Modeling Trust in Critical Systems with Möbius

KEN KEEFE
SENIOR SOFTWARE ENGINEER
LEAD MOBIUS DEVELOPER

Session Outline

- Security Modeling Introduction
- ADversary VIew Security Evaluation (ADVISE) Models
  - Attack Execution Graph
  - Adversary Profile
  - Model Execution

Security Metrics Motivation

- Security metrics were an important problem in the 2005 INFOSEC Research Council Hard Problems List
- New security metrics that are linked to the business were ranked first among six key security imperatives developed by over twenty Fortune 500 firms
- New regulatory requirements of Sarbanes-Oxley and the Basel II Accord have created more urgency for metrics that integrate security risk with overall business risk
- Almost every critical infrastructure roadmap lists security metrics as a critical challenge

Security Metrics Truths

- Security is no longer absolute (if it ever was?)
- Trustworthy computer systems/networks must operated through attacks, providing proper service in spite of possible partially successful attacks
- If security is not absolute, quantification of the "amount" of security that a particular approach provides is essential
- Quantification can be useful in:
  - A *relative* sense, to choose amount alternate design alternatives
  - In an *absolute* sense, to provide guarantees to users

Contrasting Approaches

Typical Situation Today

- Process:
  - Rely on a trusted analyst (wizard?) that examines situation, and gives advice based on experience, or
  - Form decision in a collective manner based on informal discussions among stakeholder experts
- Limitations:
  - No way to audit decision process
  - No quantifiable ranking of alternative options

Goals for the Future

- Usable tool set that enables diverse stakeholders to express
  - Multi-faceted aspects of model
  - Multiple objectives
- Way for diverse stake-holders to express concerns and objectives in common terminology
- Quantifiable ranking of alternate security policies and architectures
- Auditable decision process

Objective

Quantitative

mission-relevant

auditable

practical

cyber security risk metrics

Model-based metrics have the potential to do this.

Quantitative Security Metrics

- What does “quantitative” mean?
- There are four main types of numerical scales
  - Nominal scale (numbers as labels) [ex: a phone number]
  - Ordinal scale (sequence or rank ordering) [ex: 4th in line]
  - Interval scale (differences between values can be compared) [ex: Celsius or Fahrenheit temperature)
  - Ratio scale (an interval scale with a fixed zero point that permits ratios) [ex: distance or weight]
  - Interval and ratio scales measure *quantitative* differences.
  - Nominal and ordinal scales measure *qualitative* differences.
  - Numerical does not automatically imply quantitative.
- Consider valid operations on different types of numbers
  - Not all mathematical operations are valid on all types of numerical data
  - For example, computing the “average” of a set of phone numbers probably doesn’t make sense

What to Measure

- System's ability to resist attack.
- System's ability to detect attacks.
- System's ability to deliver service in the presence of attacks
- System's ability to recover from a attack (either restoration of service or a graceful degrade performance).

ADVISE Attack Execution Graph

An attack execution graph is defined by

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

where

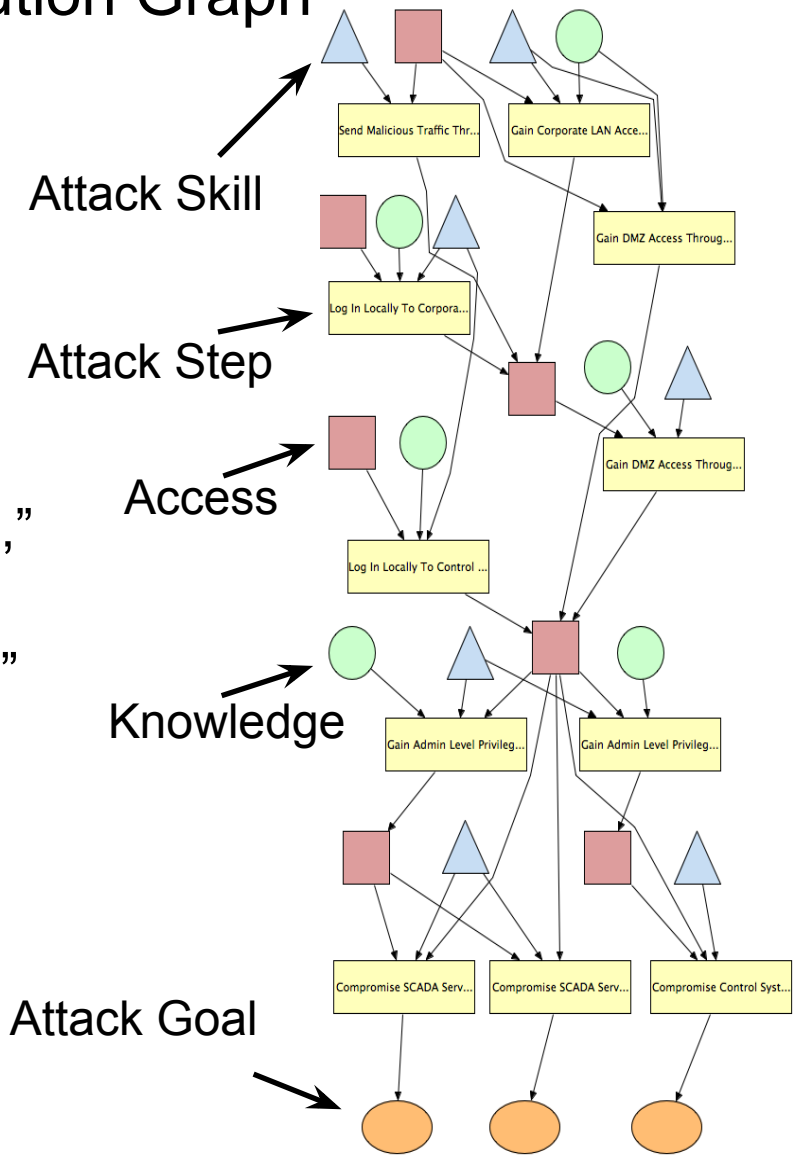
A is the set of attack steps,
e.g., “Access the network using the VPN,”

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 a goals,
e.g., “View contents of network.”



Attack Step Definition

An attack step a_i is a tuple:

$$a_i = <B_i, T_i, C_i, O_i, Pr_i, D_i, E_i>$$

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

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

T_i: $X \times R^+ \rightarrow [0, 1]$ is the time to attempt the attack step,
e.g., 5 hours.

C_i: $X \rightarrow R^{\geq 0}$ 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: $X \times Oi \rightarrow [0, 1]$ is the probability of outcome $o \in O_i$ occurring,
e.g., if (VPN exploit skill > 0.8) {0.9, 0.1} else {0.5, 0.5}.

D_i: $X \times Oi \rightarrow [0, 1]$ is the probability of the attack being detected when outcome $o \in O_i$ occurs, e.g., {0.01, 0.2}.

E_i: $X \times Oi \rightarrow X$ is the next-state that results when outcome $o \in O_i$ 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 Adversary Profile

The adversary profile is defined by the tuple

$<s_0, L, V, w_C, w_P, w_D, U_C, U_P, U_D, N>$,

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 Model State

The model state, $s \in X$, reflects the progress of the adversary in attacking the system and is defined by the tuple

$$s = \langle R_s, K_s, G_s \rangle$$

where

$R_s \in R$ is the set of access domains that the adversary can access,

$K_s \in K$ is the set of knowledge items that the adversary possesses, and

$G_s \in G$ is the set of attack goals the adversary has achieved.

ADVISE 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 Model Execution Algorithm

1: Time ← 0
2: State ← s_0 — **Simulation time and model state initialization**
3: **while** Time < EndTime **do**
4: Attack$_i$ ← β^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
of those N attack steps

(the **expected path cost** and **expected path detection**).

Goal-Driven Adversary Decision Function

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.

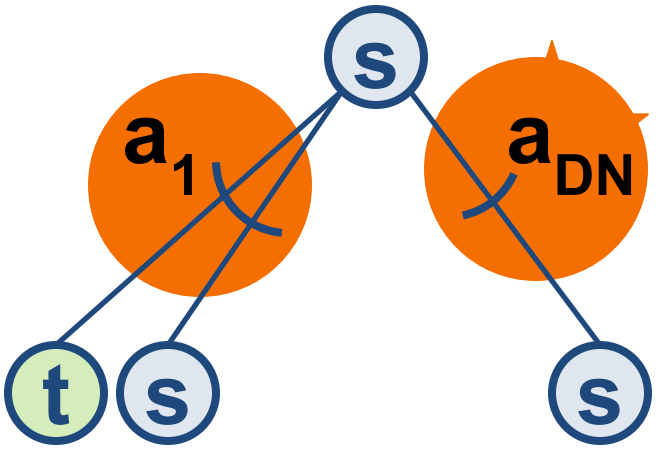
E[D] = Expected Path Detection to get to a state N attack steps away via attack step a_i.

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

Attractiveness of an attack step a_i to an adversary with planning horizon N=

$$U_C(E[C]) * w_c + U_P(E[P]) * w_p + U_D(E[D]) * w_d$$

Attractiveness Calculation Example - Planning Horizon = 1



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

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

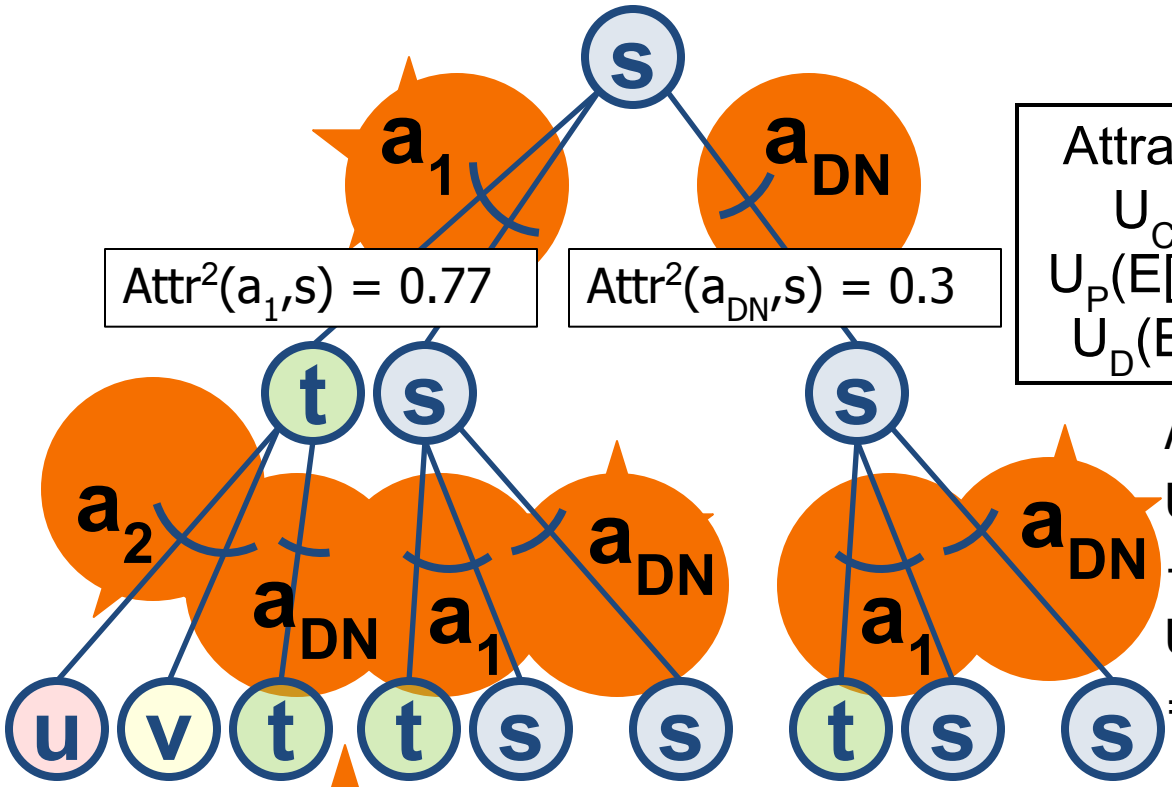
$Attr(a_{DN}) = 0.3$

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

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

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

Attractiveness Calculation Example - Planning Horizon = 1



Attractiveness of attack step a_i =
U_C(E[**path cost** of a_i]) * w_c +
U_P(E[**horizon payoff** of a_i]) * w_p +
U_D(E[**path detection** of a_i]) * w_d

$Attr^2(a_1,s)$ =
$U_C(\$500*0.9+\$0*0.1+\$1000) * w_c$
+ $U_P(\$8000*0.9 + \$0*0.1) * w_p$ +
$U_D(0.038*0.9 +0.1*0.1) * w_d$
= 0.77

$Attr^2(a_{DN},s)$ =
$U_C(\$0) * w_c$ +
$U_P(\$0) * w_p$ +
$U_D(0) * w_d$
= 0.3

$$\beta^2(s) = a_1$$