COORDINATED SCIENCE LABORATORY

ENGINEERING AT ILLINOIS

Making Sound Design Decisions Using Quantitative Security Metrics

Bill Sanders

ITI.ILLINOIS.EDU

INFORMATION TRUST INSTITUTE

INFORMATION TRUST INSTITUTE

The Problem: Assessing Security and Resilience

- Systems operate in adversarial environments
  - Adversaries seek to degrade system operation by affecting the confidentiality, integrity, and/or availability of the system information and services
  - "Resilient" systems aim to meet their ongoing operational objectives despite attack attempts by adversaries
- System security is not absolute
  - No real system is perfectly secure
  - Some systems are more secure than others
  - *But which ones are more secure?*
  - *And how much more secure are they*?

INFORMATIONTRUST INSTITUTE

Practical Applications of Security Metrics

Organizational-level Metrics

Questions the CIO cannot answer:

- How much risk am I carrying?
- Am I better off now than I was this time last year?
- Am I spending the right amount of money on the right things?
- How do I compare to my peers?
- What risk transfer options do I have?

(From CRA, Four Grand Challenges in Trustworthy Computing, 2003)

Technical Metrics

Questions the design engineer cannot answer:

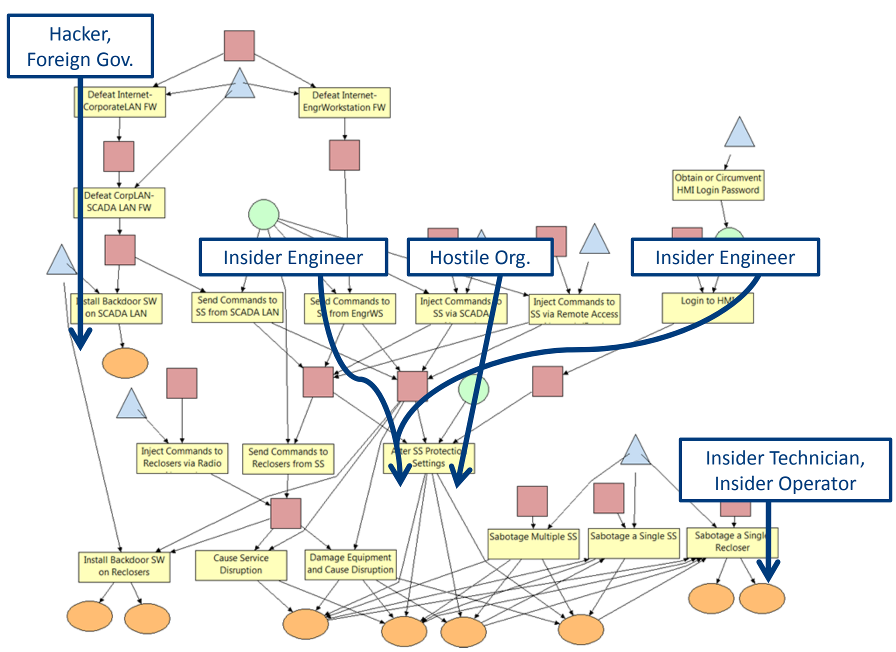
- Is design A or B more secure (confidentially, integrity, availability, privacy)?
- Have I made the appropriate design trade off between timeliness, security, and cost?
- How will the system, as implemented, respond to a specific attack scenario?
- What is the most critical part of the system to test, from a security point of view?

A Question neither can answer:

- How do the technical metrics impact the organizational-level security metrics?

3

INFORMATIONTRUST INSTITUTE



Related Work Motivating ADVISE

- Model-based security analysis
  - Attack Trees
  - Attack Graphs and Privilege Graphs
- Adversary-based security analysis
  - MORDA (Mission-Oriented Risk and Design Analysis)
  - NRAT (Network Risk Assessment Tool)

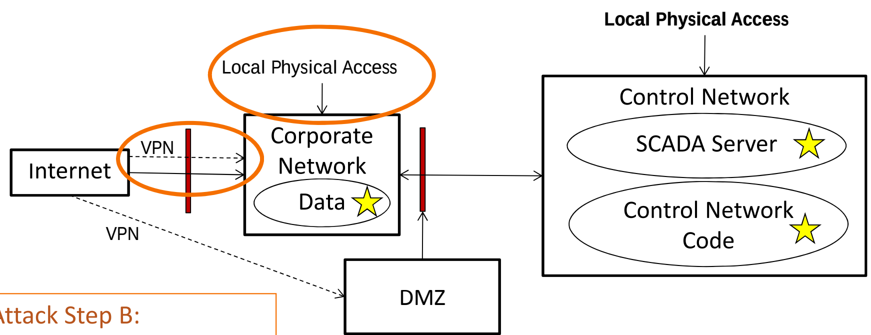
ADVISE integrates the benefits of both
model-based and adversary-based security analysis

ADversary VIew Security Evaluation (ADVISE) approach

- Adversary-driven analysis
  - Considers characteristics and capabilities of adversaries
- State-based analysis
  - Considers multi-step attacks
- Quantitative metrics
  - Enables trade-off comparisons among alternatives
- Mission-relevant metrics
  - Measures the aspects of security important to owners/operators of the system

Example: SCADA System Attack

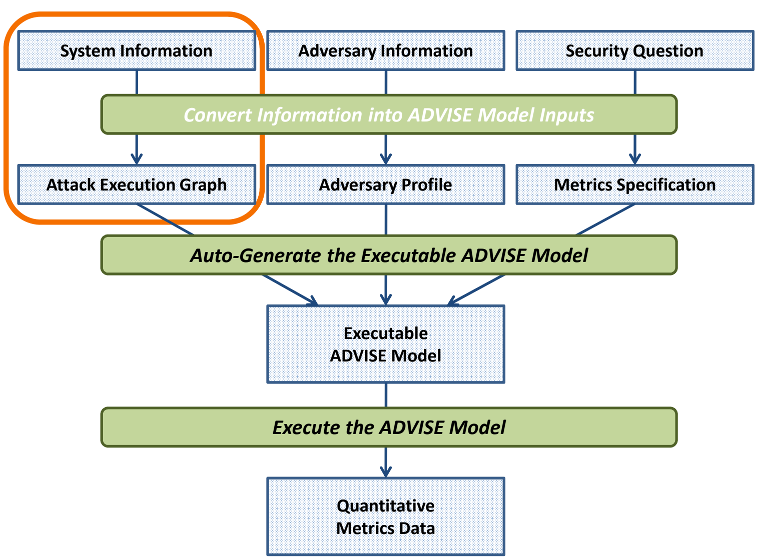
Attack Step A:
Gain Corporate Network Access
Through Local Physical Access



Attack Step B:
Gain Corporate Network
Access Through VPN

☆ = Attack Target

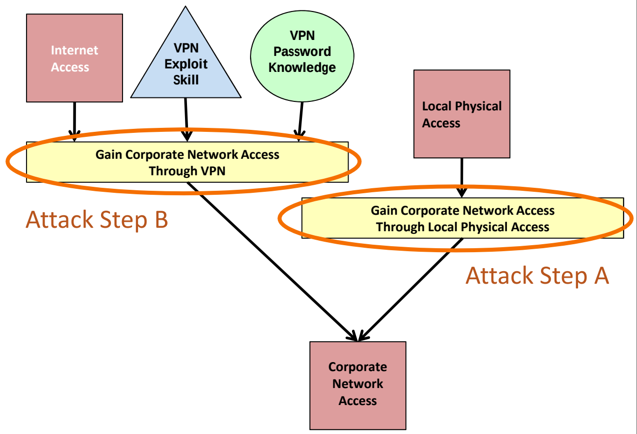
ADVISE Method Overview



INFORMATION TRUST INSTITUTE

Representing Attacks Against the System

An "attack execution graph" describes potential attack vectors against the system from an attacker point of view. Attempting an attack step requires certain skills, access, and knowledge about the system. The outcome of an attack can affect the adversary's access and knowledge about the system.



INFORMATION TRUST INSTITUTE

ADVISE System Information: Attack Execution Graph

An attack execution graph is defined by

<A, R, K, S, G>,

where

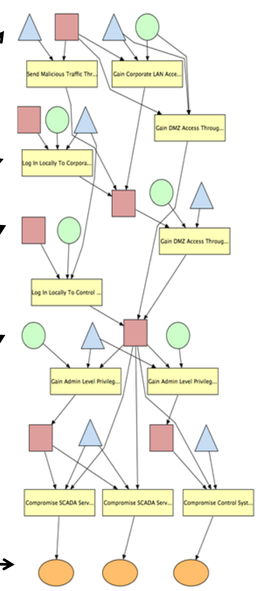
A is the set of attack steps,
e.g., "Access the network using the VPN,"

R is the set of access domains,
e.g., "Internet access," "Network access,"

K is the set of knowledge items,
e.g., "VPN username and password"

S is the set of adversary attack skills,
e.g., "VPN exploit skill," and

G is the set of adversary attack goals,
e.g., "View contents of network."



Attack Step Definition

An attack step ai is a tuple:
ai = <Bi, Ti, Ci, Oi, Pri, Di, Ei>

Note: X is the set of all states in the model.

B_i: *X* → *{True, False}* is a Boolean precondition,
e.g., (Internet Access) AND ((VPN account info) OR (VPN exploit skill)).

T_i: *X x* R^+ → *[0, 1]* is the distribution of the time to attempt the attack step,
e.g., normally distributed with mean 5 hours and variance 1 hour.

C_i: *X* → $R^{\geq 0}$ is the cost of attempting the attack step, e.g., $1000.

O_i is a finite set of outcomes, e.g., {Success, Failure}.

Pr_i: *X x Oi* → *[0, 1]* is the probability of outcome o ϵ Oi occurring,
e.g., if (VPN exploit skill > 0.8) {0.9, 0.1} else {0.5, 0.5}.

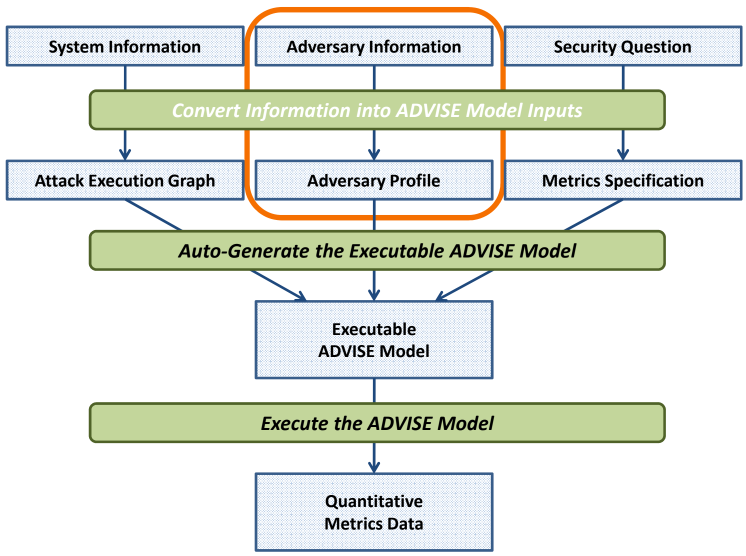
D_i: *X x Oi* → *[0, 1]* is the probability of the attack being detected when outcome o ϵ Oi occurs, e.g., {0.01, 0.2}.

E_i: *X x Oi* → *X* is the next-state that results when outcome o ϵ Oi occurs,
e.g., {gain Network Access, no effect}.

The "Do-Nothing" Attack Step

- Contained in every attack execution graph
- Represents the option of an adversary to refrain from attempting any active attack
  - The precondition $B_{DoNothing}$ is always true.
- For most attack execution graphs,
  - the cost $C_{DoNothing}$ is zero,
  - the detection probability $D_{DoNothing}$ is zero, and
  - the next-state is the same as the current state.
- The existence of the "do-nothing" attack step means that, regardless of the model state, there is always at least one attack step in the attack execution graph whose precondition is satisfied

ADVISE Method Overview



ADVISE Adversary Information: Adversary Profile

The adversary profile is defined by the tuple

<s0, L, V, wC, wP, wD, UC, UP, UD, N>,

where

$s_0 \in X$ is the initial model state, e.g., has Internet Access & VPN password,

L is the attack skill level function, e.g. has VPN exploit skill level = 0.3,

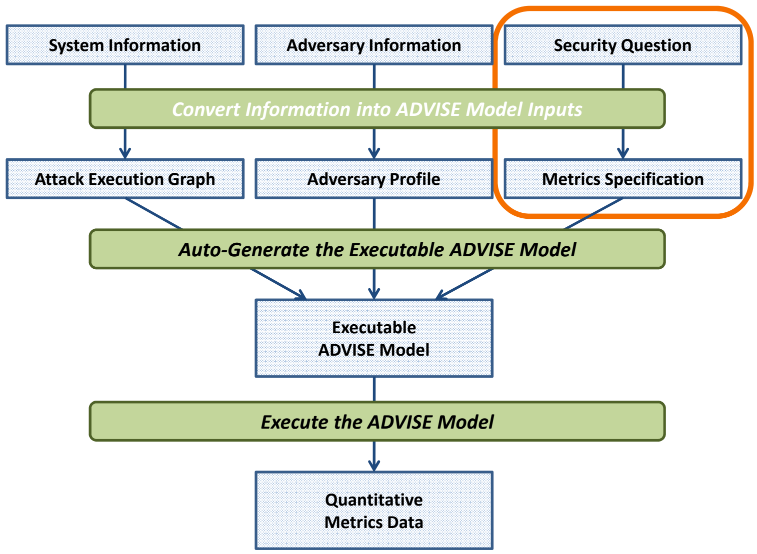
V is the attack goal value function, e.g., values "View contents of network" at $5000,

w_C, w_P, and w_D are the attack preference weights for cost, payoff, and detection probability, e.g., $w_C = 0.7$, $w_P = 0.2$, and $w_D = 0.1$,

U_C, U_P, and U_D are the utility functions for cost, payoff, and detection probability, e.g., $U_C(c)=1 - c/10000$, $U_P(p)=p/10000$, $U_D(d)=1 - d$, and

N is the planning horizon, e.g., N = 4.

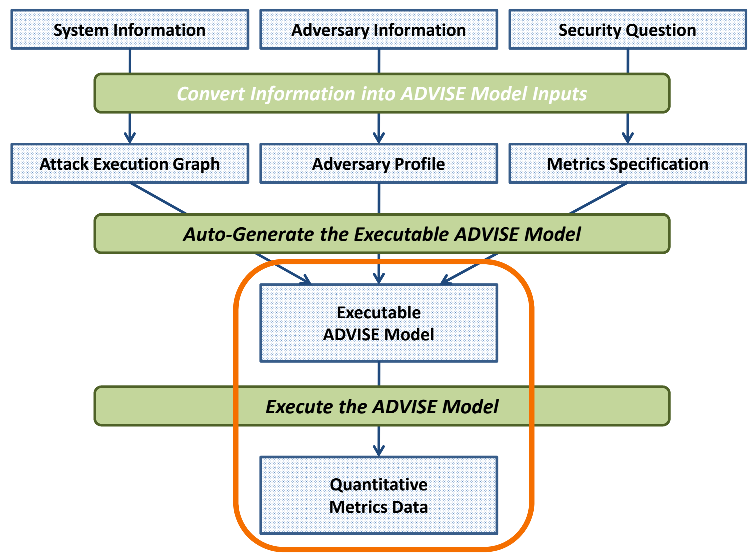
ADVISE Method Overview



ADVISE Security Question: Metrics Specification

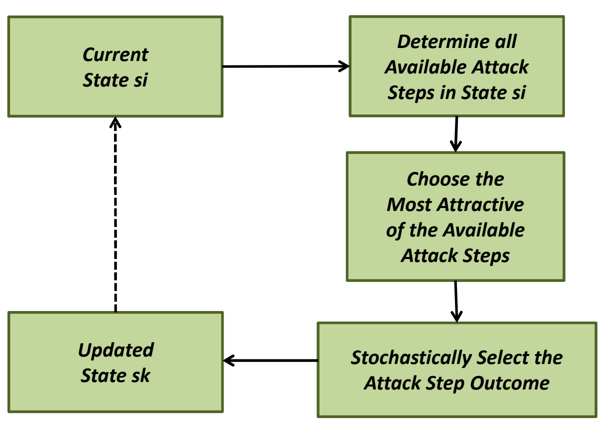
- State metrics analyze the model state
  - State occupancy probability metric (probability that the model is in a certain state at a certain time)
  - Average time metric (average amount of time during the time interval spent in a certain model state)
- Event metrics analyze events (state changes, attack step attempts, and attack step outcomes)
  - Frequency metric (average number of occurrences of an event during the time interval)
  - Probability of occurrence metric (probability that the event occurs at least once during the time interval)

ADVISE Method Overview



Model Execution: the Attack Decision Cycle

- The adversary selects the most attractive available attack step based on his attack preferences.
- State transitions are determined by the outcome of the attack step chosen by the adversary.



ADVISE Model Execution Algorithm

1: Time ← 0
2: State ← s_0 — Simulation time and model state initialization
3: **while** Time < EndTime **do**
4: $Attack_i \leftarrow \beta^N(State)$ — Adversary attack decision
5: Outcome ← o, where o ~ $Prob_i(State)$ — Stochastic outcome
6: Time ← Time + t, where t ~ $T_i(State)$ — Time update
7: State ← $E_i(State, Outcome)$ — State update
8: **end while**

$\beta^N(s)$ selects the most attractive available attack step in model state *s* using a planning horizon of *N*

Goal-driven Adversary Decision Function

When the planning horizon N is greater than 1, the
attractiveness of an available next step
is a function of
the payoff in the expected states
N attack steps from the current state
(the **expected horizon payoff**)
and
the expected cost and detection probability
of those N attack steps
(the **expected path cost** and **expected path detection**).

Goal-driven Adversary Decision Function

Attractiveness of an attack step ai
to an adversary with planning horizon N =
UC(E[C]) * wc + UP(E[P]) * wp + UD(E[D]) * wd

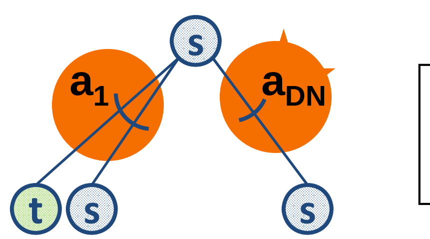
E[C] = Expected Path Cost to get to a state N attack steps away via attack step a_i.

E[P] = Expected Horizon Payoff in a state N attack steps away via attack step a_i.

E[D] = Expected Path Detection to get to a state N attack steps away via attack step a_i.

E[C], E[P], and E[D] are computed using a State Look-Ahead Tree.

Consider an adversary attack decision in state s with N = 1



C_1 = \$1000
$Pr_1(s,1) = 0.9$
$Pr_1(s,2) = 0.1$
$D_1(s,1) = 0.01$
$D_1(s,2) = 0.1$
Payoff(t) = \$0
Payoff(s) = \$0

$Attr(a_1) = 0.28$

C_{DN} = \$0
$Pr_{DN}(s,1) = 1$
$D_{DN}(s,1) = 0$
Payoff(s) = \$0

$Attr(a_{DN}) = 0.3$

Attractiveness of attack step a_i = $U_C(\text{cost of } a_i) * w_c + U_P(E[\text{payoff of } a_i]) * w_p + U_D(E[\text{detection of } a_i]) * w_d$

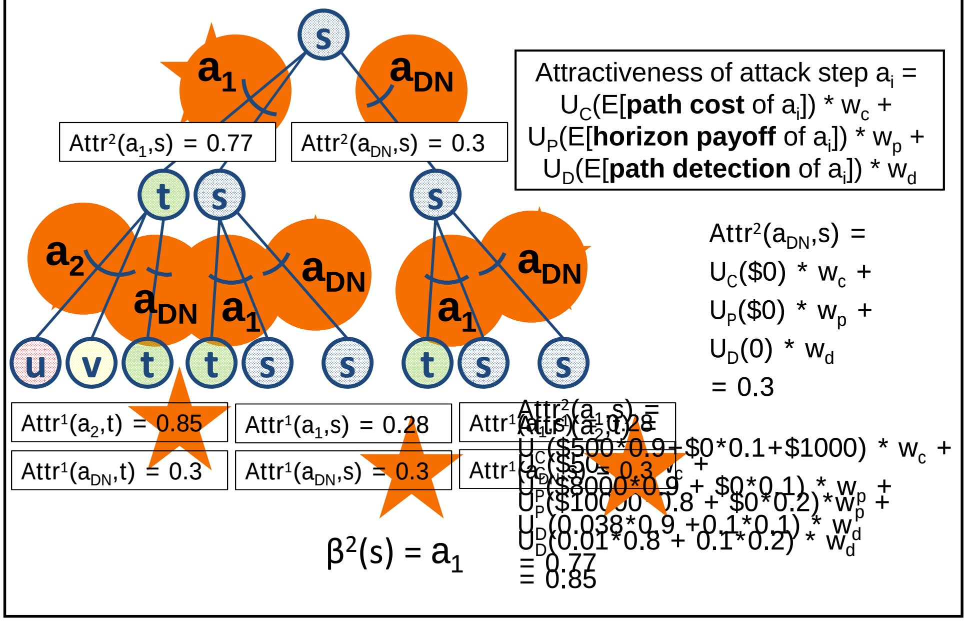
$Attr(a_{DN}) = U_C(\$0) * w_c + U_P(\$0*1) * w_p + U_D(0*1) * w_d = 0.3$

$Attr(a_1) = U_C(\$1000) * w_c + U_P(\$0*0.9 + \$0*0.1) * w_p + U_D(0.01*0.9 + 0.1*0.1) * w_d = 0.28$

$\beta^1(s) = a_{DN}$

INFORMATION TRUST INSTITUTE

Consider an adversary attack decision in state s with N = 2



INFORMATION TRUST INSTITUTE

Optimality of the Original ADVISE Decision Rule

- ***Bellman's Principle of Optimality***

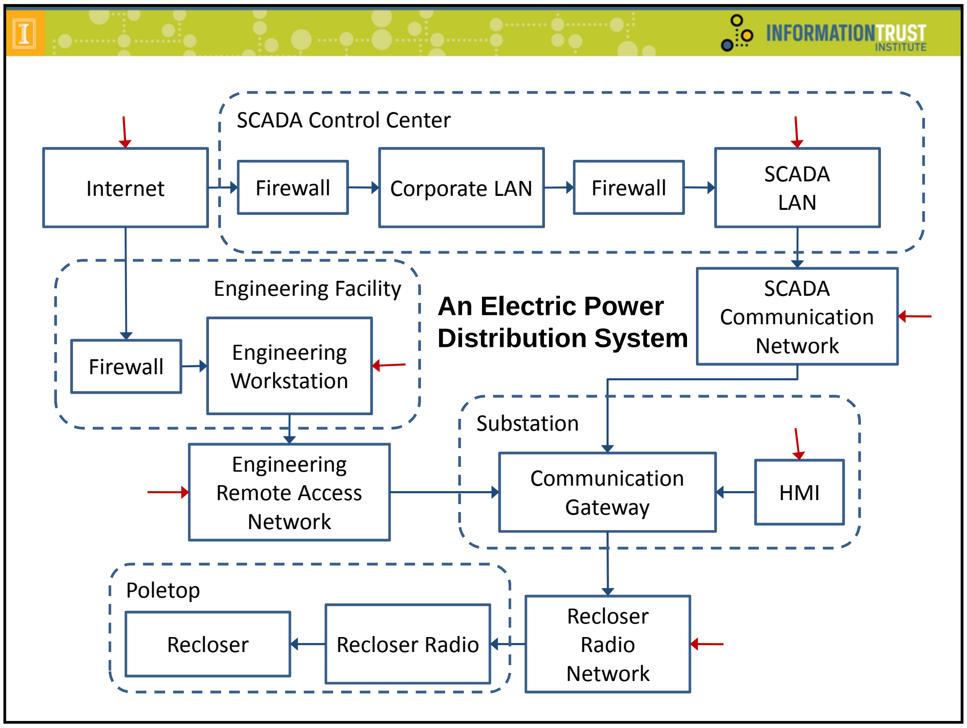
  "an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision"
- The original ADVISE decision rule implements a **provably optimal policy** when the attractiveness function is
  - wholly linear (cost and payoff only) **OR**
  - wholly multiplicative (detection only).
- The original ADVISE decision rule does ***not*** always produce an optimal decision when the decision rule combines
  - additive rewards (cost and/or payoff) **AND**
  - multiplicative rewards (detection).

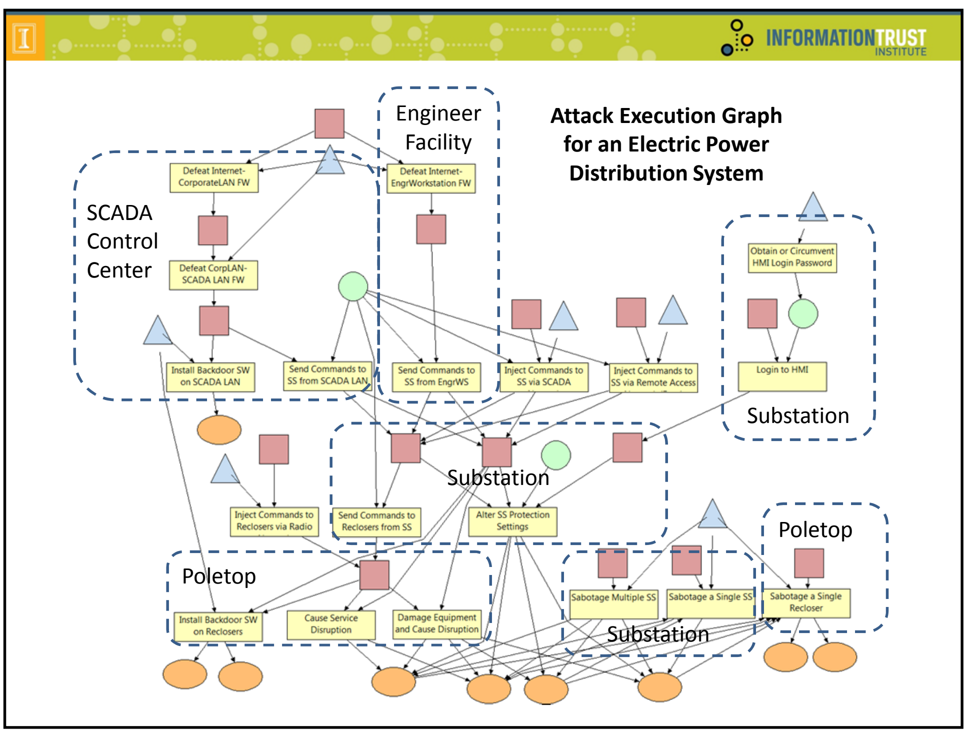
Practical Implications of Algorithm Optimality

- Adversaries modeled using this algorithm exhibit "worst case" behavior, that is, they always select a next attack step that is best for them considering
  - Adversary attack preferences
  - Adversary planning horizon
  - Available attack steps
  - Attractiveness function definition

Case Study

- Investigates the effects of architectural changes on the security of an electric power distribution system
- In particular, analyze the security impact of adding radio communication between substations and poletop reclosers


INFORMATION TRUST INSTITUTE
SCADA Control Center
Internet
Firewall
Corporate LAN
Firewall
SCADA LAN
Engineering Facility
Firewall
Engineering Workstation
An Electric Power Distribution System
SCADA Communication Network
Substation
Engineering Remote Access Network
Communication Gateway
HMI
Poletop
Recloser
Recloser Radio
Recloser Radio Network


INFORMATION TRUST INSTITUTE
Engineer Facility
Attack Execution Graph for an Electric Power Distribution System
SCADA Control Center
Defeat Internet-CorporateLAN FW
Defeat Internet-EngrWorkstation FW
Defeat CorpLAN-SCADA LAN FW
Obtain or Circumvent HMI Login Password
Install Backdoor SW on SCADA LAN
Send Commands to SS from SCADA LAN
Send Commands to SS from EngrWS
Inject Commands to SS via SCADA
Inject Commands to SS via Remote Access
Login to HMI
Substation
Substation
Inject Commands to Reclosers via Radio
Send Commands to Reclosers from SS
Alter SS Protection Settings
Poletop
Poletop
Install Backdoor SW on Reclosers
Cause Service Disruption
Damage Equipment and Cause Disruption
Sabotage Multiple SS
Sabotage a Single SS
Sabotage a Single Recloser
Substation

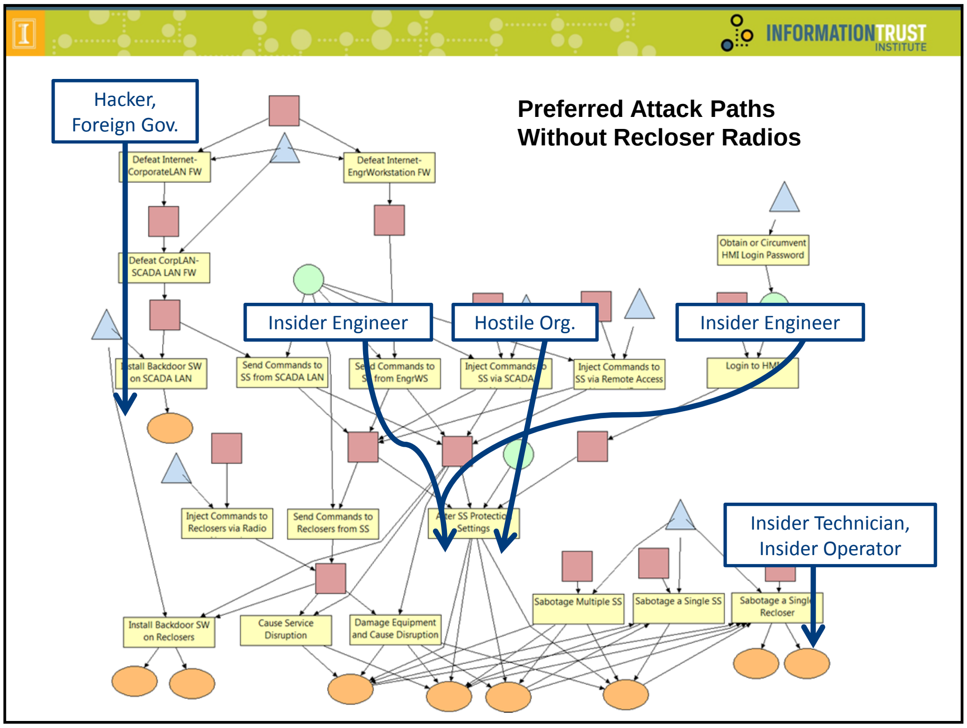
Adversary Profiles: Decision Parameters

	Foreign Government	Hacker	Hostile Organization	Insider Engineer	Insider SCADA Operator	Insider Remote Technician
Cost Preference Weight	0	0.2	0.05	0.2	0.2	0.2
Detection Preference Weight	0.5	0.4	0.2	0.1	0.1	0.1
Payoff Preference Weight	0.5	0.4	0.75	0.7	0.7	0.7

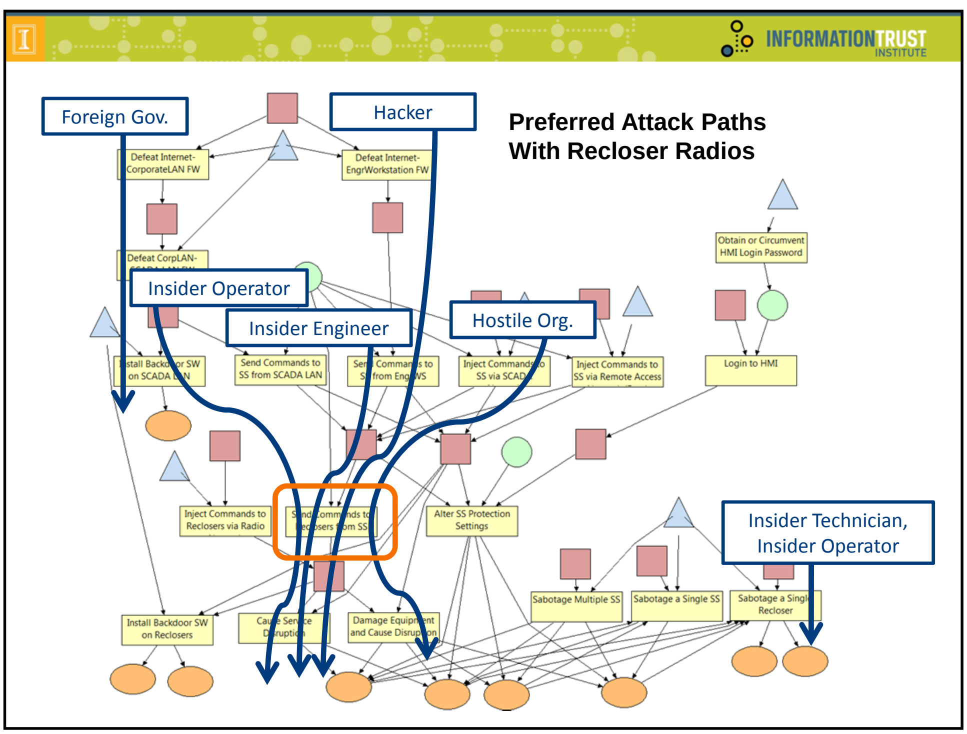
- The Foreign Government adversary is very well-funded but risk-averse.
- The Hacker is resourced-constrained.
- The Hostile Organization is moderately well-funded and more driven by payoff than the others.
- The Insider Engineer, Insider Technician, and Insider Operator are resource-constrained but willing to take risks.

Security Metrics

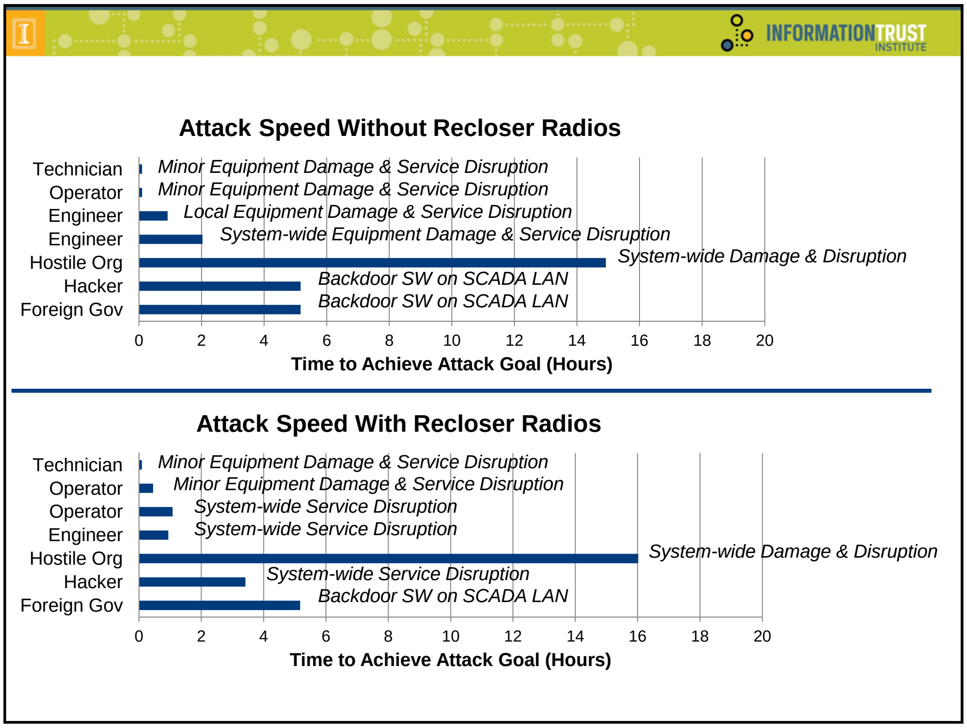
- Average Number of Attempts
  - Report for each attack step
  - Gives insight on preferred attack path of adversary
- Probability of Attack Goal Achieved at End Time
  - Report for each attack goal
  - Gives insight on what goals the adversary is actively pursuing and reaching
- Average Time-To-Achieve-Goal
  - For attack goals where the above probability metric is 1 (or close to 1)
  - Gives insight on the speed of the adversary's attack

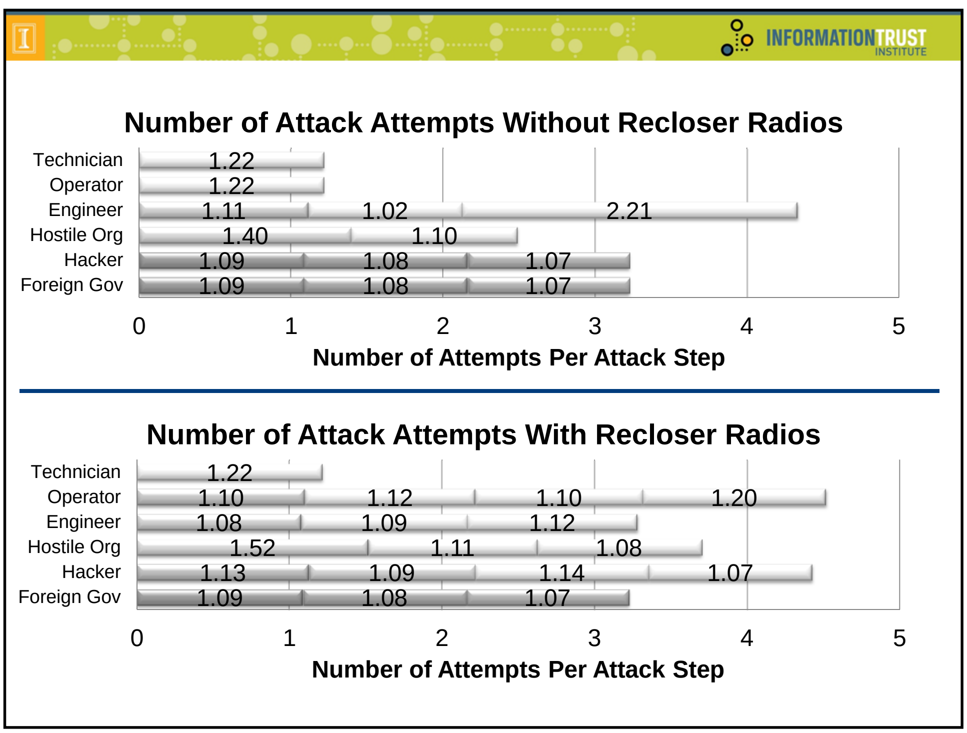


INFORMATIONTRUST INSTITUTE
Preferred Attack Paths
Without Recloser Radios
Hacker, Foreign Gov.
Insider Engineer
Hostile Org.
Insider Engineer
Insider Technician, Insider Operator
Defeat Internet-CorporateLAN FW
Defeat Internet-EngrWorkstation FW
Defeat CorpLAN-SCADA LAN FW
Obtain or Circumvent HMI Login Password
Install Backdoor SW on SCADA LAN
Send Commands to SS from SCADA LAN
Inject Commands to SS via Remote Access
Login to HMI
Inject Commands to Reclosers via Radio
Send Commands to Reclosers from SS
Sabotage Multiple SS
Sabotage a Single SS
Install Backdoor SW on Reclosers
Cause Service Disruption
Damage Equipment and Cause Disruption

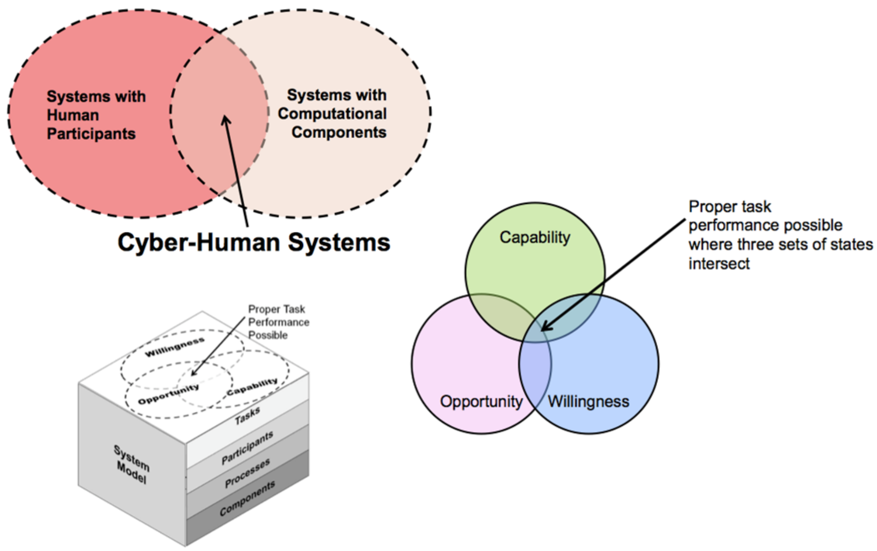


INFORMATIONTRUST INSTITUTE
Preferred Attack Paths
With Recloser Radios
Foreign Gov.
Hacker
Insider Operator
Insider Engineer
Hostile Org.
Insider Technician, Insider Operator
Defeat Internet-CorporateLAN FW
Defeat Internet-EngrWorkstation FW
Obtain or Circumvent HMI Login Password
Send Commands to SS from SCADA LAN
Inject Commands to SS via Remote Access
Login to HMI
Inject Commands to Reclosers via Radio
Alter SS Protection Settings
Sabotage Multiple SS
Sabotage a Single SS
Sabotage a Single Recloser
Install Backdoor SW on Reclosers


INFORMATION TRUST INSTITUTE
Attack Speed Without Recloser Radios
Technician
Operator
Engineer
Engineer
Hostile Org
Hacker
Foreign Gov
Minor Equipment Damage & Service Disruption
Minor Equipment Damage & Service Disruption
Local Equipment Damage & Service Disruption
System-wide Equipment Damage & Service Disruption
System-wide Damage & Disruption
Backdoor SW on SCADA LAN
Backdoor SW on SCADA LAN
0 2 4 6 8 10 12 14 16 18 20
Time to Achieve Attack Goal (Hours)
Attack Speed With Recloser Radios
Technician
Operator
Operator
Engineer
Hostile Org
Hacker
Foreign Gov
Minor Equipment Damage & Service Disruption
Minor Equipment Damage & Service Disruption
System-wide Service Disruption
System-wide Service Disruption
System-wide Damage & Disruption
System-wide Service Disruption
Backdoor SW on SCADA LAN
0 2 4 6 8 10 12 14 16 18 20
Time to Achieve Attack Goal (Hours)


INFORMATION TRUST INSTITUTE
Number of Attack Attempts Without Recloser Radios
Technician 1.22
Operator 1.22
Engineer 1.11 1.02 2.21
Hostile Org 1.40 1.10
Hacker 1.09 1.08 1.07
Foreign Gov 1.09 1.08 1.07
0 1 2 3 4 5
Number of Attempts Per Attack Step
Number of Attack Attempts With Recloser Radios
Technician 1.22
Operator 1.10 1.12 1.10 1.20
Engineer 1.08 1.09 1.12
Hostile Org 1.52 1.11 1.08
Hacker 1.13 1.09 1.14 1.07
Foreign Gov 1.09 1.08 1.07
0 1 2 3 4 5
Number of Attempts Per Attack Step

Next Challenge: Modeling Cyber-Human Systems



ADVISE Team

University of Illinois Urbana-Champaign

Mike Ford

Ken Keefe

Elizabeth LeMay

Bill Sanders

Cyber Defense Agency, Inc.

Carol Muehrcke

Case study collaborators

Bruce Barnett and Michael Dell'Anno,
GE Research

Research sponsored by Science and Technology Directorate, Department of Homeland Security, GE Research, NSA Science of Security Center

Conclusions

- Since system security cannot be absolute, quantifiable security metrics are needed
- Metrics are useful event if not perfect; e.g., relative metrics can aid in critical design decisions
- The ADVISE formalism, and its implementation in Mobius-SE
  - Is rich enough to adversary, user, and system behavior
  - Natural for security analysts
  - Semantically precise
- Mobius-SE is in alpha-test, and has been distributed to 10 organizations (industry, govt., & academics) who are using it in real case studies
- Work is on going on modeling human user behavior

Thank you!

Bill Sanders
perform.csl.illinois.edu
whs@illinois.edu

38