SAND2017-3879PE

On Trust Analysis for Microelectronics-Based Systems

April14, 2017

Brandon Eames, PhD
Technical Staff
Sandia National Laboratories
bkeames@sandia.gov

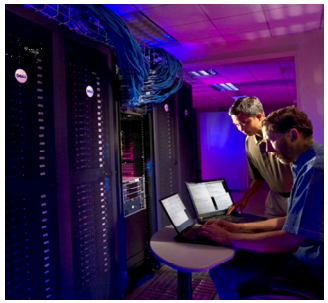
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

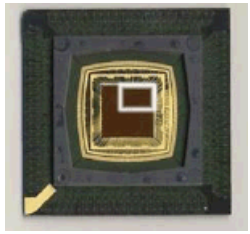
Project GUNMAN

- Based on a tip from a foreign government, in 1984 the USG quietly and quickly replaced 10 tons of electronic equipment in the US Moscow embassy
- Subsequent evaluation of replaced equipment revealed a sophisticated bug in a small number of IBM Selectric typewriters

Trust in Microelectronics Based Systems

- Society relies on microelectronics-based systems for safety, security, entertainment, travel, etc.
  - Military systems, satellites, cyber infrastructure, critical infrastructure (e.g. power grid), etc.




- Can adversaries manipulate these systems as they are developed? What would the impact be?
- Can these systems be ***trusted*** to perform their intended function?

How vulnerable are systems to development-time manipulation?

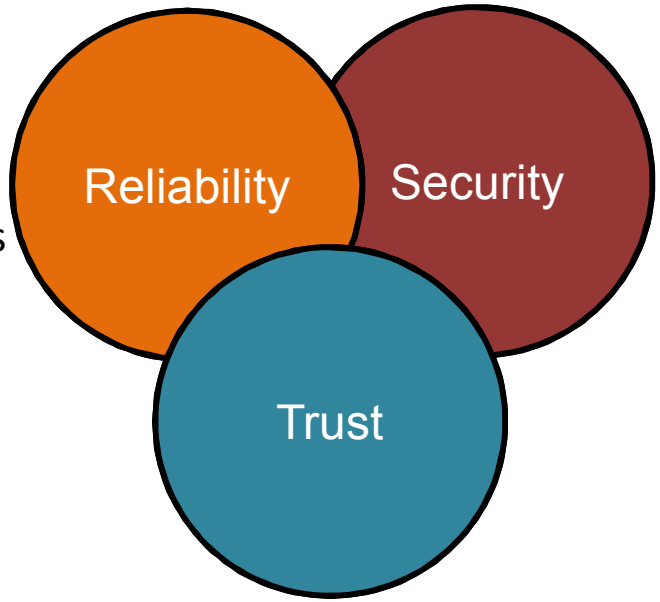
Reliability vs. Security vs. Trust

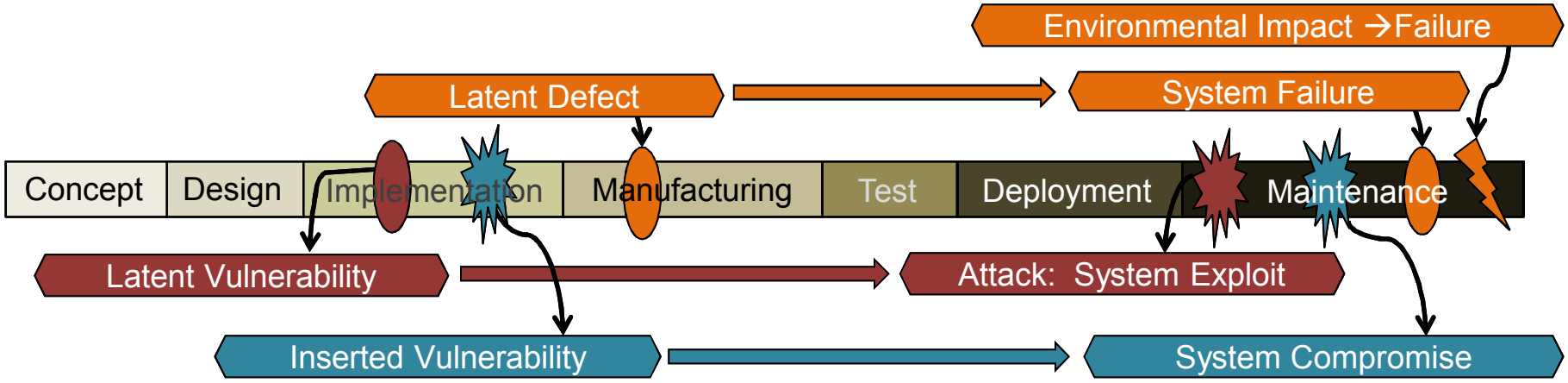
- **Reliability**: The probability that an item will perform a required function under stated conditions for a stated period of time
  - Premature System Failure → Design for Reliability

$$R \approx 1 - \left[\left(1 - \prod_{i=1}^{5}(1-J_i)\right) + \sum_{i=1}^{7} K_i^2 + 2K_2\left(K_3 + \sum_{i=5}^{7} K_i\right) \right]$$

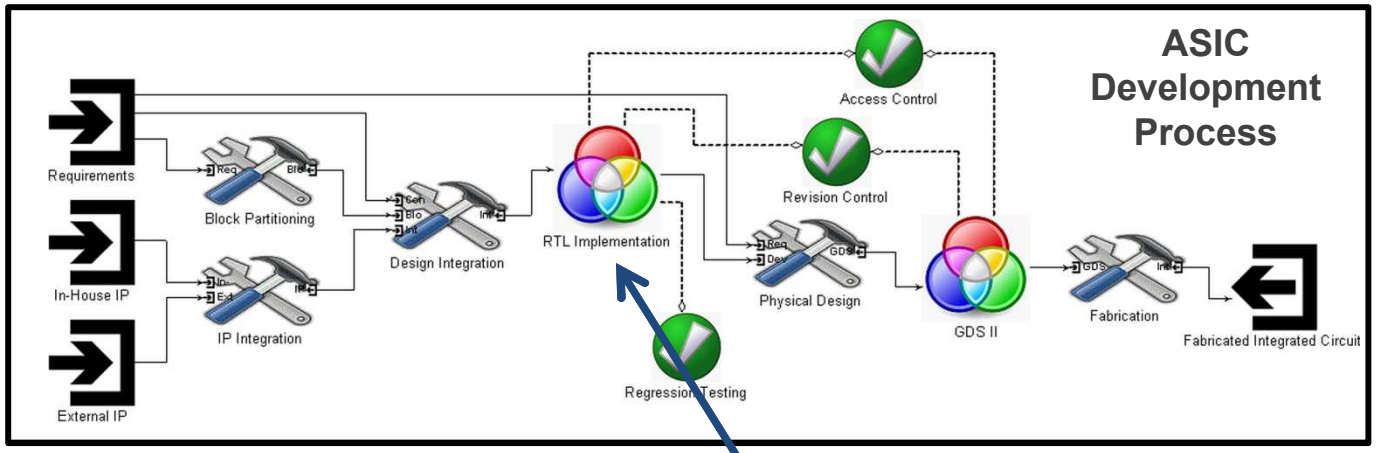
- **Security**: The protection of systems from theft or damage …, as well as from disruption … of the services they provide.
  - System Exploitation → Design for Security

- **Trust**: The confidence in … secur[ing] national security systems by assessing the integrity of the people and processes used to design, generate, manufacture, and distribute … [systems]
  - System Compromise → **Design for Trust**

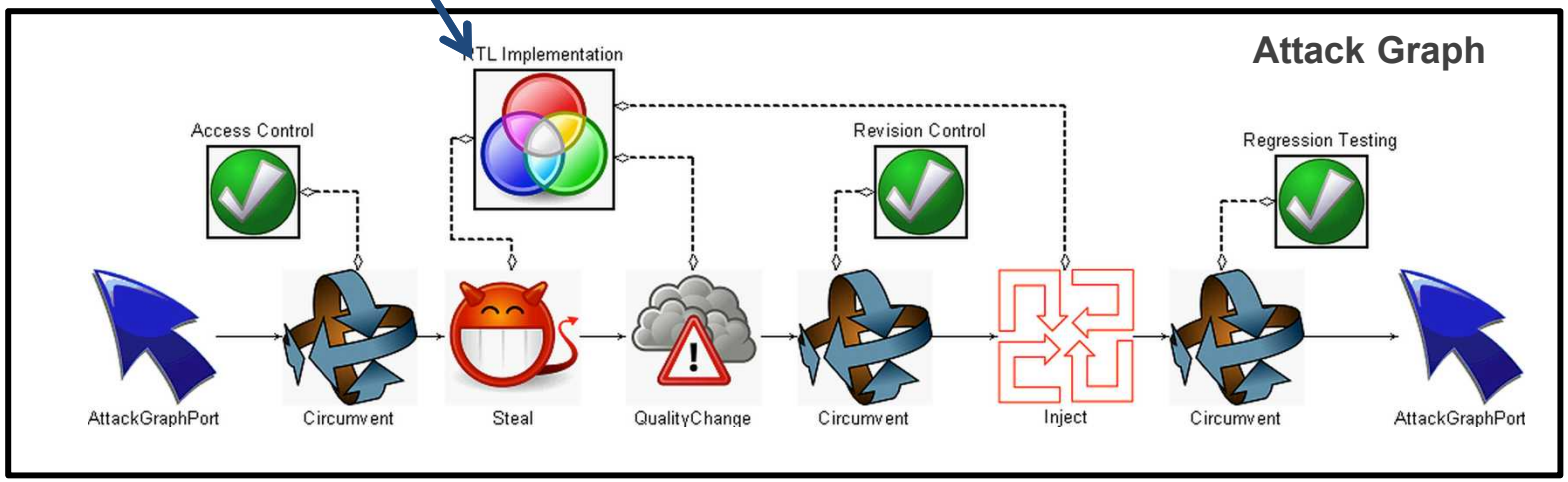




Example System: Custom ASIC

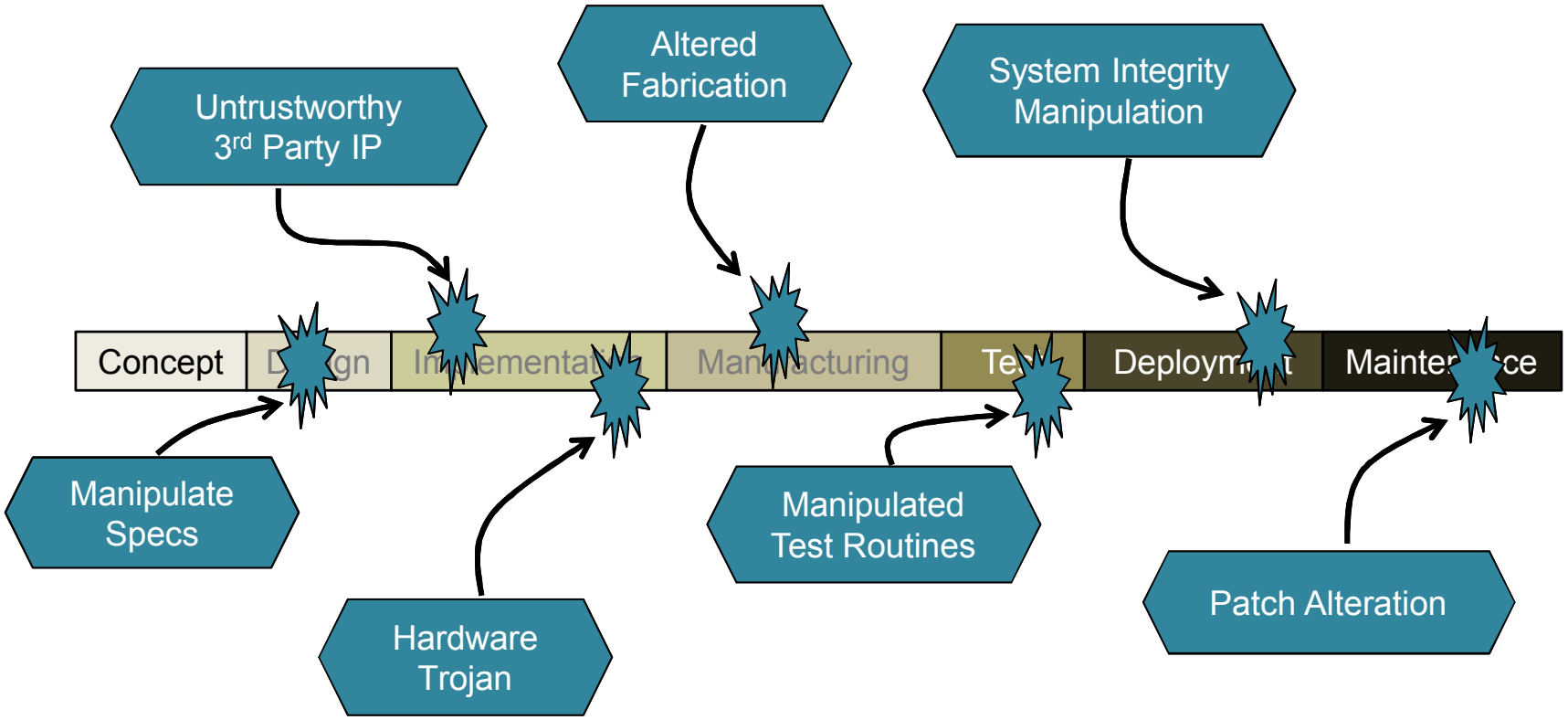


- ASIC being developed for USG program
- RTL files maintained on access-controlled, internet connected servers



- Attacker seeks to modify ASIC source RTL files without being detected
- Attack must complete before RTL is handed off to Physical Design team

Where Trust Breaks Down



Adversaries can potentially manipulate development at any point

The Challenge With Trust

Sandia National Laboratories

Is my system Trustworthy?

How much “trust” do I need?

Cost vs. Benefit of trust?

Trust vs. SWaP?

What should I really be worried about?

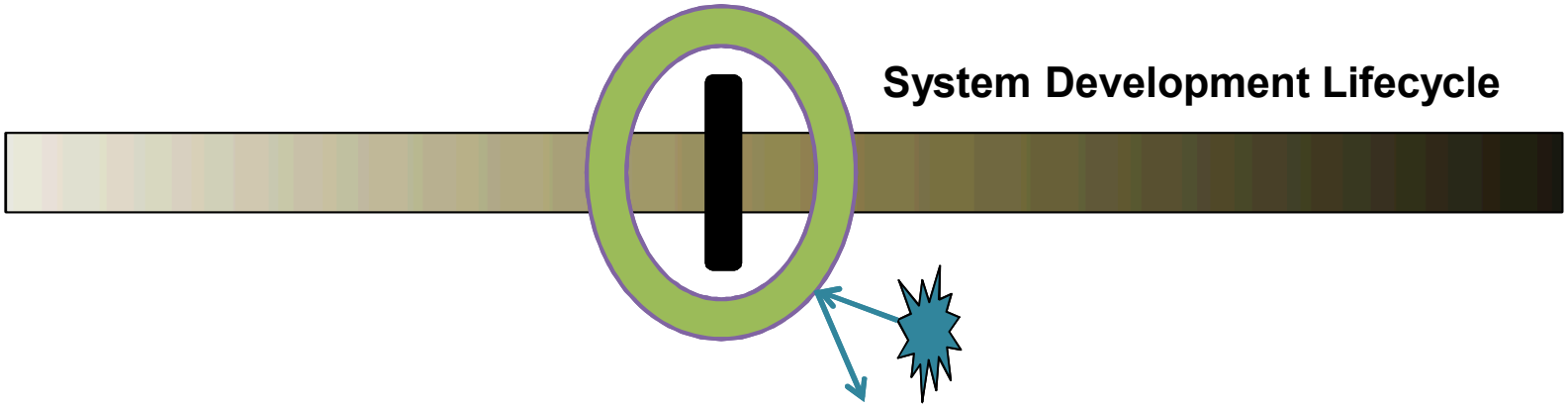
What mitigations make sense?

How do I engineer Trustworthy systems?

Is Trusted manufacturing enough?

When are we done?

Current Approach to Trusted System Development

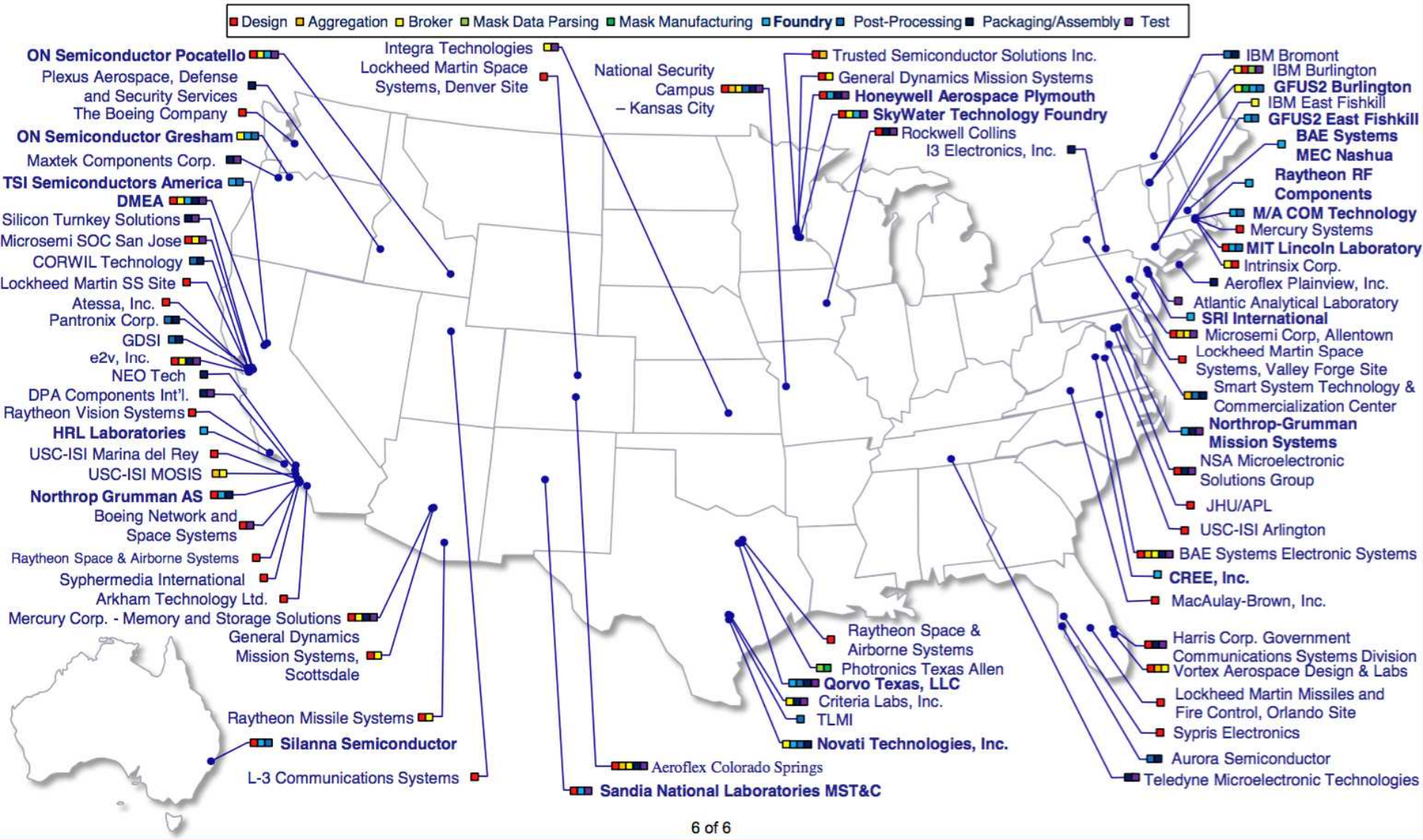


Isolate Development Process to Prevent Attacks

- Keep the attacker from manipulating the system / development process
- Process-based approaches: control information flow, control supply chain, isolated manufacturing etc.

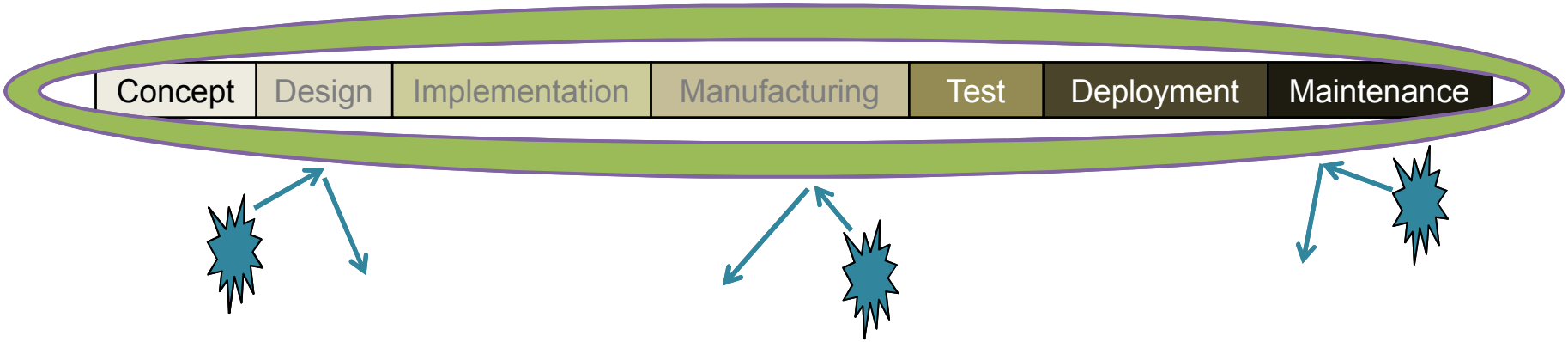
- Examples:
  - Trusted Foundry Program: Certification process to establish domestic, isolated microelectronics fabrication
    - Ensure integrity, availability of microelectronics fabrication
  - Isolated computer networks
  - Vetted design teams

78 Trusted Suppliers



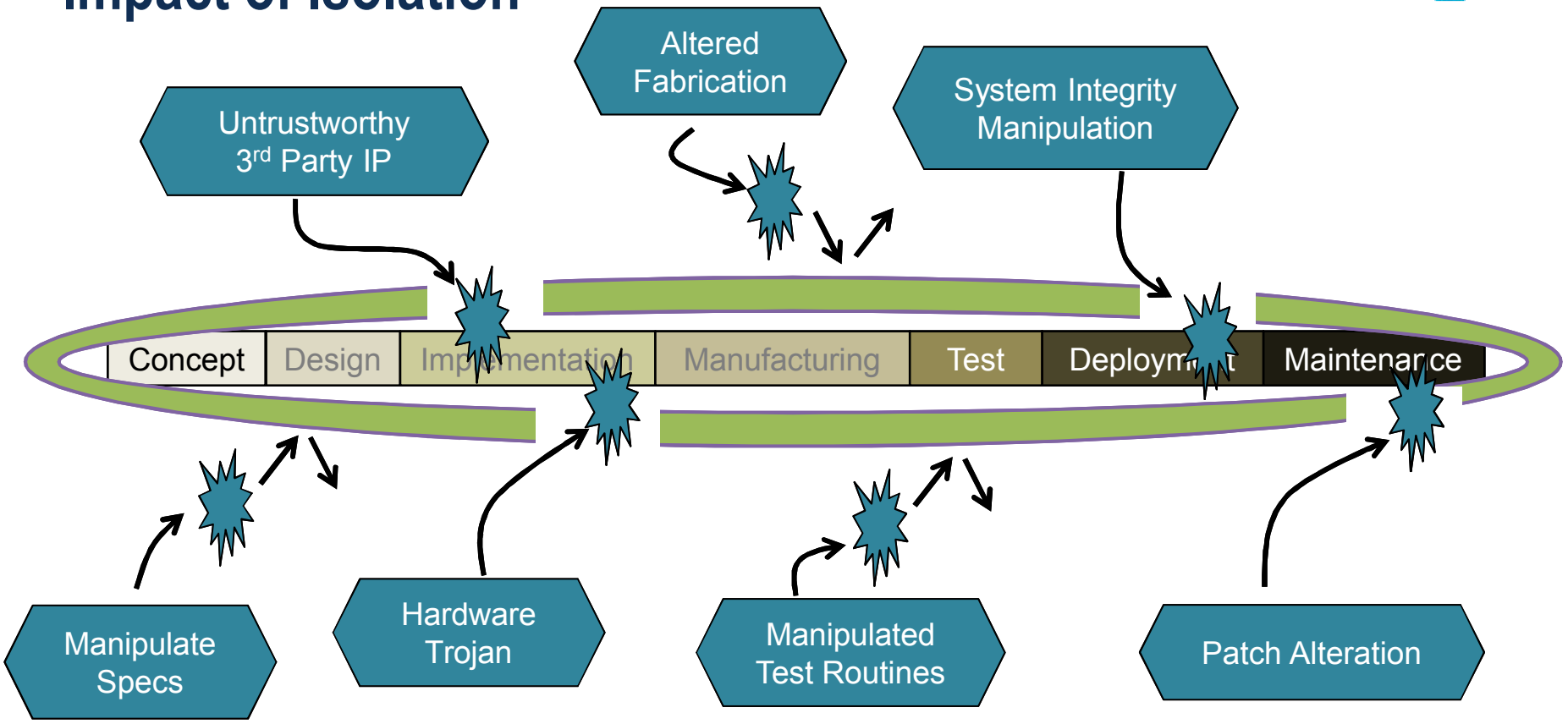
As of 15 March 2017

Impact of Isolation



- Isolation can be highly effective as an adversarial deterrent
- Can we fully isolate the complete system development lifecycle?
  - Captive fabrication (trusted foundry) addresses only one aspect of the development process
  - Completely isolated development processes are VERY expensive
    - Consider cost of leading edge microelectronics fabrication facility
  - Systems use COTS components, development tools
  - Insider threat?

Impact of Isolation



- Currently identified isolation techniques can be highly effective at deterring many paths of adversary access
- ***Gaps Remain: Practicality of real system development precludes complete isolation***

The Challenge With Trust

Sandia National Laboratories

Is my system Trustworthy?

How much “trust” do I need?

Cost vs. Benefit of trust?

Trust vs. SWaP?

What should I really be worried about?

What mitigations make sense?

How do I engineer Trustworthy systems?

Is Trusted manufacturing enough?

When are we done?

Risk-Informed Management of Enterprise Security

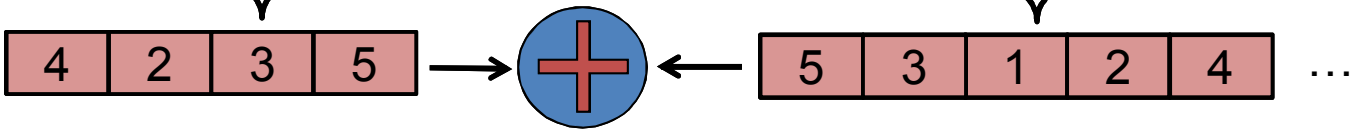
- Domain-independent approach for identifying and evaluating risk
  - Subject Matter Experts (SMEs) address:
    - What specific **potential risks** (attacks) does a system face?
    - How **difficult** is each attack to carry out?
    - What is the **consequence** of success for each attack?

Identify Attacks

Assess Consequence

Assess Difficulty

5	Path	PathID	Governance & Mission Impacts	Psychological	Economic Disruption	Public Safety	Net Consequence	Level of Engaged Outside Attack Participants	Training & Expertise Required for Active Participants	Support Structure Required	Tools Availability	Level of Contributing Insiders Required	Insider Contribution Level of Access Required	Ingenui
6	__Outsourced Implementation_Stea	0	5	3	1	7	7.483314774	5	4	5	2	3	4	
7	__Outsourced Implementation_Stea	1	3	4	4	3	5.099019514	1	3	5	2	2	2	
8	__Outsourced Implementation_Stea	2	2	3	4	3	4.242640687	4	2	4	2	3	5	
9	__Outsourced Implementation_Stea	3	2	4	2	2	3.464101615	2	3	5	3	3	4	
10	__Outsourced Implementation_Stea	4	2	1	1	4	3.16227766	3	4	3	5	3	4	
11	__Outsourced Implementation_Stea	5	2	3	5	1	4.582575695	3	1	2	1	5	3	



(Net Difficulty, Net Consequence)

Risk Informed Management of Enterprise Security

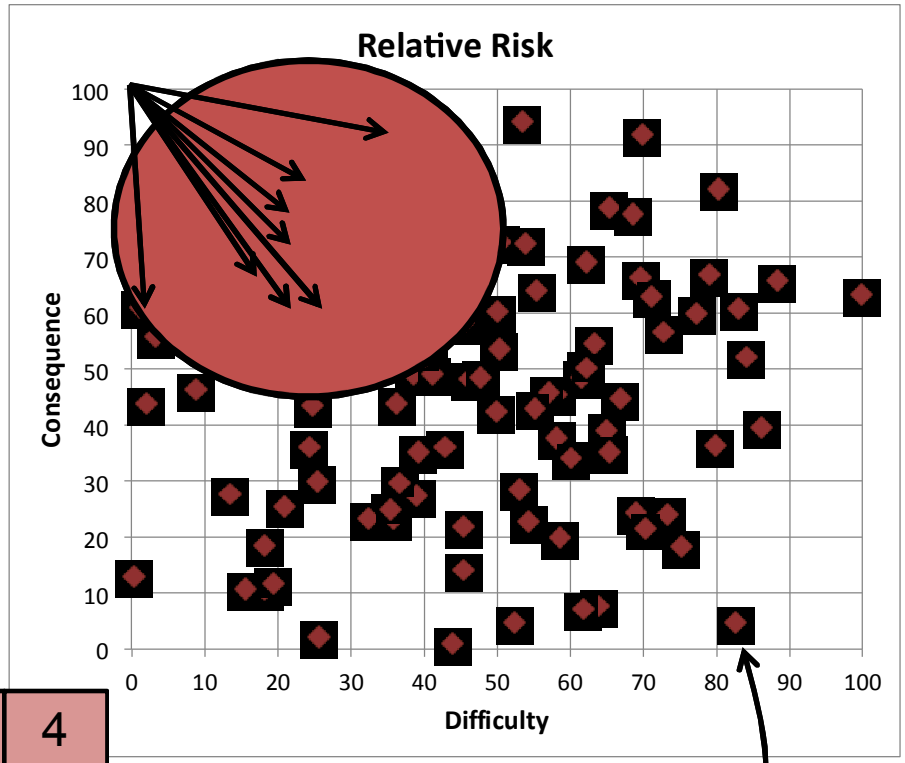
- **RIMES Assessment**
  - Domain-independent means of evaluating risk
  - 13 orthogonal dimensions for difficulty assessment, each assigned 1-5 ranking
  - 4 dimensions for consequence

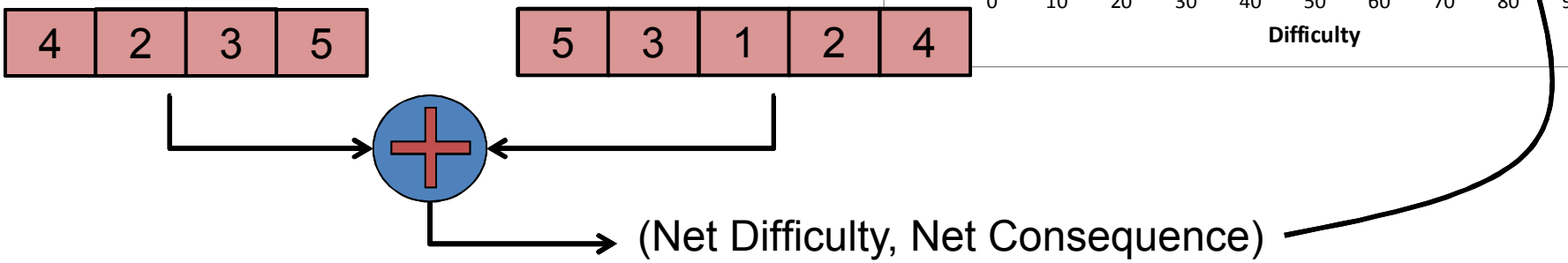
  - Attack Preparation:
    - # engaged outside participants
    - Training & expertise required
    - Support structure required
    - Tools availability
    - # contributing insiders required
    - Insider level of access required
    - Ingenuity required

  - Attack Execution:
    - Situational understanding & exploit requirements
    - Stealth/covertness required
    - Outsiders: dedication required
    - Insiders: engagement & risk
    - Operational composition / risk
    - Flexibility required

Risk Informed Management of Enterprise Security

- **RIMES Rubric**
  - Published tables guide SMEs on assignment of rankings
  - Net difficulty and net consequence of each attack guide prioritization:
    - Low difficulty, High consequence attacks are of most concern
  - Mitigation decisions driven by highest priority risks





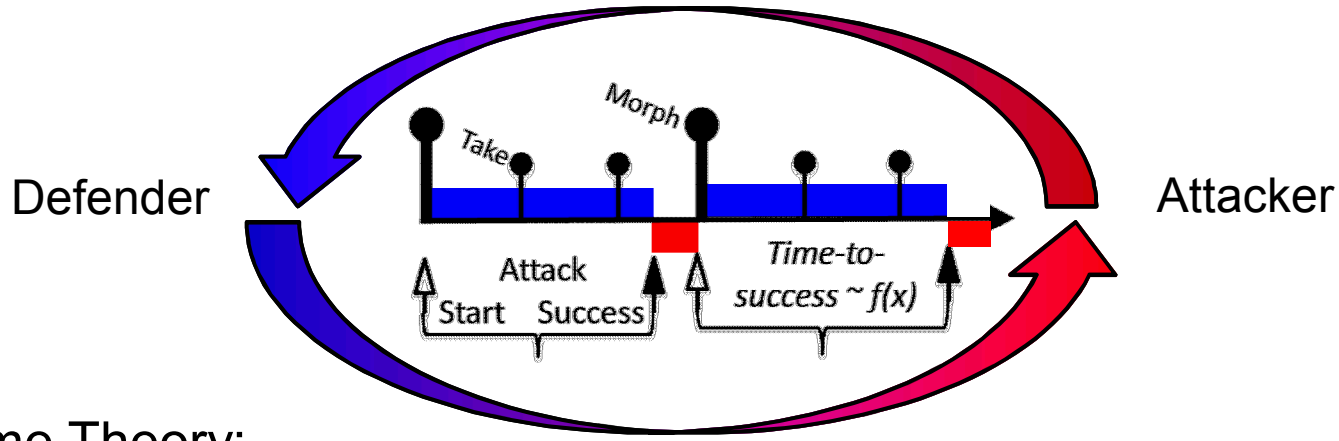
Reflections on RIMES

- RIMES offers structure to the issue of risk assessment
  - Repeatable assessment process
  - Quantified approach for ranking potential issues
- Applications:
  - Evaluation of risks associated with FPGAs in USG systems
  - Physical security of USG installations

- Challenge: Subjectivity
  - Subject matter experts have differing expertise & perspectives
  - Repeated assessments may render different results

- Can we assess system trustworthiness objectively?

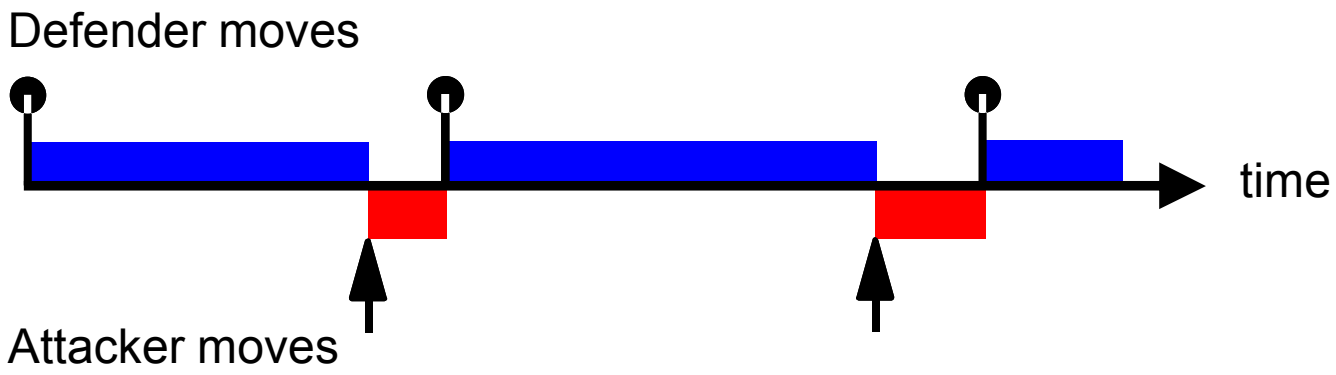
Game Theoretic Analysis: Why?



- Game Theory:
  - "The study of ***mathematical models of conflict and cooperation*** between intelligent, rational decision-makers"[1]
  - Initially developed by von Neumann and Morgenstern in 1944
  - Nobel Prizes awarded for work on game theory: 2014, 2007, 2005, 1996, 1995, 1994, 1972, 1970
- Why Game Theory for Trust?
  - Trust is concerned with the **risk of potential interaction** between **adversaries and system developers** and development processes
  - Game Theory allows explicit representation and **evaluation of dynamic interaction** between attacker and defender

1. Myerson, B. "Game Theory: Analysis of Conflict". Harvard University Press

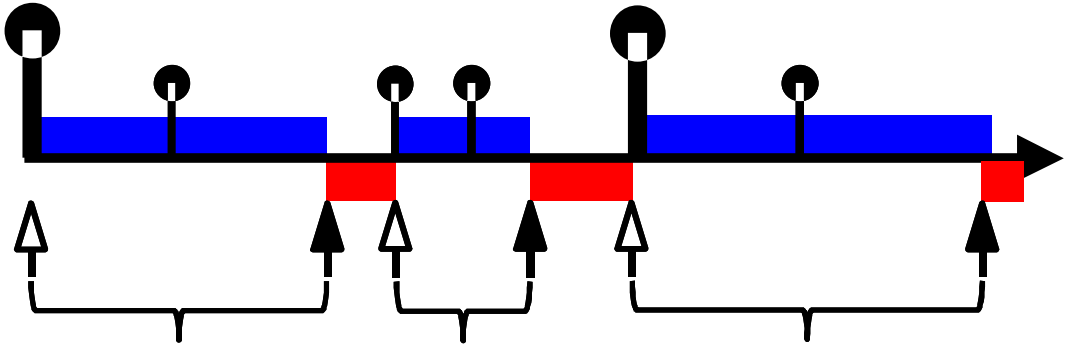
FlipIt: A Game Theoretic Model to Investigate Cyber Defense Effectiveness



FlipIt Constructs

- Two players (defender and attacker)
- A single contested resource
- Player moves seize the resource
- Moves incur a cost
- Strategy consists of move timing
- Single defender move (take)
- Limited player information
- Utility = Control Time - Cost

Probabilistic, Learning Attacker, Dynamic Defender (PLADD) Model



PLADD Model for Analysis

- Represent Attacker-Defender interaction as contention for a single resource
- Defender executes periodic actions
  - Each action wrests control from attacker
- Attacker actions wrest control from defender, after a random period of time
- Attack cost: fixed to initiate + variable cost proportional to time-to-success
- As attacker repeats attacks, they become more efficient.
- Special defender "morph" move resets attacker learning
- Goal: determine defender strategies that drive attacker costs to be prohibitive

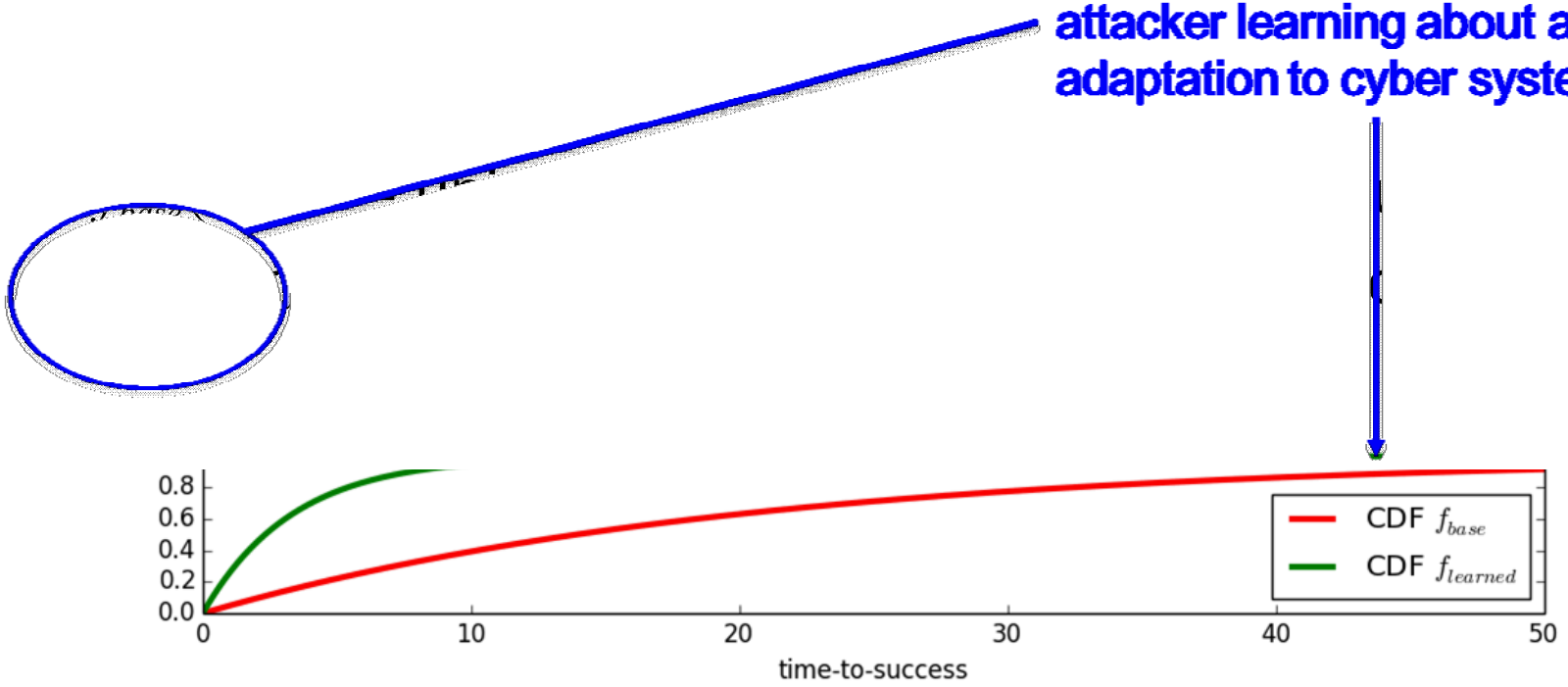
PLADD Parameters

- α = cost to start an attack
- β = cost per unit time to continue an attack
- C_{take} = cost for a take move
- C_{morph} = cost for a morph move
- N = number of take moves per morph move
- $f_{base}(x)$ = uninformed time-to-success distribution
- $f_{learned}(x)$ = informed time-to-success distribution

PLADD Parameters

- α = cost to start an attack
- β = cost per unit time to continue an attack
- C = cost for a take move

Model representation for attacker learning about and adaptation to cyber system

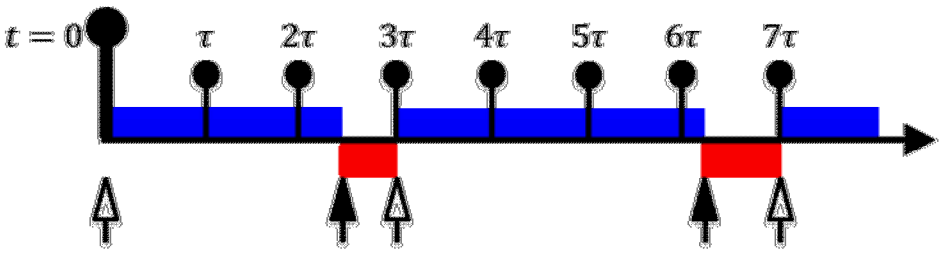


Complex Systems Attributes: Acquisition of knowledge – attacker can learn & adapt.

Mathematical Formulation

- Utility

$$u(x,S)=-\alpha-\beta x+\left(\min_{t_i\in S}\left(t_i:t_i\geq x\right)-x\right)$$

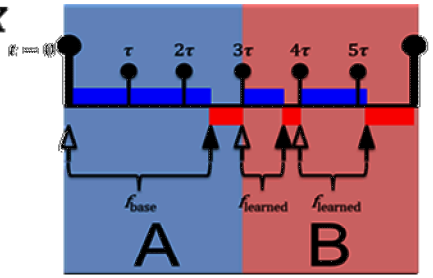


- Infinite time horizon

$$S=\{t_0,t_1,\cdots\}$$

$$E\left[u(X,S)\right]=-\alpha-\beta\int_0^\infty xf(x)dx+\int_0^\infty\left(\min_{t_i\in S}\left(t_i:t_i\geq x\right)-x\right)f(x)dx$$

- Finite time horizon



$$S=\{t_o,t_1,\cdots,t_{N+1}\}$$

$$E_{N+1}\left[u(X,S)\right]=0$$

$$E_j\left[u(X,S)\right]=-\alpha-\beta\left(t_{N+1}-t_j\right)\int_{\tau_{N+1}}^\infty f_{learned}(x-t_j)dx+\sum_{i=j+1}^{N+1}\int_{t_{i-1}}^{t_i}f_{learned}(x-t_j)\left(t_i-x-\beta(x-t_j)+E_i\left[u(X,S)\right]\right)dx$$

$$E\left[u(X,S)\right]=-\alpha-\beta t_{N+1}\int_{t_{N+1}}^\infty f_{base}(x)dx+\sum_{j=1}^{N+1}\int_{t_{j-1}}^{t_j}f_{base}(x)\left(t_j-x-\beta x+E_j\left[u(X,S)\right]\right)dx$$

Key Result

Drive Attacker Out of the Game (no MTD):

Theorem 1. In an infinite game, for any $\alpha, \beta > 0$ and for any continuous f with a valid first moment, there exists a periodic defender strategy $S^* = \{\tau^*, 2\tau^*, \ldots\}$ with take moves of period τ^*, such that:

$$E[u(X, S^*)] = 0$$

Sketch of the proof: We prove this by showing that $E[u(X, S)]$ is a continuous and differentiable function of τ that is negative when τ is small enough and positive when τ is large enough, and by applying the intermediate value theorem.

Corollary 1: In an infinite game, there exists a τ^- and a defender strategy $S^- = \{\tau^-, 2\tau^-, \ldots\}$, such that $E[u(X, S^-)] = 0$ where for any defender strategy $S' = \{\tau', 2\tau', \ldots\}$ with $\tau' < \tau^-$ the attacker utility is always negative:

$$E[u(X, S')] < 0$$

Implications:

- It is always possible to push a rational attacker out of the game.
- Pushing the attacker out need not be cost-beneficial to the defender.

Key Result

Drive Attacker Out of the Game (with MTD):

Theorem 2. In a finite game of duration $T > 0$ (MTD deployed at T) and any $\alpha, \beta > 0$ and for any continuous f_{base} and $f_{learned}$ that have valid first moments, there exists a defender strategy $S^* = \{t_1, t_2, \dots, t_N\}$ with $t_N < T$, such that:

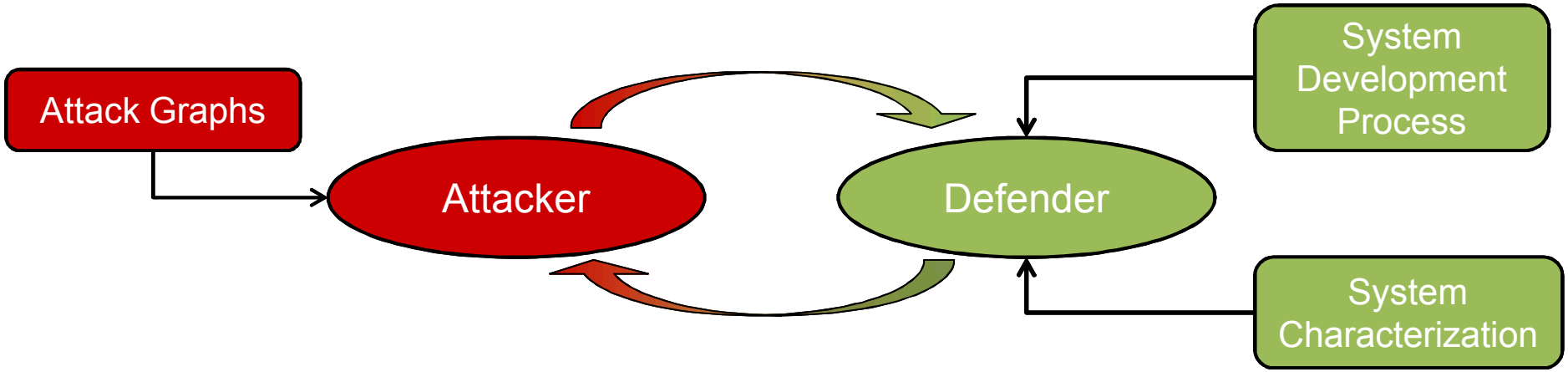
$$E[u(X, S^*)] \leq 0$$

Sketch of the proof: by construction and recursion. Start with the last step of the game and proceed backwards. In short: given the attacker fixed costs, as in Theorem 1 for infinite game, the defender can always play quickly enough to ensure negative attacker utility.

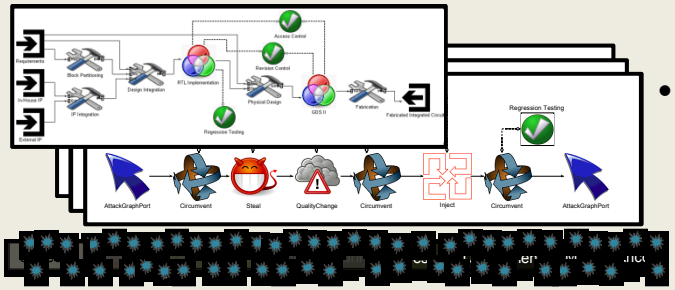
Implication: if T is small enough, then the attacker expected utility may be always negative.

Trust analysis using Game Theory

- Amalgamation of *game theory* with *relative risk assessment* to model full lifecycle trust concerns, and objectively evaluate system trustworthiness
  - Incorporate game theory, risk assessment, resiliency analysis, optimization and supply chain analytics
  - Apply PLADD to trust analysis
- Goal: Empower decision makers to make quantitative, science-based tradeoff decisions about trust



PRESTIGE: PRactical Evaluation and Synthesis of Trust in Government systEms

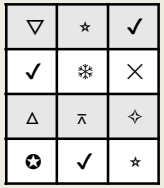


IDENTIFICATION

- Robust modeling tools
  - Characterize development processes
  - Model potential attacks

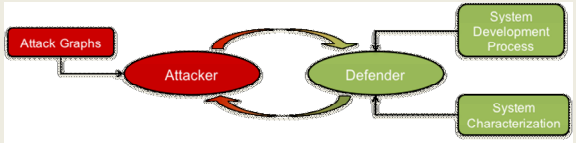
TRADEOFF ANALYSIS

- Constraints driven risk mitigation analysis
- Mathematical characterization of mitigation impact



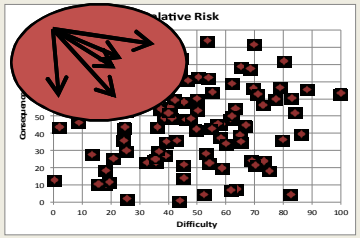
EXEMPLAR

SPACE SYSTEMS



ANALYSIS

- Game theory-based risk analysis
- Tractable attack evaluation



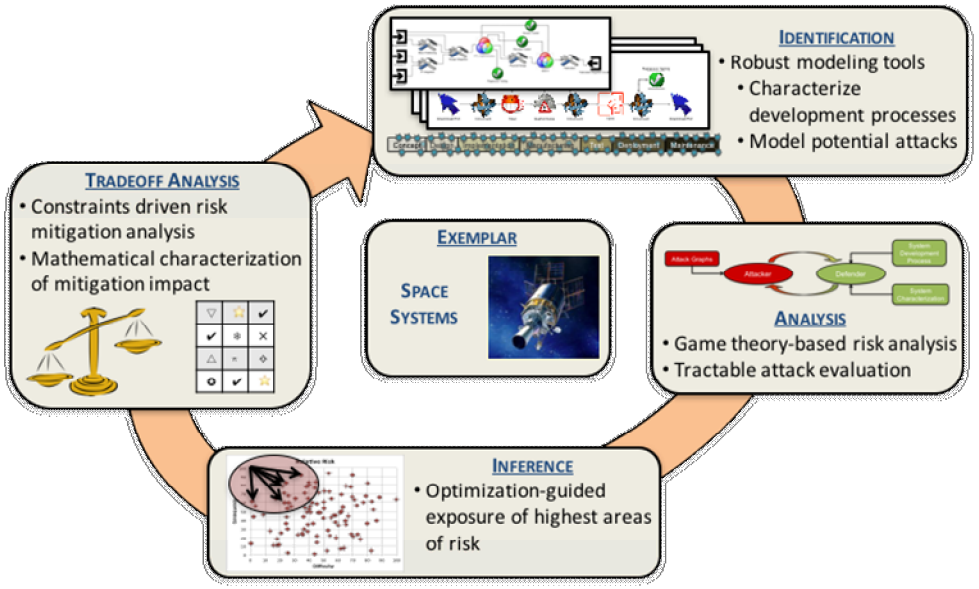
INFERENCE

- Optimization-guided exposure of highest areas of risk

Utilize mathematics as a framework for performing quantitative trust analysis, empowering engineering of trustworthy systems

Trust Analysis for Microelectronics Based Systems

- **Assertion:** Trust in microelectronics-based systems is a spectrum of levels of confidence to be quantified and verified
  - Trust is not an ethereal, "squishy" unquantifiable concept
  - Trust can be quantified, even though we deal with an unknown, motivated adversary
- **Vision:** Enable cohesive, full-lifecycle trust engineering of microelectronics based systems
  - Develop trusted and trustworthy systems from untrusted components and tools
  - Empower developers to make engineering tradeoff decisions



- **Innovative Approach:** Mathematics-driven system trust evaluation using game theory
  - Develop **game theoretic modeling** approach for characterizing attack, defense effectiveness
  - Architect **modeling and analysis tools** for conducting trust evaluation
  - **Exemplar-based validation** targeting real systems