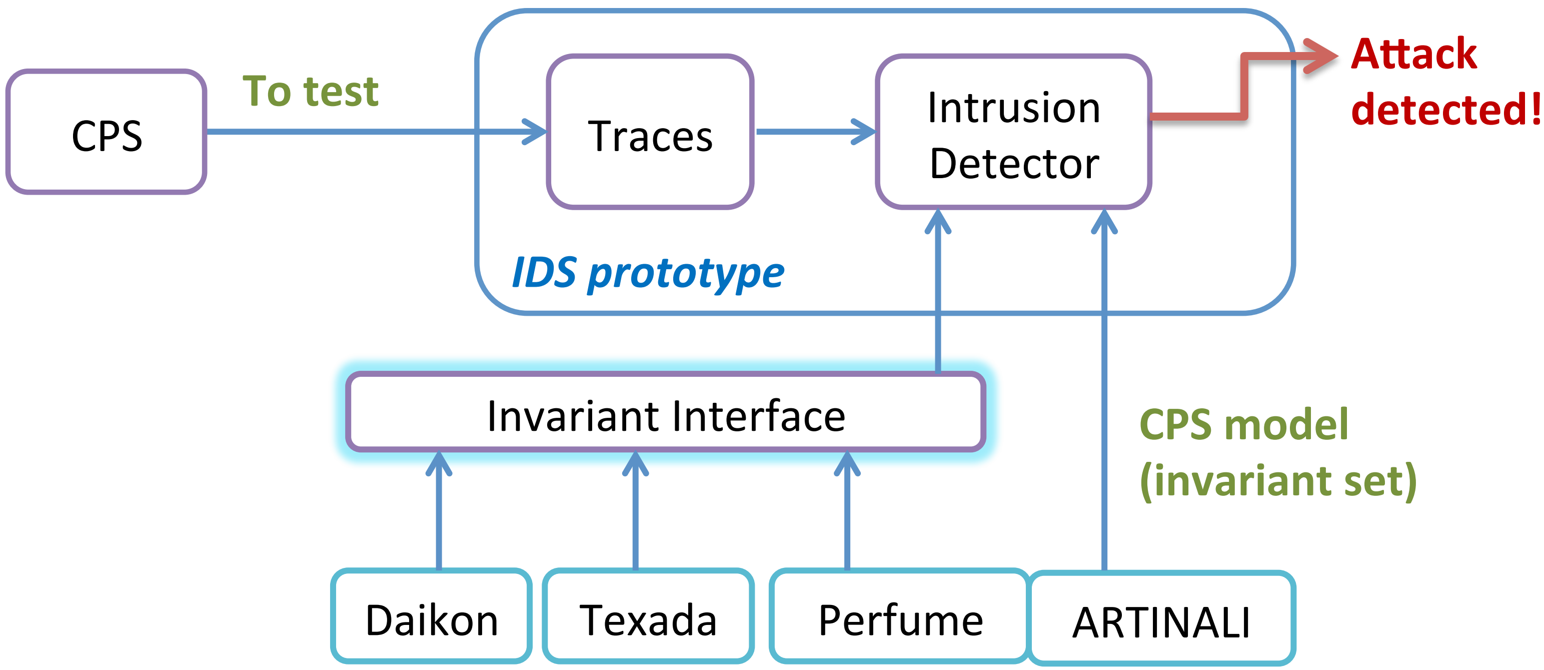
- Advanced metering infrastructure (AMI)
  - SEGMeter
    - http://smartenergygroups.com



- Smart Artificial Pancreas (SAP)
  - OpenAPS
    - https://openaps.org/



Experimental Setup



Targeted Attacks

CPS Platform	Targeted attack	Attack entry point
AMI (SEGMeter)	Meter spoofing [ACSAC2010]	Deception on A
	Sync. Tampering [ACSAC2010]	Deception on D
	Message dropping [CCNC2011]	DoS on A
SAP (OpenAPS)	CGM spoofing [Healthcom2011]	Deception on A
	Stop basal injection [BHC2011]	Deception and DoS on C
	Resume basal injection [BHC2011]	Deception and DoS on C

Take away :

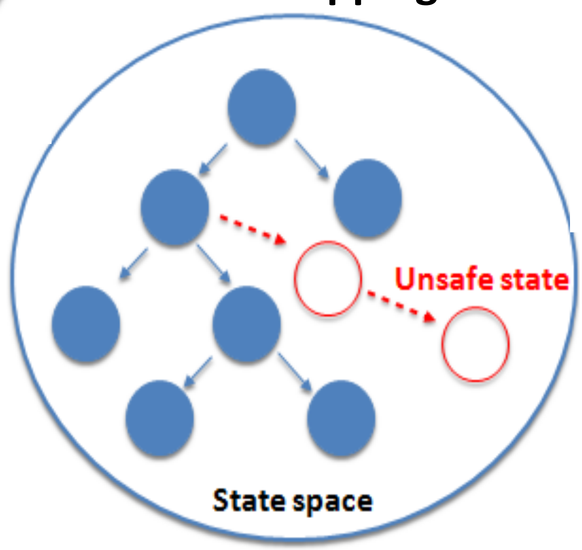
ARTINALI detected all the targeted attacks

Arbitrary Attacks

Data mutations

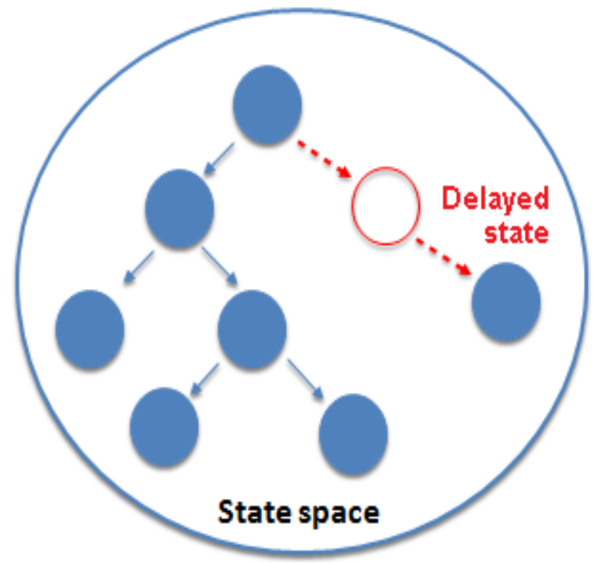
Smart facial recognition system (CVE-2016-1516)

Branch flipping



CGM spoofing in SAP, [BHC2011]

Artificial delay insertion



Synchronization tampering in smart meter, [ACSAC2010]

Accuracy Metrics

- False Negative Rate (FNR)

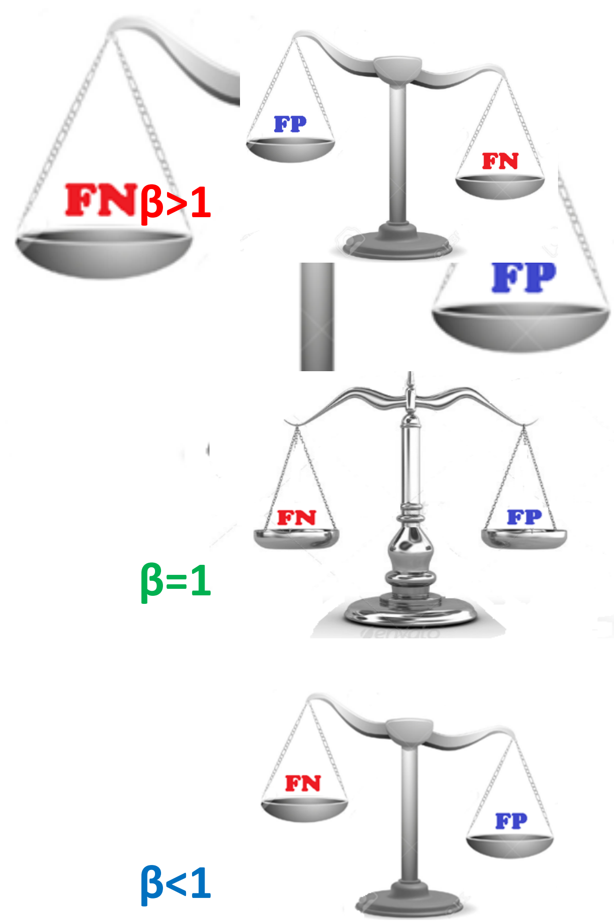
$$\frac{\textit{Number of detected attacks}}{\textit{Total number of injected attacks}} \times 100$$

- False Positive Rate (FPR)

$$\frac{\textit{Number of raised alarms}}{\textit{Total number of attack-free tests}} \times 100$$

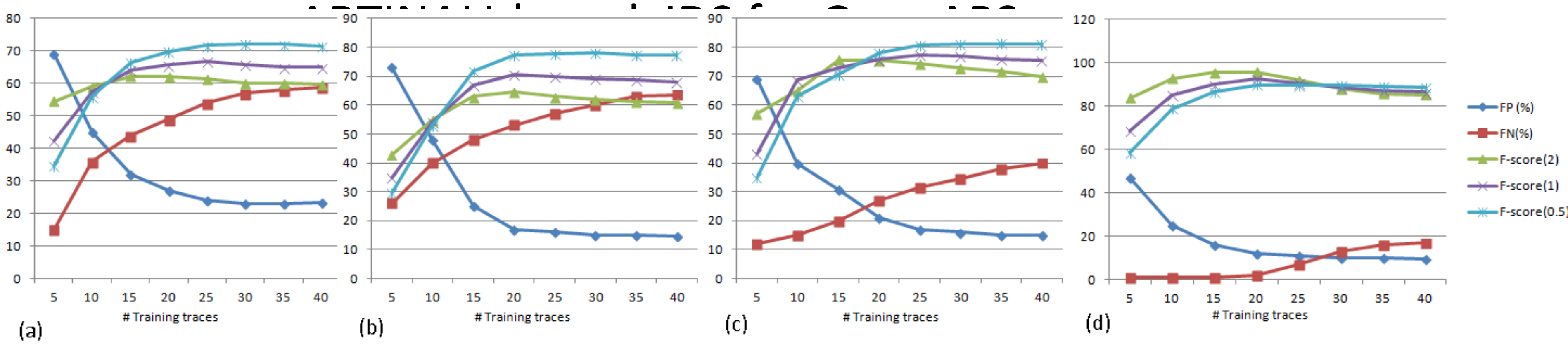
- F-Score(β)

$$\frac{(1+\beta^2)\times TP}{(1+\beta^2)\times TP+\beta^2\times FN+FP}$$

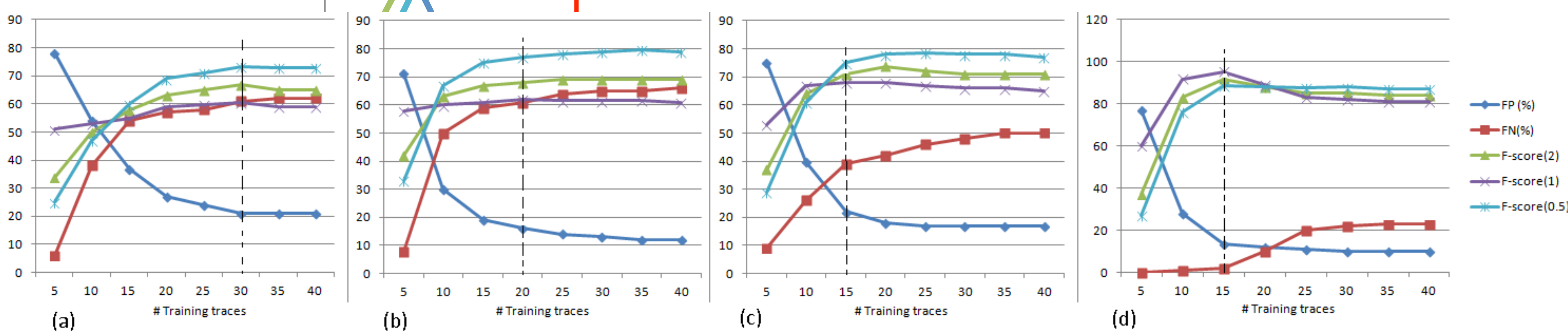


Parameter Tuning

SEGMeter



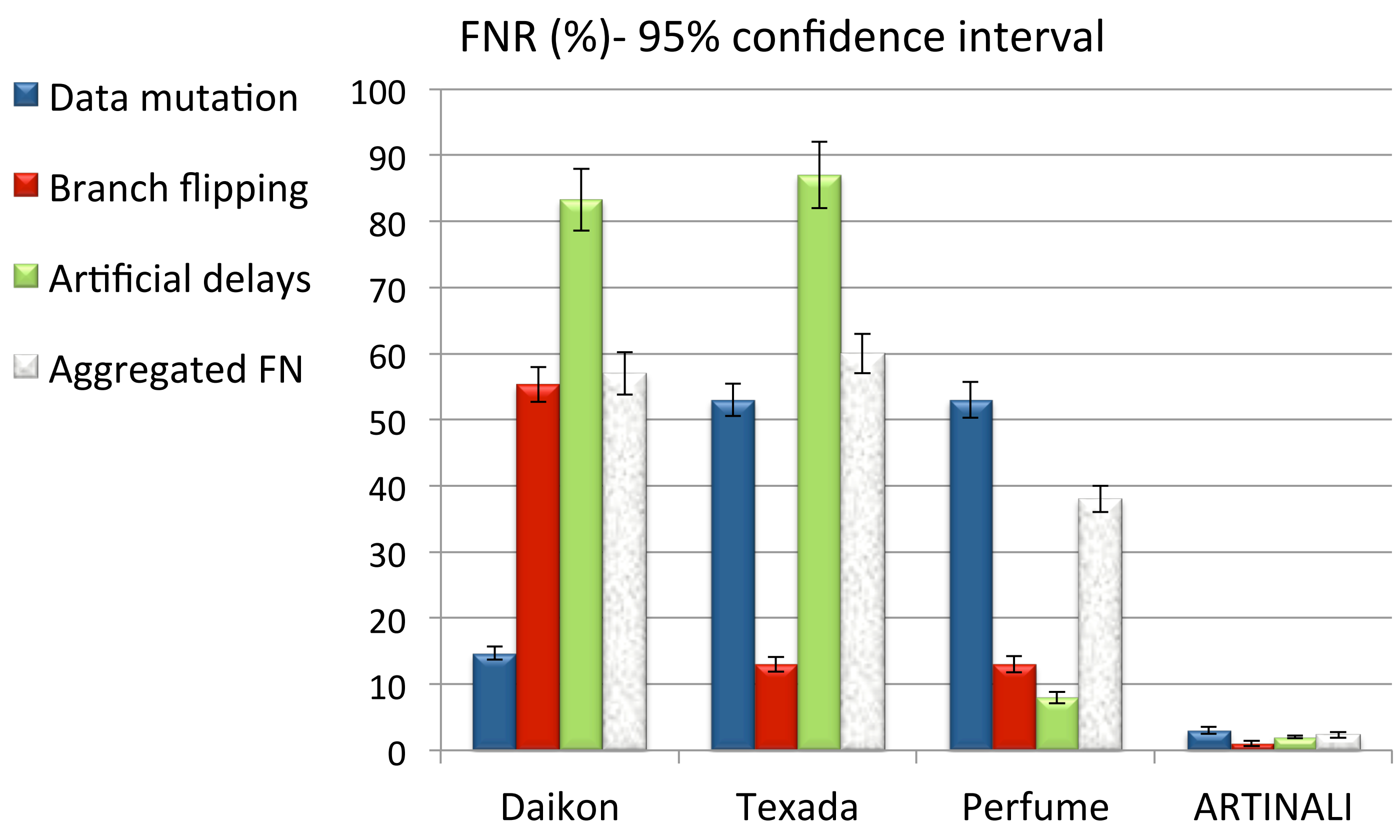
OpenAPS



(a) Daikon (b) Texada (c) Perfume (d) ARTINALI

False Negative (FN) Rate

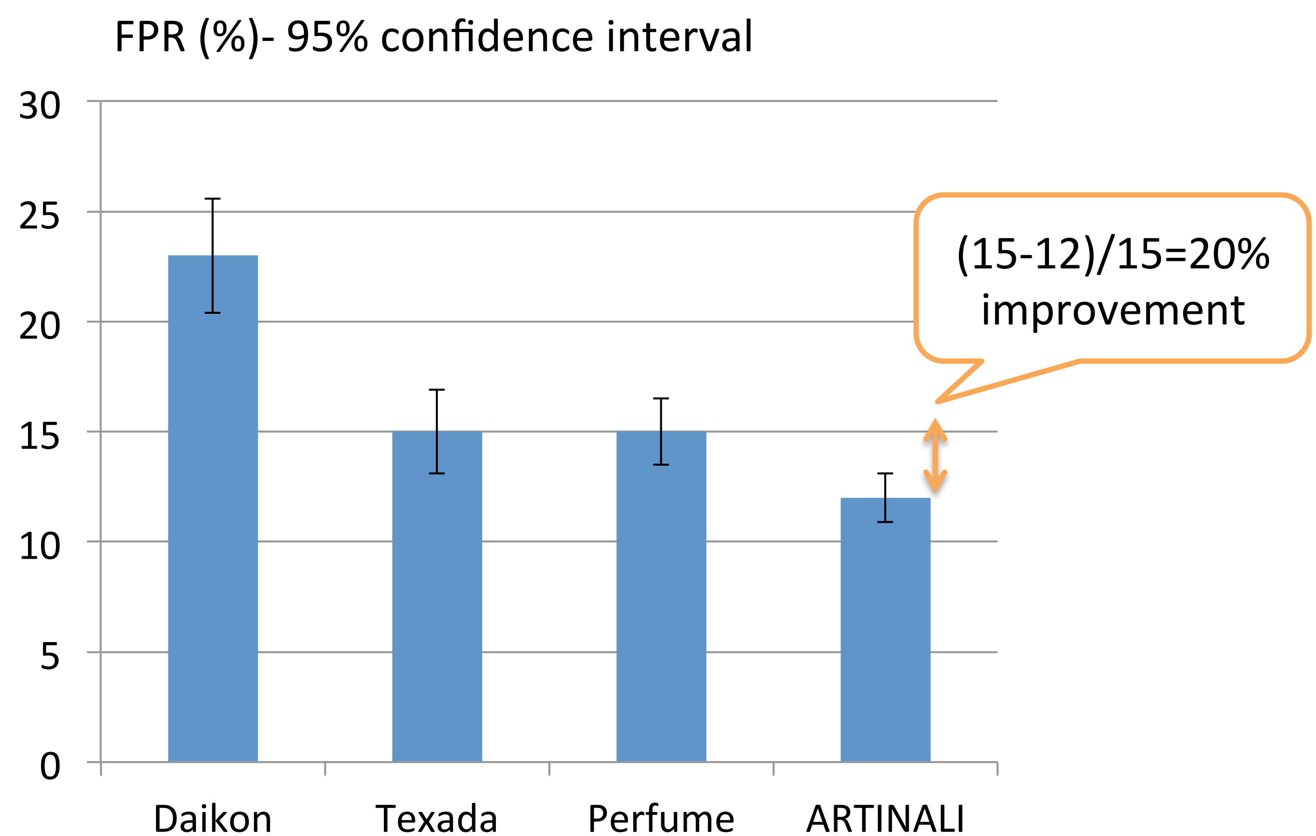
- ARTINALI-based IDS reduces the ratio of FN by 89 to 95% compared with the other tools across both platforms.
  - SEGMeter



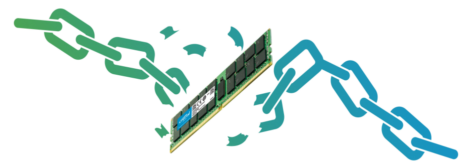
False Positive (FP) Rate

- ARTINALI-based IDS reduces the ratio of FP by 20 to 48% compared with the other tools across both platforms.

- SEGMeter

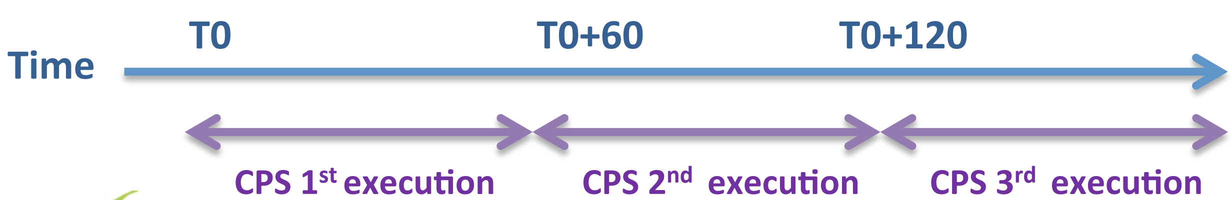


Performance and Memory



SEGMeter

	Performance Overhead (%)	Detection time (sec)	Memory usage
Daikon	27.3	16.63	1.24 MB
Texada	23.7	14.45	3.21 MB
Pefume	32.08	19.57	3.94 MB
ARTINALI	31.6	19.25	2.96 MB



Summary of ARTINALI

- ARTINALI: A Multi-Dimensional model for CPS
  - Captures *data-event-time* interplay
  - Introduces *Real-time data invariants*
  - Increases the *coverage* of IDS
  - Decreases the rate of *false positives*
  - Imposes comparable *overheads*

- Examine generalizability of ARTINALI
  - Unmanned Aerial Vehicle (UAV)

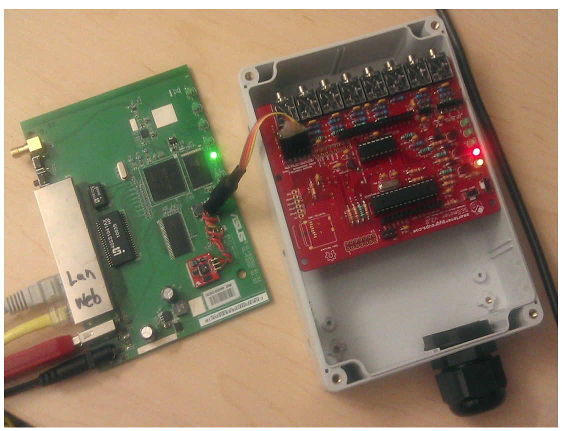
- https://github.com/karthikp-ubc/Artinali

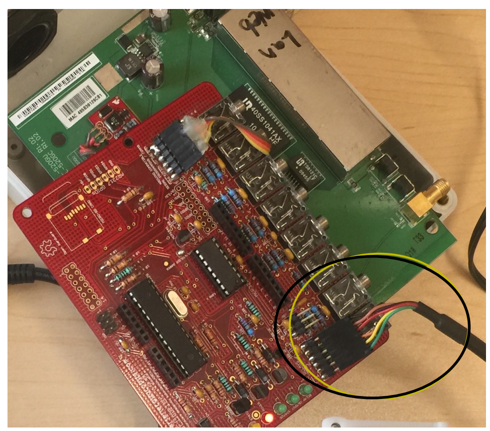
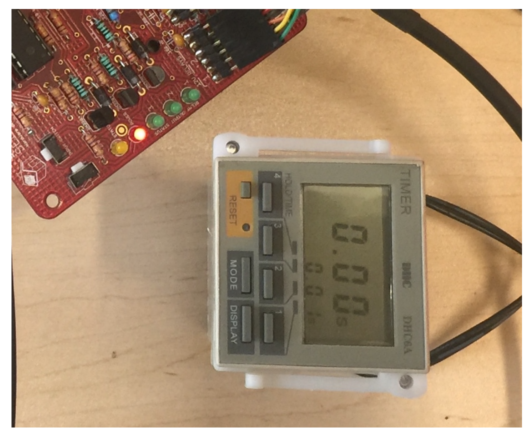
This Talk

- Motivation
- Resilience of Deep Neural Networks in Self-Driving Cars from Soft Errors [SC’17 – to appear]
- Intrusion Detection Systems for Smart Embedded Devices using Dynamic Invariants [FSE’17]
- Ongoing work and conclusion

Ongoing Work: Formal Analysis

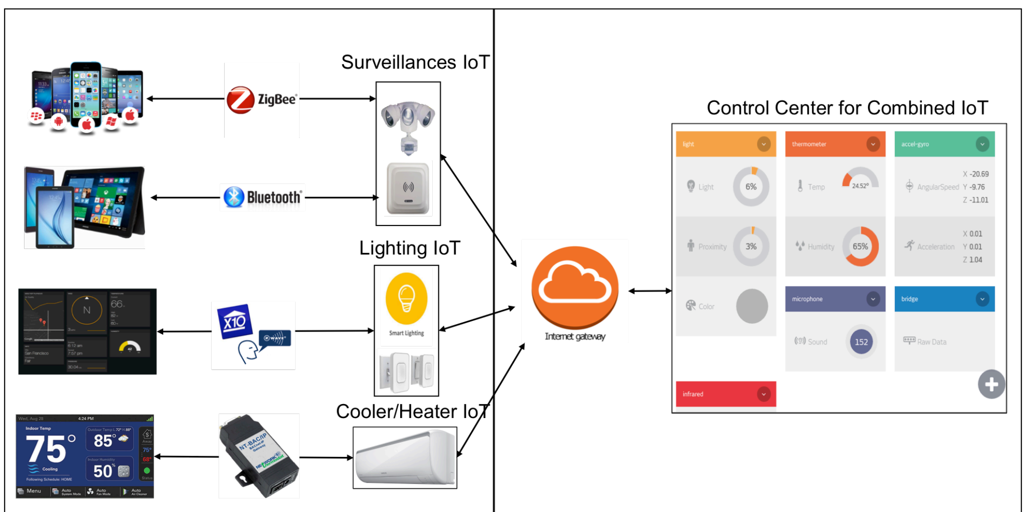
- Formally model the states of the CPS
- Combine with formal attacker models
- Model-check the system for security invariants
  - Identify unsafe states and paths to unsafe states
  - Automatically mount the attacks on the system



Ongoing Work: SmartJS

- **SmartJS: Smart JavaScript-based Runtime System for programming IoT systems**
  – Security and Performance constraints
  – Dynamic code migration to satisfy constraints



Current IoT: Devices from various manufacturers works separately

SmartJS: Brings devices from various manufacturers into a single unified platform.

sensor.js

```
// ...

// Connect
pubsub.connect(function() {

    // Repeat every second
    setInterval(function() {

        // Read temperature from GPIO pin
        var temperature = GPIO.readPin(12);

        // Publish temperature
        pubsub.publish("smartsensor/temperature", {
            id: mySensorId,
            temperature: temperature
        });

    }, 1000);

});
```

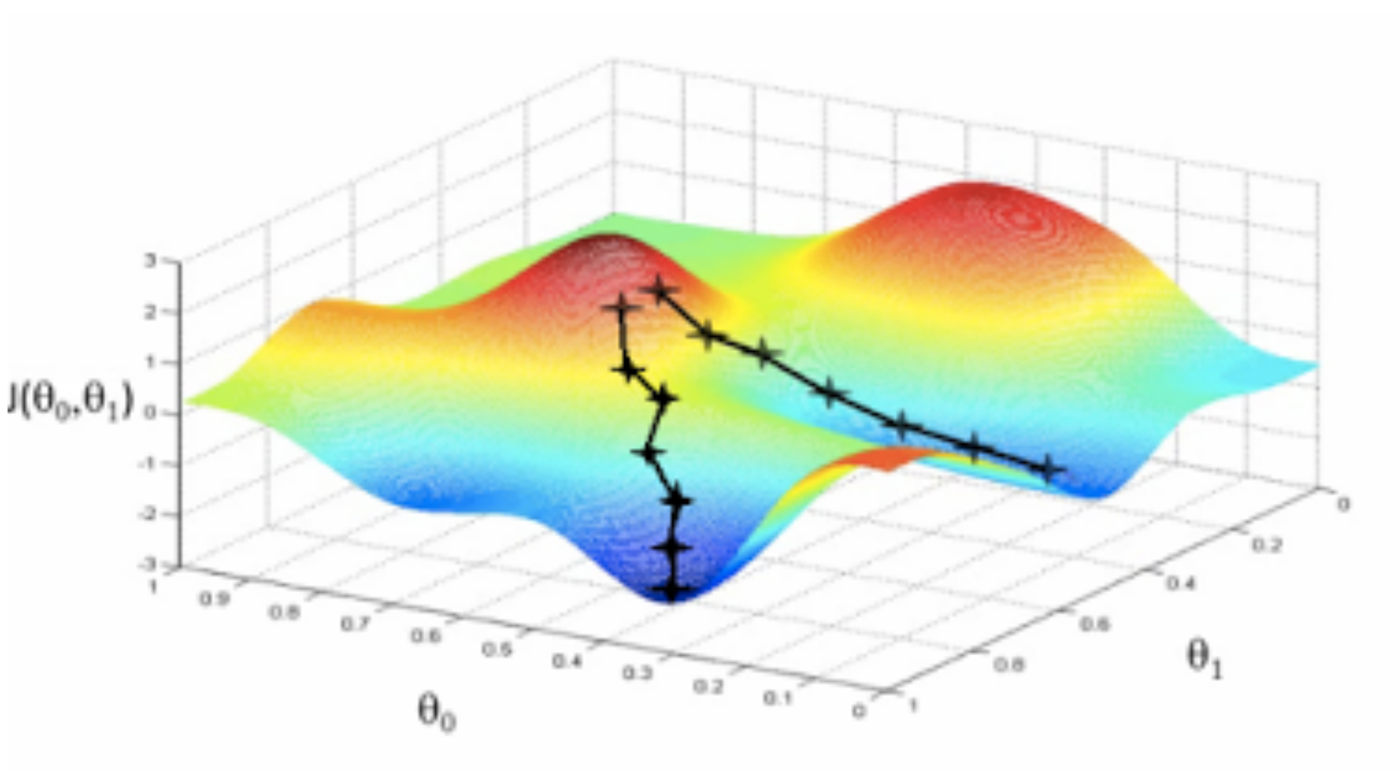
regulator.js

```
// ...

// Connect
pubsub.connect(function() {

    // Subscribe to temperature messages
    pubsub.subscribe("smartsensor/temperature", function(d) {

        if (d.temperature > threshold) {
            pubsub.publish("smartsensor/actuation", {
                id: d.id,
                powerVariation: -5
            });
        } else if (d.temperature < threshold) {
            pubsub.publish("smartsensor/actuation", {
                id: d.id,
                powerVariation: 5
            });
        }

    });
});
```

Ongoing Work: Resilient ML

Deriving ML algorithms resilient to perturbations

- Small changes → Similar outputs
- Convergence properties



Conclusion

CPS systems resilience and security are important challenges

Two systems for resilience and security

1. Deep Neural Network Accelerators for Self-Driving Cars
2. Invariant monitoring for embedded system security

Future work

1. Formal analysis for CPS
2. Smart runtimes for IoT
3. Resilient Machine Learning

Questions? **karthikp@ece.ubc.ca**