Poster: Using Causal Inference to Extract Hidden Information from Dependency Modelling

Ayodeji Rotibi
Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
rotibiao@cardiff.ac.uk

Neetesh Saxena
Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
saxenan4@cardiff.ac.uk

Pete Burnap
Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
burnapp@cardiff.ac.uk

***Abstract*—Dependency Modelling is an established Probabilistic Risk Analysis method that is frequently used to identify and quantify cyber risks in complex environments, such as Industrial Control Systems. The method is useful for examining the interrelationships between different variables, but the limited data exposure in the modelling restricts its ability to analyse multiple independent variables simultaneously or sequentially. In response to this limitation, we present a new technique that leverages the Bayesian Network method to draw inferences from unrelated events and uncovers hidden insights that Dependency Modelling may overlook. We conducted an evaluation of our proposed technique using lab-generated data that mimics Colonial pipeline operations. Our results demonstrated that the proposed technique exposes previously undetected aspects of the dependency model, providing business and asset owners with a more comprehensive understanding of their cyber risks and facilitating better decision-making. Our technique represents a significant advancement and is the first to apply this inference method to Dependency Modelling.**

***Index Terms*—Cyber risks, Dependency Modelling, Bayesian Network, Variable Elimination**

I. Introduction

The industrial technology landscape is continually evolving, resulting in an increased connection of processes and components that enhance productivity and bottom-line impact. However, this transformation also brings new risks to operational technology (OT) systems and operations, increasing complexity and posing significant cybersecurity challenges [1].

Despite continuous efforts by industries and governments to enhance cybersecurity, major industrial cyber breaches remain as likely today as they did ten years ago. Recent cyber attacks on Colonial Pipeline and JBS Foods have highlighted the consequences of cyber threats and the vulnerabilities of exchanging data and dependencies in enterprise systems [2]. Successful attacks can lead to a complete system failure, emphasising the need to evaluate alternative approaches to mitigate cyber risks in complex systems.

Dependency Modelling (DM) provides a comprehensive framework for establishing links between system events, processes, and dependencies, enabling accurate risk assessments to support informed decision-making and enhance cybersecurity [3]. Despite its capabilities, DM's limitations prevent it from providing sufficient insights to fully understand a system's complexity beyond conventional approaches, highlighting the need for alternative methods.

Contribution: Our proposed technique introduces causal inference into Dependency Modelling (DM) which allows the analysis of multiple independent nodes and accounting for simultaneous or sequential changes within the model. This multi-nodal analysis increases the identification of cyber risks that are synonymous with the tight coupling characteristics phenomena in complex systems where multiple events can fail synchronously. We believe this enhancement positions DM as a preferred method to identify cyber risks in complex environments, including Industrial Control Systems (ICS).

II. Related Work

The potential of Bayesian Networks (BN) as an adaptable and effective tool in handling incomplete or uncertain information has been recognised by researchers [4]. Previous studies, such as [5], [6], have demonstrated the suitability of BN in detecting intrusion and insider threats in system networks, showcasing its efficiency in mitigating cyber risks.

However, the use of BN in identifying cyber risks within large and complex systems, including ICS, remains limited. Existing research does not account for the impact of simultaneous or sequential failure within complex system networks, nor have effective techniques for enhancing cyber risk identification in such systems been proposed.

III. Approach

Our approach utilises Directed Acyclic Graphs (DAGs), which model the conditional dependencies between variables in a probabilistic model. We implement Bayesian Networks (BN) to perform statistical inference on DM and calculate the *conditional probabilities* of unknown variables based on their observed values. While both BN and DM are usually causally constructed, BN assumes that most variables are independent of their preceding variables, whereas DM assumes that all variables may be directly impacted by their predecessors. This property makes BN advantageous, enabling the identification of a subset of preceding parameters for each parameter in

This work has been supported by the Knowledge Economy Skills Scholarships (KESS2) - a major pan-Wales operation supported by European Social Funds (ESF) through the Welsh Government, and Thales UK

turn, which allows us to use Variable Elimination (VE) for causal inference to identify hidden or previously unknown risks within the system [7].

The VE probabilistic inference algorithm calculates the marginal probabilities of a target variable by recursively eliminating irrelevant network variables that do not impact the target variable. This efficiency in handling large and complex BN makes VE the preferred algorithm over others like the Junction Tree (JT) and Monte Carlo Markov Chain.

To retrieve hidden data from the model, we construct an inference query in the form of $P(Y|E = e)$, where Y and E are disjoint variables in the model, and E is an observed variable with a value of e [7].

IV. Validation

We aim to determine if the causal inference technique can reveal changes in the model's sensitivity when considering the combination of multiple independent nodes. The traditional 3-Point Sensitivity (3PS) approach used in DM can only assess the sensitivity impact of a single *leaf* node at a time.

To validate our approach, we employed a case study that mimics Colonial pipeline operations in an Industrial Control System (ICS) environment, named PipelineX. We focused on the communication between the IT and OT networks involved in the shipping process to track product delivery. After receiving an order from a customer, an operator at the enterprise network verifies product availability via the production network before initiating the shipping process. The shipping process generates a trigger to load the product for delivery. Figure 1 illustrates the business process description. Additional information includes the following:

- Remote login to the IT network is available via a secured Virtual Private Network (VPN) infrastructure, managed by the Enterprise Access Control.
- There is no network segmentation infrastructure between the IT and the OT networks.
- A loss of availability on the IT network due to an attack could disrupt production on the OT network.

Our data is an adaptation from an existing manufacturing environment with 67 nodes in the model. Each node has three attributes: name, dependencies (name of parent node), and the percentage probability of being in a desired state. Each node is numbered (ref) from 0 (the goal/root node) to 66. We have included some node names and descriptions in Table I.

TABLE I
NODE WITH REFERENCE NUMBER

Ref	Node Name	Description
0	Secure and Safe Production	This is the goal of the business
9	Enterprise Access Control	Access Policies are implemented
34	Wireless Protocols	Protocols are updated and secured
40	Background Checks	Security check conducted on users
41	Roles and Responsibilities	Clearly defined and assigned
42	Training	Appropriate training conducted
43	Specialised Training	Training specific to functions
44	Security Awareness	Basic requirements for users
45	Security Responsibilities	Assigned and owned
46	Event and Incident Mgt	Logged and reviewed

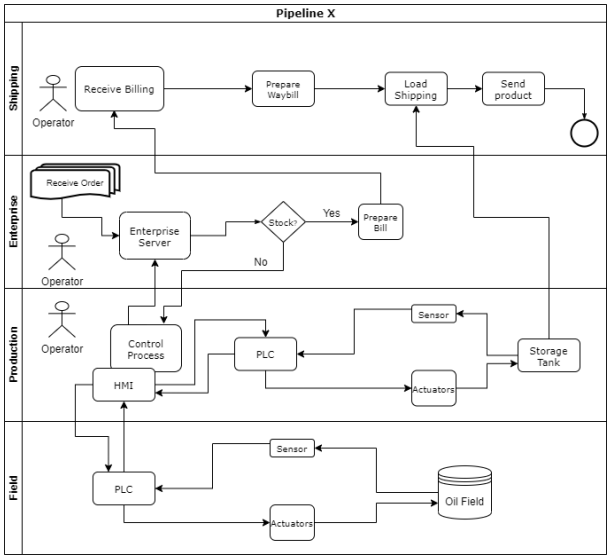

Fig. 1. Shipping Process Flow in Pipeline X

Conventional behaviour expectation dictates that if a node fails due to an event, its probability is set to 0 (or 0%). This event results in a negative impact on the overall model. Conversely, setting a node to 1 (or 100%) due to an improvement in a certain event positively impacts the model. To identify the node with the highest impact (sensitivity), we set each node to 0 and 1, respectively, and performed causal inference to obtain new probability values for the root node. This process is repeated with combinations of two and three nodes.

The resulting sensitivity scores are presented in Tables II, III, and IV, where each table lists the top-5 sensitivity scores. The *Node* column indicates the node number, corresponding to its name in Table I. The *Probability* column displays the current marginal probability of the overall goal. The last two columns reveal the sensitivity values, representing the difference between the marginal probability and the computed probability when the node is turned off (E=0) and when it is turned fully on (E=1). As an example, in Table II, Row 1 Column *E=0* displays the result of 0.16361779 - 0.042172706, while Column *E=1* is derived from *E=1*, i.e 0.167503027 - 0.16361779.

TABLE II
CAUSAL INFERENCE FOR SINGLE EVENT

Node	Probability	E=0	E=1
[40]	0.16361779	0.121445081	0.00388524
[41]	0.16361779	0.121445081	0.00388524
[43]	0.16361779	0.107265991	0.003431626
[44]	0.16361779	0.107265991	0.003431626
[45]	0.16361779	0.107265991	0.003431626

Figure 2A, B and C show the 3PS plots for each table. A short bar indicates low sensitivity, while a longer bar represents higher sensitivity. The colour of each bar corresponds to its influence, where red-coloured bars suggest a negative impact and green-coloured bars exhibit a positive influence. The junction between the bars indicates the sensitivity level

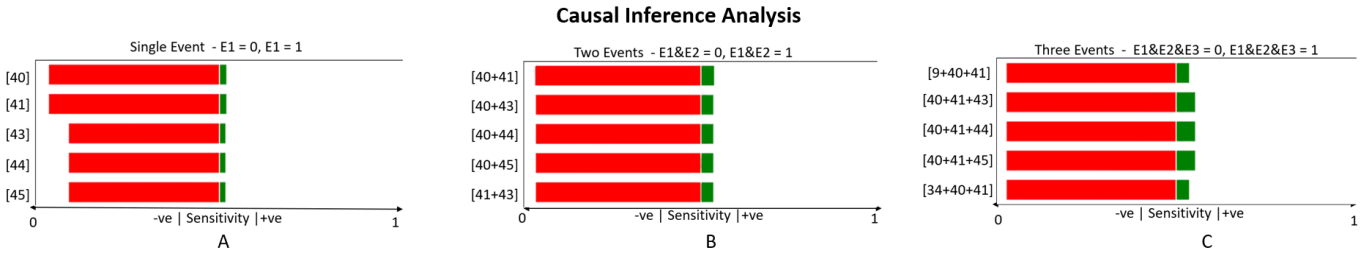

Fig. 2. 3-Point Sensitivity Using Causal Inference Analysis

TABLE III
CAUSAL INFERENCE FOR TWO (COMBINED) EVENTS

Node	Probability	E1=E2=0	E1=E2=1
[40+41]	0.163617787	0.125440583	0.007890686
[40+43]	0.163617787	0.124978108	0.007423034
[40+44]	0.163617787	0.124978108	0.007423034
[40+45]	0.163617787	0.124978108	0.007423034
[41+43]	0.163617787	0.124978108	0.007423034

TABLE IV
CAUSAL INFERENCE FOR THREE (COMBINED) EVENTS

Node	Probability	E1=E2=E3=0	E1=E2=E3=1
[9+40+41]	0.163617787	0.125560776	0.007904915
[40+41+43]	0.163617787	0.125559031	0.011537933
[40+41+44]	0.163617787	0.125559031	0.011537933
[40+41+45]	0.163617787	0.125559031	0.011537933
[34+40+41]	0.163617787	0.125553184	0.007904016

concerning the overall goal or how far it is from the probability of the goal. We observed that setting the probabilities of three nodes to zero (E1 = E2 = E3 = 0) resulted in longer red bars than only setting the probabilities of two nodes to zero (PE1 = E2 = 0). This indicates that the model is more sensitive, with a higher negative impact with more nodes. Conversely, setting the probabilities of three nodes to one (E1 = E2 = E3 = 1) resulted in a higher positive influence, with longer green bars than only setting the probabilities of two nodes to one (E1 = E2 = 1).

From the information obtained in the model, we conducted a frequency analysis to identify the nodes with the most influence, which is a function of how many times they occur in combination with other nodes. As shown in Figure 3A, Node 41 is the most influential when we perform a two-nodal causal inference. However, in a three-nodal causal inference, nodes 40 and 41 are of equal influence as both of them appeared five times, as shown in Figure 3B. The interpretation is that the asset owner may want to pay closer attention to these nodes.

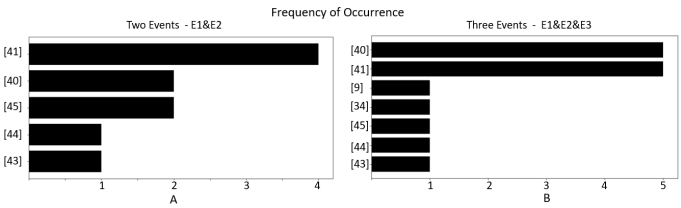

Fig. 3. Node Frequency Analysis

We validated our technique by checking the consistency of the sensitivity pattern among the three graphs. As the number of nodes in the causal inference calculation increases, the bars in both directions 3PS become longer, indicating an increased sensitivity (impact) and decreased probability of success if multiple nodes fail simultaneously. From Tables III and IV, we observed that node 40 is more critical in a 2-nodal inference while node 41 is most critical in a 3-nodal inference. This suggests that our technique can uncover critical nodes that could potentially prevent attacks, as shown by the case of the Colonial Pipeline cyber attack. Our technique enables us to discover more information from DM than it currently provides, increasing the potential for system owners to proactively manage and mitigate risks.

V. CHALLENGES AND FUTURE WORK

BN learning requires extensive computation to process causal queries for two or more nodal combinations, creating scalability issues for larger models with many nodes. To overcome these challenges, our consideration is limited to a system-driven model with a focus on processes and the interaction between processes. In the future, we hope to leverage DM's new capacity to model complex systems and to develop predictive models that can forecast future cyber risk trends, based on past data.

REFERENCES

[1] S. McLaughlin et al., 'The cybersecurity landscape in industrial control systems', Proceedings of the IEEE, vol. 104, no. 5, pp. 1039–1057, 2016.

[2] S. M. Kerner, "Colonial pipeline hack explained: Everything you need to know," WhatIs.com, https://tinyurl.com/393s2m2j (accessed Jan 26, 2023).

[3] A. O. Rotibi, N. Saxena, P. Burnap, and A. Tarter, 'Extended Dependency Modeling Technique for Cyber Risk Identification in ICS', IEEE Access, vol. 11, pp. 37229–37242, 2023.

[4] D. Koller and N. Friedman, Probabilistic graphical models: principles and techniques. MIT press, 2009.

[5] P. G. Bringas, 'Intensive use of Bayesian belief networks for the unified, flexible and adaptable analysis of misuses and anomalies in network intrusion detection and prevention systems', in 18th International Workshop on Database and Expert Systems Applications (DEXA 2007), 2007, pp. 365–371.

[6] A. Wall and I. Agrafiotis, 'A Bayesian approach to insider threat detection', Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, vol. 12, no. 2, 2021.

[7] C. J. Butz, W. Yan, P. Lingras, and Y. Y. Yao, 'The CPT Structure of Variable Elimination in Discrete Bayesian Networks', in Advances in Intelligent Information Systems, Z. W. Ras and L.-S. Tsay, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 245–257.