Toward Models for Forensic Analysis

Sean Peisert (UC San Diego)
Matt Bishop (UC Davis)
Sid Karin (UC San Diego)
Keith Marzullo (UC San Diego)

SADFE'07
April 11, 2007

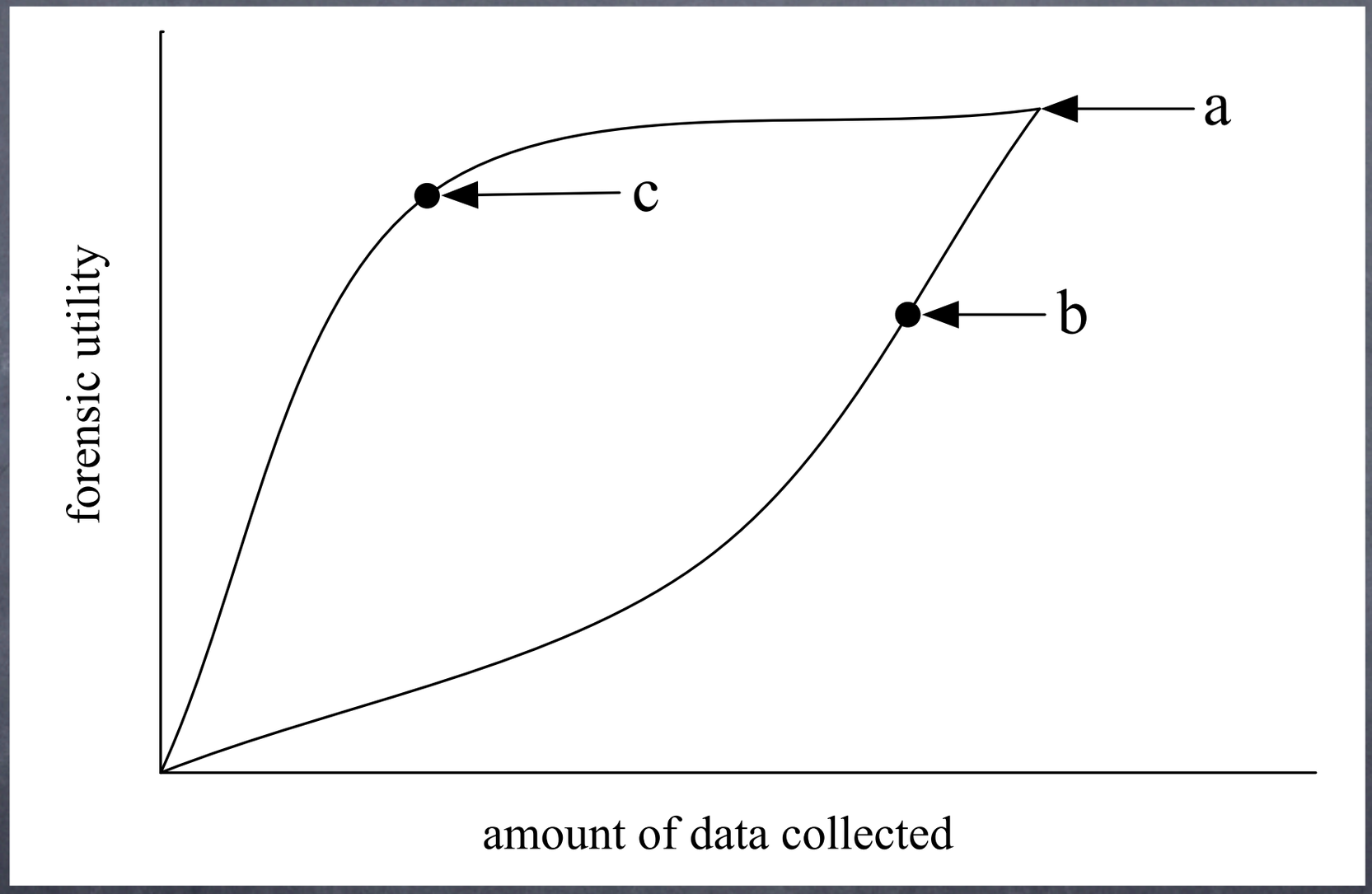
What is Forensic Analysis?

- Forensic analysis is the process of answering the questions:
  - How did an event take place?
  - What was the nature of the event?
  - What were the effects of the event?
- Forensic analysis applies to arbitrary events. This can include attacks (which we will focus on in this talk), but is not limited to attacks.
- Forensic analysis is not intrusion detection.
  - The goal of intrusion detection is to determine **whether** an attack occurred.

The Forensic Problem

- **The problem** with forensics today is that it is poorly understood what data is necessary to analyzing previous events, and there is no general solution to find that data.
- Data is often redundant, missing, vague, or misleading.
- We seek a model of forensics: a systematic approach to determining what information to log.
- In this talk, **our goal** is to present qualities that we believe that a good model should possess, and an example of a model that possesses those qualities.
- **In the future**, we will apply this model to a build an automated system to perform logging and auditing and to extend the model to other forensic domains.

Illustration of Data vs. Utility



(a) represents logging everything, with perfect utility. Many approaches have attempted or approached (b).(c) represents our goal.

Three Types of Existing Approaches

1. Approaches used by forensic analysis that were not intended for use in forensics.
2. Approaches that are "ad hoc." "Ad hoc" means "for a particular purpose." Thus, these approaches that had a purpose other than solving the forensic problem.
3. Approaches based on models of forensics, for which the goals of the model were something other than **determining what to log to analyze attacks** (which is our goal).

Related/Existing Work Using Ad Hoc Approaches

- Solutions Intended for Non-Forensic Purposes
  - Syslog
  - Process Accounting
  - IDS Alerts
- Ad Hoc Solutions Aimed at Forensics
  - Network
    - TCPwrappers [Venema92]
  - Disk/Filesystem
    - Coroner's Toolkit
    - Tripwire [Spaf94], LAFS [Wee95], other file audit [Bishop88]
  - "toolbox approach" [e.g. Farmer & Venema 2004]
  - Kernel/Filesystem
    - Sun Basic Security Module (BSM)
    - BackTracker [King06]
  - Our Early Work w/Function Calls [Peisert, et al., TDSC 2007]

Related/Existing Work Using Models

- Model of Auditing and Logging [Bishop89]
  - Looks at what is **possible** to audit & log.
- Analysis of Computer Intrusions [Gross97]
  - Focuses on **analyzing** data that is there, not determining what should be logged.
- Model of Security Monitoring [Kuperman04]
  - Focuses on **performance** issues.

Towards a Better Solution

What Should be Logged?

[Peisert, et al. in NSPW'05]

- Our evaluation of shortcomings in existing approaches resulted in 5 principles, that if followed, should ameliorate the shortcomings in future approaches if a forensic model were to use them.
  - Principle 1: Consider the entire system
  - Principle 2: Log information without assumptions about attacks and attackers
  - Principle 3: Consider effects, not just actions
  - Principle 4: Consider context to assist understanding
  - Principle 5: Present and process actions and results in a way that can be understood by a human

Attributes for Forensic Models

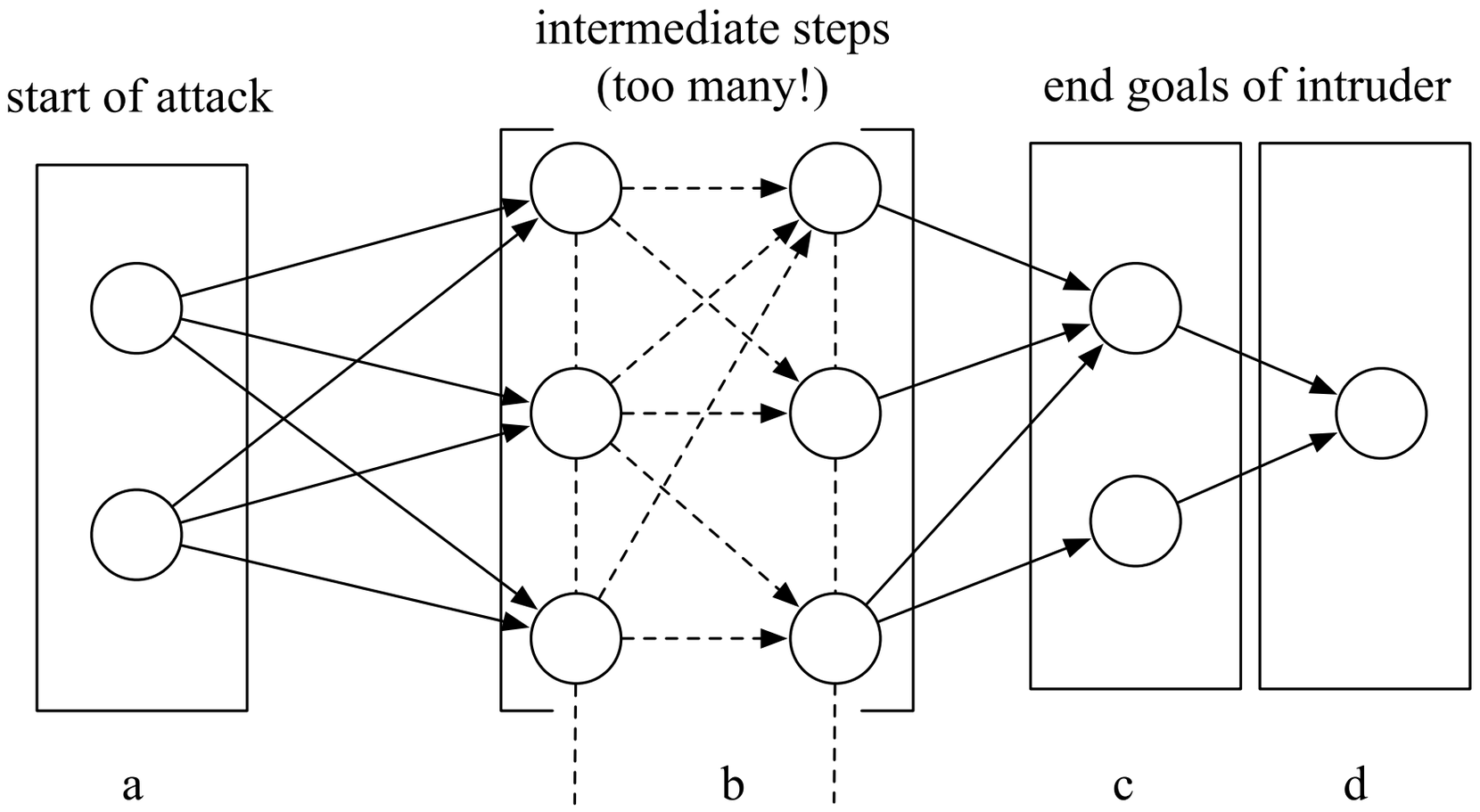
- Indicate the information to log and let the analyst choose whether to record information
- Provide tuning parameters
- Automated metrics could help
- Consider both pre-conditions and post-conditions
- Place bounds on unknown stages of attacks
- Consider the context surrounding an event
- Make the data well-formed
- Associate discrete events to analyze larger attacks
- Make logged events and actual events one-to-one

Our Model: Laocoön

\Lā — ä — kō — än\

- The model generalizes and formalizes an approach for determining what should be logged.
- Forensics is the process of analyzing **any past event**. We consider "attacks" as a means of illustration of the model, but our model should be applicable to other events, too.
- The model builds upon forensic principles & attributes.
- Uses the Requires/Provides Model [Templeton2000]
- Builds upon formalization of multi-stage attacks [Zhou07] using attack trees made up of intruder goals.

Analyzing Attacks with Attack Graphs



"attack": sequence of events that violates a security policy (could be internal, as in the insider problem)

"goal": to achieve a particular result or violation (defined using "capabilities")

"attack graph": Multiple goals linked together in dependency order

Methodology

1. Start with an attack graph representing attacker goals to achieve a set of results
2. Working backward from ultimate goal, build capability pairs for each goal.
3. Place bounds on the "unknown," intermediate goals.
4. From the capability pairs, extract the information to log.
5. Instrument the system to collect the values described by the model.
6. Log the specified information during execution.
7. Conduct the forensic analysis after an attack is suspected.

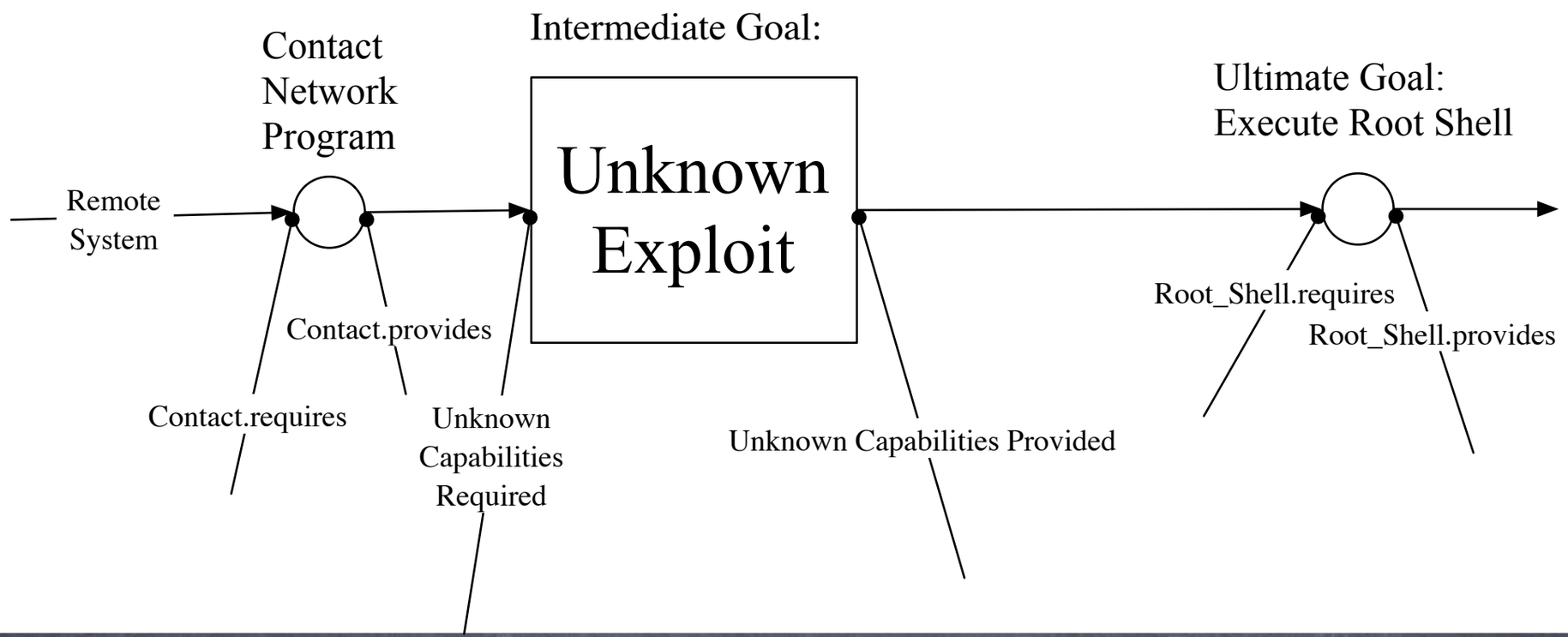
Step 1: Building Attack Graphs

- Currently manual. Eventually automated.
- Possibility: Based on "policies" [Bishop, Wee, & Frank 1996] or safety properties that can be detected in bounded time.

Step 2: Building Capability Pairs

- A goal has two sets of capabilities: requires & provides. Each set contains one or more capabilities.
- capability: a 6-tuple (based on [Zhou07])
  - source/destination address
  - credentials
  - actions
  - services
  - properties

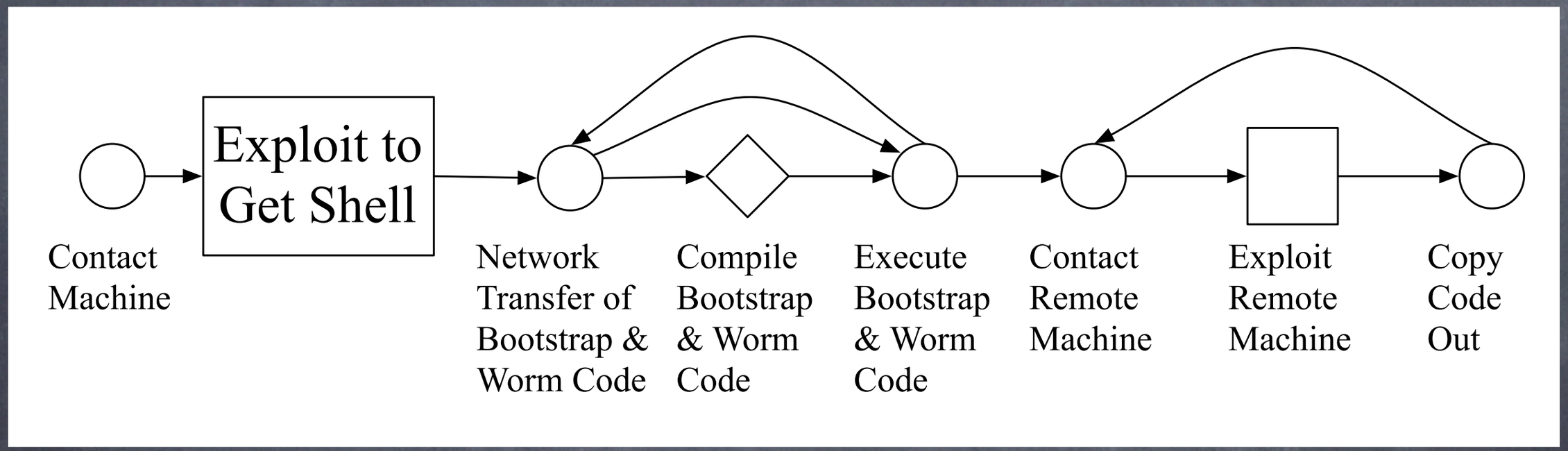
Step 3: Bounding Intermediate Goals



Step 4: Analyzing Goals to Determine what to Log

- Filter by **source**, **dest**, and **credential**
- Determine logging point (e.g. kernel, hardware) and the data to log (e.g. syscall, syscall params, assembly code, environment) by **action**, **service**, and **property.**
- Sometimes, the model may require information to be logged that involves large amounts of resources. This is a benefit of the model because it gives a choice and indicates what is missing.

Example: An Internet Worm



Future Work

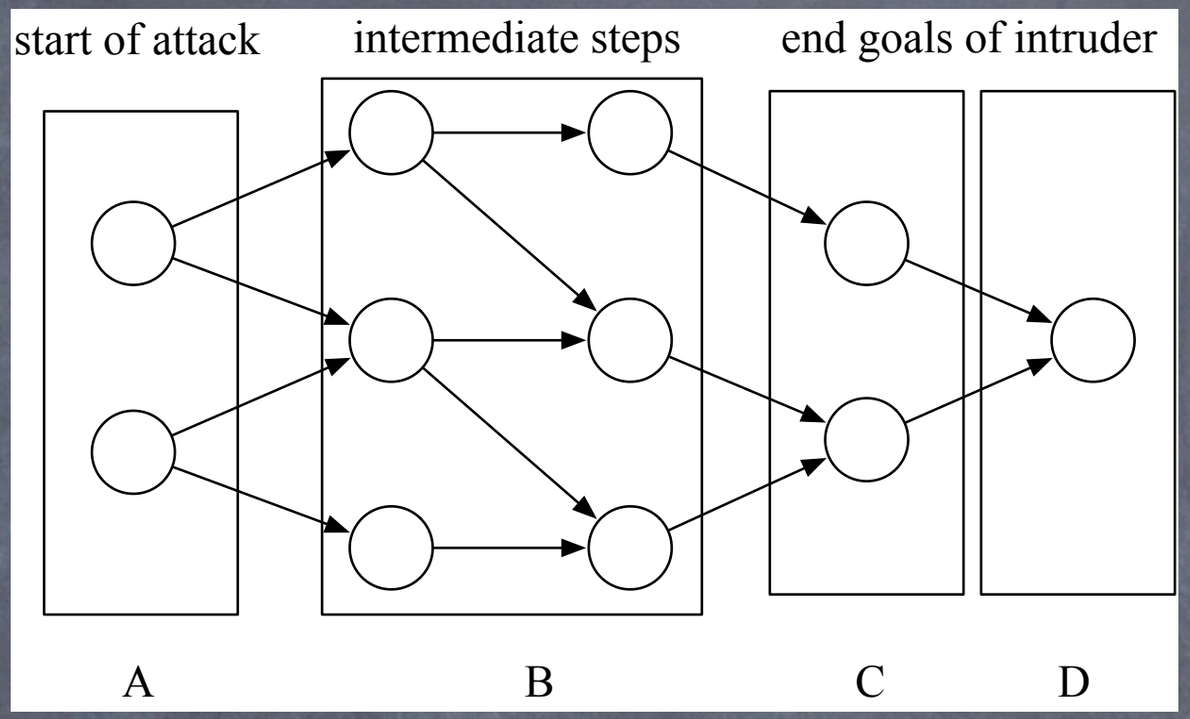
- Generalize beyond "attacks"
- Legal admissibility
- More experiments [Peisert & Bishop, WISE'07]
- Taking Laocoön from a Model to a System:
  - Automated implementation
  - Active state awareness
  - Scaling attack graphs
  - Reducing necessary paths
  - Concurrent attacks & relative time
  - Universal path ID to associate & minimize data.
  - Human interface
  - Automated graph generation ("policy discovery")

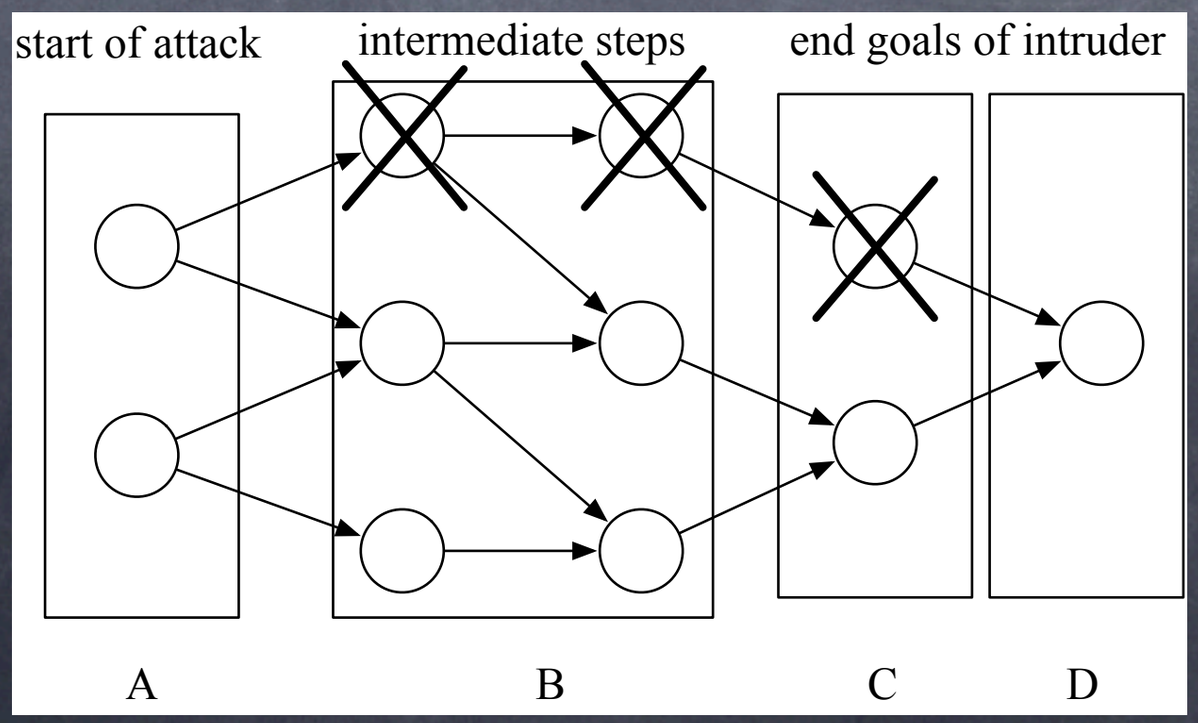
Future Work: Policy Discovery

[Bishop & Peisert, 2006 UC Davis Tech Report]

- Policy discovery is the process of reverse-engineering system configurations into human-understandable, higher-level security policies. The reverse is policy enforcement.
- Open research includes methods of using it...
  - ...to generate attack graphs
  - ...to do automated translation of capability pairs to data necessary to log and where to log it
  - ...to make logged data & events 1:1
  - ...to prove completeness of model

Future Work: Reducing Paths





Attributes for Forensic Models

- Indicate the information to log and let the analyst choose whether to record information
- Provide tuning parameters
- Automated metrics could help
- Consider both pre-conditions and post-conditions
- Place bounds on unknown stages of attacks
- Consider the context surrounding an event
- Make the data well-formed
- Associate discrete events to analyze larger attacks
- Make logged events and actual events one-to-one

Conclusions

Forensics is currently ad hoc; a rigorous model of forensics is desirable and achievable. We have presented one: Laocoön.

- Guidelines of a forensic model seem complete.
- Laocoön adheres to the guidelines
- Experiments have shown that Laocoön is effective and the data output from the model is sufficient & necessary.
- Laocoön can be used to analyze past events quickly.
- Laocoön is mindful of not recording too much or too little data, or just the wrong data.
- Laocoön can also be a foundation of many future steps that can formalize, automate, and improve forensic analysis.