FANCI: Identifying Malicious Circuits

Adam Waksman
Matthew Suozzo
Simha Sethumadhavan

Computer Architecture & Security Technologies Lab
Department of Computer Science
Columbia University

Do You Trust Hardware?

Cyber-attack concerns raised over Boeing 787 chip's 'back door' [1]

Researchers claim chip used in military systems and civilian aircraft has built-in function that could let in hackers

NSA Subverts Most Encryption, Works With Tech Organizations For Back-Door Access, Report Says [2]

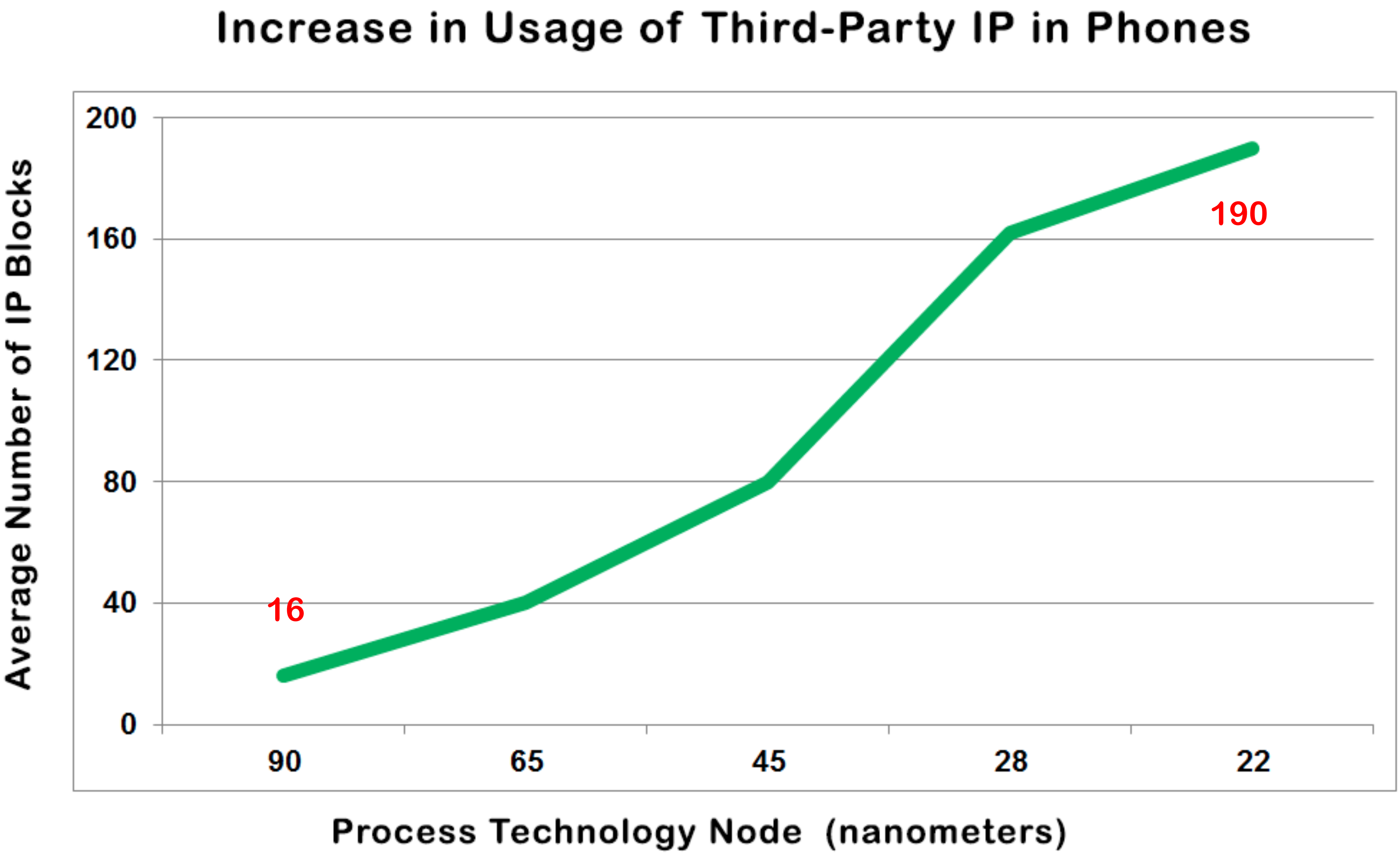
Western spooks banned Lenovo PCs after finding back doors [3]

Report suggests 'Five Eyes' alliance won't work with Chinese PCs

NSA's Own Hardware Backdoors May Still Be a "Problem from Hell" [4]

1) The Guardian 2012, 2) New York Times 2012, 3) The Register 2013, 4) Tech Review 2013

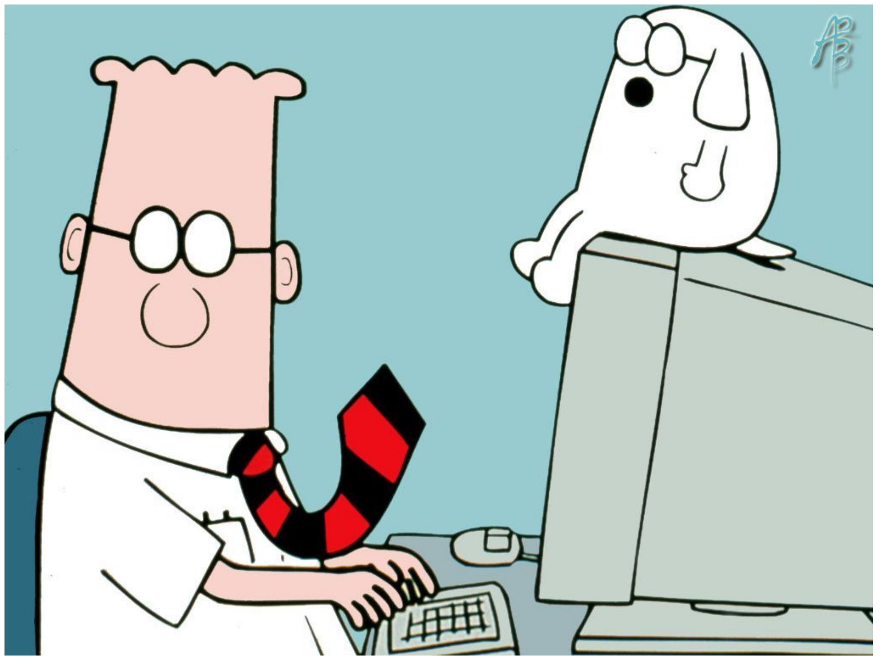
The Problem of Third-Party IP



(International Business Strategies, 2012)

Our Solution

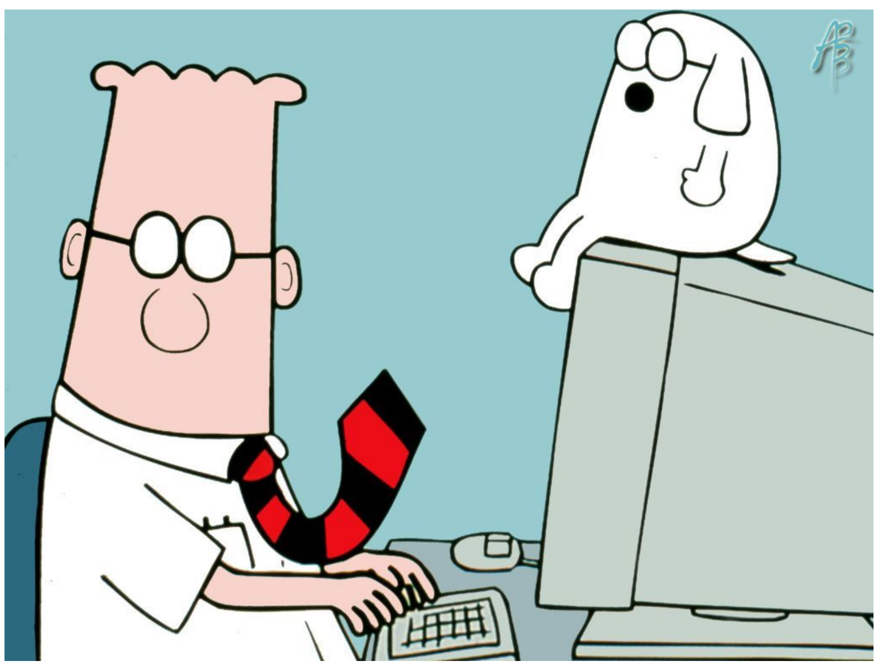
- **Automatically identify malicious circuits in third-party hardware design IP**



```
.
.
.
assign bus_x87_i = arg0 & arg1;
always @(posedge clk) begin
 if (rst) data_store_reg7 <= 16'b0;
 else begin
  if (argcarry_i37 == 16'hbacd0013) begin
   data_store_reg7 <= 16'd7777;
  end
  else data_store_reg7 <= data_value7;
 end
end
assign bus_x88_i = arg2 ^ arg3;
assign bus_x89_i = arg4 | arg6 nor arg5;
.
.
.
```

Our Solution

- **Automatically identify malicious circuits in third-party hardware design IP**
  - Engineers read few lines instead of thousands or millions



```
.
.
.
assign bus_x87_i = arg0 & arg1;
always @(posedge clk) begin
 if (rst) data_store_reg7 <= 16'b0;
 else begin
  if (argcarry_i37 == 16'hbacd0013) begin
   data_store_reg7 <= 16'd7777;
  end
  else data_store_reg7 <= data_value7;
 end
end
assign bus_x88_i = arg2 ^ arg3;
assign bus_x89_i = arg4 | arg6 nor arg5;
.
.
.
```

Currently Undergoing Testing

AIR FORCE RESEARCH LABORATORY

CSAW

CYBER SECURITY AWARENESS WEEK

@ POLYTECHNIC INSTITUTE OF NYU

Overview

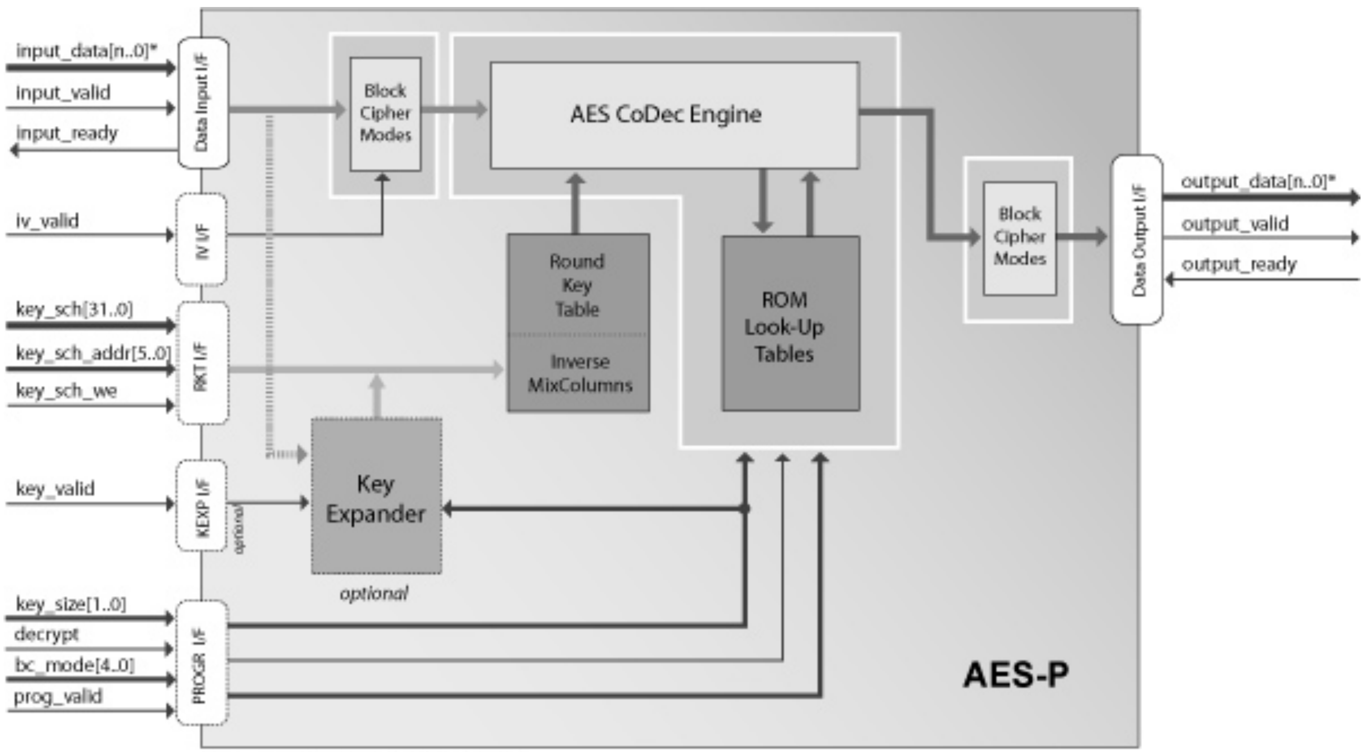
- Motivation
  - Hardware can be evil, don't live in denial
- Key Observation
  - Evil hardware is stealthy
- Algorithm
  - Rank gates by degree of stealth
- Results
  - No false negatives, pragmatic and effective
- The Future of FANCI
  - How would we attack our own tool?
- Conclusions
  - Can we really use this tool today? (Spoiler: Yes)

Backdoors: Fact #1

Backdoor = Trigger + Payload

AES Key Stealing | Ciphertext | Key Exfiltration

0xba5eba11



d(0xba5eba11)

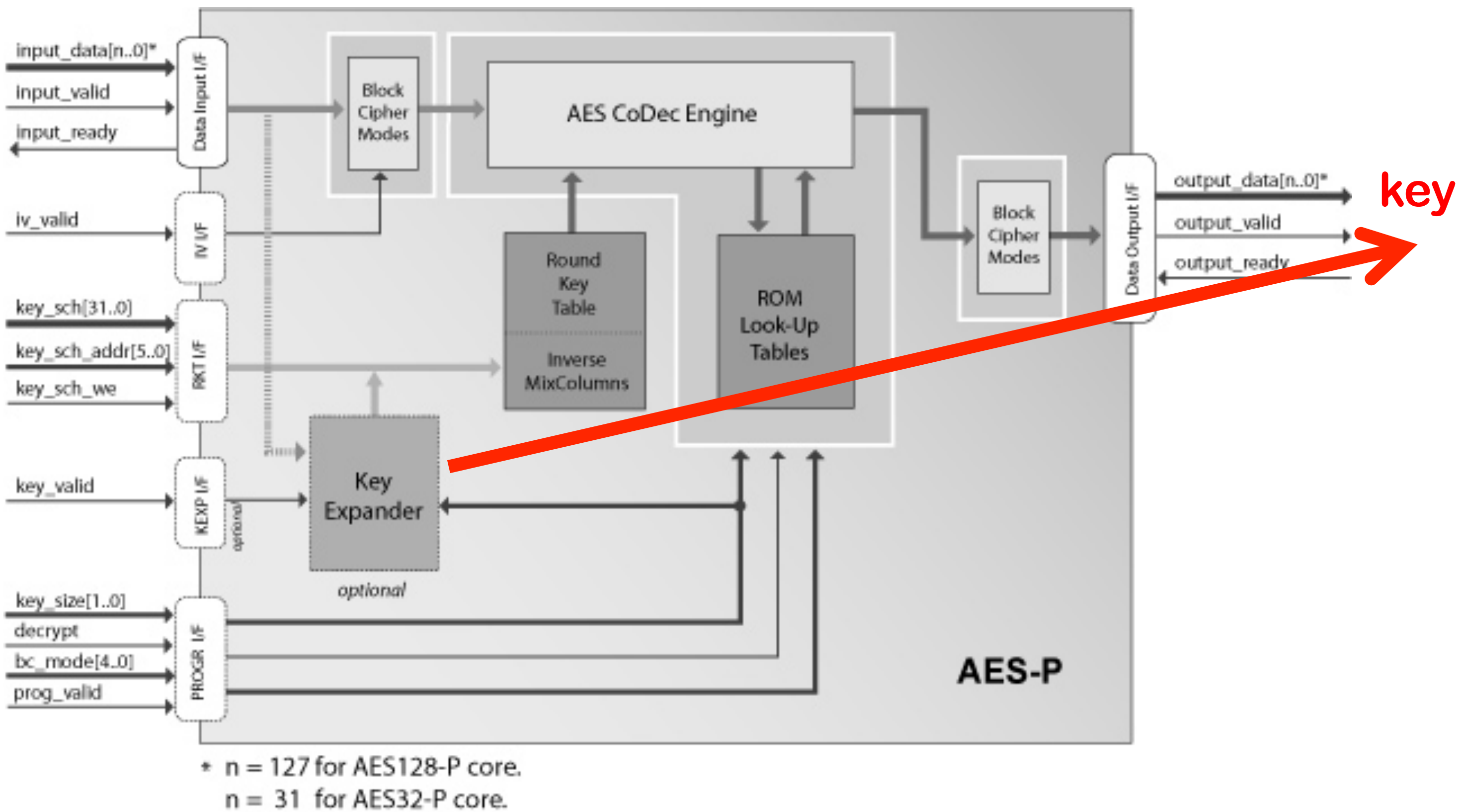
* n = 127 for AES128-P core.
n = 31 for AES32-P core.

Backdoors: Fact #1

Backdoor = Trigger + Payload

AES Key Stealing | Ciphertext | Key Exfiltration

0xba5eba11

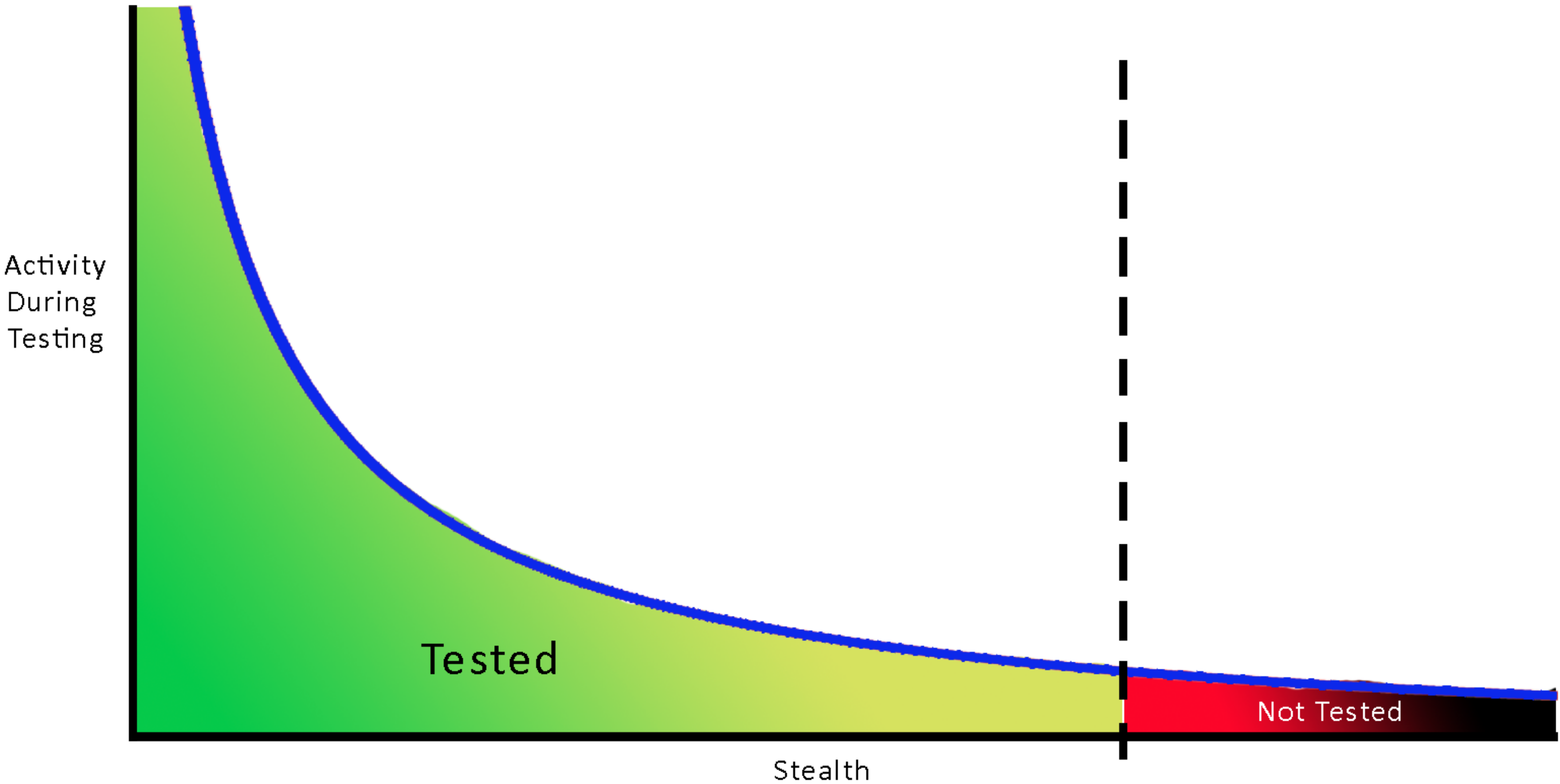


Backdoors: Fact #2

Stealth = Power

Backdoors: Fact #3

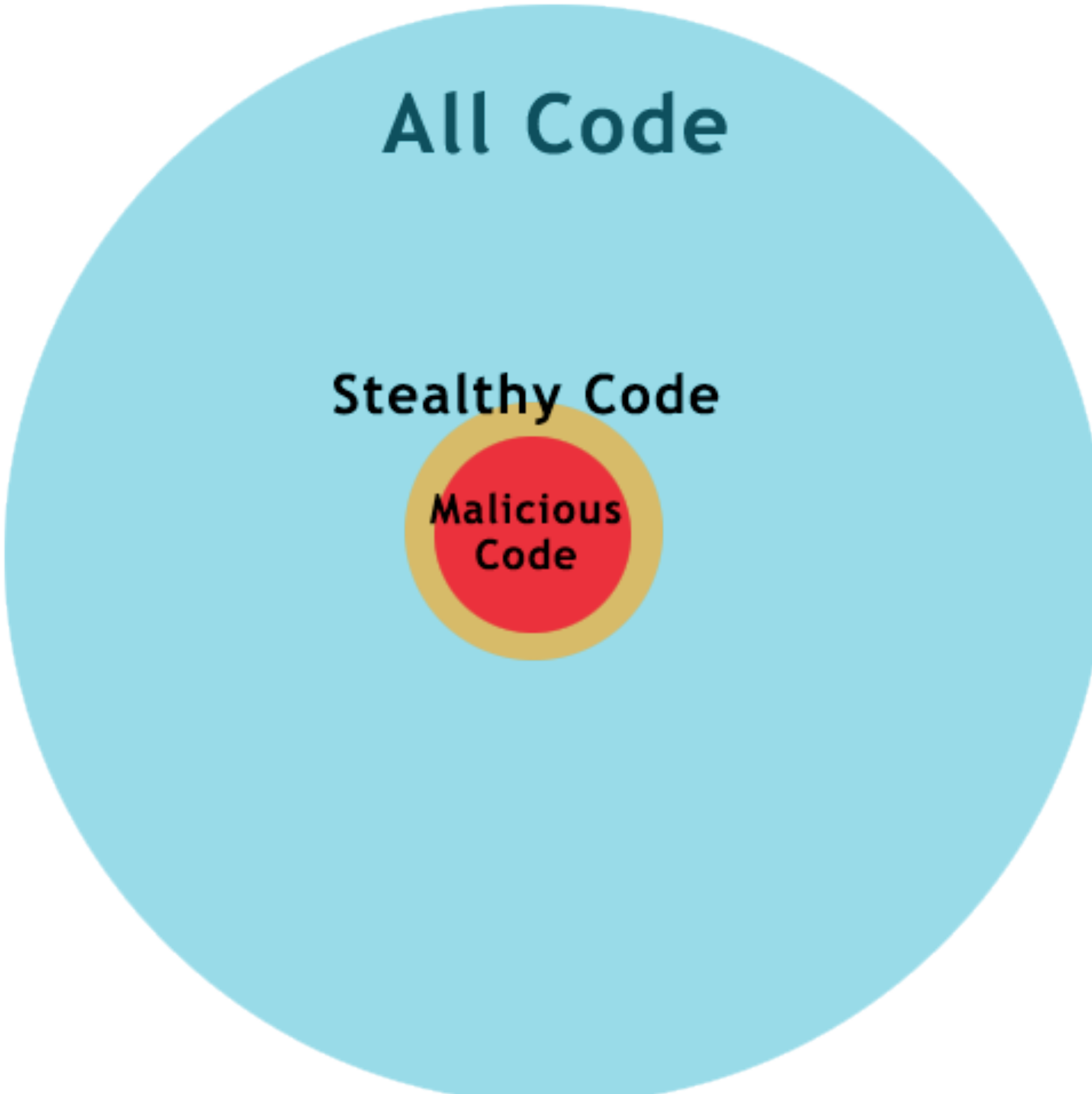
Validation != Security



What FANCI Does

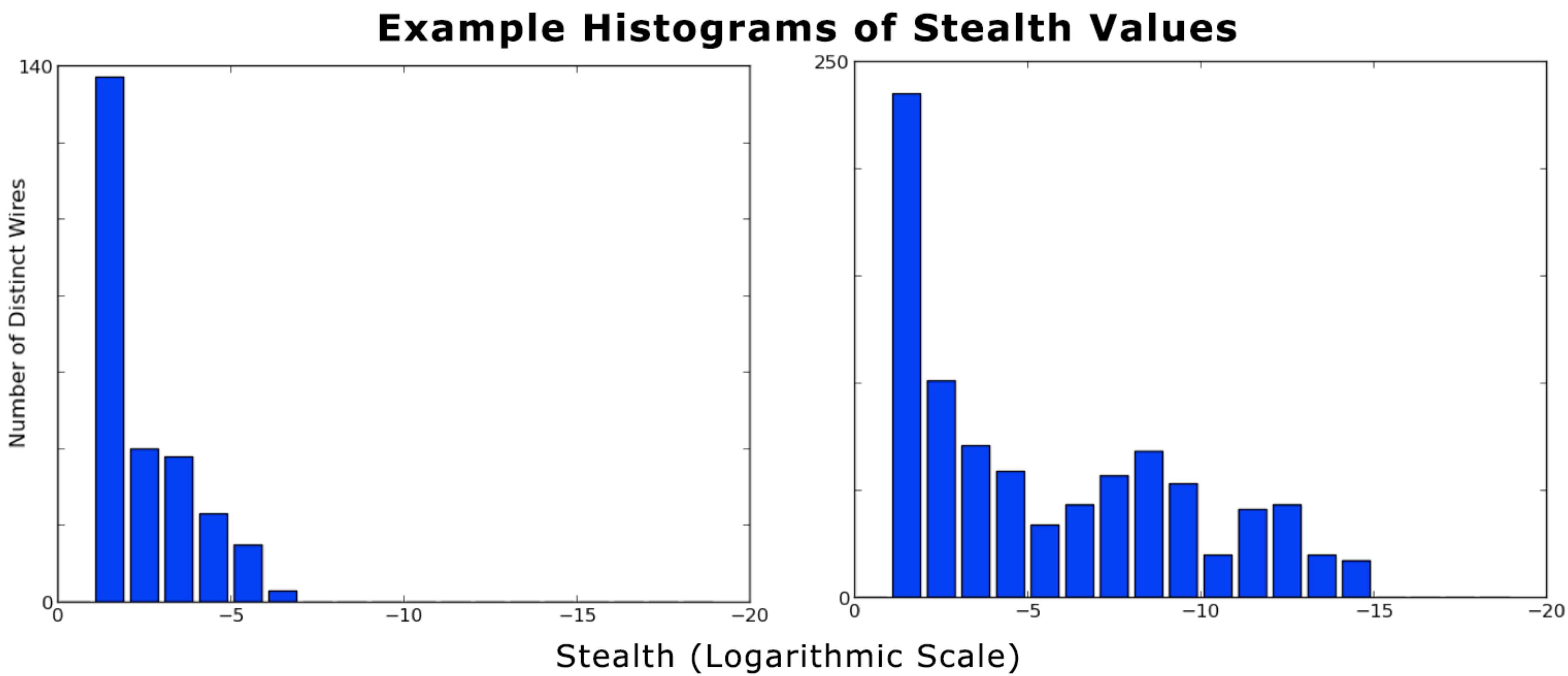
- **We need to catch stealthy circuits that validation is not able to catch**

What FANCI Does



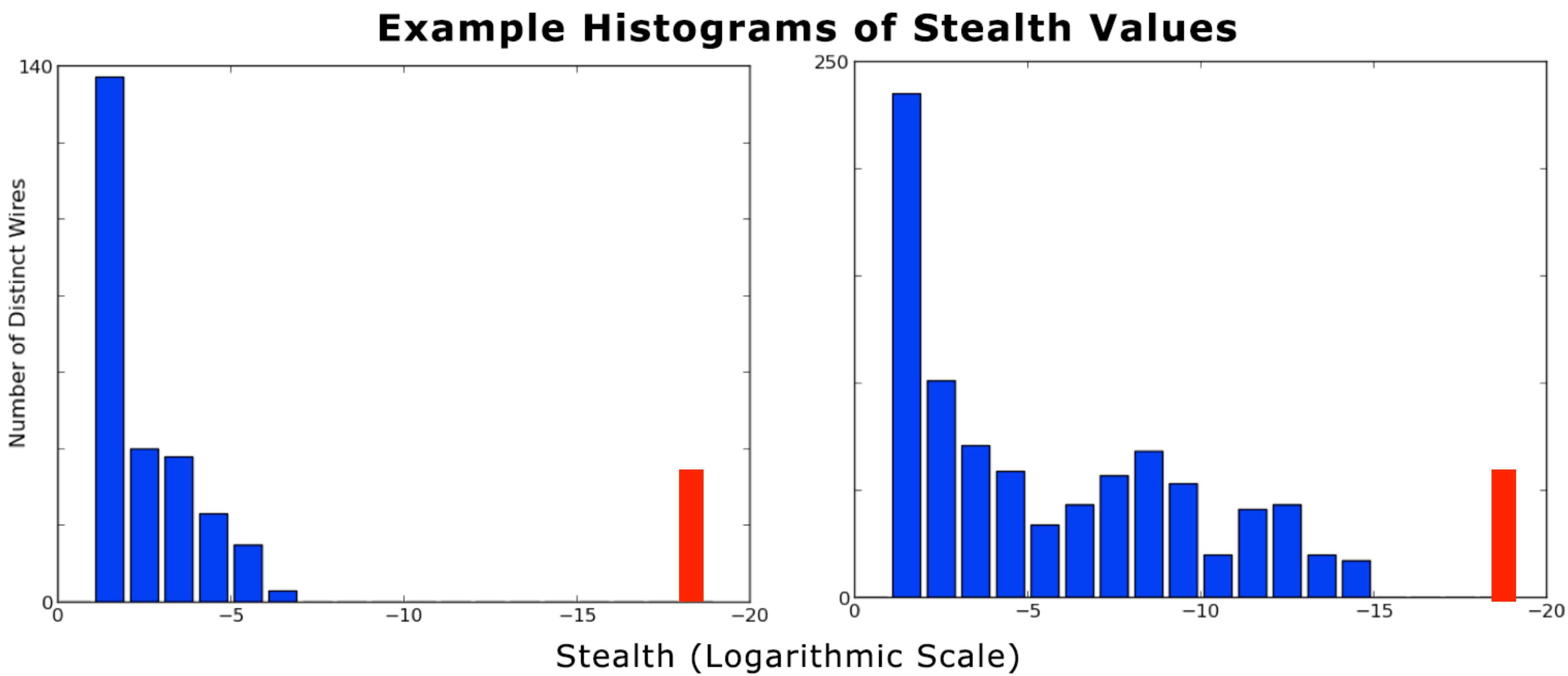
Identifying Stealthy Code

- We propose a new quantitative measure of stealth
  - We rank wires in a circuit by stealth value
- Any wire is connected to many other wires
  - *Stealth* value is computed from the *control* values of all the wires its connected to



Identifying Stealthy Code

- We propose a new quantitative measure of stealth
  - We rank wires in a circuit by stealth value
- Any wire is connected to many other wires
  - ***Stealth*** value is computed from the ***control*** values of all the wires its connected to



Defining Control

How often does an input matter?

A	B	C	OUT
1	1	1	1
1	1	0	0
1	0	1	1
1	0	0	1
0	1	1	0
0	1	0	0
0	0	1	0
0	0	0	1

Out = f(A, B, C)

How often does an input matter?

A	B	C	OUT	C Matters?
1	1	1	1	YES
1	1	0	0	
1	0	1	1	NO
1	0	0	1	
0	1	1	0	NO
0	1	0	0	
0	0	1	0	YES
0	0	0	1	

How often does an input matter?

Control = #Observed / Total = 2/4 = 0.5

A	B	C	OUT	C Matters?
1	1	1	1	YES
1	1	0	0	
1	0	1	1	NO
1	0	0	1	
0	1	1	0	NO
0	1	0	0	
0	0	1	0	YES
0	0	0	1	

The effect of C on OUT is 0.5

Larger Circuits

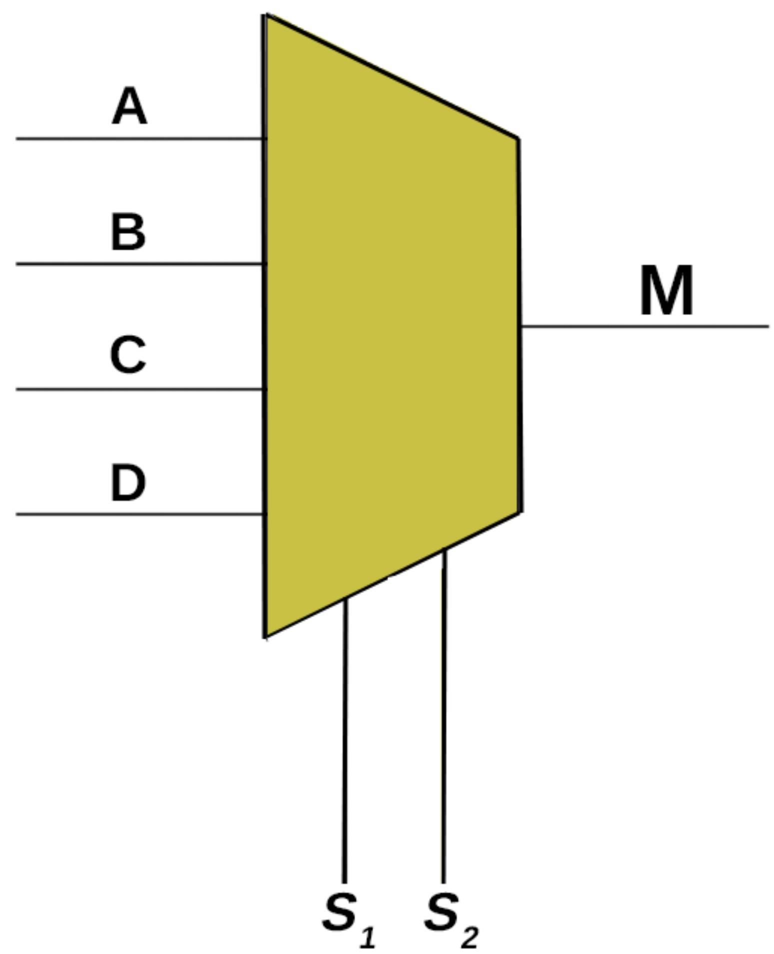
Control = #Observed / Total = 2/16 = 0.125

A	B	C	D	E	OUT	E Matters?
1	1	1	1	1	1	YES
1	1	1	1	0	0	
1	1	1	1	1	1	NO
1	1	1	0	0	1	
⋮					⋮	
0	0	0	0	1	0	YES
0	0	0	0	0	1	

32 Rows
16 Pairs

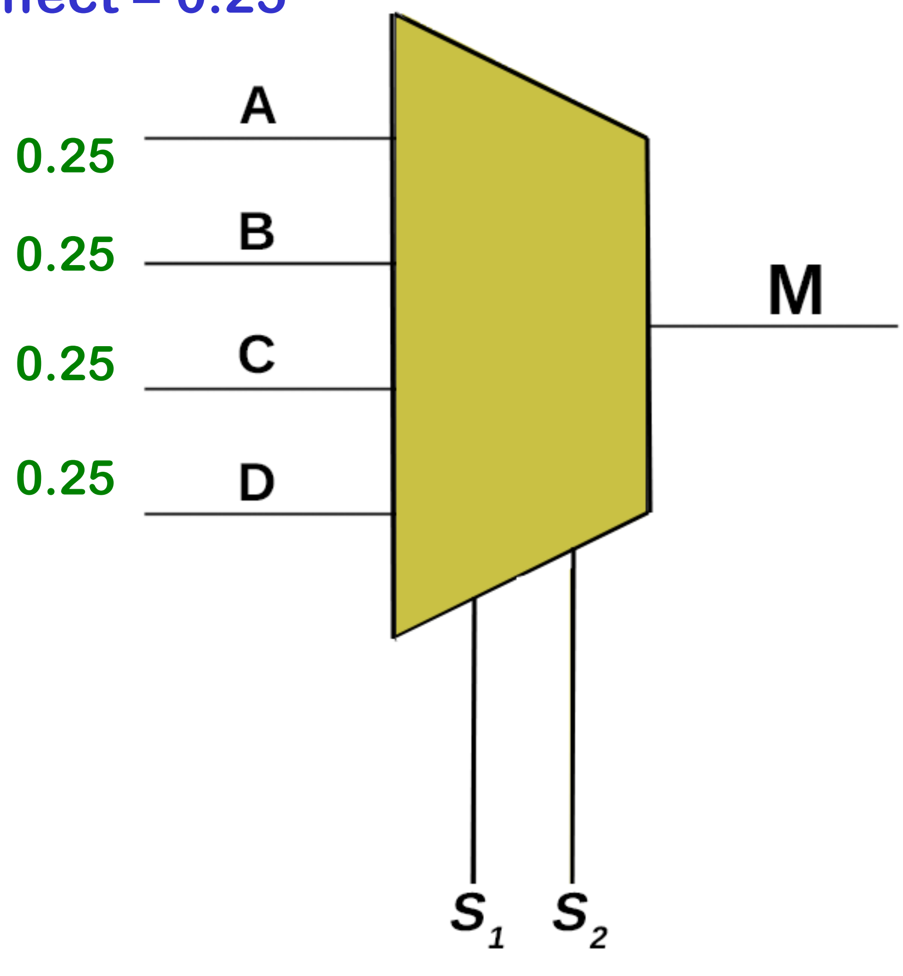
Example: 4-to-1 Mux

- Consider a real circuit (4-to-1 multiplexer)
  - How can we measure control?



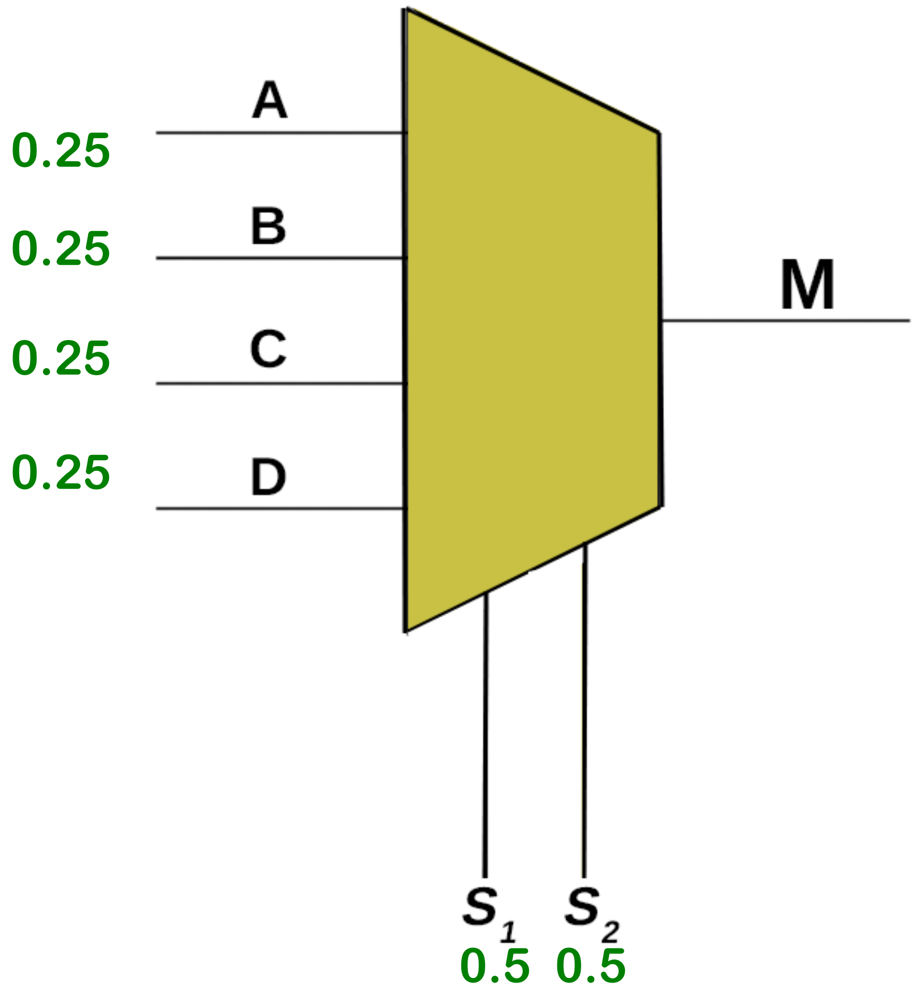
Example: 4-to-1 Mux

- **When is M dependent on A?**
  - When $S_1 = S_2 = 0$ (one fourth of cases)
  - Total effect = 0.25



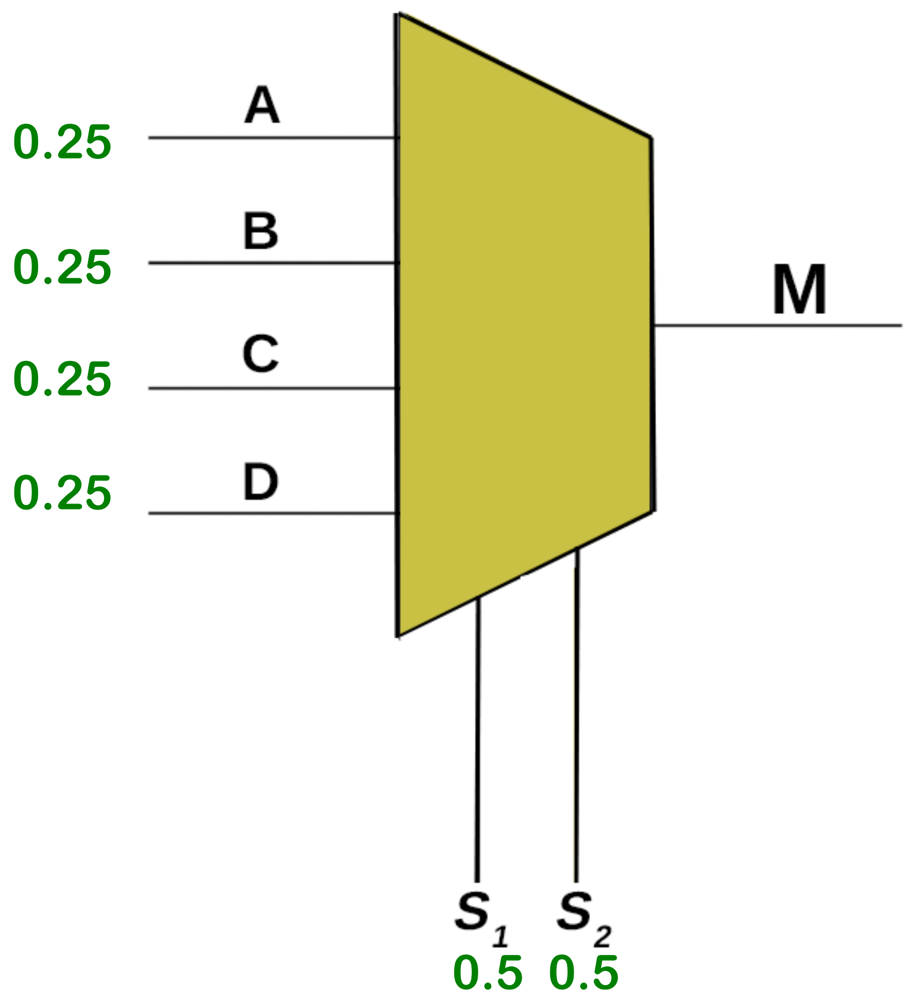
Example: 4-to-1 Mux

- M is dependent on S_1 and sometimes affected
  - When A is different from C (and $S_2 = 0$)
  - When B is different from D (and $S_2 = 1$)
  - One half of cases (total effect = 0.5)



Does This Look Suspicious?

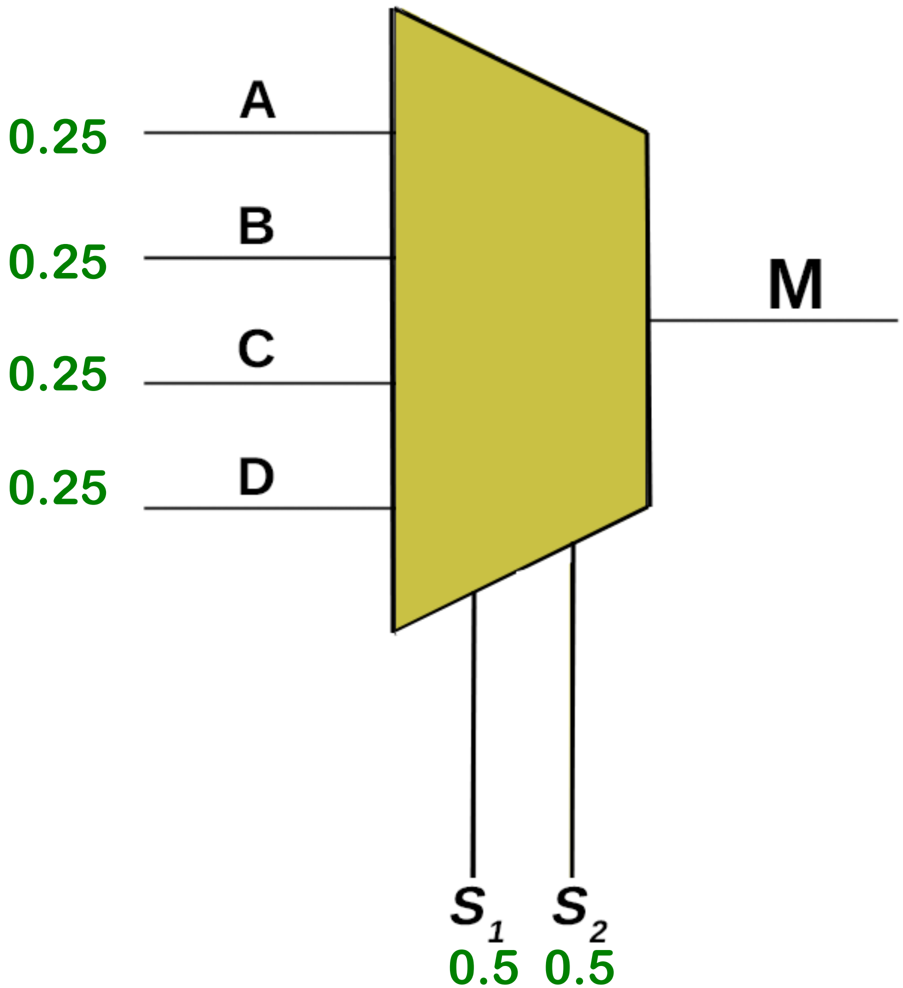
	A	B	C	D	S_1	S_2
M	0.25	0.25	0.25	0.25	0.50	0.50



Does This Look Suspicious?

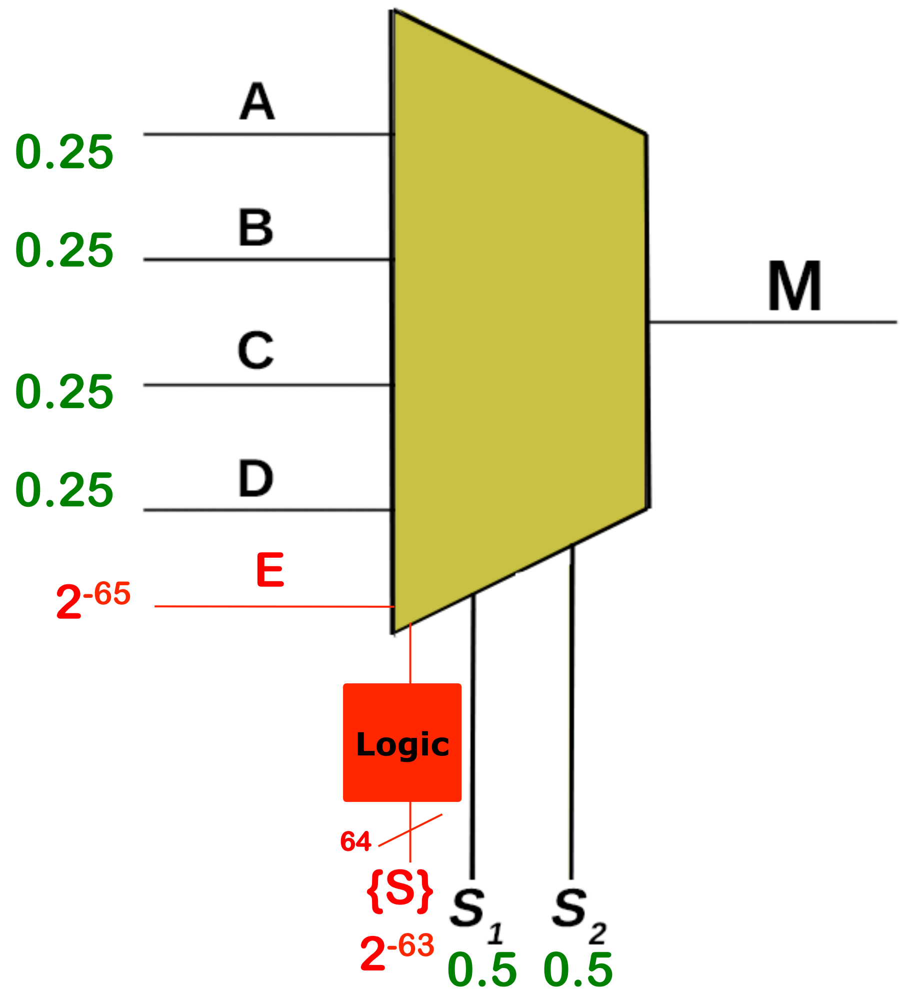
	A	B	C	D	S_1	S_2
M	0.25	0.25	0.25	0.25	0.50	0.50

Definitely not



Does This Look Suspicious?

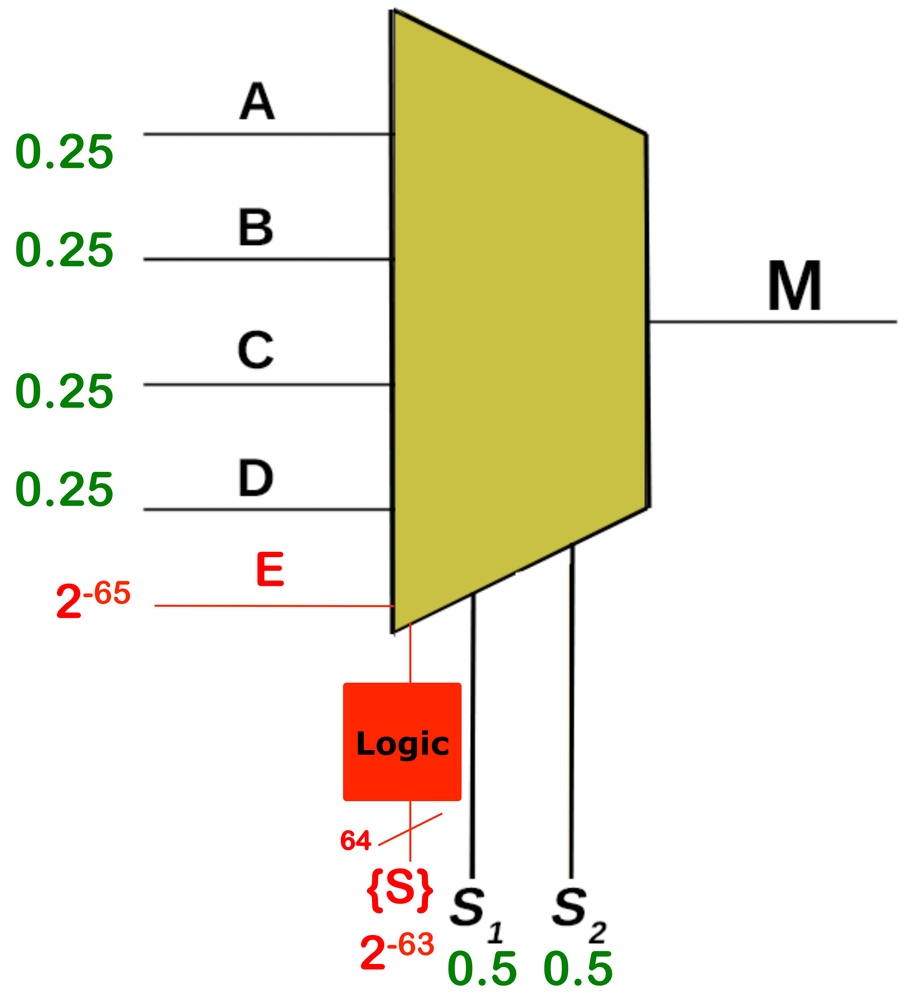
	A	B	C	D	E	S_1	S_2	$\{S_{3-66}\}$
M	0.25	0.25	0.25	0.25	2^{-65}	0.50	0.50	2^{-63}



Does This Look Suspicious?

	A	B	C	D	E	S_1	S_2	$\{S_{3\text{-}66}\}$
M	0.25	0.25	0.25	0.25	2^{-65}	0.50	0.50	2^{-63}

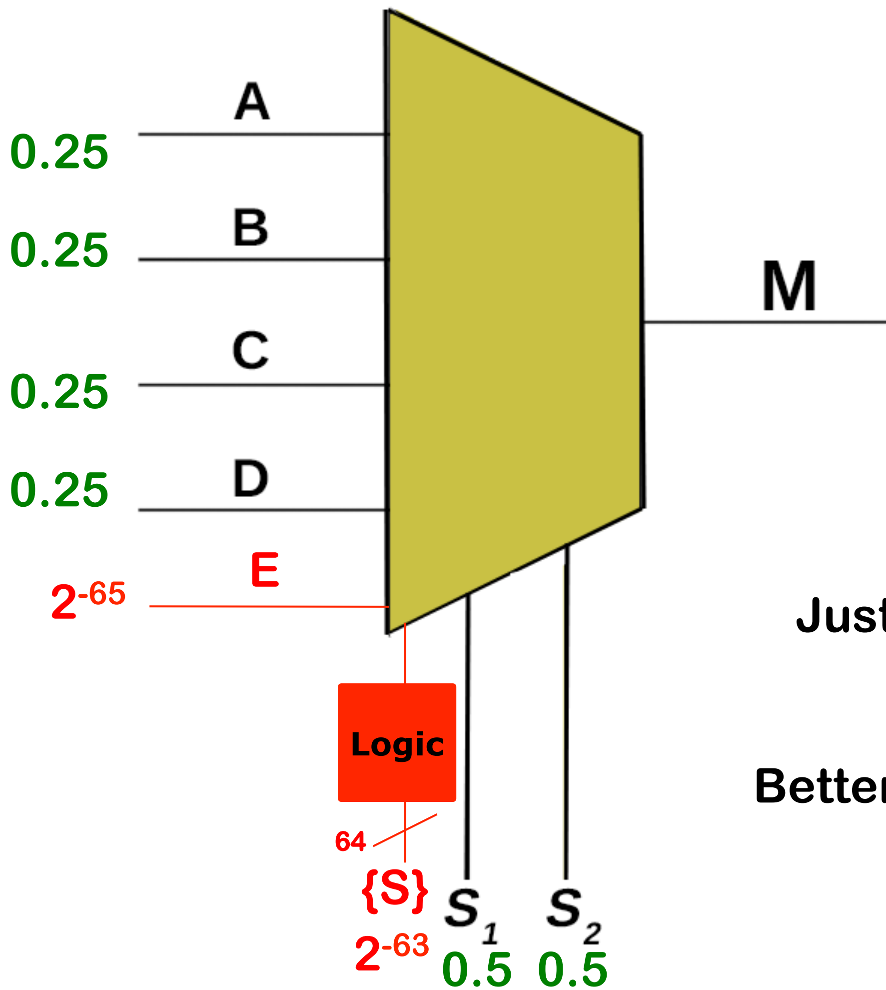
Definitely yes



Does This Look Suspicious?

	A	B	C	D	E	S_1	S_2	$\{S_{3-66}\}$
M	0.25	0.25	0.25	0.25	2^{-65}	0.50	0.50	2^{-63}

Definitely yes

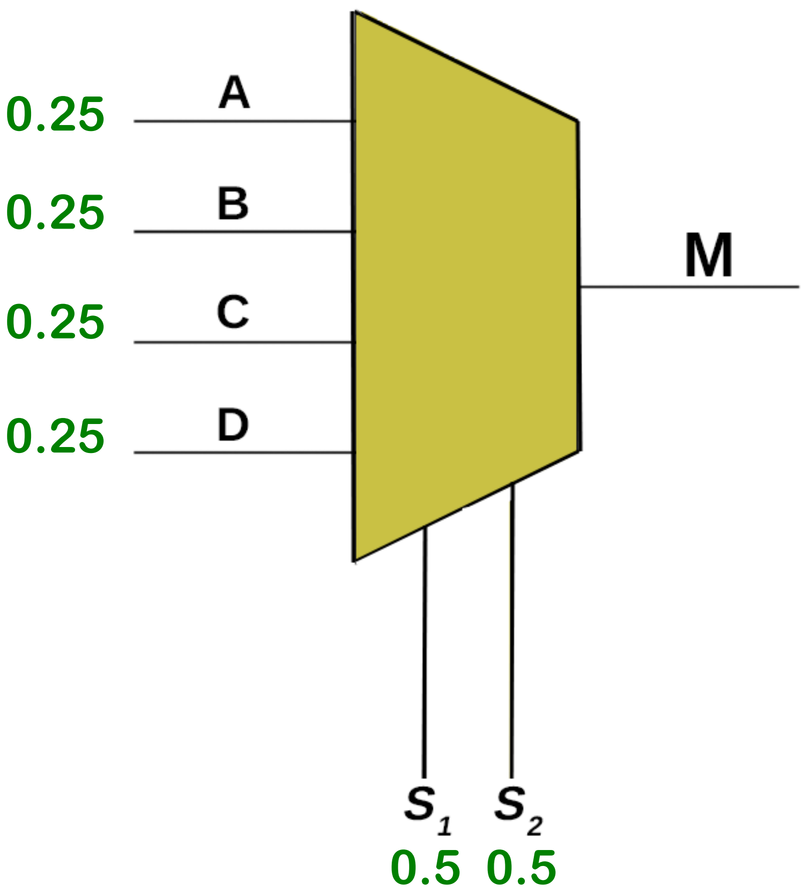


Just checking the min value
is often not enough.

Better heuristics are needed to
evaluate the vector.

Computing Stealth From Control

	A	B	C	D	S1	S2
M	0.25	0.25	0.25	0.25	0.50	0.50



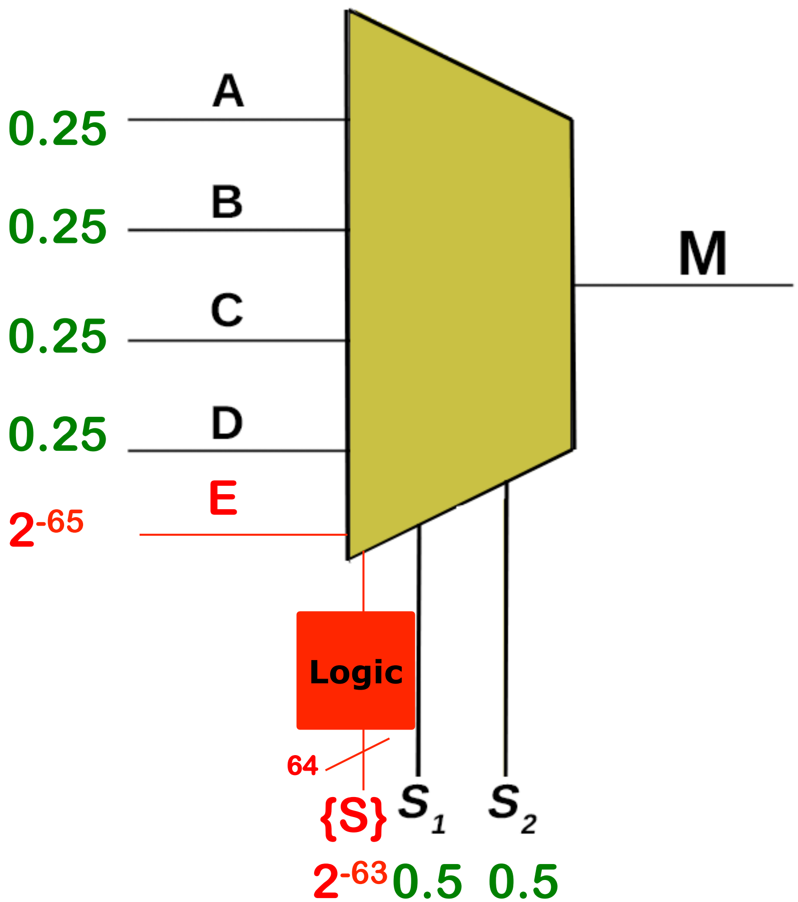
Mean(M) = (2.0 / 6) = 0.33

Median(M) = 0.25

Triviality(M) = 0.50

Computing Stealth From Control

	A	B	C	D	E	S1	S2	$\{S_{3\text{-}66}\}$
M	0.25	0.25	0.25	0.25	2^{-65}	0.50	0.50	2^{-63}



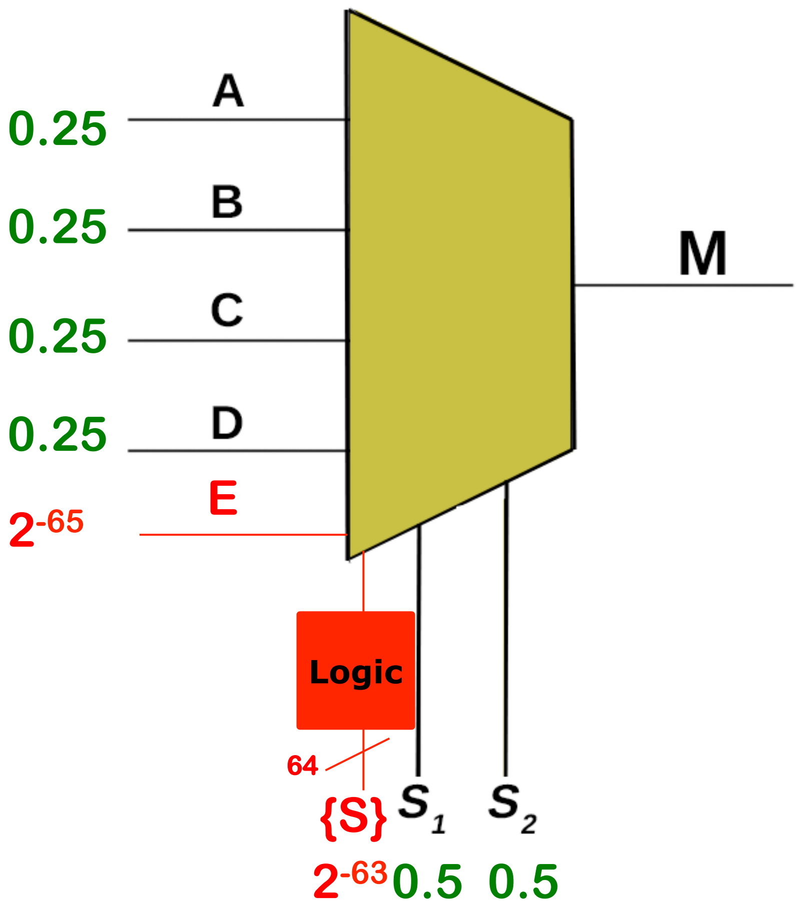
Mean(M) = (2.0 / 71) = 0.03

Median(M) = 2^{-63}

Triviality(M) = 0.50

Computing Stealth From Control

	A	B	C	D	E	S1	S2	$\{S_{3-66}\}$
M	0.25	0.25	0.25	0.25	2^{-65}	0.50	0.50	2^{-63}



Mean(M) = (2.0 / 71) = 0.03

Median(M) = 2^{-63}

Triviality(M) = 0.50

Triviality detects more triggers.
Mean/median detect more payloads.

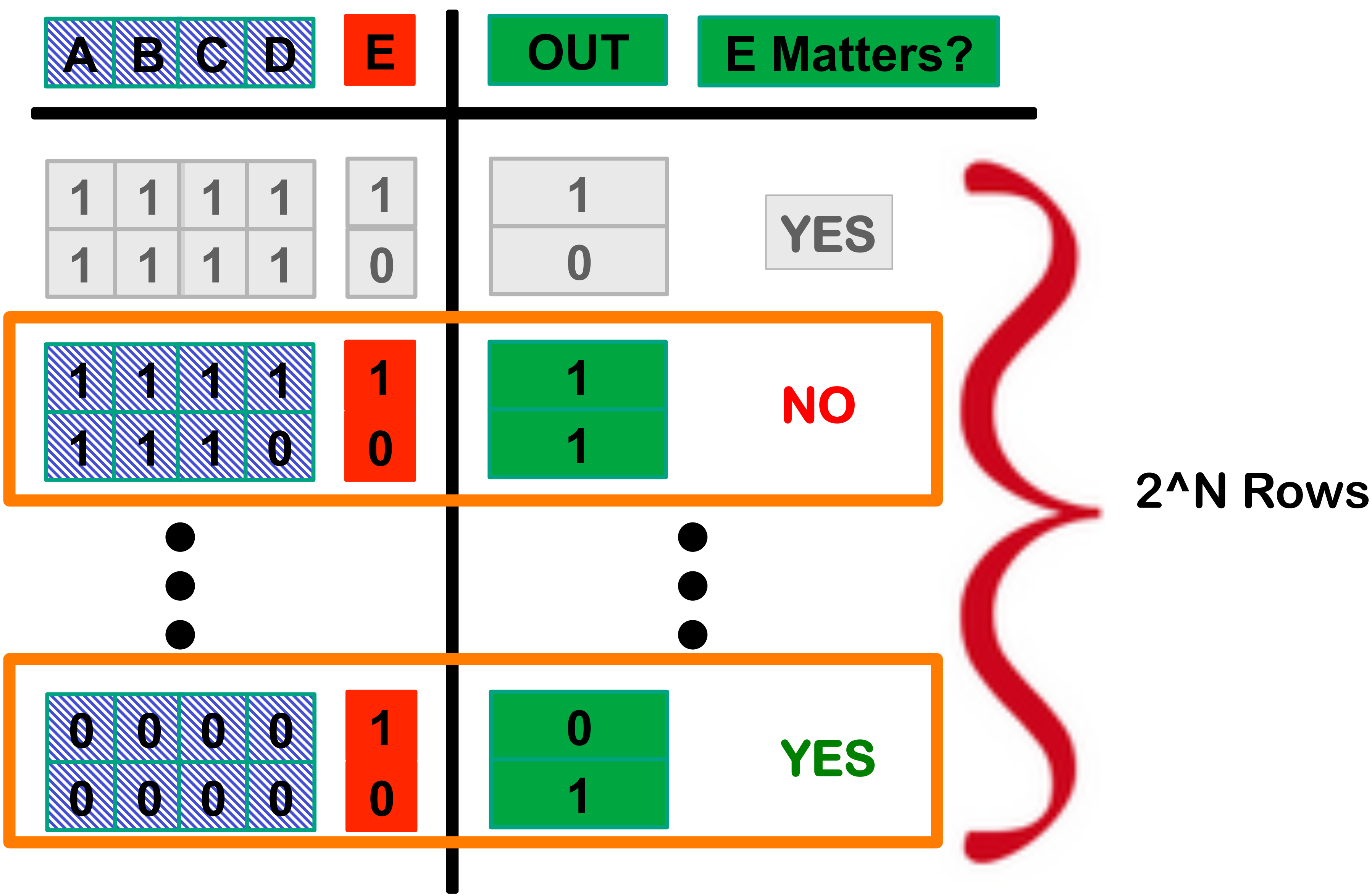
Optimization: Sampling

A	B	C	D	E	OUT	E Matters?
1	1	1	1	1	1	YES
1	1	1	1	0	0	
1	1	1	1	1	1	NO
1	1	1	0	0	1	
⋮					⋮	
0	0	0	0	1	0	YES
0	0	0	0	0	1	

2^N Rows

Optimization: Sampling

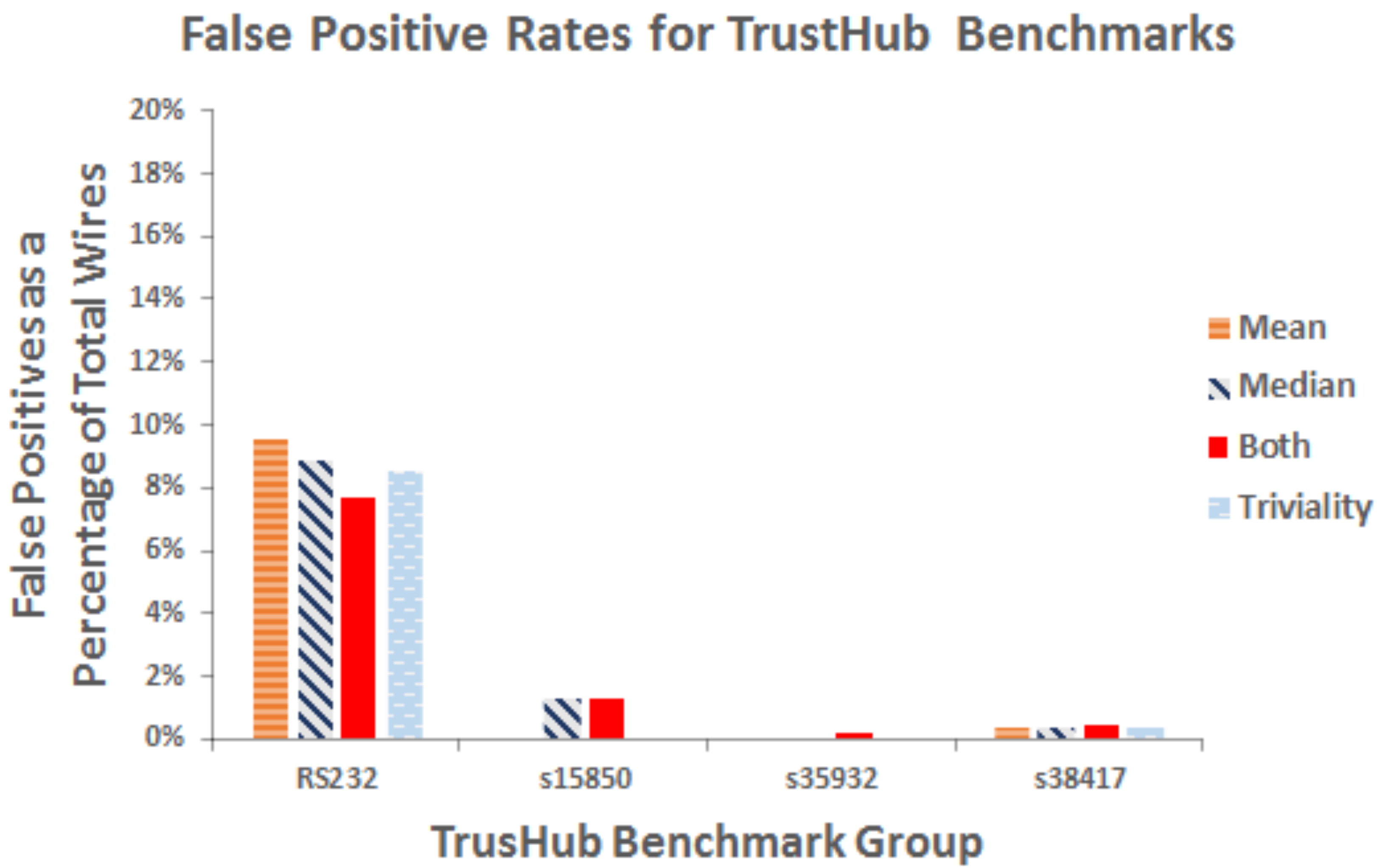
Approx. Effect = #Observed / #Sampled = ½ = 0.5



Results

- Stealth metrics are effective for existing benchmarks
  - No false negatives for TrustHub benchmarks
- Effective even on large designs
  - Able to process full (academic) microprocessor cores
- Efficient enough for modern designs
  - About 1 day to process an average sized module
- Can catch well-hidden backdoors
  - 100% coverage against “stealthy, malicious backdoors” (SSP 2011)

Effectiveness On TrustHub



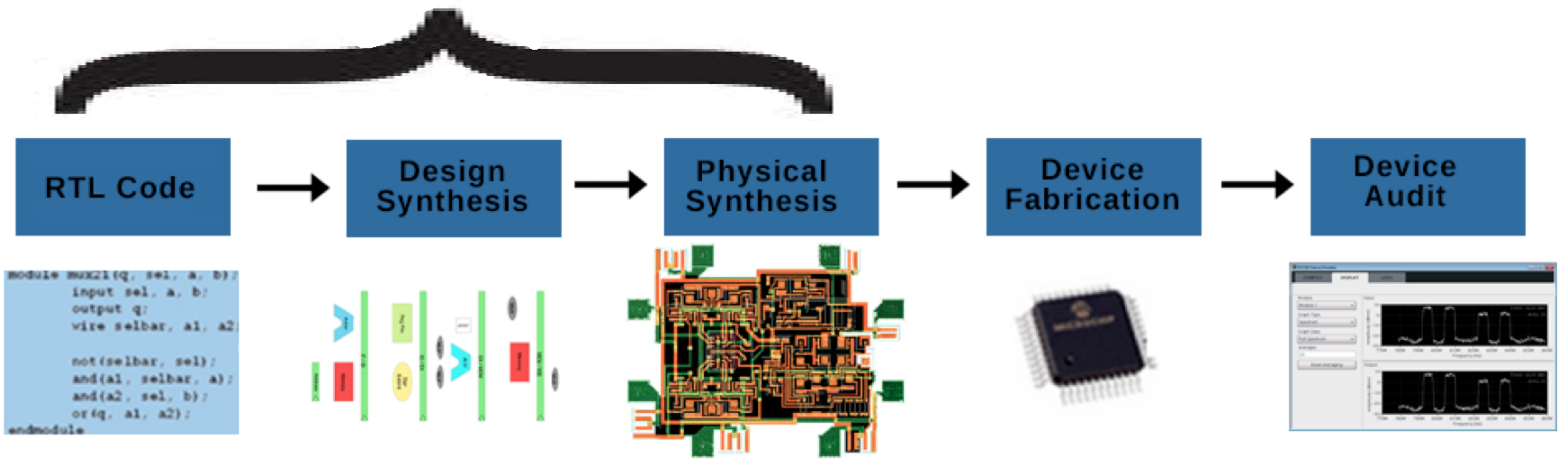
How Would We Attack FANCI?

- Frequent-Action Backdoor
  - No stealth, requires incompetent/non-existent validation engineers
- False Positive Flooding
  - Contrived design, requires naïve integration engineer
- Pathological Pipeline (State Explosion) Backdoor
  - Contrived design, requires naïve integration engineer
- Foundry (Physical/Parametric) Backdoor
  - Malicious device from benign design, requires malicious foundry

Security Assurances

- Zero false negatives ***so far***
  - Mathematical connection exists between stealth and validation
- FANCI flags wires if and only if they are stealthy
  - Static and not probabilistic or dynamic
- Can operate on digital, synchronous design IP
  - Source code or gatelists
- Can achieve design-side security with minimal validation
  - Works well with current state of practice

The Big Picture: Hardware Security



The Big Picture: Hardware Security

- Design Attacks
  - Insiders
    - Hicks et al., 2010, Waksman et al., 2010 and 2011
  - Third-Party IP
    - This Talk
- CAD Tool Attacks
  - Automated Malicious Design IP
    - This Talk
- Foundry Attacks
  - Counterfeiting
    - Chakraborty et al., 2008, Rajendran et al., 2012
  - Malicious Injections
    - Agrawal et al., 2007, Banga et al., 2008, Salmani et al., 2009, Next talk

Conclusions

- Hardware backdoors: A serious, immediate threat
  - Currently no way to certify trustworthiness
  - Causes tech. localization (increased costs)
- FANCI: Static analysis to identify suspicious circuits
  - Zero false negatives so far
  - Minimal reliance on validation personnel
- Current Status
  - Practical, ready for modern designs (e.g., AFRL, CSAW)
  - First hardware certification tool for trustworthy IP