ISSN: 2093-5374 / E-ISSN: 2093-5382

A Bayesian-Network Approach for Assessing Security and Process Safety in the Petroleum Industry

Nawaf Abdualaziz Almolhis[1*]

[1*]Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia. naalmolhis@jazanu.edu.sa,
https://orcid.org/0009-0004-7558-7165

Received: August 02, 2024; Revised: September 13, 2024; Accepted: October 10, 2024; Published: December 30, 2024

Abstract

One of the conditions for the petroleum (gas and oil) industry's continued existence in the future may be a continuation of the trend towards fewer personnel at offshore installations and more data flow between the two locations. Malicious attacks, also known as security attacks, often target offshore gas and oil installations. These attacks can trigger a series of events such as the issue as well as spread of harmful materials and/or fires, blasts as well as energy, resulting in harm to people, the environment, and property. These effects might be just as devastating as those from big accidents caused by more traditional safety-related factors. For evaluating the achievement of physical security attacks probability, the current research uses a Bayesian Network (BN) technique that combines Natural Language Processing (NLP). The process also considers the single structure of the offshore gas and oil sector. In order to dynamically conduct risk assessments for safety and security, Bayesian networks are quickly becoming a popular tool. These networks adjust the previous disaster probability values to account for original data. Another benefit of Bayesian networks is their ability to describe conditional dependency between events, as well as their handling of variables with many states.

Keywords: Bayesian Networks, Offshore Gas and Oil Industry, Safety, Security and Security Attack.

1 Introduction

Process industries are typically a significant source of concern for safety because they handle hazardous materials at high pressure and temperatures. Accidents might happen in these units as a result of poor design, malfunctioning systems, operational faults, worker ineptitude, improper maintenance, disregard for safety protocols, etc. These safety accidents could potentially have negative impacts on individuals, the environment, and property (George & Renjith, 2021). To address the documentation of main accident situations that could be achieved by malicious manipulation of the plant's physical components through the control and safety instrumented systems, a methodology called Process Hazard Analysis of Remote Manipulations through the Control System (PHAROS) was developed for this study (Iaiani et al., 2021).

A study first discusses the development history of the safety barrier concept. This work then elaborates on the comprehensive review of the definition, categorization, assessment, performance evaluation, and management of safety barriers in the chemical process industries (Yuan et al., 2022).

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 4 (December), pp. 335-347. DOI: *10.58346/JOWUA.2024.I4.022*

*Corresponding author: Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia.

Process safety management (PSM) is a structure that shows a business's dedication to process safety, improved comprehension of risks and hazards, thorough risk assessment and management, and improved experience-based learning to improve overall safety and operational performance. To implement PSM, businesses often employ an incident data reporting system (Sattari et al., 2021).

To test the integrated model-based method for figuring out the security risk of Cyber Physical Systems (CPS), this uses a CPS testbed that has real industrial controllers and communication protocols. The testbed monitors and controls a real-time simulation of an exothermic continuous stirred tank reactor (CSTR) (Tantawy et al., 2020). For system designers and risk analysts, however, the contemporary elements of CPSs remain unclear, particularly when taking into account the role of people and the interplay between safety and security (Ismail, 2024). Accidental failures and human errors are no longer the exclusive causes of safety risks. The system and its environment are at risk of physical injury from cyber safety attacks, which may now escalate into safety attacks (Carreras Guzman et al., 2020).

Another work, examine the cyberattacks that have affected the downstream, upstream as well as middle segments of the gas and oil industry from the early 1990s until 2020. This record and examine confirmed attacks for every domain, drawing from real-world complaints and publicly available system demo attacks. To understand frequent and covert assault pathways against gas and oil key activities, classify as well as map the attack types used in each scenario (Stergiopoulos et al., 2020). Another study provides an overview of the growth of process safety, risk management strategies, along with techniques throughout that forty-five-year span. Additionally, the study includes a brief evaluation of the previous symposia contributions, highlighting a few notable contributions or breakthroughs (Pasman & Fabiano, 2021).

Physical security challenges in offshore gas and oil installations have received attention despite the highlighted panorama (Esfandiari et al., 2023).

- Most approaches have been qualitative or semi-quantitative in nature.
- Using the Bayesian Network (BN) method as its base, this study tries to come up with a methodical, quantitative way to figure out the conditional probability of achievement of physical security attacks that are attempted.
- This BN model combines expert knowledge and textual information. An NLP-based model is used in both the Conditional Probability Tables (CPT) estimate and the BN architecture design (Prema et al., 2022).
- The model expands what it offers by automating the setting of its hyperparameters. Designed for use in the unique offshore gas and oil sector, the process incorporates both proactive and reactive security intervention tactics.

Here's a detail of the work organization: A review of the safety and control systems and security in the gas and oil industry are covered by Section 2. Section 3 discusses the proposed technique. Section 4 includes details on the research findings, along with some limitations. Section 5 concludes the text, followed by the references.

2 Literature Review

Abou el Kalam, (2021) defined Supervisory Control and Data Acquisition (SCADA) systems that control organizations such as oil pipelines, water treatment, chemical manufacturing facilities, as well as smart grids among others (Salkić et al., 2020). Any intrusion, whether intentional or accidental, might result in significant harm to people, property, and business. As a result, the SCADA's security is critical

for maintaining the resilience and integrity of processes and activities, as well as availability against cyberterrorist attacks along with hostile. To address this problem, this work examines security risks and vulnerabilities, with an emphasis on more recent ones. Next, to determine the appropriate security methods for these critical systems, this work design a comprehensive approach.

Amyotte et al. summarized the authors' joint research efforts to integrate intrinsic safety design (ISD) into various process activities, applications, and safety systems (Amyotte & Khan, 2021). This illustrates the four basic intrinsic safety principles including minimization, simplicity, substitution, and moderation—with examples. Over the course of a typical process life cycle, it examines the ISD features and performance indicators during both the design and operational stages. Dust explosions and their domino effects are examples of undesirable phenomena that show how an intrinsic safety approach might successfully avoid or limit. This also address the value of analyzing incident-specific case studies developed by ISD.

Cormier & Ng, (2020) discussed how process hazard analysis (PHA) could include cybersecurity vulnerability analysis in order to protect the process control network from cyber threats. This updates a layer of protection analysis (LOPA) to assess potential vulnerabilities also safeguard that important application safeguards can resist cyberattacks. It demonstrates how to strengthen the plant's defenses against cyber and conventional threats by integrating cyber security into process safety management (PSM) components such as hazard and risk assessments.

Zwetsloot et al., (2020) described in 2018, 19 significant hazard companies in the Rotterdam area had their safety cultures evaluated. Four industrial sectors actively engaged were the petrochemical industry, chemical warehousing, logistics, refineries, and bulk storage. It used a pragmatic assessment method that focused on fourteen characteristics and followed a normative approach, utilizing a safety culture development scale ranging from 1 to 5. This settled on a 3.0 as the bare minimum for big-hazard businesses. This may compare the safety cultures of different firms as well as sectors using the safety culture maturity ratings that has gathered.

Shi, (2021) highlighted that the petroleum industry significantly boosts Chinese jobs and the economy. The Chinese Petroleum Industry, is associated with high risk, meaning that there are possible hazards. Due to the difficulty in halting the spread of accidents after they have occurred under the social-technical system, it is essential to monitor the safety behaviour of employees, especially those working directly with the public in the petroleum business. Therefore, it plans to employ safety-specific transformational leadership along with high-quality connections as predictors of mindful safety practices, a subset of safety behavior that the safety atmosphere mediates. In this study, mindful safety practices will serve as the dependent variable.

Pasman et al., (2020) examined the resilience of process plants, which include those involved in power production, chemical manufacturing, gas and oil refining, as well as many more. Over the years, plant safety has evolved from reactive to proactive procedures. Safety is critical from numerous perspectives, including the protection of the local community and workers, but it's also critical from an economic and sustainable standpoint. Proactive action necessitates an anticipatory understanding of potential process failures due to external or internal disruptive disturbances. To achieve that goal, a lot of work has been put into developing risk management and assessment techniques over the years. However, a number of unknowns, including missed or unidentified dangers, have proven risk assessment to be flawed. This takes steps to guard against such unsettling dangers up to a point.

Ahmad et al., (2022) explored the possible applications and advantages of blockchain technology for managing the supply chain, production, exploration, and logistics of the gas and oil industry. This is

because blockchain technology provides properties like auditability, immutability, traceability, decentralized trust, and transparency. This examines state-of-the-art blockchain-based schemes, commercial operations, case studies, and research initiatives to show how blockchain may work in the gas and oil industry (Sungur & Atan, 2020). This article, highlights some of the possible uses of blockchain technology. Blockchain-based smart contracts automates essential tasks including protecting international trade documents, tracking and tracing petroleum products, and coordinating the purchase and bidding process for oil exploration rights to industry companies. Table 1 describes the advantages and limitations of some existing works reviewed.

Table 1: Existing Work Review

Papers and Authors	Method	Advantages	Limitations
Amyotte & Khan, (2021)	ISD	Manage hazards to trying to eliminate or decrease hazards at their source.	Still lack in more safer designs.
Pasman et al., (2020)	Risk assessment	Examines the resilience of process plants, which include those that produce electricity, manufacture chemicals, refine gas and oil, and many more.	Only less real-world examples are considered.
Ahmad et al., (2022)	blockchain technology	Manage the operations of the gas and oil supply chain adequately.	Only some obstacles are explored in the gas and oil sectors.

3 Methods and Tools

Formulation of Security Risk

Typically, in the safety field, hazard is termed as a set of potential outcomes along with their corresponding probabilities or uncertainties. Here, "frequentist probability" is the time fraction that an event occurs along with its frequency of recurrence over time. Without specifically mentioning a probabilistic component, risk is often characterized in the security domain as the trifecta of asset/value, threat, and vulnerability. However, security risk can also be characterized by events, consequences, and uncertainty. Furthermore, utilizing probabilities effectively communicates the uncertainties. The following equation (1) is a definition of the security risk based on the factors discussed.

$$R^i = f\left(P_1^i, P_2^i, C^i\right) \qquad (1)$$

Where P_1^i represents an attempted attack probability on an asset based on scenario i, and R^i denotes the security risk associated with that scenario; Given the attempt in scenario i, P_2^i is the conditional probability of successfully executing the attack; C^i is the predicted outcome of the attack in scenario i. Data, or modelling about the objectives, traits, skills, tactics of adversaries, along with the sociopolitical environment of the target facility, are necessary for the quantification of P_1^i. Because of this, risk analysts aren't the best people to fill out the necessary background information; instead, sociologists, political analysts as well as intelligence analysts, are more appropriate. The methods used to assess P_1^i are often cross-disciplinary and can apply to any vital infrastructure, as these evaluations encompass more than just the facility's industry.

Quantifying P_2^i requires determining the various cyber-attack paths as well as physical attack paths that attackers may use to harm the target. Often, this requires expertise in fields including physical security, process safety, cyber security as well as control systems engineering. So, unlike P_1^i, P_2^i evaluation is entirely based on how the IT-OT (Information Technology - Operational Technology) as well as Physical Protection System (PPS) network are set up, which can be very different (particularly

the PPS for facilities in diverse industrial sectors, where the attack paths are naturally different from those aimed at an onshore process facility). Lastly, C^i quantification calls on the kind of skills often found in process risk analysts as well as safety: the ability to simulate scenarios involving the release of hazardous chemicals, toxic dispersions, fires along with explosions.

BN Architecture

Figure 1 displays the flow chart of the proposed BN-based quantitative procedure. It looks at different security intervention strategies (both preventative and mitigation) to figure out the chance of a physical security attack succeeding given an attack attempt P_2. As a result, it aids in the vulnerability valuation step of SV/RA (Security Vulnerability/ Risk Assessment) techniques and provides an important component for estimating a facility's overall security risk (see eq. (1)).

- The PPS layout of the analyzed facility, includes physical areas along with physical barriers.
- The Security Intervention Strategies (SITS) of each organization is considered.
- In the PPS under examination, all detection, assessment, and communication components provide the quantitative data on the probabilities of detection, proper evaluation of input data, and alarm communication.

The flowchart illustrates the five phases of the proposed method, which denote the offshore gas and oil industry is the intended context for its design.

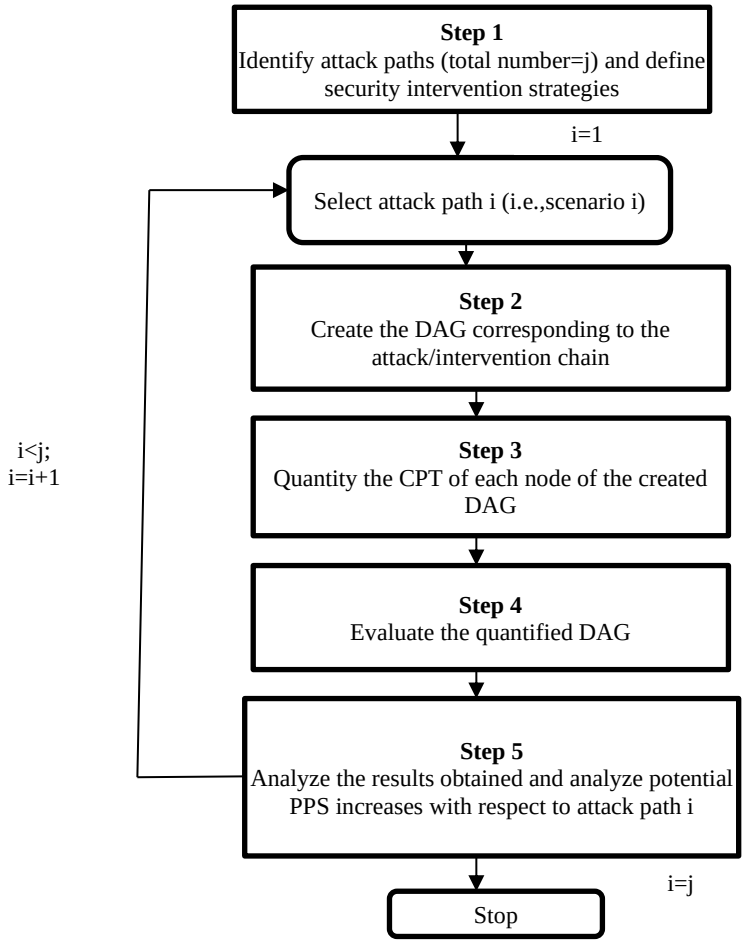

Figure 1: The Proposed Bayesian Network Flowchart

STEP 1: The initial stage is the documentation of the possible attack paths, or the exact physical steps an attacker would need to take within the PPS to carry out task. This stage also entails identifying potential security intervention methods to lessen the impact of each identified attack route within the investigated organization. If an attack occurs, it can implement a security intervention plan to either stop it in its tracks or at least lessen its effects.

STEP 2: This step includes creation of a DAG (Directed Acyclic Graphs) that depicts the attack/intervention chain for each attack path that is believed to be effective. This chain includes detecting adversaries, assessing the intrusion alert, communicating the intrusion, and acting on the detection. Every node has two states where one is for the ideal and one is for the worst-case scenario. In the absence of mitigation security intervention methods, the AT-node can only take one of two possible states i.e. successful or unsuccessful. However, when mitigation security strategies are in place, the node can take one of three possible states: successful with mitigated consequences, successful without mitigated consequences, or unsuccessful.

STEP 3: With the goal of giving marginal probabilities for root nodes as well as conditional probabilities for non-root nodes, the proposed approach seeks to quantify the CPT of every node in the DAG. These are derived from relevant previous research or extensive field data that can be collected and analyzed. This stage requires extra caution because the accuracy of the CPT determines the quality of the proposed potential outcomes.

STEP 4: This step uses the DAG results from step 3 to calculate the probabilities for the nodes of interest.

STEP 5: The proposed procedure focuses on analyzing Step 4's results in terms of the facility's vulnerability to the attack path being considered (Iaiani et al., 2023).

Probabilities Estimation in the CPT

Data and expert knowledge are used to estimate the probabilities of BN CPTs. This study exploits textual information to estimate CPT probabilities using only (Process Safety Event) PSE repository R data. Building an NLP method to get keywords and choosing model parameters are the first steps in figuring out the chances of things happening in the CPTs reports in R repository. First, it needs to extract keywords from the reports so that it can discover the most significant data in the repository free-text areas. By simplifying the BN model, this work enhances the results understandability also minimize the conditional probabilities count that require estimation. To do this, narrow focus to states with at least one relevant keyword. Reducing words to count enhances the robustness of the model parameter estimation.

Identification of the Keywords

The aim of this assignment is to find a set of N keywords, v_k, where $k = 1, \dots, N$, from the vocabulary of V tokens, $\{v_j, j = 1, \dots, V\}$, using N<V (Valcamonico et al., 2021) describes the creation of a method termed as Term Frequency Inverse Document Frequency (TFIDF) to accomplish this goal. TFIDF turn the list of tokens in every report $\tilde{d}_i$, where $i = 1, \dots, D$, into a set of numerical vectors h^i. Each generic element h_j^i is connected to a degree of the semantic value of the token v_j in report d_i. Next, the tokens that meet the condition $h_j^i \geq thresh_h$ are considered keywords retrieved from report d_i. Here, $thresh_h$ is the method's hyperparameter that stipulates the minimal value of semantic significance required for a token to be considered as a keyword. In the end, the tokens chosen for at least one report d_i, $i = 1, \dots, D$,

make up the whole set of keywords, $\{v_k, k = 1, \dots, N\}$. It is worth noting that the taxonomy automatically knows the states $\left\{\{I_g^m, m = 1, \dots, M_g\} g = 1, \dots, G\right\}$ linked with the keywords since the vocabulary tokens are organised in it.

Model Hyperparameters Setting

The model's two thresholds serve as the hyperparameters for identifying n-grams ($thresh_{NPMI}$) and keywords $thresh_h$. Since the list of linked keywords represents a report, it is necessary to find an illustration that is both are comprehensive and synthetic. This need completeness for the PSE-describing influencing factors (IF) to ensure don't miss important details, but also need synthesis to make the overall model understandable, by avoiding overcomplication, and prevent the spread of information among multiple redundant keywords.

This should optimize both the total number of keywords ($metric\ o_1^{com}$) and the amount of IF states with at least one related keyword in one repository report ($metric\ o_2^{com}$) to achieve the maximum values for completeness metrics. The representation should also be taxonomically balanced, meaning the number of keywords and vocabulary tokens for an IF state should be equal. Notably, it should minimize ($metric\ o_3^{com}$), which is the standard deviation across dissimilar IFs of the keywords count ratio mentioning a state to the quantity of tokens mentioning the vocabulary same state. ($metric\ o_4^{com}$), on the other hand, is the standard deviation across dissimilar IFs of the average ratio across all repository reports of the quantity of keywords mentioning a state to the tokens count mentioning the vocabulary same state. The representation synthesis incorporates two metrics: $\left(metric\ o_1^{syn}\right)$, or the keywords average number employed to characterize a report, should not be too small to yield representations that lack sufficient richness, and $\left(metric\ o_2^{syn}\right)$, or the conditional probabilities count in the CPTs of the BN model that require setting, should be kept to a minimum due to its connection to the knowledge as well as computational efforts required to estimation the CPTs. This study uses a triangle function with a centre at 6.5 and a range is used for the o_1^{syn} metric in equation (2).

$$F = \sum_{i=1}^{4} w_i^{com} o_i^{com} + \sum_{i=1}^{2} w_i^{syn} o_i^{syn} \qquad (2)$$

This set the weights of the metrics for completeness $w^{com} = \sum_{i=1}^{4} w_i^{com} o_i^{com}$ to match the weights of the metrics for synthesis $w^{com} = \sum_{i=1}^{2} w_i^{syn} o_i^{syn}$ so that $w^{com} = w^{com} = 0.5$. When setting the weights, it considers the following factors:

i. Instead of having large keywords count in the taxonomy (o_4^{com}), which may include synonyms and uninformative terms, it is better to have a reasonable keywords distribution in the taxonomy states (o_1^{com}). Consequently, w_4^{com} is set slightly bigger than w_1^{com}.

ii. To avoid a taxonomy with many states having extremely imbalanced numbers of keywords, it is preferable to have a balanced distribution of the keywords of the states in the taxonomy (o_2^{com}) rather than a reduction in the number of states without any keywords associated with them (o_3^{com}). Hence, w_2^{com} is intentionally designed to be somewhat bigger than w_3^{com}.

iii. w_2^{syn} is set somewhat less than w_1^{syn} since it is less desired to reduce the computational effort $\left(o_2^{syn}\right)$ than to have a suitably big number of keywords in the reports$\left(o_1^{syn}\right)$.

CPTs Conditional Probabilities Estimation

Using the R repository information, it estimates the probabilities in the BN's CPTs. According to reference (Rohmer, 2020), the conditional probability element $P\left(\sigma_s^q \sigma_s^{q-}\right)$ of the CPT linked to the node η_q is calculated.

$$P\left(\sigma_s^q \sigma_s^{q-}\right) = \frac{\rho\left(\sigma_s^q \sigma_s^{q-}\right)}{\rho\left(\sigma_s^{q-}\right)}$$

How many reports have the node η_q in state σ_s^q along with its immediate parent nodes in state η_{q-}, along with how many reports have the immediate parent nodes in state σ_s^{q-}, $\rho\left(\sigma_s^{q-}\right)$ is the number of such reports. For the CPTs that connect the nodes x and y, as well as y and z, representing the repository's category fields, the process is simple. For CPTs that have an IF as their parent node, estimate the probability of the state $x^a\left(y^b, z^c\right)$ of the node $x(y,z)$ conditional to the state I_g^m of the node ψ_g by separating the number of reports related with $x^a\left(y^b, z^c\right)$ as well as covering at least one keyword allocated to the state I_g^m,i.e.$\rho\left(x^a, I_g^m\right)\rho\left(y^b, x^a\right), \rho\left(z^c, y^b\right)$, by the number of reports associated with $\rho\left(I_g^m\right)$ as well as covering at least one keyword allocated to the state I_g^m.

4 Case Study

Petrochemical systems consider a PSE repository report spanning from 2016 to 2020. This do not disclose the exact number of accessible reports (D) here for reasons of confidentiality. The following components make up a report:

a. a free text d prepared by a system operator in English that describes the event as well as the elements affecting the PSE;

b. the cause x, which may be x^1= "Human" indicates a PSE resulting from human error or misuse of prepared equipment; x^2= "Equipment" signifies a PSE resulting from system faults or component malfunctions; and x^3="External" indicates a PSE resulting from events outside the system, including sabotage or natural disasters, or a combination of the two: x^4="Human_Equipment", x^5="Human_External", x^6="Equipment_External", x^7="Human_Equipment_External";

c. the incident kind, denoted by y, which can be y^1="fire/explosion" when a gas or liquid is ignited, y^2="gas leak" when gas is lost, or y^3="spill" when liquid is lost;

d. The tier level, represented by z, can range from z^1="1" for PSEs with strong LOPC, z^2="2" for PSEs with moderate LOPC, and z^3="3" for PSEs with less severe LOPC (API RP 754, 2021).

System specialists have created the taxonomy of the IFs. "Context," "Risk event," "Condition," "Impact," and "Barrier" are the five IFs represented by G = 5 with ψ_g ,$g = 1, .., G$. Experts have suggested and refined the matching states, $\left\{\left\{I_g^m, m = 1, \dots, M_g\right\} g = 1, \dots, G\right\}$, taking into account the repository's vocabulary. This determined the lowest level of the taxonomy—the tokens level —by seeing the repository's vocabulary and allocating the tokens to the respective states. It may see the attained taxonomy and the tokens quantity for every state. The taxonomy of the IF "barrier," falls into the "technical" or "operational" categories, as well as includes a few tokens associated with these categories. Once it constructs the taxonomy, finalize the model by delineating the causal interactions between the components. In order to validate the process, create a BN model that takes into account a single IF. This

makes it easier to understand the findings. This have specifically studied the IF "barrier" due to its significance in QRA (Quantitative Risk Assessment).

This specify eight alternative states that considers every possible barrier combination from the "technical," "operational," and "organizational" categories in relation to the node barrier, given that the development of a single PSE may be influenced by a variety of barriers. The PSE model, with the "cause" node conditionally dependent on the "barrier" node, displays the BN. This is because the performance of various barriers affects the frequency of PSEs associated with precise causes; for instance, the "no training" in "organizational" influences the "human" cause. A total of three time periods have been taken into account for the model updating: [2016, 2018], which includes D_1 reports; [2016, 2019], which comprises D_2 reports; as well as [2016, 2020], which has the whole database of D_3 =D reports.

5 Result

This work used the approach to automatically set the parameters $thresh_{NPMI}$ and $thresh_h$, and the values are presented in Table 2. One thing that stands out is:

a. The metric (o_1^{com}), consistently displays values across all three time periods, indicating a comparable use of tokens and keywords.

b. Since the metric (o_2^{com}), always equals 1 in all period, at least one keyword is constantly related with each state of the IF "Barrier."

c. This can observe that the taxonomy's IF "Barrier" states exhibit an even distribution of keywords due to the high values for metrics (o_3^{com}), and (o_4^{com}), across all three periods.

d. The metric $\left(o_1^{syn} \right)$, slightly declines with the release of more reports. The reason for this decline is that, while the percentage of tokens chosen as keywords metric (o_1^{com}) remains constant, the length of reports has increased from 50.7 in [2016, 2018] to 57.1 in [2016, 2020], resulting in an average selection of more keywords.

e. There has been no notable change to the computational effort metric, $\left(o_2^{syn} \right)$, which has not changed significantly.

Once retrieve the keywords, it can use the BN model, which estimates the parameters of the CPTs, to determine the tier levels probability. In the three periods, Table 3 reports the tier levels unconditional probabilities.

The likelihood of the occurrence of the type y^1=fire/explosion, $P(z^1/y^1)$, has almost doubled between [2016, 2018] and [2016, 2019], while the probabilities of the events y^2=gas leak and y^3=spill tend to decrease with time. For the time periods [2016, 2018] and [2016, 2019], the researchers looked into this by analyzing the particular causes probabilities conditional on the event occurrence of type "Fire/Explosion," $P(x^a/y^1)$, where $a = 1, \dots, A$ in Table 4.

Table 2: Default Values for the Parameters $thresh_{NPMI}$ and $thresh_h$.

		Periods		
		[2016,2018]	[2016,2019]	[2016,2020]
Parameter	$thresh_{NPMI}$	0.65	0.7	0.5
	$thresh_h$	0.5	0.5	0.6

Table 3: Three Time Periods where the Tier Levels have Unconditional Probabilities

Probability	Periods		
	[2016, 2018]	[2016, 2019]	[2016, 2020]
P (z^1)	0.099	0.077	0.064
P (z^2)	0.176	0.178	0.156
P (z^3)	0.724	0.745	0.780

Table 4: Should an Event of the "Fire/Explosion" Type Occur, the Probability of its Causes Increases

		Periods	
Cause	Probability	[2016,2018]	[2016,2019]
Human _Equipment_External	P ($x^7|y^1$)	0	0
Human_External	P ($x^5|y^1$)	0	0
Human _Equipment	P ($x^4|y^1$)	0	0.069
External	P ($x^3|y^1$)	0	0
Human	P ($x^1|y^1$)	0.476	0.414
Equipment	P ($x^2|y^1$)	0.524	0.517
External_ Equipment	P ($x^6|y^1$)	0	0

Table 5: The PSE Probability of Cause "Human_Equipment" Barrier States Conditionally Occurrence

Periods			
Cause	Probability	[2016,2018]	[2016,2019]
Technical_ Organizational	P ($I_5^{barrier}x^4$)	0	0.030
Technical_Operational_Organizational	P ($I_7^{barrier}x^4$)	0.020	0.340
Technical_Operational	P ($I_4^{barrier}x^4$)	0.100	0.190
Operational_Organizational	P ($I_6^{barrier}x^4$)	0.220	0.090
No_barrier_keywords	P ($I_8^{barrier}x^4$)	0.220	0.120
Operational	P ($I_2^{barrier}x^4$)	0.220	0.150
Organizational	P ($I_3^{barrier}x^4$)	0.040	0.030
Technical	P ($I_1^{barrier}x^4$)	0.180	0.060

The "human_equipment" cause, explains an equipment malfunction occurrence along with an improper action by system operators, has undergone the most significant alteration. Looking at the odds of the IF state "Barrier" if an event with the associated cause "Human Equipment" happens, where $P(I_m^{barrier}|x^4), m = 1, ..., M_{barrier}$, for the years [2016, 2018] and [2016, 2019] (Table 5), it can figure out which barrier has had the most significant effect on these changes.

The "Technical_Operational_Organizational" state, which signifies the simultaneous presence of organisational barriers as well as technical, operational, exhibits the most notable variation. System specialists have validated the existence of difficulties linked to technical, operational, and organisational barriers based on reports gathered between 2016 and 2019 of type "Fire/Explosion" at Tier level 1 (Valcamonico et al., 2024).

Limitations

This section gives a summary of the problems with the Bayesian method for risk assessment:

- In order to properly capture system behaviour when using BN as standalone techniques, their structural creation needs the assistance of professionals with deep system understanding.

- To understand the interdependencies of the various components of a BN model and construct conditional probability tables, in-depth system knowledge is required.
- Static BN can identify the most important components and efficiently assess the system's security risk level, but they can't grasp the temporal relationships between fundamental occurrences. Dynamic BN, both temporally and dynamically may capture the interaction between system variables.
- The limited validation of BN hinders user adoption of the Bayesian method.
- Feedback is part of most real-world systems, but static BNs can't represent it because they're acyclic (George & Renjith, 2021).

6 Conclusion

This research has created a technique that integrates NLP and BNs to aid QRA in the examination of PSEs in oil and gas assets. The approach enables the extraction of information from textual reports about the variables determining PSEs and the quantification of the probabilities of the repercussions' severity. This have subjected the suggested approach to a repository of reports of PSEs occurring in hydrocarbon plants. Based on the results, this research is capable of performing the following tasks:

1. Find key words for barriers to mitigation and prevention that will have a big effect on the severity of PSE effects.
2. Make sure that its parameters are always up-to-date with new information.
3. Look at how changes to system design as well as management actions affect the severity of accidents by watching how their chances change over time and figuring out the types of PSEs, their causes, as well as the states of the barriers that were most affected.

The next step in improving the methodology is to combine expert knowledge and data from other sources with the parameters and probabilities calculated in the CPTs using PSE reports.

References

[1] Abou el Kalam, A. (2021). Securing SCADA and critical industrial systems: From needs to security mechanisms. *International Journal of Critical Infrastructure Protection*, *32*, 100394. https://doi.org/10.1016/j.ijcip.2020.100394

[2] Ahmad, R. W., Salah, K., Jayaraman, R., Yaqoob, I., & Omar, M. (2022). Blockchain in oil and gas industry: Applications, challenges, and future trends. *Technology in society*, *68*, 101941. https://doi.org/10.1016/j.techsoc.2022.101941

[3] Amyotte, P. R., & Khan, F. I. (2021). The role of inherently safer design in process safety. *The Canadian Journal of Chemical Engineering*, *99*(4), 853-871. https://doi.org/10.1002/cjce.23987

[4] API RP 754. (2021). American petroleoum association (API) recommended practice process safety performance indicators for the refining and petrochemical industries. api.org

[5] Carreras Guzman, N. H., Wied, M., Kozine, I., & Lundteigen, M. A. (2020). Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis. *Systems Engineering*, *23*(2), 189-210. https://doi.org/10.1002/sys.21509

[6] Cormier, A., & Ng, C. (2020). Integrating cybersecurity in hazard and risk analyses. *Journal of Loss Prevention in the Process Industries*, *64*, 104044. https://doi.org/10.1016/j.jlp.2020.104044

[7] Esfandiari, A., Khoddami, S. A., & Samimi, A. (2023). Investigating the Process of Different Structures of Gas Hydrate Formation. *International Academic Journal of Innovative Research, 2*(2), 01–07.

[8] George, P. G., & Renjith, V. R. (2021). Evolution of safety and security risk assessment methodologies towards the use of bayesian networks in process industries. *Process Safety and Environmental Protection*, *149*, 758-775. https://doi.org/10.1016/j.psep.2021.03.031

[9] Iaiani, M., Tugnoli, A., Bonvicini, S., & Cozzani, V. (2021). Major accidents triggered by malicious manipulations of the control system in process facilities. *Safety science*, *134*, 105043. https://doi.org/10.1016/j.ssci.2020.105043

[10] Iaiani, M., Tugnoli, A., Cozzani, V., Reniers, G., & Yang, M. (2023). A Bayesian-network approach for assessing the probability of success of physical security attacks to offshore Oil&Gas facilities. *Ocean Engineering*, *273*, 114010. https://doi.org/10.1016/j.oceaneng.2023.114010

[11] Ismail, W. S. (2024). Threat Detection and Response Using AI and NLP in Cybersecurity. *Journal of Internet Services and Information Security, 14*(1), 195-205. https://doi.org/10.58346/JISIS.2024.I1.013

[12] Pasman, H. J., & Fabiano, B. (2021). The Delft 1974 and 2019 European Loss Prevention Symposia: Highlights and an impression of process safety evolutionary changes from the 1st to the 16th LPS. *Process Safety and Environmental Protection*, *147*, 80-91. https://doi.org/10.1016/j.psep.2020.09.024

[13] Pasman, H., Kottawar, K., & Jain, P. (2020). Resilience of process plant: what, why, and how resilience can improve safety and sustainability. *Sustainability*, *12*(15), 6152. https://doi.org/10.3390/su12156152

[14] Prema, M., Raju, V., & Ramya, M. (2022). Natural Language Processing for Data Science Workforce Analysis. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 13*(4), 225-232. https://doi.org/10.58346/JOWUA.2022.I4.015

[15] Rohmer, J. (2020). Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: A comprehensive review. *Engineering Applications of Artificial Intelligence*, *88*, 103384. https:// doi.org/10.1016/j.engappai.2019.103384

[16] Salkić, Z., Lugović, B., & Babajić, E. (2020). Petrography and Mineral Chemistry of Oligocene Shoshonitic Dacites from the Central Bosnia. *Archives for Technical Sciences, 1*(22), 1–10. https://doi.org/10.7251/afts.2020.1222.001S

[17] Sattari, F., Macciotta, R., Kurian, D., & Lefsrud, L. (2021). Application of Bayesian network and artificial intelligence to reduce accident/incident rates in oil & gas companies. *Safety Science*, *133*, 104981. https://doi.org/10.1016/j.ssci.2020.104981

[18] Shi, H. (2021). The influence of safety-specific transformational leadership and high-quality relationships on mindful safety practices through safety climate: a study in Chinese petroleum industry. *Journal of Applied Security Research*, *16*(3), 328-344. https://doi.org/10.1080/19361610.2020.1761744

[19] Stergiopoulos, G., Gritzalis, D. A., & Limnaios, E. (2020). Cyber-attacks on the Oil & Gas sector: A survey on incident assessment and attack patterns. *IEEE Access*, *8*, 128440-128475. https://doi.org/10.1109/ACCESS.2020.3007960

[20] Sungur, Ş., & Atan, M. M. (2020). Effect of active packaging films containing natural antioxidant essential oils on the oxidative stability of the African catfish (Clarias gariepinus). *Natural and Engineering Sciences*, *5*(3), 155-166. https://doi.org/10.28978/nesciences.832984

[21] Tantawy, A., Abdelwahed, S., Erradi, A., & Shaban, K. (2020). Model-based risk assessment for cyber physical systems security. *Computers & Security*, *96*, 101864. https://doi.org/10.1016/j.cose.2020.101864

[22] Valcamonico, D., Baraldi, P., & Zio, E. (2021, November). Natural Language Processing and Bayesian Networks for the Analysis of Process Safety Events. In *2021 5th International*

Conference on System Reliability and Safety (ICSRS) (pp. 216-221). IEEE. https://doi.org/10.1109/ICSRS53853.2021.9660733

[23] Valcamonico, D., Baraldi, P., Zio, E., Decarli, L., Crivellari, A., & La Rosa, L. (2024). Combining natural language processing and bayesian networks for the probabilistic estimation of the severity of process safety events in hydrocarbon production assets. *Reliability Engineering & System Safety*, *241*, 109638. https://doi.org/10.1016/j.ress.2023.109638

[24] Yuan, S., Yang, M., Reniers, G., Chen, C., & Wu, J. (2022). Safety barriers in the chemical process industries: A state-of-the-art review on their classification, assessment, and management. *Safety science*, *148*, 105647. https://doi.org/10.1016/j.ssci.2021.105647

[25] Zwetsloot, G. I., van Middelaar, J., & Van der Beek, D. (2020). Repeated assessment of process safety culture in major hazard industries in the Rotterdam region (Netherlands). *Journal of Cleaner Production*, *257*, 120540. https://doi.org/10.1016/j.jclepro.2020.120540

Author Biography

Nawaf Abdualaziz Almolhis, received B.Sc. in Computer Engniring from Albaha Gollege. He got M.S in Information Technology from Kettreing University, USA. He received his Phd degree in Information Security and Forensics from University of Idaho, USA. He is a faculty member at CS department, in College of Computer Science and Information Technology, at Jazan University. His research interests include, cyber security, social network analytics, machine learning, and IoT forensics.