A Quantitative Methodology for Security Monitor Deployment

Problem

- We want to be able to determine monitors to deploy in terms of fusion and intrusion detection goals
  - Monitors must collect sufficient information to meet goals
  - Should support intrusion detection even when monitors are taken out

- Monitoring is expensive!
  - Price of monitors

    NIKSUN NetDetector/NetVCR 2005 – DPI, analytics, and alerting appliance
    Price: $10,000 to $100,000[1]
  - Data storage costs

    Large enterprises can generate hundreds of GBs of logs *per day*
  - Cost of data analysis and response

    Salary of a computer forensic analyst: upwards of $80,000[3]

[1] Jerry Shenk, "NetDetector/NetVCR 2005 Traffic Analyzer," SANS Institute, Whitepaper, Aug. 2007.
[2] "Salary: Network Security Administrator," *Glassdoor*. [Online]. Available: http://www.glassdoor.com/Salaries/network-security-administrator-salary-SRCH_KO0,30.htm.
[3] "Salary: Computer Forensic Analyst," *Glassdoor*. [Online]. Available: http://www.glassdoor.com/Salaries/computer-forensic-analyst-salary-SRCH_KO0,25.htm.

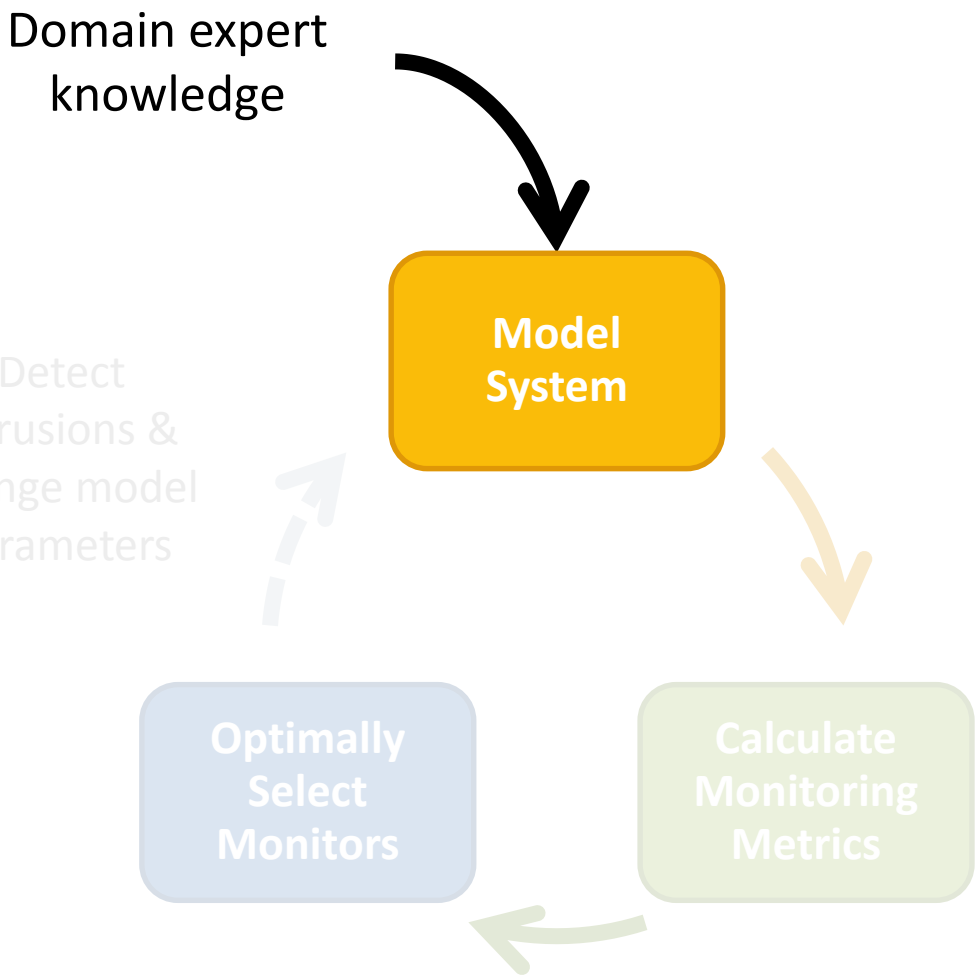
Our contribution

We have developed a **cost-sensitive** methodology for monitor selection that **meets fusion and intrusion detection goals**

Guiding principles

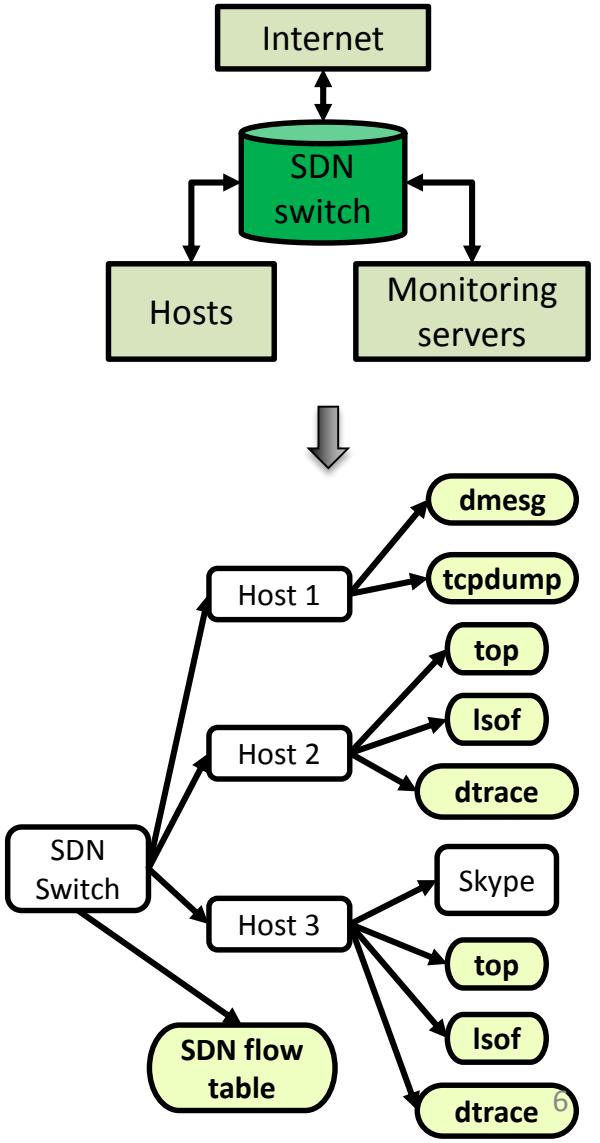
- Monitors and computing assets can be compromised
  - Monitor compromise can affect ability to detect intrusions
- Monitors should be deployed with redundancy such that the effect of compromise or unavailability is mitigated

Outline


Domain expert knowledge
Model System
Calculate Monitoring Metrics
Optimally Select Monitors
Detect intrusions & change model parameters

Model: System model

- **Assets**: computing components in the system to monitor and protect
- **Monitors**: sensors that can be deployed to provide information about system behavior and intrusions
- **Asset-asset dependence**: directed dependency between assets
- **Monitor-asset dependence**: dependence relationship between monitors and assets



Model: System model

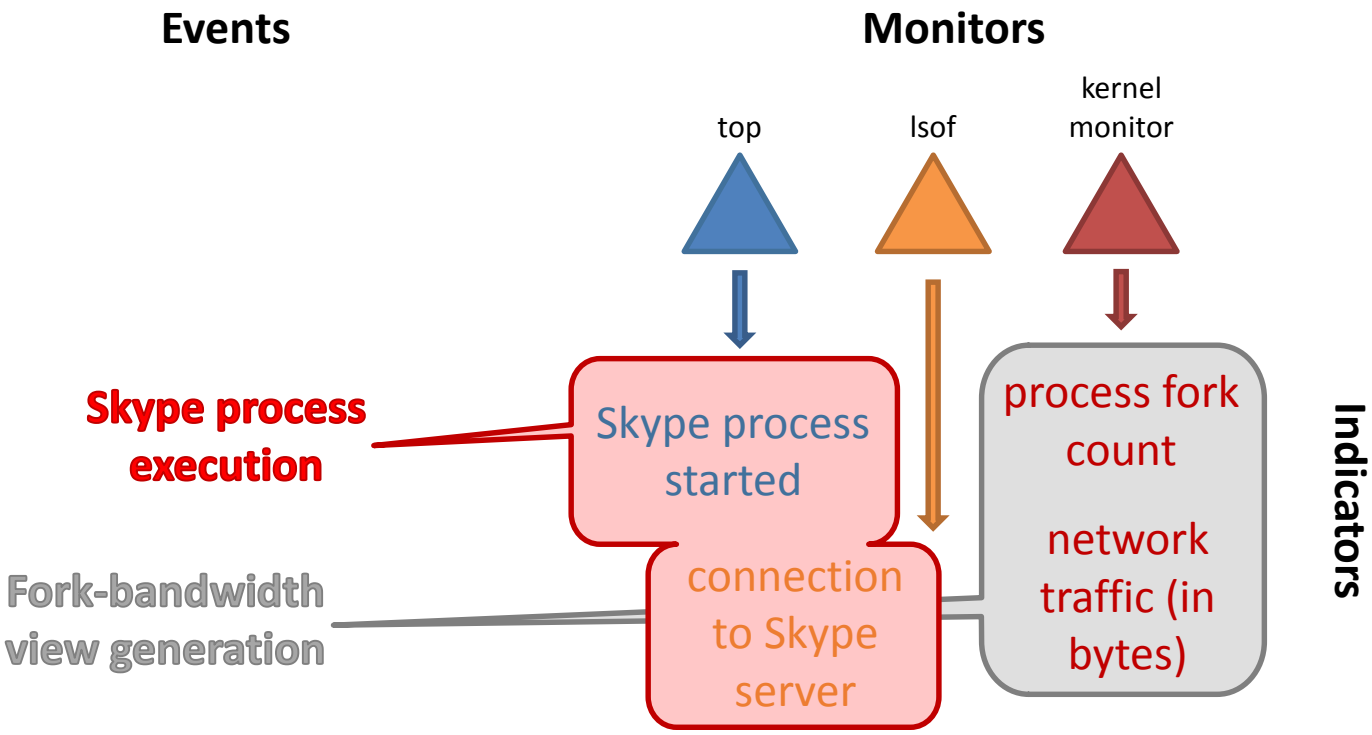
- **Truthfulness of monitors:** what proportion of a monitor's output can be trusted
  - Measure of compromisability of monitors or the assets on which they depend
- **Resource costs of assets:** monetary cost of dedicating computing resources to monitoring

Model: Data model

- **Indicators**: primitives representing information provided by monitors about events in the system
- **Events**: occurrences in the system we would like to observe
  - E.g., malicious activity, network resource usage profile
  - Events are detectable using sets of indicators (similar to an IDS signature)

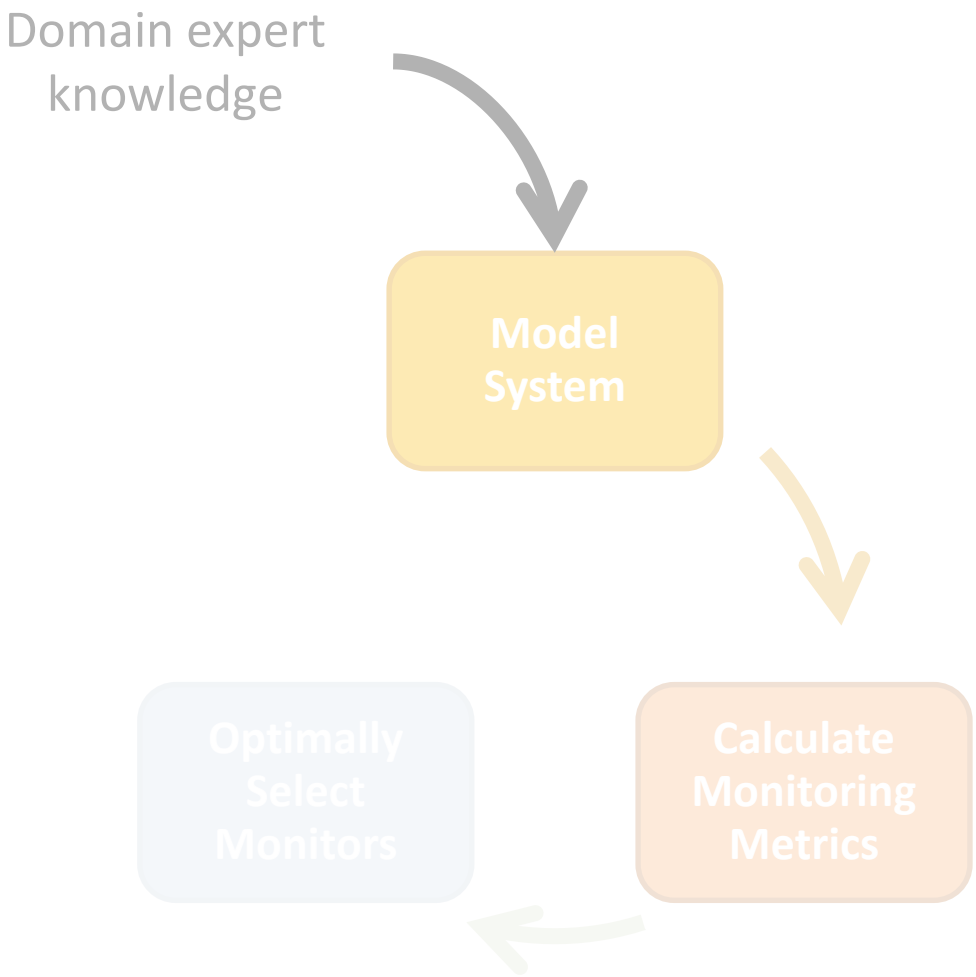
- **Detectability**: an event is *detectable* if at least one of its indicator sets is observable given the set of selected monitors

Model: Data model illustration



- Events can map to multiple sets of indicators

Outline


Domain expert knowledge
Model System
Calculate Monitoring Metrics
Optimally Select Monitors

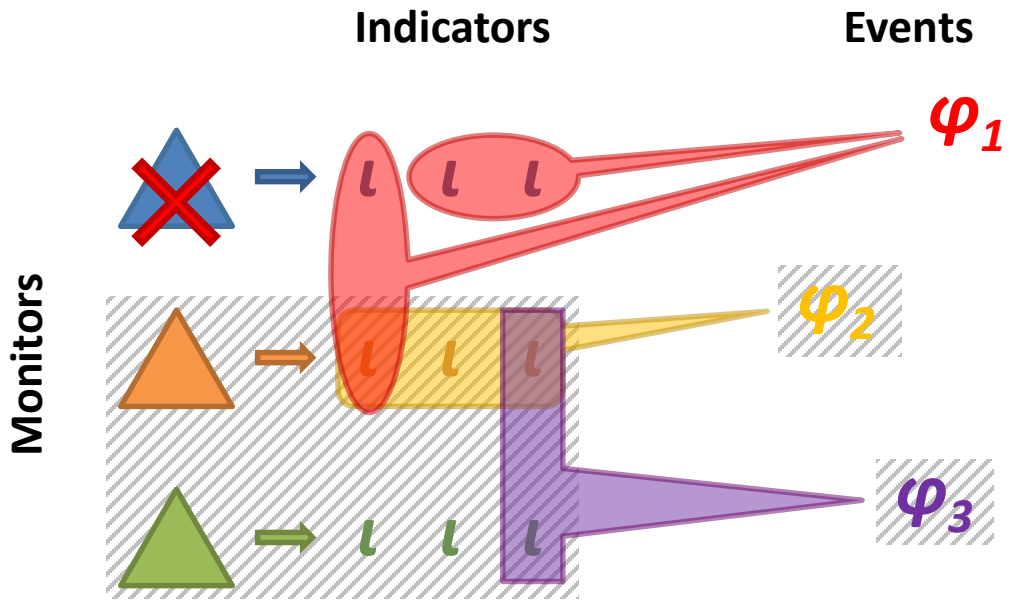
Monitoring metrics

- **Goal of metrics**: *quantify* utility and cost of monitors in supporting fusion and intrusion detection
- Three monitor utility metrics:
  - Coverage
  - Redundancy
  - Confidence
- One cost metric:
  - Monitor cost

- Developing other metrics for specific analyses

Metrics: Coverage

- **Definition:** overall fraction of select events that can be detected given a set of monitors

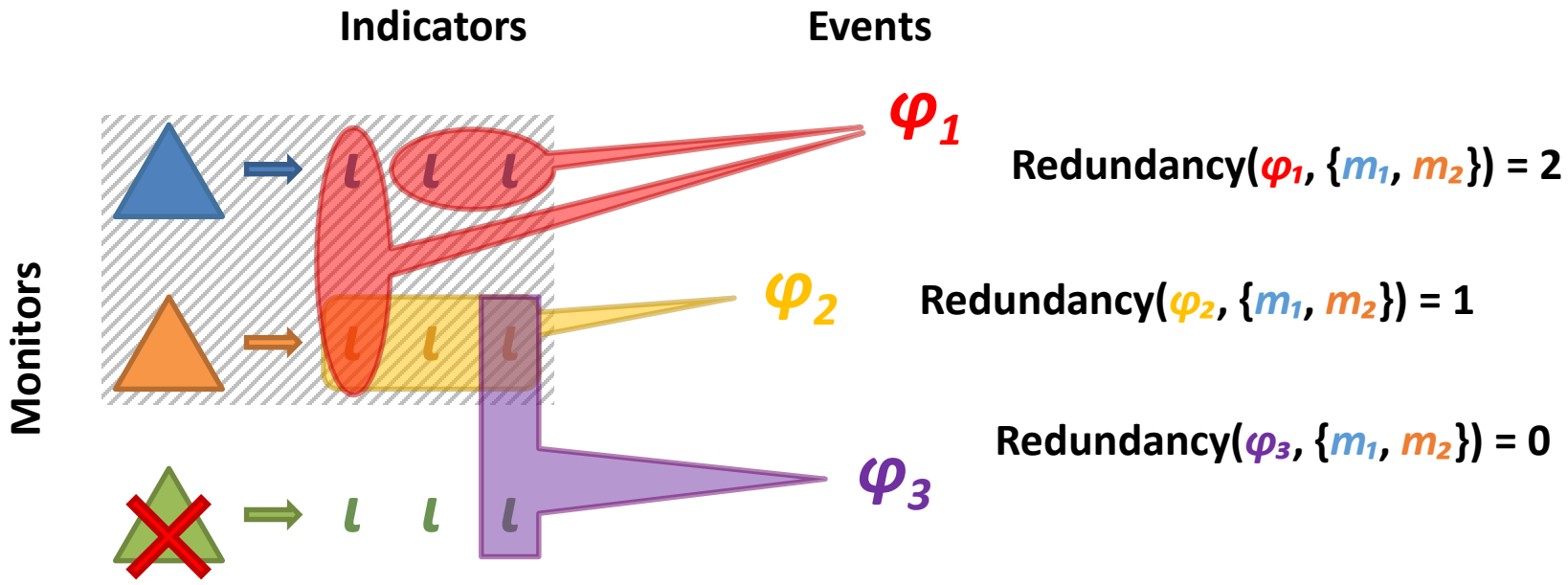


Coverage({φ_1, φ_2, φ_3}, {m_2, m_3}) = 67%

$$\text{Coverage}\left(\Phi, M_d\right) = \frac{\left|\left\{\phi : \delta\left(\phi, M_d\right) \wedge \phi \in \Phi\right\}\right|}{\left|\Phi\right|}$$

Metrics: Redundancy

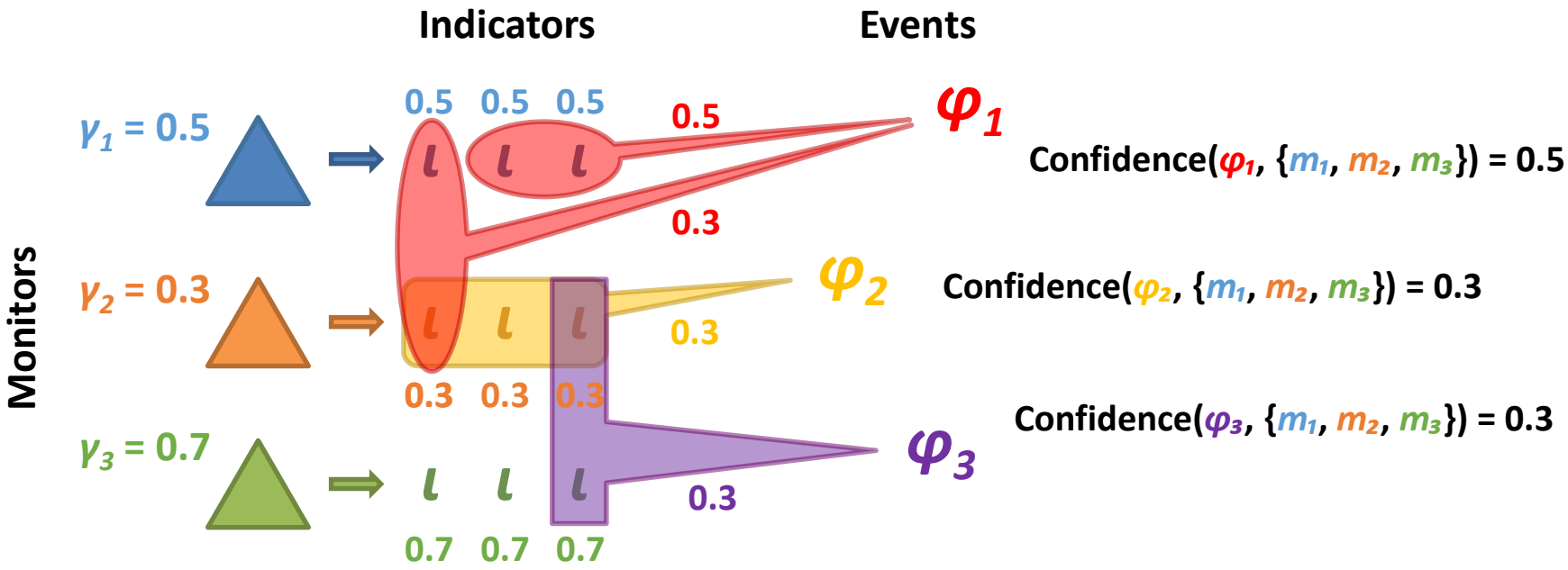
- **Definition:** the number of ways an event can be detected given a set of monitors



$$\text{Redundancy}(\phi, M_d) = \sum_{\sigma \in \varsigma(\phi, M_d)} \min_{\iota \in \sigma} \left| \{ m : m \in M_d, \iota \in \alpha(m) \} \right|$$

Metrics: Confidence

- **Definition**: belief in the ability of the monitors to detect events accurately, even when monitors are compromised



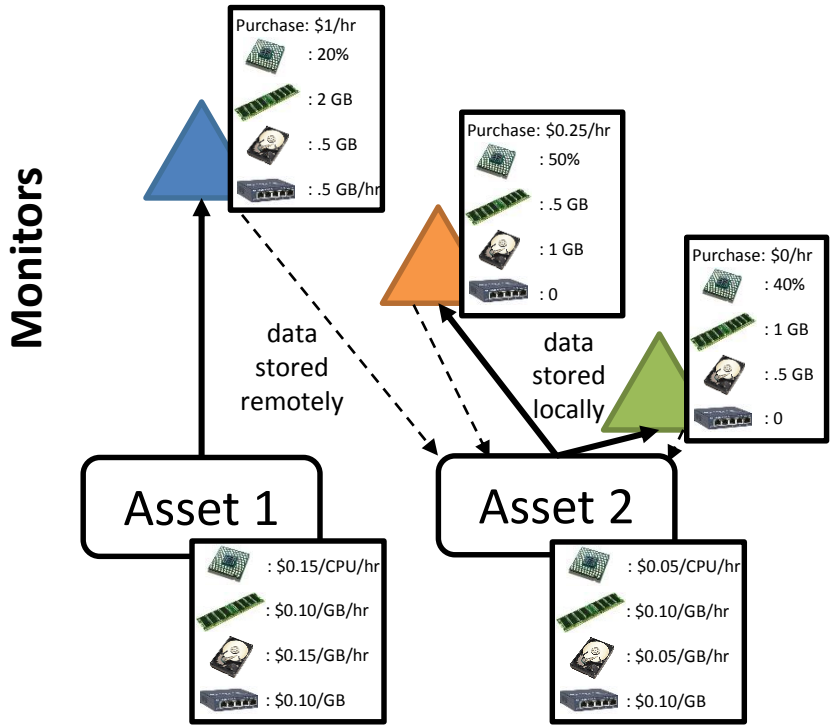
$$\text{Confidence}\left(\phi, M_d\right) = \max_{\sigma \in \beta(\phi)} \min_{\iota \in \sigma} \gamma_I\left(\iota, M_d\right)$$

Metrics: Cost model

- Monitor cost is based on the following resources (units are in parentheses):
  - CPU utilization (e.g., per CPU core per hour)
  - Memory utilization (e.g, per GB per hour)
  - Disk storage (e.g., per TB per hour)
  - Network communication (e.g., per GB)
- Monitors also have an amortized purchase price and recurring maintenance cost

Metrics: Cost

- **Definition:** overall value of the computing resources consumed by monitors that are deployed in the system per unit time



Cost(m_1) = $1.31/hr

Cost(m_2) = $0.38/hr

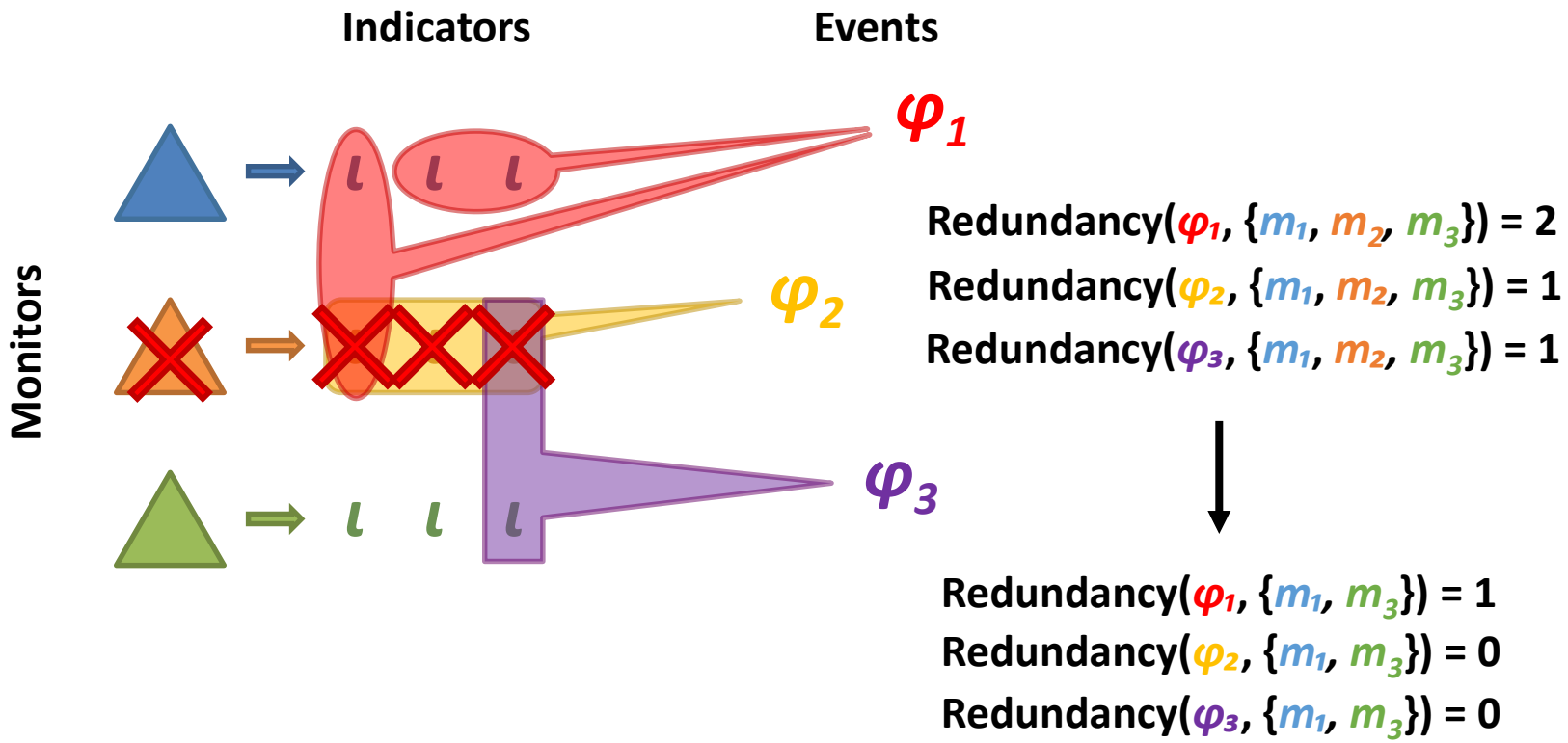
Cost(m_3) = $0.15/hr

Cost({m_1, m_2, m_3}) = $1.84/hr

$$\mathrm{Cost}(M_d) = \sum_{m \in M_d} \Big(P(m) + \mathrm{CPUCost}(v)\,\mathrm{CPUUtilization}(m) + \mathrm{MemoryCost}(v)\,\mathrm{MemoryUtilization}(m) + \mathrm{NetworkCost}(v)\,\mathrm{NetworkUtilization}(m) + \mathrm{DiskCost}(\mathrm{LogsTo}(m))\,\mathrm{DiskUtilization}(m) \Big)$$

Proposed Metric: Sensitivity

- **Definition:** What is the change in coverage, redundancy, or confidence when an asset is taken out?



Outline

Domain expert knowlege
Model System
Calculate Monitoring Metrics
Optimally Select Monitors

Optimal selection methodology

- **Goal**: to be able to use methodology to answer a variety of monitor selection questions
  - What is the minimum set of monitors that can detect a given attack/set of attacks (assuming no compromise)?
  - What is the minimum set of monitors required to generate a fused system view with a specified redundancy per field?
  - Under cost constraints, what set of monitors will maximize my ability to detect a high-priority attack?

Optimal selection methodology:

Constrained-cost monitor selection

Given :

$S = (V, E), \Phi, I, \alpha, \beta, \gamma_M$

$\mathbf{w}_{\text{Coverage}}, \mathbf{min}_{\text{Coverage}}$

$\mathbf{w}_{\text{Redundancy}_\phi}, \mathbf{w}_{\text{Confidence}_\phi}, \mathbf{min}_{\text{Redundancy}_\phi}, \mathbf{min}_{\text{Conf}}$

maxCost

Objective function: monitoring utility, defined as weighted sum of metric values

- Parameterized by user-specified weight parameters

Maximize :

$$\mathbf{w}_{\text{Coverage}}\text{Coverage}(\Phi, M_d) + \sum_{\phi \in \Phi}\left(\mathbf{w}_{\text{Redundancy}_\phi}\text{Redundancy}(\phi, M_d)\right.$$

$$+ \mathbf{w}_{\text{Confi}}$$

Constraints:

- Cost function to minimize
- User-specified minimum detection metric requirements

Subject to :

$\text{Cost}(M_d) \leq \mathbf{maxCost}$

$\text{Coverage}(\Phi, M_d) \geq \mathbf{min}_{\text{Coverage}}$

$\text{Redundancy}(\phi, M_d) \geq \mathbf{min}_{\text{Redundancy}_\phi} \quad \forall$

$\text{Confidence}(\phi, M_d) \geq \mathbf{min}_{\text{Confidence}_\phi} \quad \forall$

$M_d \in \{0,1\}^{|M|}$

0-1 integer nonlinear programming problem, with monitors as input variables

Optimal selection methodology:

Unconstrained cost monitor selection

Given :

$S = (V, E), \Phi, I, \alpha, \beta, \gamma_M$

$\mathbf{min}_{\text{Coverage}}$

$\mathbf{min}_{\text{Redundancy}_\phi}, \mathbf{min}_{\text{Confidence}_\phi} \quad \forall \phi \in \Phi$

Minimize :

$\text{Cost}(M_d)$

Objective function: cost metric

Subject to :

$\text{Coverage}(\Phi, M_d) \geq \mathbf{min}_{\text{Coverage}}$

$\text{Redundancy}(\phi, M_d) \geq \mathbf{min}_{\text{Redundancy}_\phi} \quad \forall \phi \in \Phi$

$\text{Confidence}(\phi, M_d) \geq \mathbf{min}_{\text{Confidence}_\phi} \quad \forall \phi \in \Phi$

$M_d \in \{0,1\}^{|M|}$

Constraints: user-specified minimum detection metrics requirements

Solving for optimal monitor selection

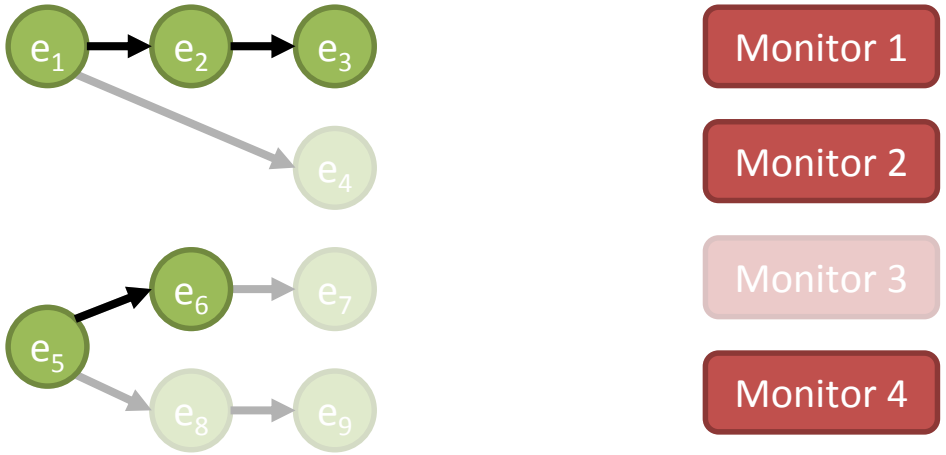
- Cannot use traditional convex optimization or mixed-integer nonlinear programming techniques
  - Objective functions and constraints are nonlinear and non-convex
  - Search space consists of binary vectors
- We solve the optimal monitor selection integer programs using branch-and-bound
  - Searches over space of possible selections, pruning suboptimal sets of monitor selections

Scalability

- Theoretical runtime complexity for the solver is $O\left(2^{|M|}|I|(|M|+|\Phi|)\right)$
  - Linear in the number of events ($|\Phi|$) and indicators ($|I|$)
  - Superexponential in the number of monitors ($|M|$)
- We are working on improving scalability to support hundreds of monitors
- Investigating:
  - Linear relaxation method – approximating the objective function using a linear function during each step of the branch-and-bound algorithm
  - Changing variables for optimization to make objective functions convex or constraints linear
  - Using heuristic approaches to find near-optimal solution

Next steps: Predictive Monitoring

- Given an intrusion detection mechanism, dynamically update monitoring to improve detectability of events in progress
- Important when finer granularity monitoring can be done, but is prohibitively expensive to deploy all of the time



Conclusions

- We quantify event detection ability using our metrics
- We represent fusion and intrusion detection requirements in terms of our metrics
- We define a methodology for selecting monitors that allows us to compute optimal monitor selection under constraints
- Our optimization problem is very expressive

Ongoing and Future Work

- Improving scalability of optimal selection algorithm
- Per-asset redundancy and sensitivity analysis
  - Can be used to evaluate “grand lie” hypothesis
- Using fusion & response to develop more metrics
- Completion of the deployment loop
  - Using intrusion detection algorithms and observed indicators to drive *predictive monitoring*