Developing Optimal Causal Cyber-Defence Agents via Cyber Security Simulation

NSE ML+Security Reading Group

Kim Hammar

kimham@kth.se

Division of Network and Systems Engineering
KTH Royal Institute of Technology

September 23, 2022

The Context and Key Points of the Paper

- The paper proposes an approach to develop automated cyber defence agents
  - Models the security problem with a **structured causal model**.
  - Computes optimal defender strategies through Dynamic Causal Bayesian Optimization (DCBO).
  - Evaluates defender strategies in a cyber security simulation.
  - The simulation is open-sourced and is called "Yawning Titan".

Outline

- **Background**
  - Causal Inference
  - The do-calculus
  - Causal diagrams
  - Causal structured models
- **The Paper**
  - Approach
  - Cyber Security Simulator (Yawning Titan)
  - Causal Model
  - Computing Optimal Defender Interventions
  - Evaluation Results
- **Conclusions**
- **Discussion**
  - Strong points
  - Limitations of the paper
  - Discussion about future work

Causal Inference

- **Causality:** cause-effect relationships among variables. Study causation to make sense of data, to guide actions and policies.
- **Causal inference:**
  - What is the effect on Y if I change X? (interventional question)
  - Example: What is the effect on the security of my system if I update this firewall rule?
  - I changed X and observed Y, what if I had changed Z instead? (counterfactual question)
  - Example: What is the probability that an attack that compromised server N_1 would still have compromised N_1 if I had used two-factor authentication instead of one-factor?
- **How to model causality mathematically?**
  - Probability theory?
  - It is not sufficient! Causation is not correlation!
  - Assume $\mathbb{P}[\text{I am ill}|\text{I went to the hospital}] > 0.5$. Does it mean going to the hospital causes illness?
  - If causation=correlation then our conclusion would be that to avoid illness we should avoid going to the hospital. Nonsense!

Causal Inference

- **Causality:** cause-effect relationships among variables. Study causation to make sense of data, to guide actions and policies.
- **Causal inference:**
  - What is the effect on Y if I change X? (interventional question)
  - Example: What is the effect on the security of my system if I update this firewall rule?
  - I changed X and observed Y, what if I had changed Z instead? (counterfactual question)
  - Example: What is the probability that an attack that compromised server N_1 would still have compromised N_1 if I had used two-factor authentication instead of one-factor?
- **How to model causality mathematically?**
  - Probability theory?
  - It is not sufficient! Causation is not correlation!
  - Assume $\mathbb{P}[\text{I am ill}|\text{I went to the hospital}] > \mathbb{P}[\text{I am ill}]$. Does it mean going to the hospital causes illness?
  - If causation=correlation then our conclusion would be that to avoid illness we should avoid going to the hospital. Nonsense!

Causal Inference

- **Causality:** cause-effect relationships among variables. Study causation to make sense of data, to guide actions and policies.
- **Causal inference:**
  - What is the effect on Y if I change X? (interventional question)
  - Example: What is the effect on the security of my system if I update this firewall rule?
  - I changed X and observed Y, what if I had changed Z instead? (counterfactual question)
  - Example: What is the probability that an attack that compromised server N_1 would still have compromised N_1 if I had used two-factor authentication instead of one-factor?
- **How to model causality mathematically?**
  - Probability theory?
  - It is not sufficient! Causation is not correlation!
  - Assume $\mathbb{P}[\text{I am ill} | \text{I went to the hospital}] > \mathbb{P}[\text{I am ill}]$. Does it mean going to the hospital causes illness?
  - If causation=correlation then our conclusion would be that to avoid illness we should avoid going to the hospital. Nonsense!

Causal Inference

- **Causality:** cause-effect relationships among variables. Study causation to make sense of data, to guide actions and policies.
- **Causal inference:**
  - What is the effect on Y if I change X? (interventional question)
  - Example: What is the effect on the security of my system if I update this firewall rule?
  - I changed X and observed Y, what if I had changed Z instead? (counterfactual question)
  - Example: What is the probability that an attack that compromised server N_1 would still have compromised N_1 if I had used two-factor authentication instead of one-factor?
- **How to model causality mathematically?**
  - Probability theory?
  - It is not sufficient! Causation is not correlation!
  - Assume $\mathbb{P}[\text{I am ill}|\text{I went to the hospital}] > \mathbb{P}[\text{I am ill}]$. Does it mean going to the hospital causes illness?
  - If causation=correlation then our conclusion would be that to avoid illness we should avoid going to the hospital. Nonsense!

Causal Inference

- **Causality:** cause-effect relationships among variables. Study causation to make sense of data, to guide actions and policies.
- **Causal inference:**
  - What is the effect on Y if I change X? (interventional question)
  - Example: What is the effect on the security of my system if I update this firewall rule?
  - I changed X and observed Y, what if I had changed Z instead? (counterfactual question)
  - Example: What is the probability that an attack that compromised server N_1 would still have compromised N_1 if I had used two-factor authentication instead of one-factor?
- **How to model causality mathematically?**
  - Probability theory?
  - It is not sufficient! Causation is not correlation!
  - Assume $\mathbb{P}[\text{I am ill}|\text{I went to the hospital}] > \mathbb{P}[\text{I am ill}]$. Does it mean going to the hospital causes illness?
  - If causation=correlation then our conclusion would be that to avoid illness we should avoid going to the hospital. Nonsense!

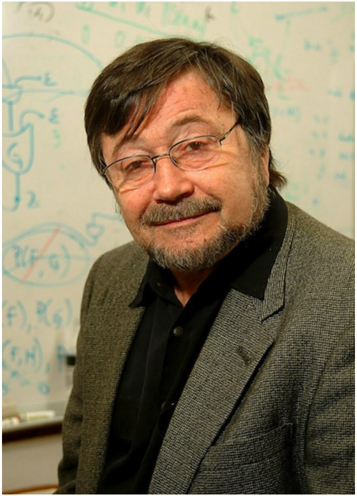
Causal Inference

- **Causality:** cause-effect relationships among variables. Study causation to make sense of data, to guide actions and policies.
- **Causal inference:**
  - What is the effect on Y if I change X? (interventional question)
  - Example: What is the effect on the security of my system if I update this firewall rule?
  - I changed X and observed Y, what if I had changed Z instead? (counterfactual question)
  - Example: What is the probability that an attack that compromised server N_1 would still have compromised N_1 if I had used two-factor authentication instead of one-factor?
- **How to model causality mathematically?**
  - Probability theory?
  - It is not sufficient! Causation is not correlation!
  - Assume $\mathbb{P}[\text{I am ill}|\text{I went to the hospital}] > \mathbb{P}[\text{I am ill}]$. Does it mean going to the hospital causes illness?
  - If causation=correlation then our conclusion would be that to avoid illness we should avoid going to the hospital. Nonsense!

The Causal Revolution (According to Pearl)

The causal revolution: causality has been transformed from a concept shrouded in mystery into a mathematical object with well-defined semantics and well-founded logic. - Pearl

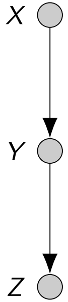
- **The "new" formal framework for causality**:
  - Causal graphs
  - Structured causal models (SCMs)
  - The do-calculus



Judea Pearl. Turing award winner 2011.

The Formal Framework of Causality

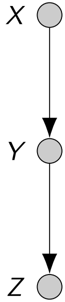
- **If causation is not correlation, then what is it?**:
  - We know that $\mathbb{P}[Y|X] \neq \mathbb{P}[Y] \;\not\Rightarrow\; X$ causes Y.
  - To denote the causal effect on Y when setting $X = x$, we use $\mathbb{P}[Y|do(X = x)]$.
  - $do(X = x)$ is the *do-operator*, representing an intervention on X.
- **The do-calculus**
  - An axiomatic system for calculating interventional distributions $\mathbb{P}[Y|do(X = x)]$.
- **Causal graphs**
  - A probabilistic graphical model.
  - A directed acyclic graph that encodes causal relationships.
- **Structured causal models (SCMs)**
  - An SCM encodes relationships among variables
  - Defined by the tuple $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$.
  - $\boldsymbol{U}$: exogeneous variables, $\boldsymbol{V}$: endogeneous variables, $\boldsymbol{F}$: functions that define causal relationships.



Causal graph.

The Formal Framework of Causality

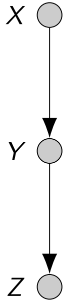
- **If causation is not correlation, then what is it?**:
  - We know that $\mathbb{P}[Y|X] \neq \mathbb{P}[Y] \not\Longrightarrow X$ causes Y.
  - To denote the causal effect on Y when setting $X = x$, we use $\mathbb{P}[Y|do(X = x)]$.
  - $do(X = x)$ is the *do-operator*, representing an intervention on X.
- **The do-calculus**
  - An axiomatic system for calculating interventional distributions $\mathbb{P}[Y|do(X = x)]$.
- **Causal graphs**
  - A probabilistic graphical model.
  - A directed acyclic graph that encodes causal relationships.
- **Structured causal models (SCMs)**
  - An SCM encodes relationships among variables
  - Defined by the tuple $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$.
  - $\boldsymbol{U}$: exogeneous variables, $\boldsymbol{V}$: endogeneous variables, $\boldsymbol{F}$: functions that define causal relationships.



Causal graph.

The Formal Framework of Causality

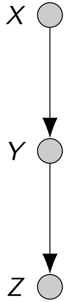
- **If causation is not correlation, then what is it?**:
  - We know that $\mathbb{P}[Y|X] \neq \mathbb{P}[Y] \not\Rightarrow X$ causes Y.
  - To denote the causal effect on Y when setting $X = x$, we use $\mathbb{P}[Y|do(X = x)]$.
  - $do(X = x)$ is the *do-operator*, representing an intervention on X.
- **The do-calculus**
  - An axiomatic system for calculating interventional distributions $\mathbb{P}[Y|do(X = x)]$.
- **Causal graphs**
  - A probabilistic graphical model.
  - A directed acyclic graph that encodes causal relationships.
- **Structured causal models (SCMs)**
  - An SCM encodes relationships among variables
  - Defined by the tuple $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$.
  - $\boldsymbol{U}$: exogeneous variables, $\boldsymbol{V}$: endogeneous variables, $\boldsymbol{F}$: functions that define causal relationships.



Causal graph.

The Formal Framework of Causality

- **If causation is not correlation, then what is it?**:
  - We know that $\mathbb{P}[Y|X] \neq \mathbb{P}[Y] \;\not\Longrightarrow\; X$ causes Y.
  - To denote the causal effect on Y when setting $X = x$, we use $\mathbb{P}[Y|do(X = x)]$.
  - $do(X = x)$ is the *do-operator*, representing an intervention on X.
- **The do-calculus**
  - An axiomatic system for calculating interventional distributions $\mathbb{P}[Y|do(X = x)]$.
- **Causal graphs**
  - A probabilistic graphical model.
  - A directed acyclic graph that encodes causal relationships.
- **Structured causal models (SCMs)**
  - An SCM encodes relationships among variables
  - Defined by the tuple $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$.
  - $\boldsymbol{U}$: exogeneous variables, $\boldsymbol{V}$: endogeneous variables, $\boldsymbol{F}$: functions that define causal relationships.



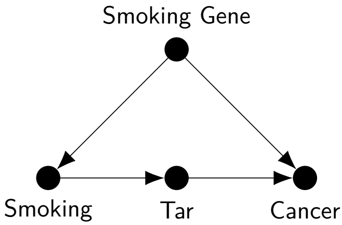
Causal graph.

Causal Diagram Example

- **Does smoking cause cancer?**
  - Historically one of the most debated questions in science. R.A Fisher argued for no (lifetime smoker).
  - Now we know that the answer is yes.
  - That smoking and cancer are correlated was shown early.
  - But how do you show that smoking causes cancer? What if there is a gene that causes cancer and also makes you love cigarettes?
  - We can answer this question by **randomized controlled trials.** Problem: involves forcing people to smoke for 40+ years (not ethical!).
  - Conclusion: Inferring causal relationships from data alone is very difficult. Generally need a causal model to express causal assumptions to make sense of the data.

Causal Diagram Example

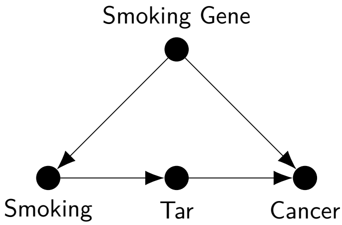
- **Does smoking cause cancer?**
  - Historically one of the most debated questions in science. R.A Fisher argued for no (lifetime smoker).
  - Now we know that the answer is yes.
  - That smoking and cancer are correlated was shown early.
  - But how do you show that smoking causes cancer? What if there is a gene that causes cancer and also makes you love cigarettes?
  - We can answer this question by **randomized controlled trials.** Problem: involves forcing people to smoke for 40+ years (not ethical!).
  - Conclusion: Inferring causal relationships from data alone is very difficult. Generally need a causal model to express causal assumptions to make sense of the data.



Possible causal graph of smoking (smoking gene is a confounder).

Causal Diagram Example

- **Does smoking cause cancer?**
  - Historically one of the most debated questions in science. R.A Fisher argued for no (lifetime smoker).
  - Now we know that the answer is yes.
  - That smoking and cancer are correlated was shown early.
  - But how do you show that smoking causes cancer? What if there is a gene that causes cancer and also makes you love cigarettes?
  - We can answer this question by **randomized controlled trials.** Problem: involves forcing people to smoke for 40+ years (not ethical!).
  - Conclusion: Inferring causal relationships from data alone is very difficult. Generally need a causal model to express causal assumptions to make sense of the data.

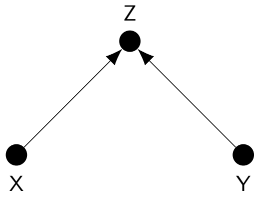


Possible causal graph of smoking (smoking gene is a confounder).

Structured Causal Model Example

- Let Z model the salary of an employee, X the number of years of education the employee has, and Y the number of years in the profession the employee has.
- An example structured causal model (SCM) M to model the causal effects of X and Y on Z:

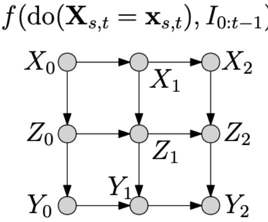
$$\begin{aligned} M &= \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle && \text{SCM} \\ \boldsymbol{U} &= \{X, Y\} && \text{exogeneous variables} \\ \boldsymbol{V} &= \{Z\} && \text{endogeneous variables} \\ \boldsymbol{F} &= \{f_Z\} && \text{causal relations} \\ f_Z : Z &= 2X + 3Y \end{aligned}$$



Causal graph for the example SCM.

Optimization and Decision Problems Based on SCMs

Given SCM $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$, where $\boldsymbol{V}$ is partitioned into a set of control variables $\boldsymbol{X}$, a set of covariates $\boldsymbol{Z}$, and an outcome variable Y, decide which variables in $\boldsymbol{X}$ to intervene on and their values to achieve the desired effect on Y.



- If the outcome variable Y is a quantity to be minimized or maximized, a causal decision problem can be formulated as a causal optimization problem:

$$\boldsymbol{X}_s^*, \boldsymbol{x}_s^* = \underset{\boldsymbol{X}_s \in \mathcal{P}(\boldsymbol{X}), \boldsymbol{x}_s \in dom(\boldsymbol{X}_s)}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$$

- If all control variables $X_i \in \boldsymbol{X}$ are discrete, it can be formulated as a causal multi-armed bandit.

Approach of the Paper

1. **Design** a cyber simulator, which will be used for experiments.
2. **Model** a scenario in the simulator with an SCM (Structured Causal Model)
3. **Compute** optimal defender interventions for the SCM through DCBO
4. **Evaluate** the convergence rate of DCBO against baselines.

Approach of the Paper

1. **Design** a cyber simulator, which will be used for experiments.
2. **Model** a scenario in the simulator with an SCM (Structured Causal Model)
3. **Compute** optimal defender interventions for the SCM through DCBO
4. **Evaluate** the convergence rate of DCBO against baselines.

Approach of the Paper

1. **Design** a cyber simulator, which will be used for experiments.
2. **Model** a scenario in the simulator with an SCM (Structured Causal Model)
3. **Compute** optimal defender interventions for the SCM through DCBO
4. **Evaluate** the convergence rate of DCBO against baselines.

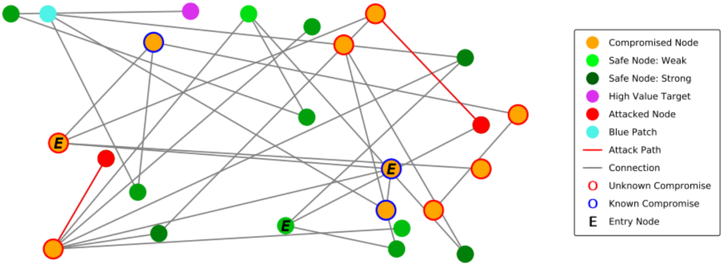
Approach of the Paper

1. **Design** a cyber simulator, which will be used for experiments.
2. **Model** a scenario in the simulator with an SCM (Structured Causal Model)
3. **Compute** optimal defender interventions for the SCM through DCBO
4. **Evaluate** the convergence rate of DCBO against baselines.

Cyber Security Simulator: Yawning Titan

Yawning Titan is an abstract, highly flexible, cyber security simulator that is capable of simulating a range of cyber security scenarios.

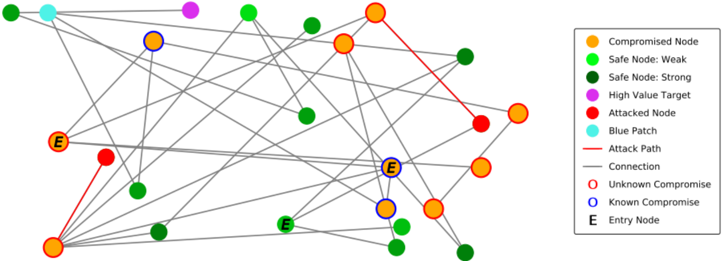
- Network is represented as a graph.
- Each node in the graph corresponds to a machine and has:
  - A vulnerability score
  - An isolation status
  - A compromised status
  - A discovered status



Cyber Security Simulator: Yawning Titan

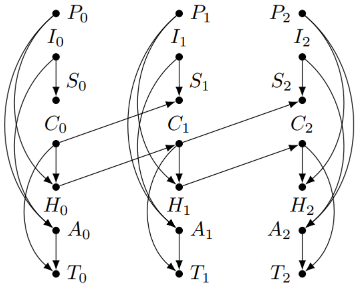
- Examples of **defender actions** in the simulation:
  - reduce the vulnerability of nodes
  - scan the network for intrusions
  - reset a node back to its initial state
  - deploy deceptive nodes
- Attacker can attack nodes.
- **Probablistic model of attack success with attacker skill level** RS **and vulnerability score** $vuln(V_i)$:

$$\frac{100 \times RS^2}{RS + (1 - vuln(V_i))} \geq u \sim \mathcal{U}(0, 100)$$



Causal Model

- Discrete time-steps $t = 0, 1, \ldots,$. At each step, both agents take one action each.
- $T_t = C_t + A_t$ is the total cost of attacks (C_t) and defensive actions (A_t) at time-step t (minimization objective).
- S_t is the attack surface and H_t is the likelihood of further compromise at time-step t.
- $P_t, I_t \in [0, 1]$ are probabilities of defender restoring and isolating a node in S_t at time-step t, respectively.



Causal diagram of the security scenario. Assumes no unobserved confounders.

The Structured Causal Model (SCM)

Recall that an SCM $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$ contains exogeneous variables ($\boldsymbol{U}$, defender actions in this case), endogeneous variables $\boldsymbol{V}$ (e.g. cost and compromised nodes), and causal relationships ($\boldsymbol{F}$).

$$
\begin{aligned}
P_t &= p_t(RES) \\
I_t &= p_t(ISO) \\
S_t &= |K_t^c \cap \phi_t^c| \\
C_t &= \left(\sum_{n=1}^{n=N} \Gamma_c [n \in K_t] \right)^{1.5} \\
H_t &= \sum_{n \in K_t} \sum_{v \in N^+(n)} (vuln(v)[v \notin \phi_t]) \\
A_t &= \begin{cases} \Gamma_{RES} & \mathcal{A}_t = RES \\ \Gamma_{ISO} & \mathcal{A}_t = ISO \end{cases} \\
T_t &= C_t + A_t
\end{aligned}
$$

The Estimated Structured Causal Model (SCM)

- The SCM $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$ presented on the previous slide depends on information that is unknown to the defender, such as the compromised nodes and the vulnerability scores.
- The defender estimates the SCM by placing Gaussian process estimators on all functions $f_i \in \boldsymbol{F}$:

$$
\begin{aligned}
P_t &= f_P(t) + \epsilon_P \\
I_t &= f_I(t) + \epsilon_I \\
S_t &= f_S(C_{t-1}, I_t) + \epsilon_S \\
C_t &= f_C(H_{t-1}) + \epsilon_C \\
H_t &= f_H(P_t, C_t) + \epsilon_H \\
A_t &= f_A(P_t, C_t) + \epsilon_A \\
T_t &= f_T(C_t, A_t) + \epsilon_T
\end{aligned}
$$

- The Gaussian processes are fitted based on data collected from running simulations with a random defender agent.

Computing Optimal Defender Interventions

- Given the estimated SCM $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$, the optimal defender interventions $do(\boldsymbol{X}_1 = \boldsymbol{x}_1), do(\boldsymbol{X}_2 = \boldsymbol{x}_2), \ldots$ (here $\boldsymbol{X}_{s,t} \subset \{P_t, I_t\}$ and $\boldsymbol{x}_{s,t} \in [0,1]^{|\boldsymbol{X}_{s,t}|}$) are obtained by solving the following optimization problem:

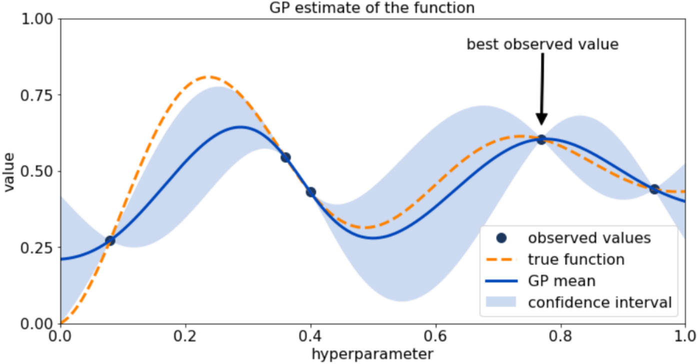
$$\boldsymbol{X}^*_{t,s}, \boldsymbol{x}^*_{s,t} = \underset{\boldsymbol{X}_{s,t} \in \mathcal{P}(\boldsymbol{X}_t), \boldsymbol{x}_{s,t} \in dom(\boldsymbol{X}_{s,t})}{\arg\min} \mathbb{E}[T_t | do(\boldsymbol{X}_{s,t} = \boldsymbol{x}_{s,t}), \mathbb{1}_{t>0} \cdot I_{0:t-1}]$$

- Focus on a network of 10 nodes with 25 time-steps for making interventions.

- Three algorithms are considered for solving the above optimization problem: DCBO, CBO, and BO.

Bayesian Optimization (BO)

- Bayesian Black-box optimization method
- Finds the optimum of a function $f(\boldsymbol{x})$ on some compact set $\boldsymbol{X}$
- Uses a probabilistic model of $f(\boldsymbol{x})$ (typically a Gaussian process)
- The model of GP is used to optimizing an acquisition function α to decide where in $\boldsymbol{X}$ to evaluate f.



Bayesian Optimization

Algorithm 1 Bayesian Optimization.

procedure BAYESIAN OPTIMIZATION
    $p(f(\boldsymbol{x})) = \mathcal{GP}(f; \mu, K)$
    $\boldsymbol{x}_1 \sim \mathcal{U}(\boldsymbol{X})$
    $\mathcal{D} = \{(\boldsymbol{x}_1, f(\boldsymbol{x}_1))\}$
    for $n = 1, 2, \ldots$ **do**
        $\boldsymbol{x}_{n+1} = \arg\max_{\boldsymbol{x}} \alpha_{EI}(\boldsymbol{x}, \mathcal{D})$
        $\mathcal{D} = \mathcal{D} \cup \{(\boldsymbol{x}_{n+1}, f(\boldsymbol{x}_{n+1}))\}$
        $p(f(\boldsymbol{x})) = p(f(\boldsymbol{x})|\mathcal{D}) = \mathcal{GP}(f; \mu_{f|\mathcal{D}}, K_{f|\mathcal{D}})$
    end for
    return $\max \mathcal{D}$
end procedure

Causal Bayesian Optimization (CBO)

- **CBO generalizes BO** to the case where causal information about the optimization problem is available in an SCM M:
  1. Causal optimization objective (selecting both which variables to intervene on and their values)

  $$\boldsymbol{X}_s^*, \boldsymbol{x}_s^* = \underset{\boldsymbol{X}_s \in \mathcal{P}(\boldsymbol{X}), \boldsymbol{x}_s \in dom(\boldsymbol{X}_s)}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$$

  2. Causal surrogate model (extends the GP to estimate interventional distributions $\mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$ based on both observational and interventional data.)
  3. Causal acquisition function. (acquisition function which only considers intervention sets in the pruned version of $dom(\boldsymbol{X}_s)$, i.e. intervention sets that are POMIS)
  4. Integrates both observational and interventional data. (CBO evaluates interventions on subsets $\boldsymbol{X}_s \subseteq \boldsymbol{X}$, including the empty intervention set $\boldsymbol{X}_s = \emptyset$, which yields observational data that is used to update the causal GP through the do-calculus.)

Causal Bayesian Optimization (CBO)

- **CBO generalizes BO** to the case where causal information about the optimization problem is available in an SCM M:
  1. Causal optimization objective (selecting both which variables to intervene on and their values)

$$\boldsymbol{X}_s^*, \boldsymbol{x}_s^* = \underset{\boldsymbol{X}_s \in \mathcal{P}(\boldsymbol{X}), \boldsymbol{x}_s \in dom(\boldsymbol{X}_s)}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$$

  2. Causal surrogate model (extends the GP to estimate interventional distributions $\mathbb{E}[Y|do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$ based on both observational and interventional data.)
  3. Causal acquisition function. (acquisition function which only considers intervention sets in the pruned version of $dom(\boldsymbol{X}_s)$, i.e. intervention sets that are POMIS)
  4. Integrates both observational and interventional data. (CBO evaluates interventions on subsets $\boldsymbol{X}_s \subseteq \boldsymbol{X}$, including the empty intervention set $\boldsymbol{X}_s = \emptyset$, which yields observational data that is used to update the causal GP through the do-calculus.)

Causal Bayesian Optimization (CBO)

- **CBO generalizes BO** to the case where causal information about the optimization problem is available in an SCM M:
  1. Causal optimization objective (selecting both which variables to intervene on and their values)

$$\boldsymbol{X}_s^*, \boldsymbol{x}_s^* = \underset{\boldsymbol{X}_s \in \mathcal{P}(\boldsymbol{X}), \boldsymbol{x}_s \in dom(\boldsymbol{X}_s)}{\arg\min} \mathbb{E}[Y|do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$$

  2. Causal surrogate model (extends the GP to estimate interventional distributions $\mathbb{E}[Y|do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$ based on both observational and interventional data.)
  3. Causal acquisition function. (acquisition function which only considers intervention sets in the pruned version of $dom(\boldsymbol{X}_s)$, i.e. intervention sets that are POMIS)
  4. Integrates both observational and interventional data. (CBO evaluates interventions on subsets $\boldsymbol{X}_s \subseteq \boldsymbol{X}$, including the empty intervention set $\boldsymbol{X}_s = \emptyset$, which yields observational data that is used to update the causal GP through the do-calculus.)

Causal Bayesian Optimization (CBO)

- **CBO generalizes BO** to the case where causal information about the optimization problem is available in an SCM M:
  1. Causal optimization objective (selecting both which variables to intervene on and their values)

$$\boldsymbol{X}_s^*, \boldsymbol{x}_s^* = \underset{\boldsymbol{X}_s \in \mathcal{P}(\boldsymbol{X}), \boldsymbol{x}_s \in dom(\boldsymbol{X}_s)}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$$

  2. Causal surrogate model (extends the GP to estimate interventional distributions $\mathbb{E}[Y|do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$ based on both observational and interventional data.)
  3. Causal acquisition function. (acquisition function which only considers intervention sets in the pruned version of $dom(\boldsymbol{X}_s)$, i.e. intervention sets that are POMIS)
  4. Integrates both observational and interventional data. (CBO evaluates interventions on subsets $\boldsymbol{X}_s \subseteq \boldsymbol{X}$, including the empty intervention set $\boldsymbol{X}_s = \emptyset$, which yields observational data that is used to update the causal GP through the do-calculus.)

Causal Bayesian Optimization (CBO)

- **CBO generalizes BO** to the case where causal information about the optimization problem is available in an SCM M:
  1. Causal optimization objective (selecting both which variables to intervene on and their values)

$$\boldsymbol{X}_s^*, \boldsymbol{x}_s^* = \underset{\boldsymbol{X}_s \in \mathcal{P}(\boldsymbol{X}), \boldsymbol{x}_s \in dom(\boldsymbol{X}_s)}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$$

  2. Causal surrogate model (extends the GP to estimate interventional distributions $\mathbb{E}[Y | do(\boldsymbol{X}_s = \boldsymbol{x}_s)]$ based on both observational and interventional data.)
  3. Causal acquisition function. (acquisition function which only considers intervention sets in the pruned version of $dom(\boldsymbol{X}_s)$, i.e. intervention sets that are POMIS)
  4. Integrates both observational and interventional data. (CBO evaluates interventions on subsets $\boldsymbol{X}_s \subseteq \boldsymbol{X}$, including the empty intervention set $\boldsymbol{X}_s = \emptyset$, which yields observational data that is used to update the causal GP through the do-calculus.)

Dynamic Causal Bayesian Optimization (DCBO)

- **DCBO generalizes CBO** to consider temporal dynamics of an SCM:
  1. Dynamic structured causal model causal model with causal structure across time $M_t = \langle \boldsymbol{U}_{0:t}, \boldsymbol{V}_{0:t}, \boldsymbol{F}_{0:t} \rangle$ where $0:t$ denotes the union of the corresponding variables or functions up to time t and $\mathcal{G}_{0:T}$ is a causal dynamic Bayesian network.
  2. Dynamic causal optimization objective (optimization objective that accounts for past interventions)

     $$\boldsymbol{X}^*_{s,t}, \boldsymbol{x}^*_{s,t} = \underset{\boldsymbol{X}_{s,t} \in \mathcal{P}(\boldsymbol{X}_t), \boldsymbol{x}_s \in dom(\boldsymbol{X}_{s,t})}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_{s,t} = \boldsymbol{x}_s), \mathbb{1}_{t>0} \cdot I_{0:t-1}]$$

     where $I_{0:t-1} = \bigcup_{i=0}^{t-1} do(\boldsymbol{X}^*_{s,i} = \boldsymbol{x}^*_{s,i})$ denotes previous interventions and $\mathbb{1}_{t>0}$ is the indicator function.
  3. Dynamic causal surrogate model. (extends the Causal Gaussian Process to estimate interventional distributions conditioned on past interventions)

Dynamic Causal Bayesian Optimization (DCBO)

- **DCBO generalizes CBO** to consider temporal dynamics of an SCM:
  1. Dynamic structured causal model causal model with causal structure across time $M_t = \langle \boldsymbol{U}_{0:t}, \boldsymbol{V}_{0:t}, \boldsymbol{F}_{0:t} \rangle$ where $0:t$ denotes the union of the corresponding variables or functions up to time t and $\mathcal{G}_{0:T}$ is a causal dynamic Bayesian network.
  2. Dynamic causal optimization objective (optimization objective that accounts for past interventions)

     $$\boldsymbol{X}^*_{s,t}, \boldsymbol{x}^*_{s,t} = \underset{\boldsymbol{X}_{s,t} \in \mathcal{P}(\boldsymbol{X}_t), \boldsymbol{x}_s \in dom(\boldsymbol{X}_{s,t})}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_{s,t} = \boldsymbol{x}_s), \mathbb{1}_{t>0} \cdot I_{0:t-1}]$$

     where $I_{0:t-1} = \bigcup_{i=0}^{t-1} do(\boldsymbol{X}^*_{s,i} = \boldsymbol{x}^*_{s,i})$ denotes previous interventions and $\mathbb{1}_{t>0}$ is the indicator function.
  3. Dynamic causal surrogate model. (extends the Causal Gaussian Process to estimate interventional distributions conditioned on past interventions)

Dynamic Causal Bayesian Optimization (DCBO)

- **DCBO generalizes CBO** to consider temporal dynamics of an SCM:
  1. Dynamic structured causal model causal model with causal structure across time $M_t = \langle \boldsymbol{U}_{0:t}, \boldsymbol{V}_{0:t}, \boldsymbol{F}_{0:t} \rangle$ where $0:t$ denotes the union of the corresponding variables or functions up to time t and $\mathcal{G}_{0:T}$ is a causal dynamic Bayesian network.
  2. Dynamic causal optimization objective (optimization objective that accounts for past interventions)

     $$\boldsymbol{X}^*_{s,t}, \boldsymbol{x}^*_{s,t} = \underset{\boldsymbol{X}_{s,t} \in \mathcal{P}(\boldsymbol{X}_t), \boldsymbol{x}_s \in dom(\boldsymbol{X}_{s,t})}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_{s,t} = \boldsymbol{x}_s), \mathbb{1}_{t>0} \cdot I_{0:t-1}]$$

     where $I_{0:t-1} = \bigcup_{i=0}^{t-1} do(\boldsymbol{X}^*_{s,i} = \boldsymbol{x}^*_{s,i})$ denotes previous interventions and $\mathbb{1}_{t>0}$ is the indicator function.
  3. Dynamic causal surrogate model. (extends the Causal Gaussian Process to estimate interventional distributions conditioned on past interventions)

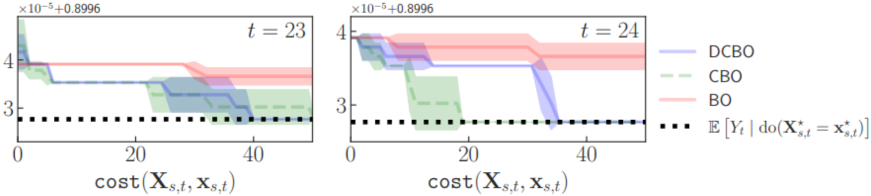
Dynamic Causal Bayesian Optimization (DCBO)

- **DCBO generalizes CBO** to consider temporal dynamics of an SCM:
  1. Dynamic structured causal model causal model with causal structure across time $M_t = \langle \boldsymbol{U}_{0:t}, \boldsymbol{V}_{0:t}, \boldsymbol{F}_{0:t} \rangle$ where $0:t$ denotes the union of the corresponding variables or functions up to time t and $\mathcal{G}_{0:T}$ is a causal dynamic Bayesian network.
  2. Dynamic causal optimization objective (optimization objective that accounts for past interventions)

     $$\boldsymbol{X}^*_{s,t}, \boldsymbol{x}^*_{s,t} = \underset{\boldsymbol{X}_{s,t} \in \mathcal{P}(\boldsymbol{X}_t), \boldsymbol{x}_s \in dom(\boldsymbol{X}_{s,t})}{\arg\min} \mathbb{E}[Y | do(\boldsymbol{X}_{s,t} = \boldsymbol{x}_s), \mathbb{1}_{t>0} \cdot I_{0:t-1}]$$

     where $I_{0:t-1} = \bigcup_{i=0}^{t-1} do(\boldsymbol{X}^*_{s,i} = \boldsymbol{x}^*_{s,i})$ denotes previous interventions and $\mathbb{1}_{t>0}$ is the indicator function.
  3. Dynamic causal surrogate model. (extends the Causal Gaussian Process to estimate interventional distributions conditioned on past interventions)

Evaluation Results



Evaluation results.

- CBO and DCBO converge faster than BO by exploiting causal structure.
- CBO actually perform sligthly better than DCBO. This indicates that the temporal structure incorporated in DCBO is not that useful for this particular problem.

Summary and Contributions

- **Summary.** Presents a novel causal optimization approach to compute optimal defender strategies:
  1. Model security problem with a structured causal model
  2. Fit model using Gaussian processes and data from a simulator
  3. Formulate optimal defender interventions as a causal dynamic otimization problem
  4. Solve the optimization problem using causal extensions to Bayesian optimization (CBO, DCBO)
  5. Evaluate obtained defender interventions in simulation.

- **Contributions.**
  1. The approach based on causal optimization and SCM is novel to the cyber domain.
  2. Presents a new cyber security simulator.

Discussion

- **Strong points**
  - The approach is novel and since the causal optimization approach has not previously been explored in the cyber domain, it lays a foundation for future work.
  - Discusses the causal approach in relation to traditional approaches based on control/game/learning/decision theory.
- **Weak points**
  - Static attacker
  - Simplistic defender model
  - Difficult to define the SCM in practice
  - No conclusions can be made from the results other than that DCBO and CBO outperforms BO in an abstract simulation.
- **Discussion points**
  - Myopic?
  - How useful are the simulation results? Effective solutions in simulated environments have been demonstrated for 15+years, are we getting closer to something that can work in practice?