Insider Threat Modeling: an Adversarial Risk Analysis approach

Chaitanya Joshi
(*with David Rios Insua & Jesus Rios*)

Department of Mathematics & Statistics,

University of Waikato, New Zealand.

31^{st} May 2019
GDRR 2019, Washington DC.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Adversarial Risk Analysis (ARA)

Improvement over Game Theory.

- Does not assume common knowledge.
- Solves the problem from the point of view of just one of the players (defender).
- Incorporates their knowledge and uncertainties about their choices, outcomes as well as their adversary's preferences, choices and outcomes.
- Can incorporate different strategic thinking models for the adversary.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Insider threat

- A significant and widespread problem that is often under reported.
- Very little data available.
- Challenging to prevent, detect and counter.

Where?

- Security and intelligence agencies.
- Finance, business.
- Data/knowledge security, cyber security, etc.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

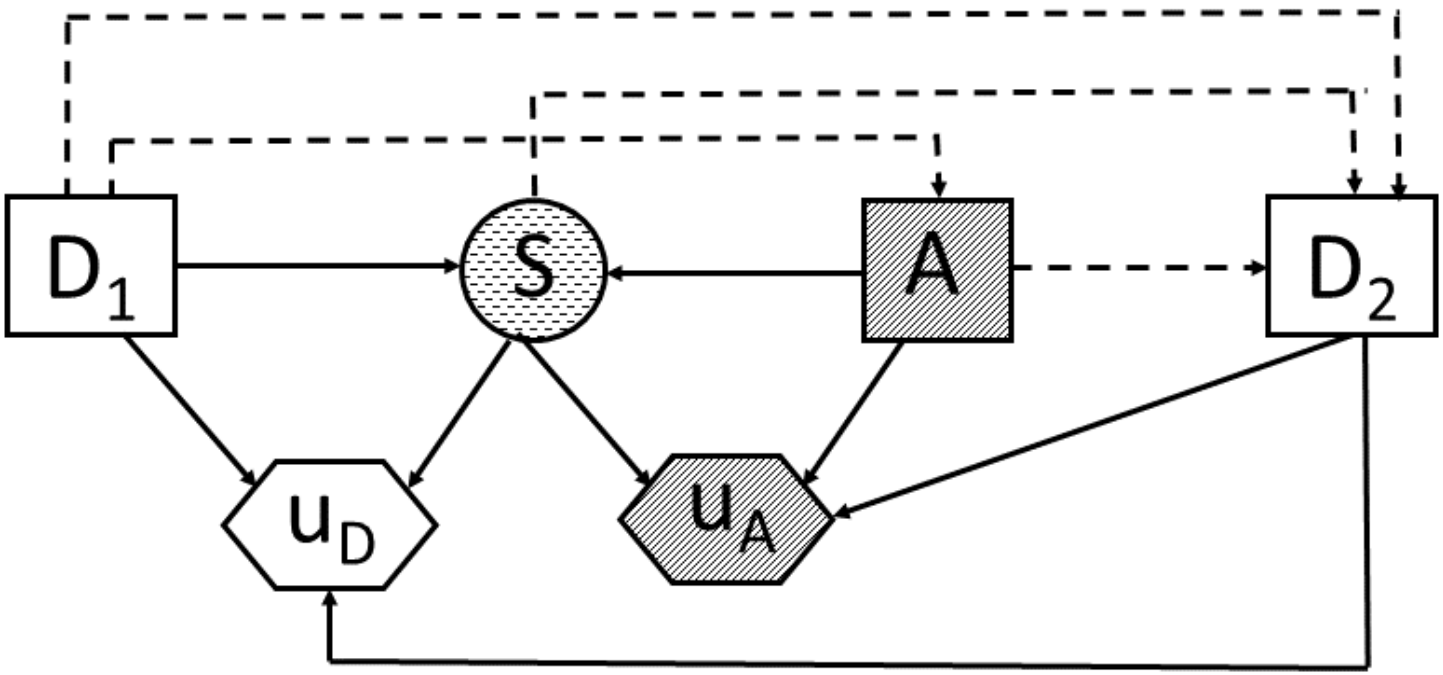


Figure: Bi-agent Influence Diagram (BAID) for the Defend-Attack-Defend insider threat game

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

- Initially, the organization must choose one of the available preventive measures d_1 in the set $\mathcal{D}_1$.
- Having observed the preventive measure taken, the employee will adopt one of the actions a in $\mathcal{A}$.
- The set $\mathcal{S}$ consists of the possible outcomes s that can occur as a result of the preventive measure d_1 and the attack a adopted.
- Once the attack has been detected, the organization will choose to carry out one of the possible actions d_2 in the set $\mathcal{D}_2$ to end the attack, limit any damage and possibly pre-empt future attacks leading to the final outcomes of both agents, respectively, evaluated through their utility functions u_D and u_A.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

For its solution, the defender must first quantify the following:

1. The distribution $p_D(a|d_1)$ modeling her beliefs about the attack a chosen at node A by the employee given the chosen defense d_1.
2. The distribution $p_D(s|d_1, a)$ modeling her beliefs about the outcome s of the attack, given a and d_1.
3. Her utility function $u_D(d_1, s, d_2)$ which evaluates the consequences associated with their first (d_1) and second (d_2) defensive actions as well as the outcome s of the attack.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

Given these assessments, the defender first seeks to find the action $d_2^*(d_1, s)$ maximizing her utility

$$d_2^*(d_1, s) = \arg \max_{d_2 \in \mathcal{D}_2} u_D(d_1, s, d_2), \tag{1}$$

leading to the best second defense when the first one was d_1 and the outcome was s. Then, they seek to compute the expected utility $\psi_D(d_1, a)$ for each $(d_1, a) \in \mathcal{D}_1 \times \mathcal{A}$ as

$$\psi_D(d_1, a) = \int u_D(d_1, s, d_2^*(d_1, s)) p_D(s|d_1, a)\, ds. \tag{2}$$

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

Moving backwards, she computes her expected utility for each $d_1 \in \mathcal{D}_1$ using the predictive distribution $p_D(a|d_1)$ through

$$\psi_D(d_1) = \int \psi_D(d_1, a) p_D(a|d_1)\, da. \tag{3}$$

Finally, the defender has to find her maximum expected utility decision $d_1^* = \arg\max_{d_1 \in D_1} \psi_D(d_1)$. This backward induction shows that the defender's optimal strategy is to first choose d_1^* and, then, after having observed s, choose $d_2^*(d_1^*, s)$.

Catch! The above analysis requires the defender to elicit $p_D(a|d_1)$. How do you elicit that???

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

Moving backwards, she computes her expected utility for each $d_1 \in \mathcal{D}_1$ using the predictive distribution $p_D(a|d_1)$ through

$$\psi_D(d_1) = \int \psi_D(d_1, a) p_D(a|d_1)\, da. \tag{3}$$

Finally, the defender has to find her maximum expected utility decision $d_1^* = \arg\max_{d_1 \in D_1} \psi_D(d_1)$. This backward induction shows that the defender's optimal strategy is to first choose d_1^* and, then, after having observed s, choose $d_2^*(d_1^*, s)$.

Catch! **The above analysis requires the defender to elicit** $p_D(a|d_1)$**. How do you elicit that???**

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

- The defender could model the attacker's strategic analysis by assuming that the attacker will perform an analysis similar to hers to find their optimal attack a^*. To do so, the defender should assess the attacker's utility function $u_A(a, s, d_2)$ and probability distributions $p_A(s|a, d_1)$ and $p_A(d_2|d_1, a, s)$.
- But these are not available to the defender - we could model her uncertainty about them through a random utility function $U_A(a, s, d_2)$ and random probability distributions $P_A(s|a, d_1)$ and $P_A(d_2|d_1, a, s)$.
- Once these random quantities are elicited, the defender solves the attacker's decision problem using backward induction. This is done by following a process similar to how they solved their own decision problem but taking into account the randomness in judgments.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

First, the defender finds the random expected utility for each $d_2 \in \mathcal{D}_2$

$$\boldsymbol{\Psi}_{\boldsymbol{A}}(d_1, a, s) = \int U_A(a, s, d_2) P_A(d_2|d_1, a, s)\, dd_2. \tag{4}$$

Then, they find the random expected utility for each pair $(d_1, a) \in \mathcal{D}_1 \times \mathcal{A}$

$$\boldsymbol{\Psi}_{\boldsymbol{A}}(d_1, a) = \int \boldsymbol{\Psi}_{\boldsymbol{A}}(d_1, a, s) P_A(s|d_1, a)\, ds, \tag{5}$$

and compute the random optimal attack $A^*(d_1)$ given the defense d_1

$$A^*(d_1) = \arg\max_{a \in \mathcal{A}} \boldsymbol{\Psi}_{\boldsymbol{A}}(d_1, a). \tag{6}$$

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

A simple *defend-attack-defend* model.

Finally, once the defender assesses $A^*(d_1)$, she is able to solve her decision problem. The desired predictive distribution by the defender about the attack chosen a given the initial defense d_1 is

$$p_D(a|d_1) = p_D(A^* = a|d_1) \text{ and}$$
$$p_D[A^* \leq a|d_1] = \int_0^a P_D(A^* = x|d_1)\, dx. \tag{7}$$

Let's get real!

Not all employees are the same!

ARA with segmented employees.

We classify the employees as A_1 (*the good*), A_2 (*the bad*) and A_3 (*the ugly*), with S_1, S_2 and S_3 being the corresponding outcome sets.

- A_1 : Correctly and promptly perform their duties including following any procedures to prevent insider attacks. Positive impact on the productivity and work culture of the organization and will correctly report any suspicious activity, thus helping the organization to protect itself.
- A_2 : While not intentionally working to harm the organization, will help to create an environment which could increase the chances of an insider attack through their accidental or deliberate actions. For e.g. misuse the defensive procedures, creating a culture of mis-trust and loss in productivity. Therefore, their actions will be negative to the organization.
- A_3 : Actively aim at harming the organization. They are the ones who intend to launch an insider attack. Their actions will therefore be very negative to the organization.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees.

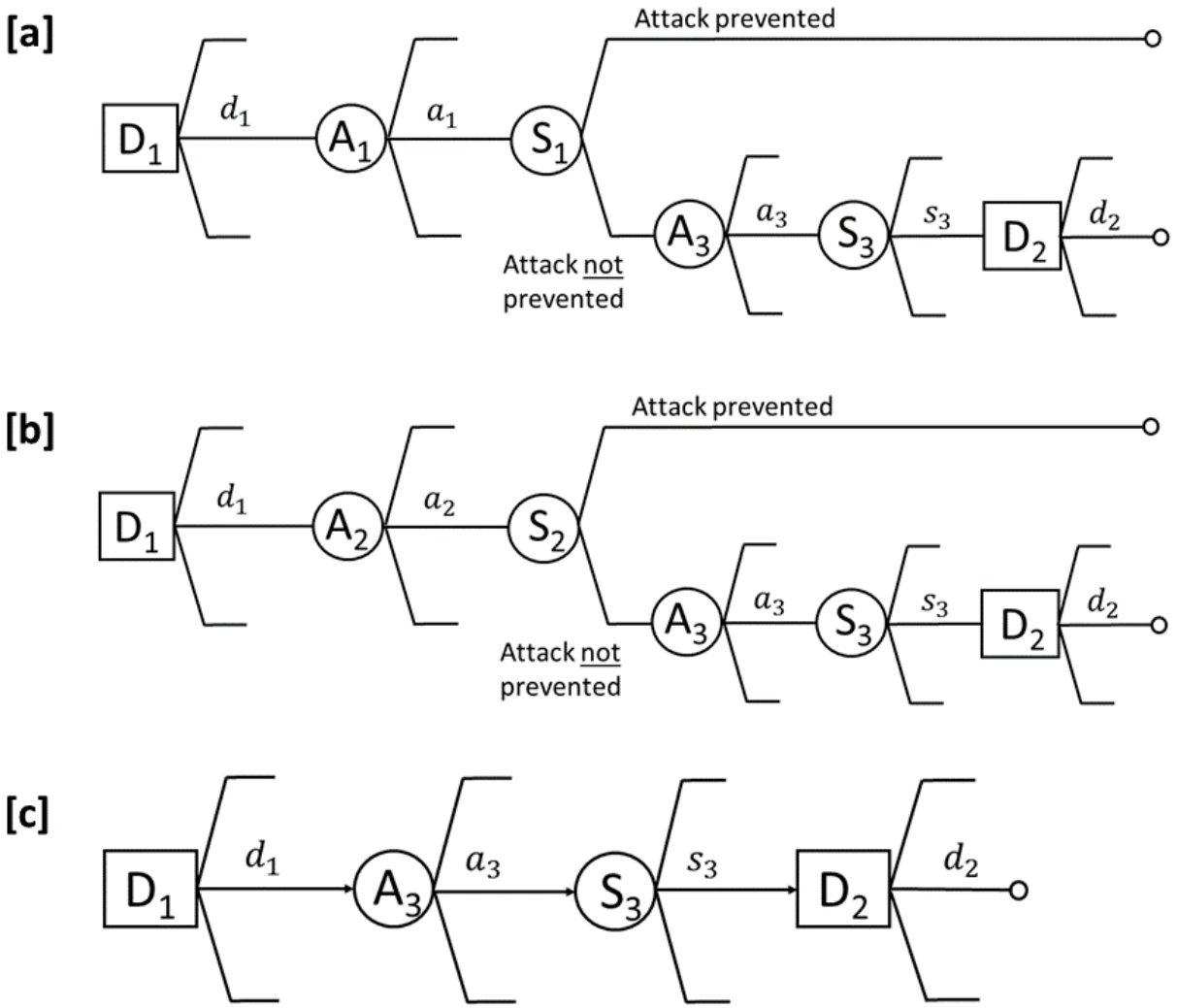


Figure: Decision trees for the three games in the insider threat

ARA with segmented employees.

- Actions by A_1 may reduce the chance of insider attacks as well as the chance of one of them succeeding. Similarly, actions by A_2 may increase the chance of an insider attack as well as the chance of one of them succeeding.
- Game [c] is identical to the simple *Defend-Attack-Defend* game.
- ARA solution to games [a] and [b] will consist of identical sets of steps. Henceforth, we use A_i, $i = 1, 2$ and S_i, $i = 1, 2$.
- Note that we differentiate between node A_i, which is uncertain, and node A_3, which is a decision node but belonging to a different decision maker, as this last one is strategic.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees.

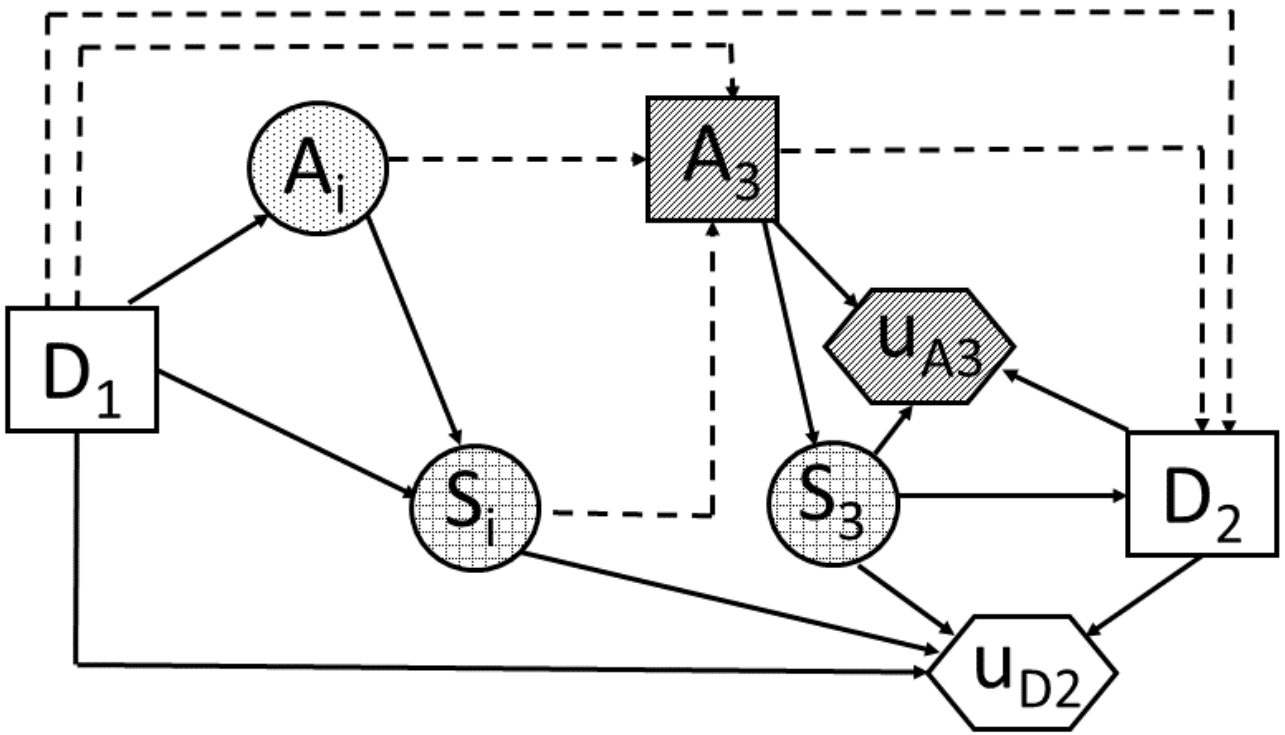


Figure: MAID for decision trees [a] and [b] in the segmented employees insider threat game.

ARA with segmented employees.

The defender must first quantify the following.

1. Her predictive distribution $p_D(a_i|d_1)$ about the action that will be chosen at node A_i given the defense d_1.
2. Her predictive distribution $p_D(s_i|d_1, a_i)$ about the outcome of such action, given a_i and d_1.
3. Her predictive distribution $p_D(a_3|d_1, a_i, s_i)$ about the attack that will be chosen at note A_3 given the outcome s_i and actions a_i and d_1.
4. Her predictive distribution $p_D(s_3|d_1, a_i, s_i, a_3)$ about the outcome of the attack, given outcome s_i, actions a_3, a_i and d_1.
5. The utility function $u_D(d_1, a_i, s_i, a_3, s_3, d_2)$ given their first and second defensive actions, the outcomes of the attack s_3 and s_i and the actions a_3 and a_i.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees.

Given these, the defender works backwards along the decision trees in Figure 2 [a] or [b]. First, they seek to find the action $d_2^*(d_1, a_i, s_i, a_3, s_3)$ maximizing their utility

$$d_2^*(d_1, a_i, s_i, a_3, s_3) = \arg \max_{d_2 \in \mathcal{D}_2} u_D(d_1, a_i, s_i, a_3, s_3, d_2). \tag{8}$$

Then, for each $(d_1, a_i, s_i, a_3) \in \mathcal{D}_1 \times \mathcal{A}_i \times \mathcal{S}_i \times \mathcal{A}_3$, they seek to compute the expected utility $\psi_D(d_1, a_i, s_i, a_3)$ through

$$\psi_D(d_1, a_i, s_i, a_3) = \int u_D(d_1, a_i, s_i, a_3, s_3, d_2^*(d_1, a_i, s_i, a_3, s_3)) \, p_D(s_3 | d_1, a_i, s_i, a_3) \, ds_3. \tag{9}$$

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees.

Next, they compute the expected utility $\psi_D(d_1, a_i, s_i)$ for each (d_1, a_i, s_i) through

$$\psi_D(d_1, a_i, s_i) = \int \psi_D(d_1, a_i, s_i, a_3) p_D(a_3 | d_1, a_i, s_i)\, da_3. \tag{10}$$

They then find the expected utility $\psi_D(d_1, a_i)$ for each (d_1, a_i), as

$$\psi_D(d_1, a_i) = \int \psi_D(d_1, a_i, s_i) p_D(s_i | d_1, a_i)\, ds_i, \tag{11}$$

and their expected utility for each $d_1 \in \mathcal{D}_1$ using their predictive distribution $p_D(a|d_1)$

$$\psi_D(d_1) = \int \psi_D(d_1, a_i) p_D(a_i | d_1)\, da_i. \tag{12}$$

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees.

- Finally, the defender finds their maximum utility decision as $d_1^* = \arg\max_{d_1 \in \mathcal{D}_1} \psi_D(d_1)$.
- Thus, the defender's optimal strategy is to first choose d_1^* and then, after having observed a_i, s_i, a_3 and s_3, choose action $d_2^*(d_1^*, a_i, s_i, a_3, s_3)$.

The above analysis requires the defender to elicit $p_D(a_3|d_1, a_i, s_i)$ and $p_D(a_i|d_1)$.

- A_i being non-strategic, $p_D(a_i|d_1)$ can be elicited using historical data/research on employee behavior.

ARA with segmented employees.

- Finally, the defender finds their maximum utility decision as $d_1^* = \arg\max_{d_1 \in \mathcal{D}_1} \psi_D(d_1)$.
- Thus, the defender's optimal strategy is to first choose d_1^* and then, after having observed a_i, s_i, a_3 and s_3, choose action $d_2^*(d_1^*, a_i, s_i, a_3, s_3)$.

The above analysis requires the defender to elicit $p_D(a_3|d_1, a_i, s_i)$ and $p_D(a_i|d_1)$.

- A_i being non-strategic, $p_D(a_i|d_1)$ can be elicited using historical data/research on employee behavior.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees.

Eliciting $p_D(a_3|d_1, a_i, s_i)$ is less straightforward.

- Can be elicited by modeling attacker's strategic thinking.
- To elicit it, the defender must assess $U_A(a_3, s_3, d_2)$, $P_A(s_3|a_3, d_1)$ and $P_A(d_2|d_1, a_3, s_3)$.
- Once elicited, the defender solves the attacker's decision problem using backward induction - similar to how they solved their own decision problem.

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

ARA with segmented employees - simultaneous actions!

- In the previous model, the employees have been assumed to act in a sequential manner and only two in any given game.
- But scenarios where any two or all three types of employees act simultaneously can also be easily modelled using ARA.

ARA with segmented employees - simultaneous actions!

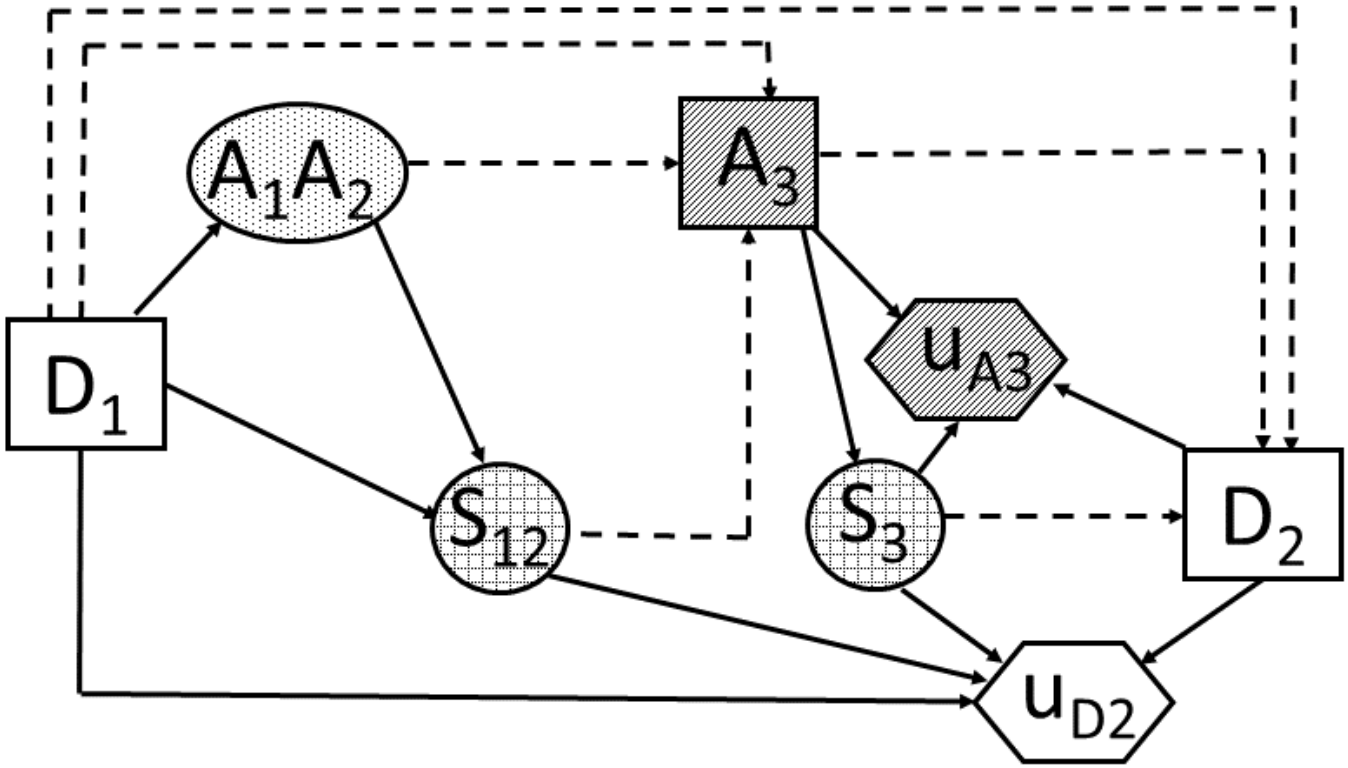
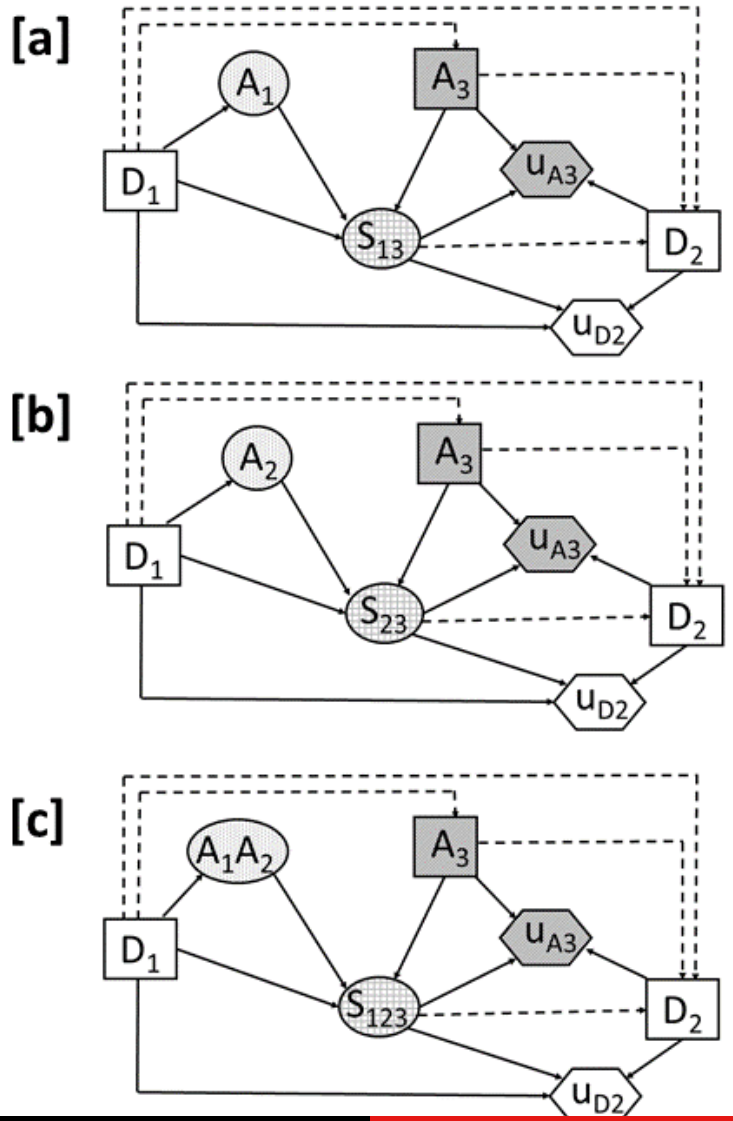


Figure: MAID for the game where A_1 and A_2 act simultaneously followed by A_3.

ARA with segmented employees - simultaneous actions!



Example: simple defend-attack-defend model

We consider an insider threat scenario for an organization dealing with information/data collection.

- It already has its sites and IT systems protected so that only authorized personnel are able to access them.
- However, anticipating attacks, the organization is considering implementing an additional security layer to defend itself.
- The defensive actions (D_1) under consideration are:
  1. anomaly detection/data provenance tools;
  2. information security measures and employee training;
  3. carrying out random audits.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Example: simple defend-attack-defend model

- The malicious insider's aim could be financial fraud, data theft, espionage or whistle blowing.
- Regardless, we assume that the attacker's options (A) refer to its scale, say *small*, *medium* or *large*.
- Assume that the attack will either fully succeed (S) or fail (F).
- We assume that once the attack is detected, the organisation can choose to carry out one of the following defensive actions (D_2):
  1. major upgrade of defenses;
  2. minor upgrade of defenses; or
  3. no upgrade.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Example: simple defend-attack-defend model

Assessing defender's utility function $u_D(d_1, a, s, d_2)$.

- We assume that u_D aggregates the monetary costs $c(d_1)$ and $c(d_2)$ associated with actions d_1 and d_2 respectively and the monetized perceived utilities associated with every (a, s) combination through

$$u_D(d_1, a, s, d_2) = c(d_1) + c(d_2) + u(a, s). \tag{13}$$

- The costs and perceived utilities are scaled from -100 to 100.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Example: simple defend-attack-defend model

Assessing defender's utility function $u_D(d_1, a, s, d_2)$.

Table: Costs associated with defensive actions d_1 and d_2.

d_1	$c(d_1)$	d_2	$c(d_2)$
Anom. det. & Data prov.	-100	Major upgrade	-100
Info. Sec.& train.	-60	Minor upgrade	-25
Random audits	-50	No upgrade	0

Table: Monetized Perceived Utility for Every Combination (a, s).

a	s	$u(a, s)$
Small	Success	-25
Small	Fail	30
Medium	Success	-50
Medium	Fail	60
Large	Success	-100
Large	Fail	80

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Example: simple defend-attack-defend model

Eliciting $p_D(a|d_1)$

Table: $p_D(a|d_1)$ elicited by defender using their beliefs

D_1	A = small	A = med.	A = large
Anom. det. & Data prov.	0.8	0.15	0.05
Info. sec. & train.	0.2	0.6	0.2
Random audits	0.5	0.4	0.1

Table: $P_D(a|d_1)$ elicited by defender modeling the strategic analysis of the attacker

D_1	A = small	A = med.	A = large
Anom.det. & Data prov.	0.112	0.102	0.786
Info.sec. & train.	0.706	0.132	0.162
Random audits	0.399	0.119	0.482

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Example: simple defend-attack-defend model

Eliciting $p_D(a|d_1)$

Table: $p_D(a|d_1)$ elicited by defender using their beliefs

D_1	A = small	A = med.	A = large
Anom. det. & Data prov.	0.8	0.15	0.05
Info. sec. & train.	0.2	0.6	0.2
Random audits	0.5	0.4	0.1

Table: $P_D(a|d_1)$ elicited by defender modeling the strategic analysis of the attacker

D_1	A = small	A = med.	A = large
Anom.det. & Data prov.	0.112	0.102	0.786
Info.sec. & train.	0.706	0.132	0.162
Random audits	0.399	0.119	0.482

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Example: simple defend-attack-defend model

Defender's expected utility $\psi_D(d_1)$ for each d_1 using Eq. (3).

Table: How $\psi_D(d_1)$ changes based on the choice of $p_D(a|d_1)$!

D_1	when $p_D(a\|d_1)$ - beliefs	when $p_D(a\|d_1)$ -strategic thinking
Anom.det. & Data prov.	$\psi_D = -69.005$	$\psi_D = -36.115$
Info.sec. & train.	$\boldsymbol{\psi_D = -29}$	$\psi_D = -39.051$
Random audits	$\psi_D = -39.75$	$\boldsymbol{\psi_D = -34.566}$

- Eliciting $p_D(a|d_1)$ differently can lead to different optimal choices (as one would expect)!
- Also true for any of the other subjective choices - utilties and probabilities - involved.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Summary & Discussion!

- ARA can be a practical tool for strategic decision making!
- Insider threat is a serious security risk and ARA model had not yet been developed!

Further Work:

- Bounded rationality.
- Evolving defend-attack-defend scenarios - using MDP.
- Robustness w.r.t priors and utilities. ARA can be based entirely on subjective beliefs (and NO DATA)!

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Summary & Discussion!

- ARA can be a practical tool for strategic decision making!
- Insider threat is a serious security risk and ARA model had not yet been developed!

Further Work:

- Bounded rationality.
- Evolving defend-attack-defend scenarios - using MDP.
- Robustness w.r.t priors and utilities. ARA can be based entirely on subjective beliefs (and NO DATA)!

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

We are hiring! (*www.waikato.ac.nz*)

THE UNIVERSITY OF WAIKATO

Lecturer/Senior Lecturer in Statistics

School of Computing and Mathematical Sciences
Division of Health, Engineering, Computing and Science

We are seeking to appoint a Lecturer or Senior Lecturer who will contribute to teaching, postgraduate supervision, administration and research. Applicants must have the ability to teach at all levels of study and in particular have broad and in-depth knowledge of Statistics. Preference will be given to candidates with interests in one or more of the following research areas: Bayesian methods, high dimensional/big data or data analytics including statistical machine learning.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato