A Unified Submodular Framework for Multimodal IC Trojan Detection

Farinaz Koushanfar, Azalia Mirhoseini, Yousra Alkabani
Electrical and Computer Engineering Department
Rice University

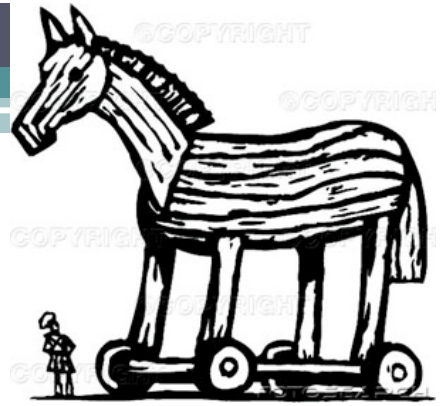
Outline

- Introduction
- Assumptions
- Related work
- Contributions
- Trojan detection
- Experimental evaluation
- Summary and conclusion

Introduction

- Horizontal business model
  - IP providers, IC designers, and foundries are separate companies
- Potential threats of this business model
  - Theft of IPs, piracy of ICs, and addition of Trojans
- Trojan component maybe inserted
  - To monitor, control, spy, or steal information from the chip

Trojan detection challenges



- Trojan insertion mechanisms not known
- Many opportunities
- Exponential growth of the number of gates vs. linear increase in the number of pins
  - Limited controllability and observability
- Foundaries with advanced technologies
  - Multiple opportunities to insert Trojans
- Increase in process variation with technology
  - Difficult to detect change due to Trojan insertions

Assumptions

- Unobtrusive timing, static power and dynamic power testing
- Adversarial model
  - Trojan IC has the same set of input/output pins as the original design
  - ICs have already passed the standard parametric and functional tests
  - The nominal profile of each gate in each modality is available

Related work

- Power signature for Trojan detection
  - Building IC fingerprints from power traces, statistical testing to detect altered ICs [Aragwal et al. S&P 2007]
  - Power measurements from multiple supply ports, and emperical sensitivity analysis [Rad et al. HOST and ICCAD 2008, IEEE trans. on VLSI 2009]
  - Partitioning the circuit into regions and testing suspected regions [Banga and Hsiao HOST 2008]
- Timing signature for Trojan detection
  - E.g., principal component analysis [Jin and Makris HOST 2008]

Related work(need abstract of the works)

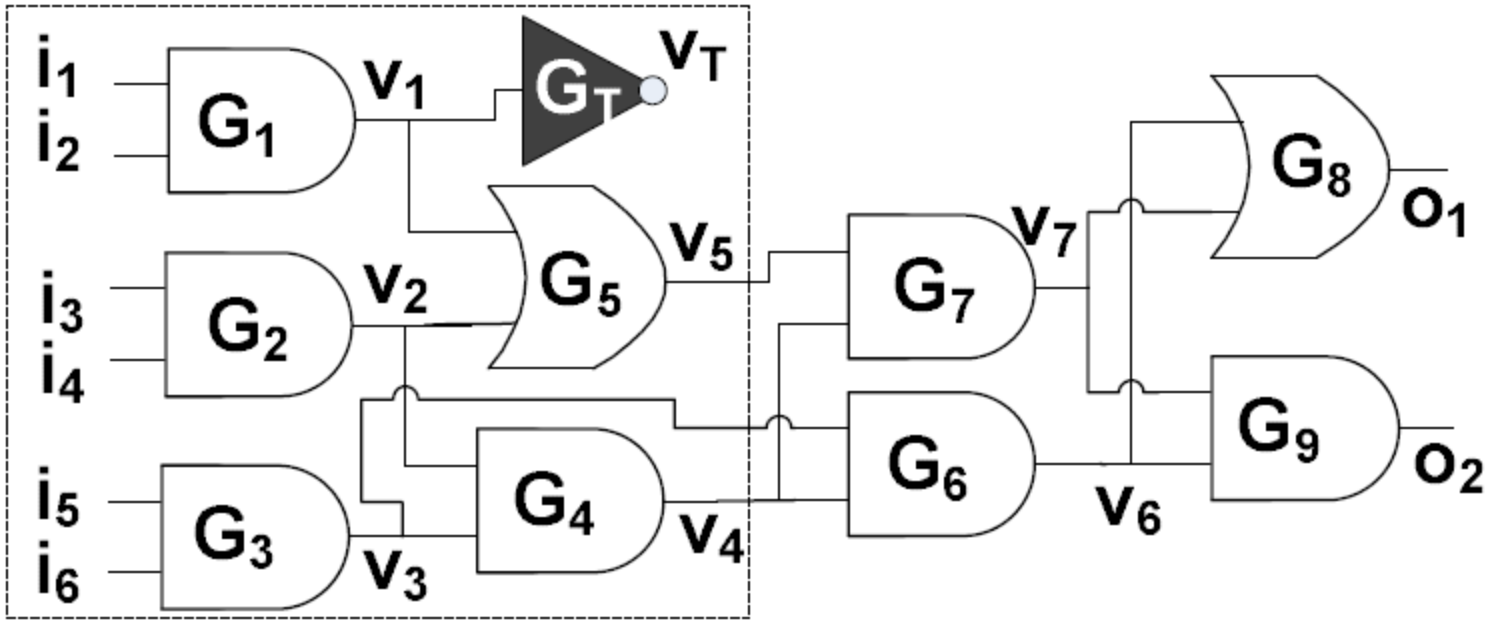
- Gate-Level Characterization
  - Gate-Level Characterization: Foundations and Hardware Security Applications[Wei et al. DAC 2010]
  - SVD-Based Ghost Circuitry Detection[Nelson et al. IH2009]
  - Characterizing leakage current of gates with a consistency based algorithm [Alkabani et al. ICCAD 2009]

Contributions

- A new unified formal framework for IC Trojan detection by noninvasive measurements
  - Unimodal anomaly detection built upon the gate level profiling (Timing , static and dynamic power modalities investigated)
  - Different multimodal Trojan detection methods for combining unimodal detection results
- A submodular objective function
  - An iterative greedy detection algorithm that achieves a near optimal solution in polynomial time
- A method to calibrate the systematic variations
  - Robust to measurement noise and process variation

Submodular property: intuition

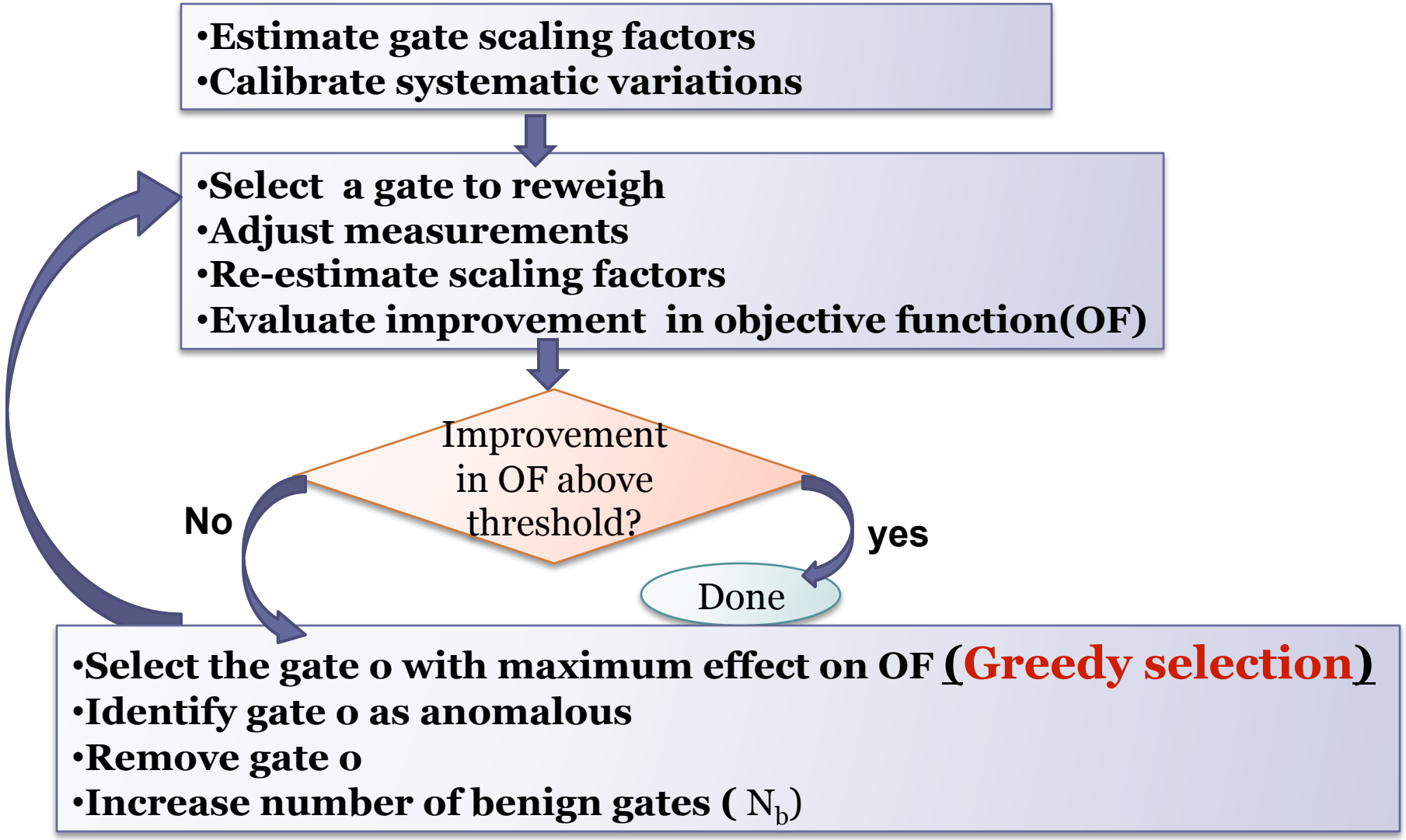
- Inserting a Trojan would have a higher impact on a small circuit than inserting the same Trojan to a larger circuit that contains the small one



Unimodal Trojan detection method

- Trojan detection method is built upon the gate profile estimation
  - Generate input vectors that enable gate profile measurements
  - Apply the measurement vectors and map the measured values to gate **scaling factors** (deviation from the nominal gate profile}
  - The anomalies are detected through an iterative hierarchical approach

Unimodal detection algorithm



Scaling factor estimation & calibration

- **Estimate scaling factors**
- **Calibrate scaling factors**

- A: matrix of nominal values
- Φ: scaling factors
- E: measurment error
- B: measured value for each input
- Solve AΦ+E=B minimizing mean square error of E to compute Φ

- Filter out systematic variations using 2D high pass filter

Reweighing scaling factors

- **Select a gate to reweigh**
- **Adjust measurements**
- **Re-estimate scaling factors**

- Reweigh using a Gaussian kernel function
- Recompute B after reweighing the scaling factors
- Solve $A\Phi+E=B$ minimizing mean square error of E to get Φ

Objective function

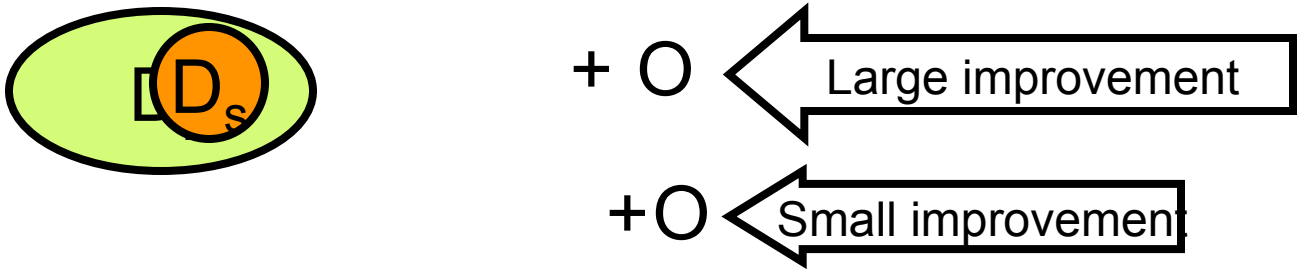
- **Evaluate improvement in OF**
- **Remove the anomalous gate o**

- Objective function: $R(\Gamma) = L(D) - L(D \backslash \Gamma)$
  - L is maximum likelihood of error or min $L_2(E)$
  - D is the set of all gates
  - Γ is the set of anomalous gates
- The objective function is submodular
- $D \backslash \Gamma$ or removing anomalous gates is to set their scaling factors equal to the nominal unity value

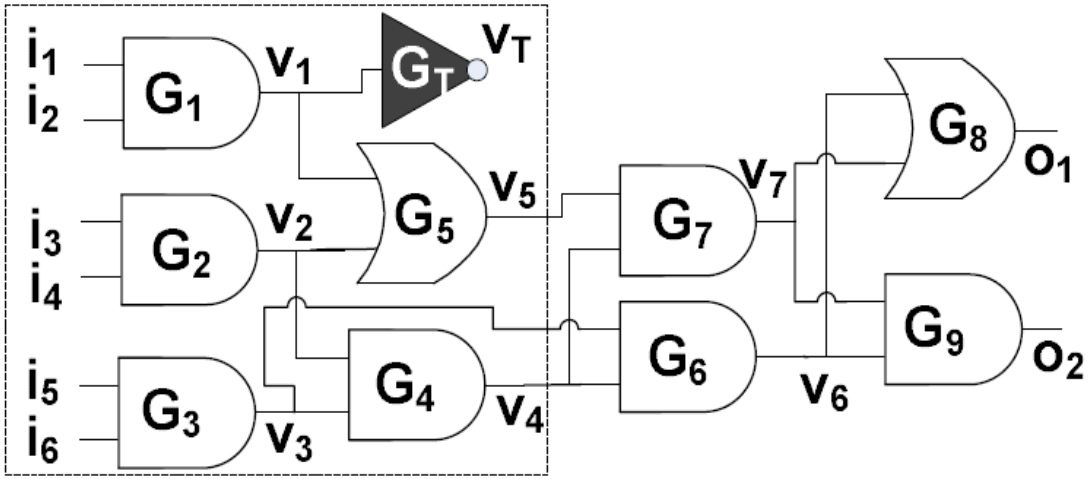
Objective function R is submodular

- $R(\Gamma) = L(D) - L(D \setminus \Gamma)$ is a penalty reduction function
- Penalty will not be reduced if we do not reweigh a new anomalous gate, i.e., $R(\varnothing) = 0$
- R is a non decreasing set function:
  - Reweighing a new anomaly could just decrease the associated penalty, i.e., $R(\Gamma 1) \leq R(\Gamma 2)$, for $\Gamma 1 \subseteq \Gamma 2 \subseteq D$
- R satisfies the diminishing return property

Diminishing returns characterization



- Reweighing a gate in a smaller set of gates D_s, improves the reward at least as much as reweighing it in a larger set of gates D_l, with $D_s \subseteq D_l$



Greedy approach: near optimal solution of submodular function

Theorem: [Nemhauser et al '78]
For a submodular function R greedy algorithm gives constant factor approximation

$$R(\Gamma_{greedy}) >= (1-1/e)\ R(\Gamma_{opt})$$

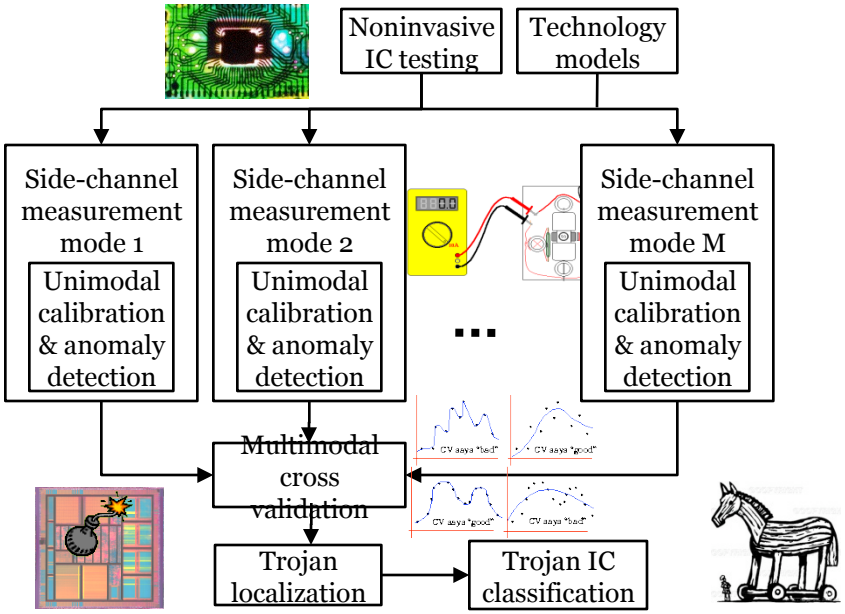
~63%

- Greedy algorithm gives near-optimal solution!
- For information gain: guarantees best possible unless P = NP! [Krause & Guestrin '05]

More on calibration

- Inter-chip variations
  - Affect the mean of the variations
  - Adjusted by shifting the mean of the extracted profile values to have a mean of unity
- Intra-chip variations
  - In form of a spatial distribution, e.g., 2D Gaussian in our model
  - Systematic intra-chip variation is slower than the rate of change because of the Trojan insertion
  - Resolved by 2D Discrete Cosine Transform (DCT) high pass filter

The unified multiomodal Trojan detection framework



Multimodal Trojan detection techniques

- Unanimous voting
  - Decreases P_D but improves P_{FA}
- Conservative voting
  - Increases P_{FA} but also increases $P_{D,}$ gives maximum achievable P_D
- Majority voting
  - Trade-off between P_D and P_{FA}
- Weighed voting
  - Tradeoff between P_D and P_{FA}

Experimental evaluation

- Setup
  - MCNC'91 benchmarks
  - ABC synthesis tool
  - HSPICE for nominal leakage computation
  - Placement by the Dragon tool
  - Matlab for the simulations and solving the quadratic programs (QPs)

Measurement setup(Please take a look at it)

- Timing:
  - Testing pattern generation method described in [Yang et al.]
- Leakage current
  - The IDDQ tests via off-chip pins by the precision measurement unit (PMU) [Sabade et al.]
  - TetraMAX ATPG is used for IDDQ test generation
- Dynamic current
  - IDDT tests by averaging methods that do not require high precision or high frequency measurement devices needed for capturing the transient signals [Jha et al.]

Gate-level characterization vs. measurment noise(Static power)

Ct	Size	i/p	o/p	3%	5%	10%
C8	165	28	18	5.6	7.0	11.6
C432	206	36	7	1.7	3.5	7.2
C1355	512	41	32	8.5	10.0	12.1
C499	532	41	32	2.9	4.5	9.0
C3450	1131	50	22	4.0	5.9	9.8

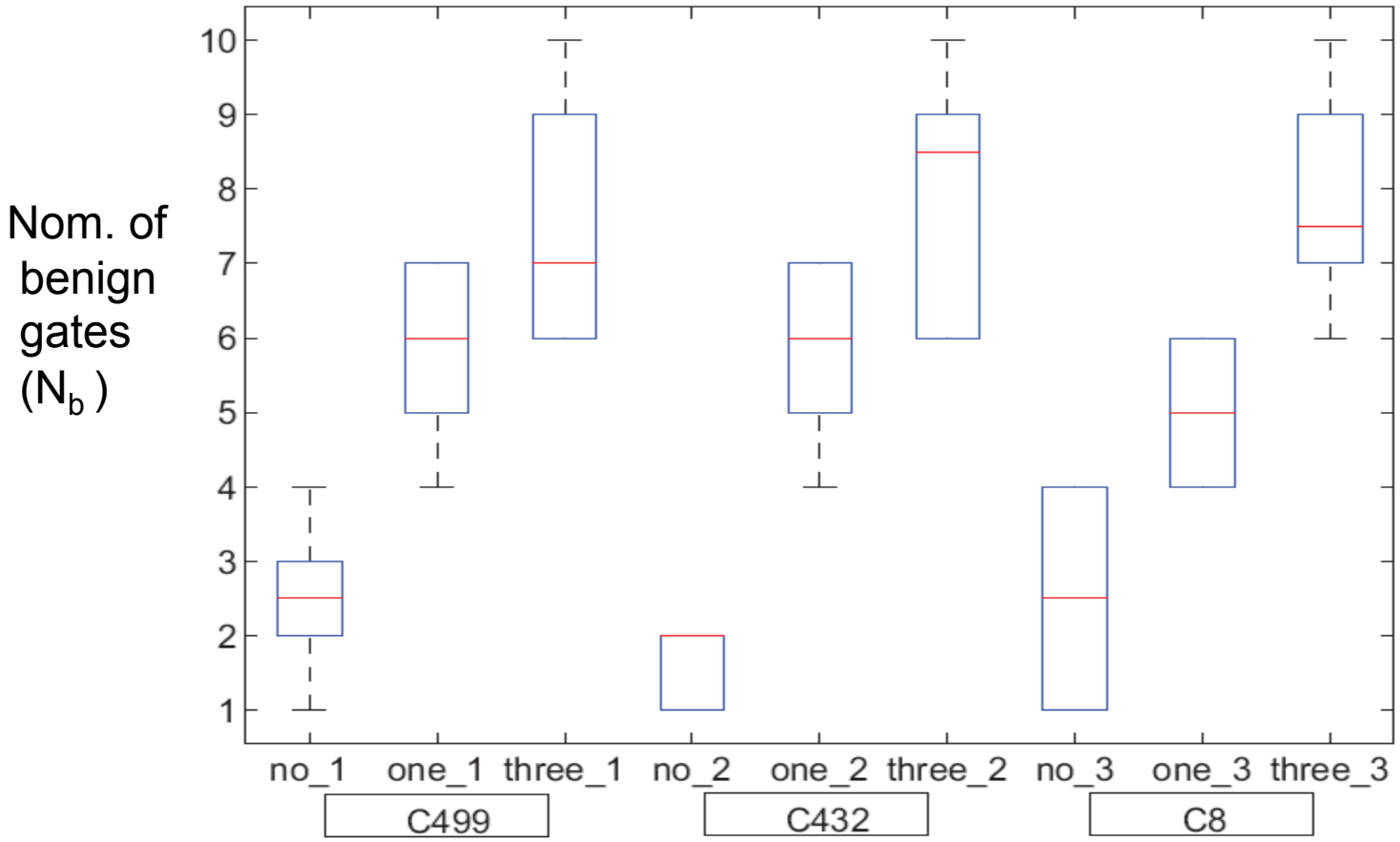
Gate-level characterization vs. measurment noise(Dynamic power)

Ct	Size	i/p	o/p	3%	5%	10%
C8	165	28	18	4.2	6.4	11.2
C432	206	36	7	1.5	3.1	6.9
C1355	512	41	32	7.8	9.1	11.5
C499	532	41	32	2.2	4.2	8.8
C3450	1131	50	22	3.5	6	9.5

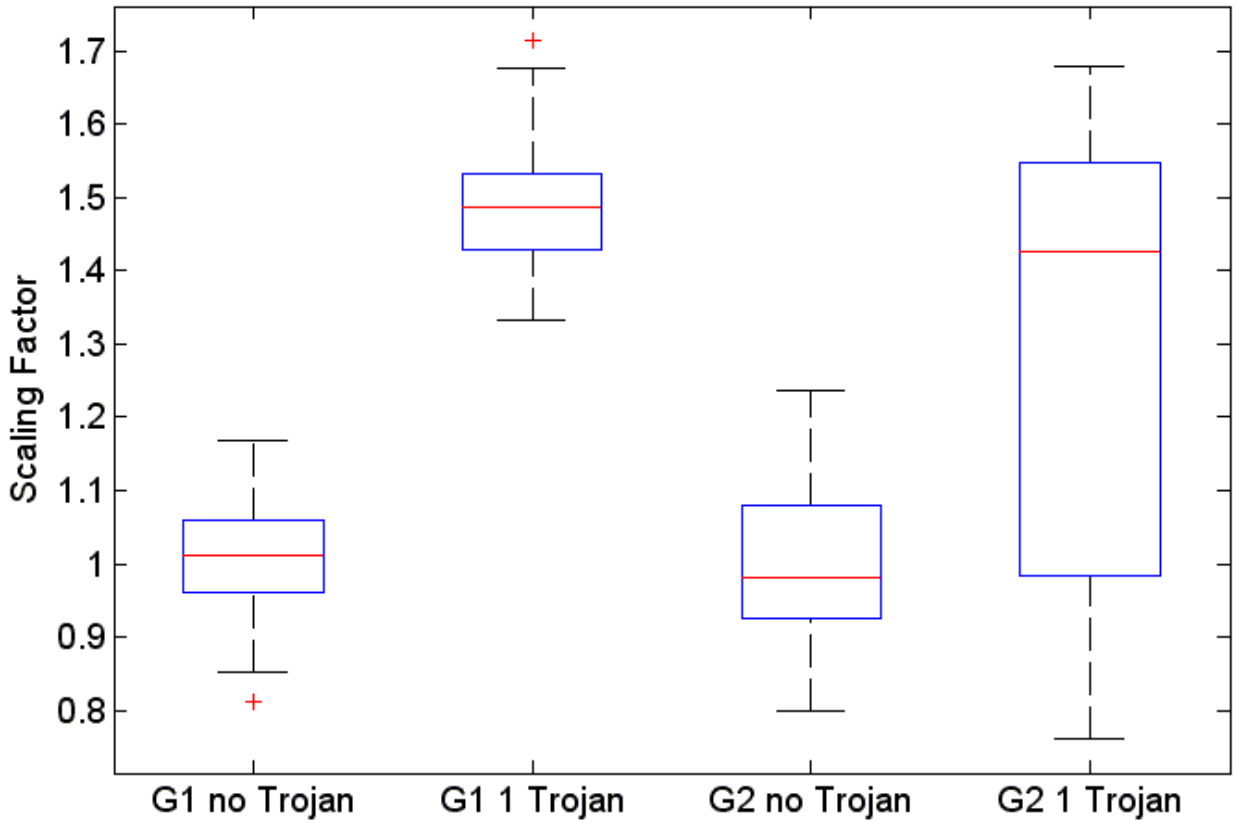
Gate-level characterization vs. measurment noise(Timing)

Ct	Size	i/p	o/p	3%	5%	10%
C8	165	28	18	5.3	7	11.5
C432	206	36	7	3.8	5.4	10.1
C1355	512	41	32	4	8	12.3
C499	532	41	32	5	6.5	12
C3450	1131	50	22	2.9	4.1	9.2

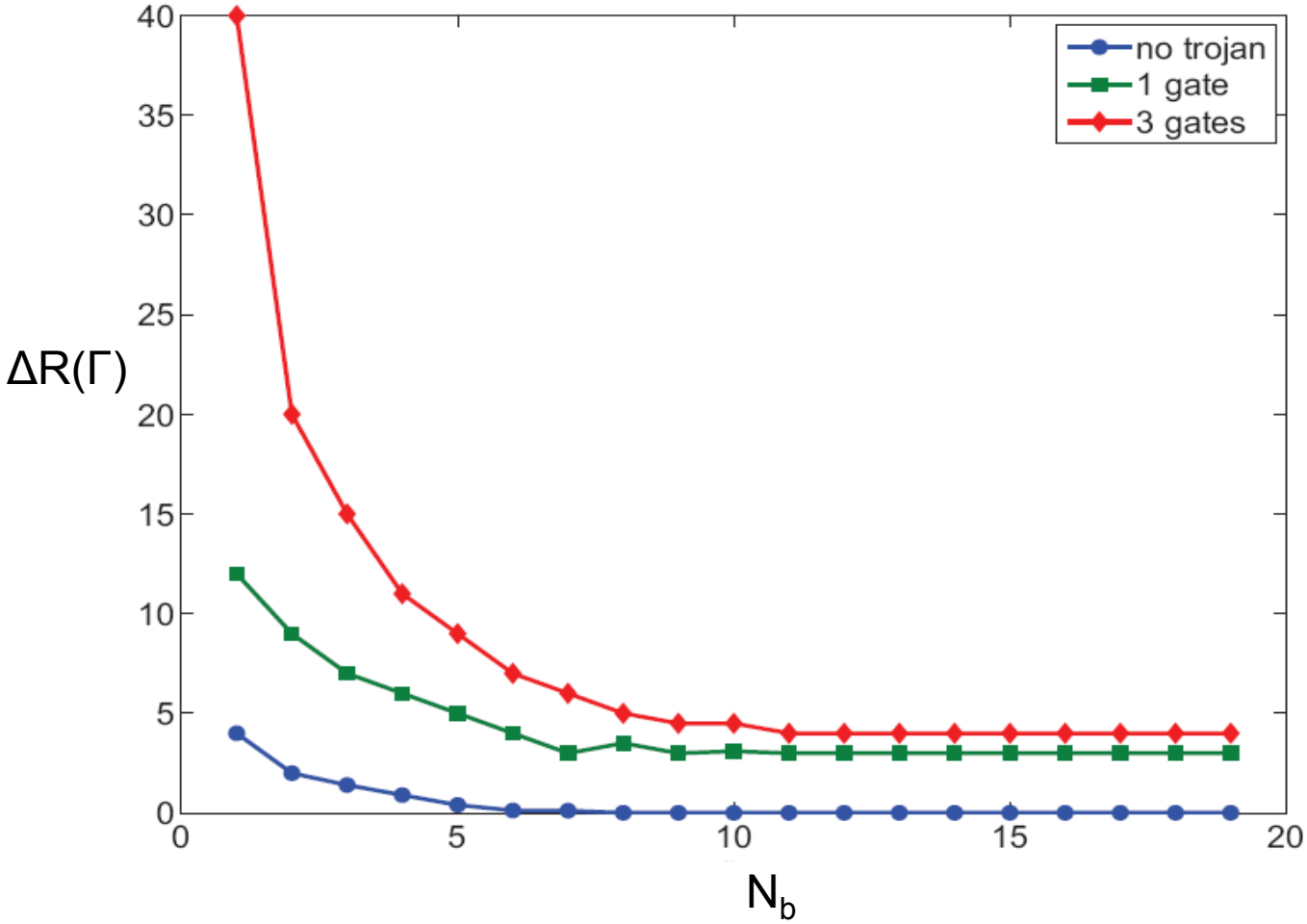
Boxplots of N_b for Trojan free, 1 Trojan, and 3 Trojan gates



Leakage scaling factors for two anomalous gates in C432



The stepwise diminishing return improvement for leakage modality



The number of gates giving false alarm in a non Trojan circuit

Ct	Unanimous	Conservative	Majority	Weighted Voting
C8	0/165	3/165	1/165	2/165
C432	1/206	2/206	1/206	1/206
C1355	0/512	4/512	2/512	2/512
C499	0/532	3/532	1/532	1/532
C3450	0/1131	3/1131	3/1131	3/1131

Summary and conclusion

- Proposing a unified noninvasive Trojan detection framework
- Formulating the optimization problem for simultaneous gate level profiles and Trojan detection for each modality
- Exploiting submodularity to achieve a near optimal solution for unimodul detection
- Devising and comparing four methods for combining the results of multiple unimodal detections.