Hybrid Risk Assessment Model based on Bayesian Networks

Francois-Xavier Aguessy, Olivier Bettan, Gregory Blanc, Vania Conan, and Herve Debar
francois-xavier.aguessy@telecom-sudparis.eu

Thales Communications & Security, Paris, France
Telecom SudParis, Institut Mines-Télécom, Évry, France

IWSEC 2016, Tokyo, September 12th, 2016

Outline

1. Introduction
2. State of the art
3. Hybrid Risk Assessment Model
4. Conclusion

Introduction

- Context:
  - Increase in the number and complexity of attacks.
  - Need means to know the attacks that can happen, are happening, and to prevent them.
- Goal: Modelling multi-step attacks for Dynamic Risk Assessment.
- Assess the level of security of an information system according to security alerts.
- Determine the attacks that are currently happening.
- Know how the attacker arrived here and what he could do next.
- Models based on attack graph.

Outline

1 Introduction

2 State of the art

- Attack Graphs
- Dynamic Risk Assessment models
- Cycle problem

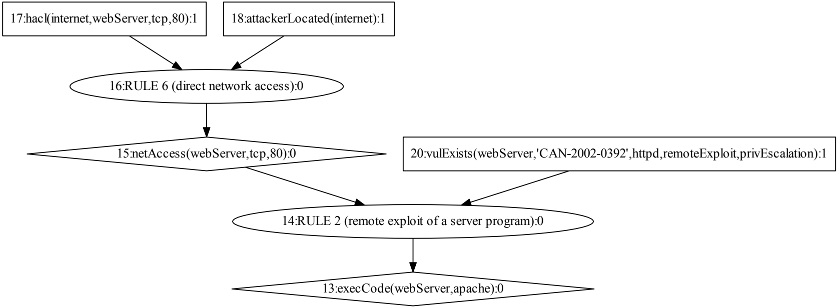
3 Hybrid Risk Assessment Model

4 Conclusion

Attack graphs

- First representation of network attacks.
- Several formalisms regrouped under the name *Attack Graph*.
- Logical attack graphs:
  - AND/OR directed graph,
  - Nodes are logical facts reachable by an attacker,
  - Leaves represent the preconditions used to achieve goals.
- Topological attack graphs:
  - Based on logical attack graphs,
  - More concise and understandable,
  - Nodes are machines or IP addresses linked by attack steps.

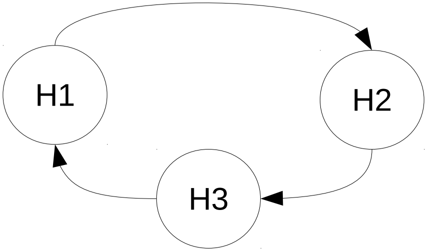
Attack graphs


17:hacl(internet,webServer,tcp,80):1
18:attackerLocated(internet):1
16:RULE 6 (direct network access):0
15:netAccess(webServer,tcp,80):0
20:vulExists(webServer,'CAN-2002-0392',httpd,remoteExploit,privEscalation):1
14:RULE 2 (remote exploit of a server program):0
13:execCode(webServer,apache):0

Attack graphs

- First representation of network attacks.
- Several formalisms regrouped under the name *Attack Graph*.
- Logical attack graphs:
  - AND/OR directed graph,
  - Nodes are logical facts reachable by an attacker,
  - Leaves represent the preconditions used to achieve goals.
- Topological attack graphs:
  - Based on logical attack graphs,
  - More concise and understandable,
  - Nodes are machines or IP addresses linked by attack steps.

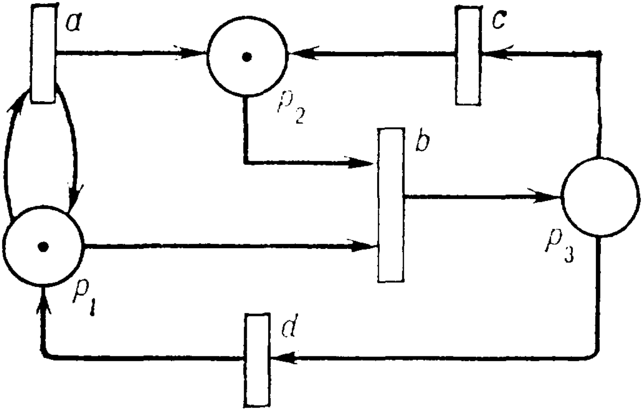
Attack graphs



Dynamic Risk Assessment models

- Attack graphs:
  - ✓ Technology mastered,
  - ✓ Contains accurate description of multi-steps attacks,
  - × Not created to model on-going attacks (no nodes for detection/alerts, no position of attacker).
- Attack nets:
  - ✓ Concurrency and progress of several attacks,
  - × Attacker can not be in several places (several privileges),
  - × Difficult to add tokens (representing alerts) during runtime.
- Bayesian attack graphs:
  - ✓ Powerful tools to compute and propagate probabilities,
  - ✓ Description of attacks more expressive (no-more AND/OR),
  - × Size of Conditional Probability Tables
  - × Management of cycles (Bayesian networks need acyclic graphs).

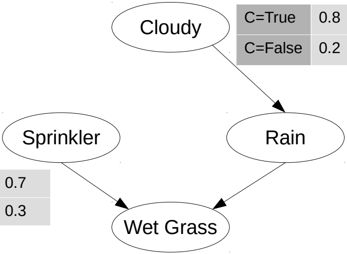
Dynamic Risk Assessment models



Dynamic Risk Assessment models

- Attack graphs:
  - ✓ Technology mastered,
  - ✓ Contains accurate description of multi-steps attacks,
  - × Not created to model on-going attacks (no nodes for detection/alerts, no position of attacker).
- Attack nets:
  - ✓ Concurrency and progress of several attacks,
  - × Attacker can not be in several places (several privileges),
  - × Difficult to add tokens (representing alerts) during runtime.
- Bayesian attack graphs:
  - ✓ Powerful tools to compute and propagate probabilities,
  - ✓ Description of attacks more expressive (no-more AND/OR),
  - × Size of Conditional Probability Tables
  - × Management of cycles (Bayesian networks need acyclic graphs).

Dynamic Risk Assessment models



C=True	0.8
C=False	0.2

Cloudy	True	False
R=True	0.8	0.1
R=False	0.2	0.9

R=True	0.7
R=False	0.3

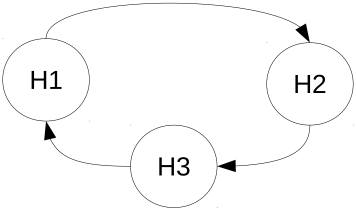
Rain	True		False	
Sprinkler	True	False	True	False
WG=True	0.99	0.90	0.90	0.0
WG=False	0.01	0.10	0.10	1.0

Dynamic Risk Assessment models

- Attack graphs:
  - ✓ Technology mastered,
  - ✓ Contains accurate description of multi-steps attacks,
  - × Not created to model on-going attacks (no nodes for detection/alerts, no position of attacker).
- Attack nets:
  - ✓ Concurrency and progress of several attacks,
  - × Attacker can not be in several places (several privileges),
  - × Difficult to add tokens (representing alerts) during runtime.
- Bayesian attack graphs:
  - ✓ Powerful tools to compute and propagate probabilities,
  - ✓ Description of attacks more expressive (no-more AND/OR),
  - × Size of Conditional Probability Tables
  - × Management of cycles (Bayesian networks need acyclic graphs).

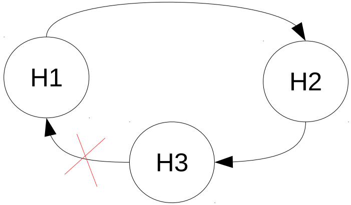
Cycles in attack graphs

A topological attack graph:



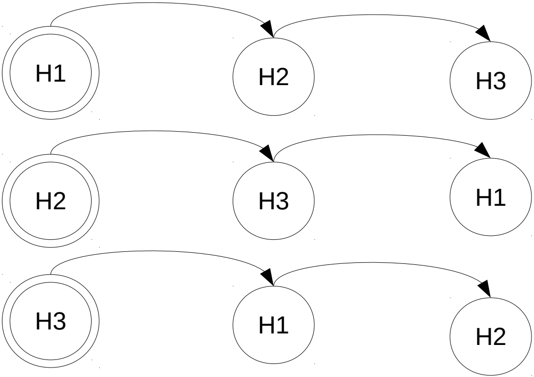
Cycles in attack graphs

Current approaches followed to build Bayesian Attack graphs from a cyclic graph (when mentioned):



Cycles in attack graphs

But there are three possible paths:

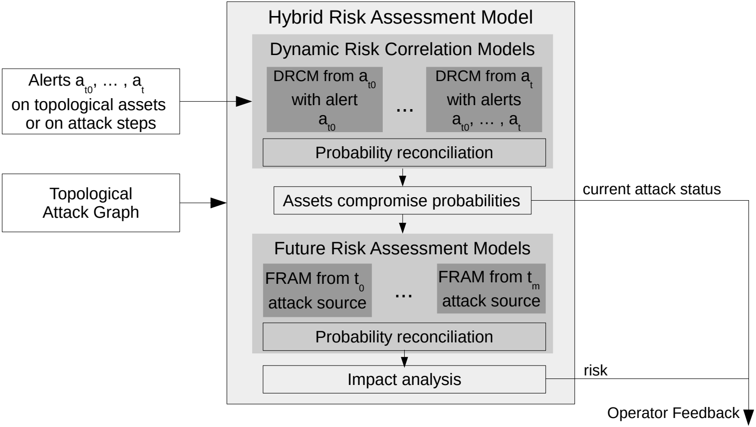


The solution we propose: enumerate the paths.

Outline

1. Introduction
2. State of the art
3. Hybrid Risk Assessment Model
   - Architecture
   - Dynamic Risk Correlation Model
   - Future Risk Assessment Model
   - Performance results
4. Conclusion

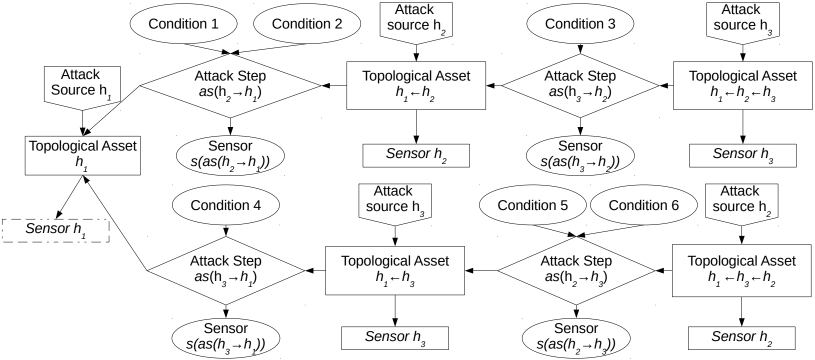
High-level model architecture



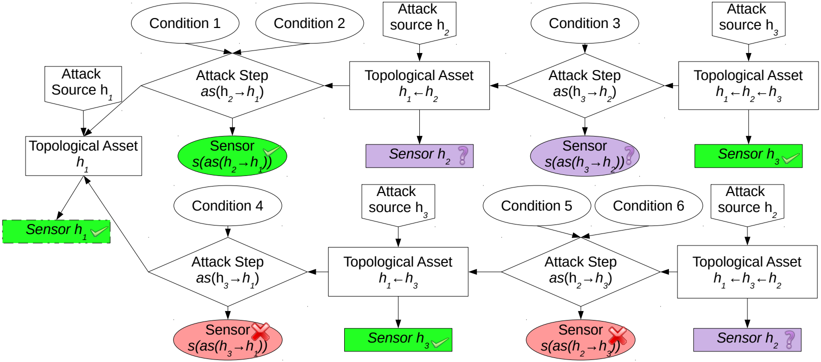
Dynamic Risk Correlation Model

- Build from a bunch of (ordered) alerts.
- To analyze how these alerts may have been produced.
- Gives attack sources and attack paths (via the Bayesian topological nodes) probabilities.

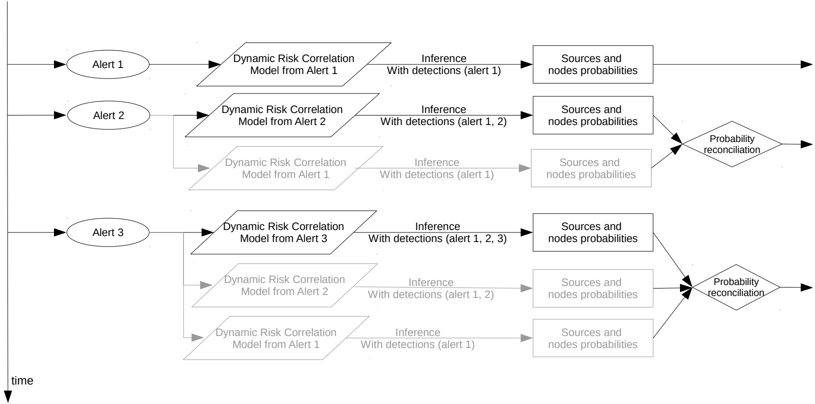
Dynamic Risk Correlation Model from alert on h_1


Condition 1
Condition 2
Attack source h_2
Condition 3
Attack source h_3
Attack Source h_1
Attack Step $as(h_2 \to h_1)$
Topological Asset $h_1 \leftarrow h_2$
Attack Step $as(h_3 \to h_2)$
Topological Asset $h_1 \leftarrow h_2 \leftarrow h_3$
Topological Asset h_1
Sensor $s(as(h_2 \to h_1))$
Sensor h_2
Sensor $s(as(h_3 \to h_2))$
Sensor h_3
Sensor h_1
Condition 4
Attack source h_3
Condition 5
Condition 6
Attack source h_2
Attack Step $as(h_3 \to h_1)$
Topological Asset $h_1 \leftarrow h_3$
Attack Step $as(h_2 \to h_3)$
Topological Asset $h_1 \leftarrow h_3 \leftarrow h_2$
Sensor $s(as(h_3 \to h_1))$
Sensor h_3
Sensor $s(as(h_2 \to h_3))$
Sensor h_2

Dynamic Risk Correlation Model from alert on h_1

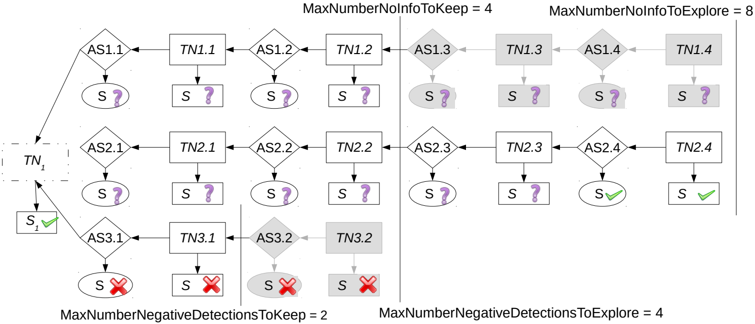

Condition 1
Condition 2
Attack source h_2
Condition 3
Attack source h_3
Attack Source h_1
Attack Step $as(h_2 \rightarrow h_1)$
Topological Asset $h_1 \leftarrow h_2$
Attack Step $as(h_3 \rightarrow h_2)$
Topological Asset $h_1 \leftarrow h_2 \leftarrow h_3$
Topological Asset h_1
Sensor $s(as(h_2 \rightarrow h_1))$ ✓
Sensor h_2 ?
Sensor $s(as(h_3 \rightarrow h_2))$?
Sensor h_3 ✓
Sensor h_1 ✓
Condition 4
Attack source h_3
Condition 5
Condition 6
Attack source h_2
Attack Step $as(h_3 \rightarrow h_1)$
Topological Asset $h_1 \leftarrow h_3$
Attack Step $as(h_2 \rightarrow h_3)$
Topological Asset $h_1 \leftarrow h_3 \leftarrow h_2$
Sensor $s(as(h_3 \rightarrow h_1))$ ✗
Sensor h_3 ✓
Sensor $s(as(h_2 \rightarrow h_3))$ ✗
Sensor h_2 ?

Build of the model according to detections


Alert 1
Dynamic Risk Correlation Model from Alert 1
Inference With detections (alert 1)
Sources and nodes probabilities
Alert 2
Dynamic Risk Correlation Model from Alert 2
Inference With detections (alert 1, 2)
Sources and nodes probabilities
Dynamic Risk Correlation Model from Alert 1
Inference With detections (alert 1)
Sources and nodes probabilities
Probability reconciliation
Alert 3
Dynamic Risk Correlation Model from Alert 3
Inference With detections (alert 1, 2, 3)
Sources and nodes probabilities
Dynamic Risk Correlation Model from Alert 2
Inference With detections (alert 1, 2)
Sources and nodes probabilities
Dynamic Risk Correlation Model from Alert 1
Inference With detections (alert 1)
Sources and nodes probabilities
Probability reconciliation
time

Performance improvement – pruning

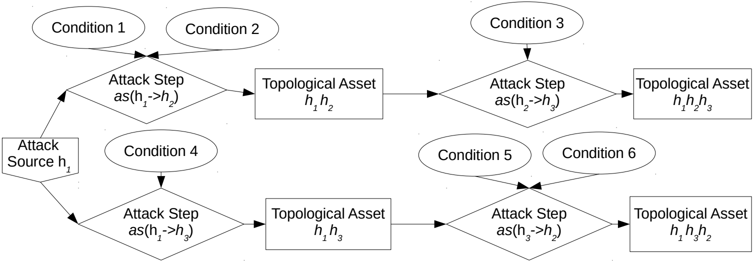
- Prune paths that do not bring information.
- Count the number of no-detection or no-information.
- Two parameters: maximum to keep, and maximum to explore.


MaxNumberNoInfoToKeep = 4
MaxNumberNoInfoToExplore = 8
AS1.1
TN1.1
AS1.2
TN1.2
AS1.3
TN1.3
AS1.4
TN1.4
S ?
AS2.1
TN2.1
AS2.2
TN2.2
AS2.3
TN2.3
AS2.4
TN2.4
S ✓
TN_1
S_1 ✓
AS3.1
TN3.1
AS3.2
TN3.2
S ✗
MaxNumberNegativeDetectionsToKeep = 2
MaxNumberNegativeDetectionsToExplore = 4

Future Risk Assessment model

- Build from an attack source with its probability.
- To analyze the most probable possible futures.
- Dynamicity by updating the probability of conditions, taking into account the context (already exploited vulnerabilities...).

Example of Future Risk Assessment model

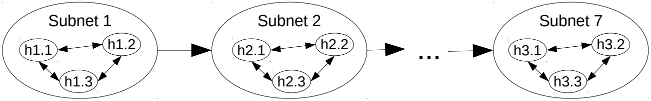

Condition 1
Condition 2
Condition 3
Attack Step $as(h_1\text{->}h_2)$
Topological Asset h_1h_2
Attack Step $as(h_2\text{->}h_3)$
Topological Asset $h_1h_2h_3$
Attack Source h_1
Condition 4
Condition 5
Condition 6
Attack Step $as(h_1\text{->}h_3)$
Topological Asset h_1h_3
Attack Step $as(h_3\text{->}h_2)$
Topological Asset $h_1h_3h_2$

Performances ?

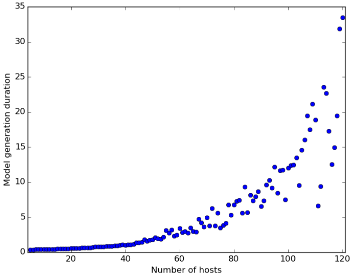
- No evidences,
- No sensors,
- Only Forward propagation.
- No need to go very far from detections / attack sources,
- Several small models in parallel.

Performances

Simulations network topology:



HRAM model generation and inference duration:



Outline

1 Introduction

2 State of the art

3 Hybrid Risk Assessment Model

4 Conclusion

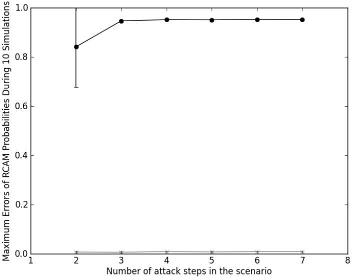
Conclusion

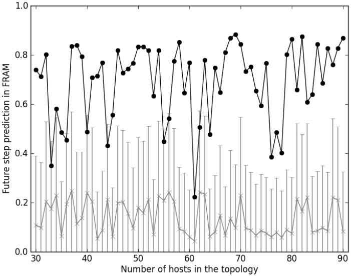
- Bayesian inference is a powerful tool to deduce the effects of several events on a global model.
- Well adapted to Dynamic Risk Assessment problem.
- To use the inference algorithms, necessary to satisfy the constraints of the formalism (acyclic, CPT size. . .).
- Definition of an hybrid model combining dynamic risk correlation models (past) with possible future models (future).
- Generation of the HRAM on topologies far bigger than the state of the art.

Thanks for your attention! Any questions?

- F.-X. Aguessy, O. Bettan, G. Blanc, V. Conan, H. Debar. Hybrid Risk Assessment Model based on Bayesian Networks. In *11th International Workshop on Security, IWSEC 2016, Tokyo, Japan, September 12-14, 2016, Proceedings*, 2016.
- ✉ `francois-xavier.aguessy@telecom-sudparis.eu`
- Slides available online @ `https://fxaguessy.fr/en/articles/hram/`

Accuracy results





Performance improvements – Polytree

- A directed graph is a polytree if its underlying undirected graph is a tree.
- Even exact inference algorithms are much more performing (Lauritzen or Pearl).
- Can do exact inference up to 25.000 nodes (whereas problems with > 500) with a normal laptop.
- Specification of the dynamic risk correlation models as polytrees.