Cloud Enterprise Dynamic Risk Assessment (CEDRA): a dynamic risk assessment using dynamic Bayesian networks for cloud environment

Dawood Behbehani (dbehbehani@outlook.com)
City, University of London
Nikos Komninos
City, University of London
Khalid Al-Begain
Kuwait College of Science and Technology
Muttukrishnan Rajarajan
City, University of London

Research Article

Keywords: Dynamic risk assessment, Quantitative risk-analysis, Decision-making

Posted Date: April 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1512376/v1

Cloud Enterprise Dynamic Risk Assessment (CEDRA): a dynamic risk assessment using dynamic Bayesian networks for cloud environment

Dawood Behbehani[a,*], Nikos Komninos[a], Khalid Al-Begain[b] and Muttukrishnan Rajarajan[a]

[a]School of Mathematics, Computer Sciences and Engineering, City, University of London, United Kingdom [b]Kuwait College of Science and Technology, Kuwait

ARTICLEINFO

:

Keywords

Dynamic risk assessment

Quantitative risk-analysis

Decision-making

ABSTRACT

Cloud computing adoption has been increasing rapidly amid COVID-19 as organisations accelerate the implementation of their digital strategies. Most models adopt traditional dynamic risk assessment, which does not adequately quantify or monetise risks to enable business-appropriate decision-making. To address this issue, we propose a Cloud Enterprise Dynamic Risk Assessment (CEDRA) model that uses CVSS, threat intelligence feeds and information about exploitation availability in the wild using dynamic Bayesian networks to predict vulnerability exploitations and financial losses. We evaluate the proposed model in a real case scenario that demonstrates the applicability of this model.

1. Introduction

1.1. Cloud Services

Cloud computing adoption has been increasing rapidly among organisations of all sizes [18]. Amid COVID-19, enterprises have accelerated their digital transformation; thus, cybersecurity has become even a bigger challenge. Consequently, technology has become more essential in both our working and our personal lives. Companies have realised the importance of adapting to the market's needs. As such, the adoption of cloud services and agile methodology has enabled the rapid delivery of services online. The mechanism of cloud services introduces many benefits for organisations, such as ease of deployment, on-demand scalability, wide accessibility and ease of management [4]. This has propelled the growth of cloud-based adoption and application. As adoption widens, so does the threat landscape. To cater to the dynamically changing landscape of threats towards the use of cloud services, effective security countermeasures should be implemented, selected based on risk and threat assessment [22].

1.2. Risks in cloud services

Risks levels in cloud environments tend to fluctuate due to time-varying factors, such as the emergence of new vulnerabilities in safety barriers, installation of new software/components and misconfigurations. It is essential to quantify these time-dependent factors and their relations via robust calculation techniques. The outcome is then used to derive quantified estimates of risk that are mapped against a pre-defined risk criterion. Traditional risk assessment methods, such as quantitative is consistently emerging [22]. Dynamic Bayesian Network (DBN), an extended version of standard Bayesian Network (BN) with the concept of time, is a Probabilistic Graphical Model (PGM) that can be used to build models from data or expert opinion using Bayes' theorem. The model is ideal for a wide range of tasks, including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction and decision making under uncertainty. BN corresponds to a set of random variables, and their relationships are signified using a directed acyclic graph (DAG). The arcs signify the causal relationship between nodes, and the nodes represent the variables. The main node is called a 'root node'. If there is an arc connected to another node, that is called a 'parent node', and if a the 'parent node' is connected to another node, that is called a 'child node', as seen in Fig. 1. All nodes in the BN are allotted an initial probability. The purpose of BN is to use the node probabilities and their relational dependencies to assess and update the distribution probabilities of the random variables based on provided evidence and posterior. The calculation of the probabilities of the 'child node' is a blend of 'parent node' probability and conditional probability tables (CPT). This is considered the main protocol of the well-known 'Bayes theorem' of conditional probabilities [reference].

1.3. Contributions

In the literature, most applied Bayesian network model in risk assessment studies are focused on areas such as aiming to predict drilling-related risks, asset failure in thermal power plant and industrial control systems, however do not take into consideration the monetary losses. To overcome this problem, this study aims to develop a cloud risk assessment framed that enables the assignment of monetary losses terms to the consequences nodes, thereby enabling experts to understand better the

risk assessment (QRA), and such methods as Fault Tree (FT), Event Tree (ET) and Whatif analysis are limited in terms of incorporating new information or evidence, as these models cannot handle data scarcity and uncertainties. Therefore, it is important to focus research on the field of dynamic risk assessment, where risk

response to cyber attacks on physical systems, tank storage, etc). To ensure the feasibility of our study, a cloud service is simulated, and one undesired event is simulated based on the capital one breach case constructed. Our risk assessment model is developed and assessed against BN. The rest of this paper is organised as follows. Section two demonstrates related work and justifies the reasons for using DBN. Our proposed risk assessment model is demonstrated in section three. The case study is then highlighted in section four to demonstrate our model. Finally, a discussion of the proposed model is presented in section five and our work is concluded in section six.

2. Related work

2.1. Risk Assessment theory

The objective of risk assessment is to identify threats and vulnerabilities in a particular area or scope to apply adequate mitigation controls that would reduce the risk to an acceptable level. The process is continuous to measure risk factors as they change and develop over a significant time [24]. The numerous impacts of risks on organisations adopting cloud services are either tangible or intangible losses, such as downtime, data loss and reputation jeopardy. However, in this study, we limit these impacts to system asset loss.

2.2. Dynamic risk assessment

There are numerous cyber risk assessment frameworks aimed at exercising risk evaluation that were developed by governmental agencies, the cyber defence industry, and academic institutions. Nevertheless, such frameworks lack the mechanism to handle dynamically changing environment and cannot adapt their countermeasures and priorities to changes happening within inter-systems and external environments [15]. Another major issue in risk management involves real-time data scarcity and uncertainties to enable adequate risk calculation [31]. Data mining to predict the outcome of a cyberattack is a challenge especially in a situation where a series of factors may be involved in launching a malicious attack to cause losses. This requires a synthetic model that entails attack knowledge and system knowledge for analysing attack vectors and cybersecurity risks [32]. For example, [19] financial risks of any consequence. In this work, BT analysis has been converted into a dynamic BN network that estimates monetary losses. It is the first time such an analysis has been applied to cyber threats in cloud environments (previous usage of this method was in proposed the Bayesian attack graph method in risk assessment for predicting potential attacks. Such a model, including other models such as fault tree analysis (FTA), event tree analysis (ETA), bow-tie analysis (BTA), Markov chain analysis (MCA) and Bayesian network (BN) require a large amount of prior knowledge about attacks [31] or referred to as epistemic, a form of a uncertainty [23]. Another uncertainty can be referred to as aleatory uncertainty, where this is considered a nondeterministic nature of the events [17]. Therefore, one could say that a dynamic risk assessment is a method that reassess risk by continuously updating probabilities of events, as new information is fed and made available [6]. Such information could be from sensors deployed at the organizations such as proportionalintegral-derivative controllers for Supervisory Control and Data Acquisition (SCADA) systems, or security information and event management (SIEM) for network security [11]. For example, [21] have proposed a dynamic risk assessment based on the 2.0 semantic version of STIX™ for cyber threat intelligence that enables fetching indicators of compromise such as malicious URL, domain, and IP addresses. A dynamic risk assessment can be further considered a technique that takes consideration the effect of nonlinear interactions within inter-processes in its operational risk estimation. Thus, this type of technique can provide a realistic estimation of the operational risk in complex processes [29]. Dynamic risk assessments that leverage expert opinion in their methodologies such as [3] tend to be incomplete and subjective, thus the assessment results may be inaccurate. They may also potentially incapable for predicting unknown attacks [8]. To address this issue, various scholars have proposed models that can compensate for the necessity of acquiring historical data. For example, the fuzzy probability Bayesian network approach replaces limited historical data with fuzzy probabilities and a fuzzy approximate dynamic inference algorithm [32]. Another example, is an automated intrusion response system that utilises dynamic risk assessment using fuzzy logic. This is merely because fuzzy logic lessens the level of uncertainty of risk factors [7].

2.3. Decision-making

One of the main objectives of risk management is to achieve an efficient decision-making process that

ultimately aims to reduce system risk and strive for cost-benefit strategies. Decision-makers pursue to advocate less resources on countermeasures and acquire more benefit from strategy execution [20].

2.4. Bayesian Network

Bayesian Networks (BN) provide a useful mechanism in the risk analysis field due to their ability to model probabilistic data. [5] states that dynamic BNs take into consideration temporal dependencies based on time. Basic BNs do not consider alterations in time or manage time-evolving environments; thus, dynamic BNs (DBNs) are ideal for handling time-dependent risk assessments. [12] demonstrate how BNs have gained popularity because of their capabilities in predictive and diagnostic analyses. However, traditional BNs can only demonstrate relationships between variables at a specific time points, or for a specific period of time. They do not indicate temporal relationships between different times. To address this issue, DBNs can be used to present changes over time and relationships between a device's current, past or future states.

2.5. Dynamic Bayesian Networks

A DBN is an extension of a BN that introduces relevant temporal dependencies to model the dynamic behaviour of variables. Numerous inference algorithms are available for DBN modelling. In [28], the forwards–backwards inference and mutual information were used to model the Bayesian inference. In [28] and [13], the authors claim this is a major barrier to acquiring the precise probability of basic events related to a system target in a situation when objective data is scarce to predict probabilities pertaining to a target system. In addition, the availability of large data samples is a necessity of deep learning. Therefore, expert judgement is deemed an appropriate approach to obtain the occurrence probabilities of events. Although judgements obtained from expert opinions are subjective and susceptible to a margin of error, they are the ideal way forward. A DBN can also be leveraged using a FT, ET and BT, as these cannot handle the dynamic nature of operational risks. For example, [13] proposed a dynamic risk assessment methodology based on FT methods that is mapped into DBNs. However, [10] claims that a DBN requires a high number of simulation runs, merely because dynamic probabilistic risk assessment models require heavy computational power and system memory for running a vast number of simulations. Therefore, they proposed the use of a DBN with clustering analysis to enable a reduced number of simulation runs and the quantification of emerging system risk in a probabilistic manner for thermal-hydraulic simulation data. The approach was demonstrated with the mean shift clustering algorithm along with the bandwidth selection method. In addition, [27] asserts that the DBN has a strong advantage when handling uncertainty, because it is a directed acyclic graph that describes the conditional probability relationship between parameters using probabilistic inference theory.

2.6. Threat Intelligence

Today's cybersecurity threats have emerged, and traditional approaches based on heuristics and signatures are not very effective against dynamically changing threats known to be evasive, persistent and complex. Organisations must gather the latest cyber threat information to deter attacks in a timely manner [25]. Threat intelligence (TI) represents an actionable defence that aims to reduce the gap between the attack and the organisation's defensive action [26]. TI can take on multiple forms and means. Numerous methods and tools offer TI feeds, such as IBM X-Force Exchange, CrowdStrike Intelligence Exchange and AlienVault OTX Pulse. In this study, we utilized the AlienVault OTX Pulse to retrieve pulse information regarding the vulnerabilities in scope of this study. This enables us to understand which vulnerabilities attackers are primarily pursuing.

3. Framework for dynamic risk assessment

The detailed process of the proposed methodology, cloud enterprise dynamic risk assessment (CEDRA), is shown in Figure 1.

1. **Construction of a bow-tie (BT) model (top event, initiating events and safety barriers)**

BT has been widely employed as a graphical approach to represent an end-to-end accident scenario, from its causes to its consequences. The top event is placed in the centre, and on the left-hand side is a fault tree that identifies the potential events causing the

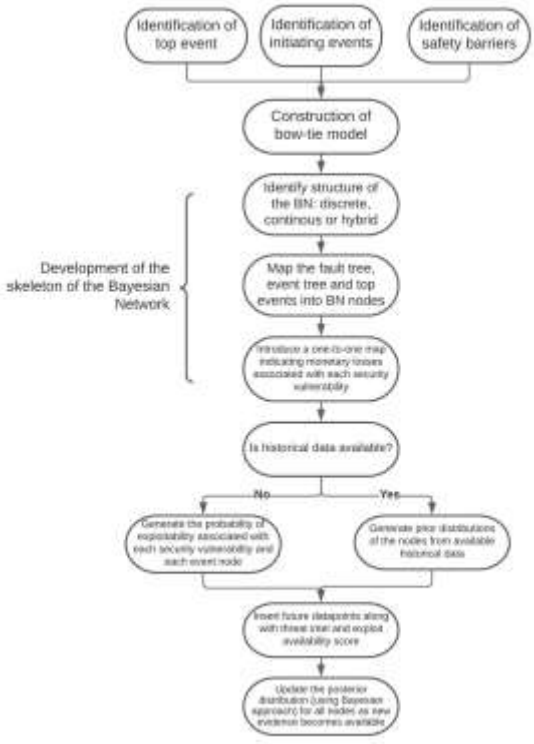

Figure 1: Schematic of the framework developed

critical event. On the right-hand side is an event tree that depicts the possible consequences of the critical events based on the safety barrier's success or failure.

2. **Developing of the skeleton of the Bayesian network from BT** Starting out with a BT, a Bayesian network (BN) needs to be constructed. A BN is a directed acyclic graph that is widely used in risk and safety analyses, inspired by probability and uncertainty. Graphically, a BN's structure is created based on a fault tree and an event tree in such a manner that the top event and the causes shown in the fault tree are demonstrated by nodes, while its relationship is depicted through arcs in the BN. OR gates and AND gates are used to present relationships between the causes and the top event in the fault tree. The process of mapping a fault tree into a BN comprises nodal representations of the barriers. Each node generally has discrete outputs of nodes. For equipment that operates continuously, nodes can take continuous values. Therefore, matters such as the probability of failure within a specified time period and the time to failure (TTF) for equipment under continuous operation are taken into consideration. The failure rate is considered constant or time dependent. When the nodes have discrete outputs, we use a discrete node BN with marginal prior probabilities, and when the nodes have continuous outputs, we use a marginal distribution for the nodes (Weibull, exponential or gamma distributions).

Asset	Value (USD)
Web Server	0
SQL Server	0
Gateway Server	10,000
Application Server	10,000

Table 1
DRAs modelling based on Formulas and Framework

3. **Incorporating monetary loss terms in the BN** This study discusses a method for incorporating monetary losses resulting from a failure (or the triggering of a top event). Each security vulnerability is associated with a monetary loss. The total financial loss associated with the breach of an event would be given by the probability of a breach of an event node multiplied by the financial loss associated with that node. The financial loss associated with a node can be calculated from the financial loss associated with each security vulnerability. The financial loss limited to asset purchasing cost has been calculated by the following formula:

$$F(t) = \sum_{j=0}^{6} P_{exploitability,d}(v_j, t).\ a(v_j)$$

where F(t) is the financial loss at time t. Pexploitability,d (considering EA and TI scores) of vulnerability vj,t) is the dynamic probability of exploitability (where j varies from 0 to 6), and a(vj) is the maximum asset loss of vulnerability vj. For simplicity, we divided the monetary losses based on asset value. We can expect higher risk consequences leading to higher monetary class nodes and vice-versa over time. Please refer to table 1.

4. **Construction of the complete BN** Once the initiating events, monetary loss nodes and top event and consequences are identified and the skeleton of the BN

is determined, the final Bayesian network can be developed. When expert opinion is not available, step 3 can be skipped, and a BN is constructed without considering the risk influence factors.

5. **Check whether historical data are available** A BN is a probabilistic graphical model in which the nodes are connected in a directed acyclic graph. Conditional dependencies are assigned in terms of conditional probability tables (CPT). Imagine a BN with n variables A1, A2, . . . , An. The joint probability distribution of the BN can be simplified as

$$P(A_1, A_2, \ldots, A_n) = \prod_{i=1}^{n} P\left(A_i \mid Parents(A_i)\right)$$

where Parents(Ai) denote the set of parent nodes of the node Ai. If A and B are two random events with a prior probability P(A) and P(B), the posterior probability of event A occurring, given that B has occurred, can be determined by Bayes' rule

$$P(A \mid B) = \frac{P(A)\,P(B \mid A)}{P(B)}$$

data are not available, we construct the initial CPT using the CVSS scores of all the security vulnerabilities used in constructing the event graph. We will use the static scores (access vector, attack vector, permission for intervention and user interaction) to create a static probability of exploitability values for the initial CPT. As more evidence is added, including dynamic data, such as threat intelligence scores and exploit availability information scores, the CPTs will be updated accordingly.

6. **Evaluation of the dynamic risk profile** When no historical data are present, the BN has no information on how to connect initiating events with consequences. As data are added, the CPT can be updated. The CPT is updated using the following:

-Dynamic CVSS scores (threat intelligence scores,

exploit availability scores)

-A Bayesian model that takes as input historic CPT values and new evidence (for example, whether a security node was breached or a top event was triggered) and uses that to update the CPT tables

3.1. Risk Assessment Model

The vulnerabilities of the cloud environment are commonly identified using numerous scanning tools and the inputs of system experts (manual assessment), including historical data on previous incidents by attackers. System vulnerabilities can also be identified on the basis of asset information from the Common Vulnerabilities and Exposures (CVE) database of asset/product information such as product names and versions. The third version of the Common Vulnerability Scoring System (CVSS) developed by the National Institute of Standards and Technology (NIST) is aimed at defining the characteristics of vulnerabilities and generating a numerical score to reflect the severity of a vulnerability. The score consists of a combination of parameters, including access vector (AV), access complexity (AC), and authentication (AU). Further information can be obtained from the CVSS v3.1 user guide [reference]. [30] proposed to assess the probability of vulnerability exploitation using the equation

$$P_{exploitability,s} = \frac{C}{F} \times AV \times AC \times PR \times UI$$

In this paper, we adopt this method while introducing two new parameters: exploit availability (EA) and threat intelligence (TI).Therefore, the successful exploitation of a specific vulnerability after the introduction of EA and TI is as follows:

$$P_{exploitability,d} = \frac{C_0}{F_0} \times AV \times AC \times PR \times UI \times TI \times EA$$

where C and C_0 denote the exploitation factor and F and F_0 the upper limit of the exploitation score. The variables AV, AC, PR and UI represent the static metrics: the access vector, the metric of the attack vector, the metric of the required permissions for intervention and the metric of the user interaction, respectively. The dynamic metrics are represented by TI and EA—the threat intel and exploit availability scores, respectively. EA was set at a value of 0.33

when available and 0.66 when not available. As for TI calculation, values < 10 were set at 0.45; between 10 and 45, at 0.50; and >45, at 0.55. To construct the initial BN at t = 0, we assume no knowledge of the dynamic BN and hence use $P_{exploitability,s}$ to generate the initial CPT values. In our Bayesian network, certain event nodes can be caused only by the occurrence of all security event nodes; in this case, we use the AND gate to describe this relationship that has been adopted from [14].

$$P\left(X_i \mid Parent\left(X_i\right)\right) = \begin{cases} 0, \ if \ \exists X_j \in Parent\left(X_i\right), \ X_j^s = 0 \\ P\left(\bigcap_{X_j^s=1} e_i\right) = \prod_{X_j^s=1} P\left(e_i\right), \ otherwise \end{cases}$$

When the occurrence of any security event out of the many possible events triggers a warning, we use the OR gates to describe the relationship (for example, the relationship between v0 and v6 going to the node 'block at firewall' is described by OR logic).

$$P\left(X_i \mid Parent\left(X_i\right)\right) = \begin{cases} 0, \ if \forall X_j \in Parent\left(X_i\right), \ X_j^s = 0 \\ P\left(\bigcup_{X_j^s=1} e_i\right) = 1 - \prod_{X_j^s=1} \left(1 - P\left(e_i\right)\right), \ otherwise \end{cases}$$

4. Case Study

This section presents a case study that applies our proposed framework. The selected scenario is based on the Capital One breach that occurred in 2019. Capital One is the fifth largest consumer bank in the U.S. and is considered one of the early banks to adopt the cloud computing environment from Amazon, which played a key role in the 2019 incident. The bank's objective was to reduce its on-premise data centre operation and expand its cloud service footprint. They also worked closely with AWS to craft a security model to achieve a robust, secure cloud operation. In 2019, the bank announced that adversaries gained unauthorised access and obtained certain types of personal information from Capital One credit card customers and individuals [16]. CloudSploit published an incident analysis report indicating that the access to the vulnerable system was achieved by executing a Server-Side Request Forgery (SSRF) attack that exploited a misconfigured web application firewall (WAF) known as "ModSecurity". In a typical SSRF attack, the attacker's objective is to initiate a unauthorised connection to internal-only services within the organisation's infrastructure in order to gain access to internal systems [2]. Based on our research, the vulnerability known as CVE-2019-2828

ID	CVE ID	AV	AC	AU/PR	UI	TI	EA	$P_{exploitability}$
v0	CVE-2019-2828	N(0.85)	L(0.77)	N(0.85)	R(0.62)	0(0.45)	No(0.5)	0.596
v1	CVE-2021-32791	N(0.85)	H(0.44)	N(0.85)	N(0.85)	0(0.45)	No(0.5)	0.467
v2	CVE-2021-1636	N(0.85)	L(0.77)	L(0.62)	N(0.85)	1(0.45)	No(0.45)	0.596
v3	CVE-2021-38639	L(0.55)	L(0.77)	L(0.62)	N(0.85)	-0.45	-0.45	0.351
v4	CVE-2021-36965	N(0.85)	L(0.77)	N(0.85)	N(0.85)	-0.45	-0.45	0.744
v5	CVE-2020-0670	L(0.55)	L(0.77)	L(0.62)	N(0.85)	0 (0.45)	No (0.45)	0.386
v6	CVE-2020-0720	A(0.62)	L(0.77)	H(0.5)	N(0.85)	0 (0.45)	No (0.45)	0.386

Table 2
Probability that one vulnerability is successfully exploited

was the cause of the exploitation. The vulnerability was added into the National Vulnerability Database few days post the incident. The vulnerability has a base score of 9.6, implying that its requires minimal effort for exploitation that enables unauthenticated attacker with network access via HTTP to compromise the WAF component [1]. The dataset used in this study was generated using a python script that fetches vulnerabilities from NVD, exploit-db.com (to retrieve information regarding exploitability availability in the wild), and AlienVault OTX Pulse to retrieve threat intelligence pulse information regarding the vulnerabilities in the scope of the study. We ran the dataset generation tool for a month to retrieve sufficient data. The vulnerability related to the Capital One incident was added synthetically onto the dataset.

The corresponding bow-tie diagram was constructed based on the components of the case study. Each of the vulnerabilities from v0 to v6 should be associated with an exploitability probability, where $P_{exploitability}$ represents the probability of successful exploitation. We are constructing a BN with dynamic risk assessment that assumes no prior information on the probability values. As vulnerabilities occur, the probability values for the different events will be recorded. As these probability values are recorded, the conditional probability tables would be updated (how to construct CPT tables is shown in [14], Tables 8-14)

Tables 2–6 show the conditional probability distributions of the different event nodes (WS: web server, SS: SQL server, GS: gateway server, AS: admin server). These nodes were taken from the topology of the test network in [9]. The irrelevant nodes that were not used in our case study were removed.

Tables 4, 5 and 6 only show a subsample of the very many possibilities. There are 32 possibilities in Table 3, 64 possibilities in Table 4 and 128 possibilities in Table 5. Here, we have shown only a small subset related to different risk scenarios. Let us consider a scenario in which the evidence chain goes via v0, v2, v3 and v6 . In

this case, the probability of the web server being compromised is P(WS = WS1 | v01, v10) = 0.571.

The probability of the gateway server being compromised is P(GS = GS1 | v01, v10, v31 , v40, v50) = 0.270. This value is obtained by $P_{exploitability}$ (v0) * $P_{exploitability}$ (v3) using the AND gate relationship.

The probability of the admin server being compromised is P(AS = AS1 | v01, v10, v31 , v40, v50 , v61) = 0.116.

WS	WS_0	WS_1
P(WS=WSs | v00, v10)	1	0
P(WS=WSs | v01, v10)	0.429	0.571
P(WS=WSs | v00, v11)	0.429	0.571
P(WS=WSs | v01, v11)	0.184	0.816

Table 3
Conditional probability distribution of the web server (WS) node

SS	SS_0	WS_1
P(SS = SSs | v00, v10, v20)	1	0
P(SS = SSs | v00, v10, v21)	1	0
P(SS = SSs | v00, v11, v20)	1	0
P(SS = SSs | v01, v10, v20)	1	0
P(SS = SSs | v00, v11, v21)	0.696	0.304
P(SS = SSs | v01, v10, v21)	0.696	0.304
P(SS = SSs | v01, v11, v20)	1	0
P(SS = SSs | v01, v11, v21)	0.566	0.434

Table 4
Conditional probability distribution of the SQL server (SS) node

GS	GS_0	GS_1
P(GS = GSs | v00, v10, v30 , v40, v50)	1	0
P(GS = GSs | v01, v10, v30 , v40, v51)	0.73	0.27
P(GS = GSs | v01, v10, v31 , v40, v50)	0.73	0.27
P(GS = GSs | v01, v11, v30 , v40, v51)	0.615	0.385
P(GS = GSs | v00, v10, v31 , v40, v51)	1	0
P(GS = GSs | v01, v10, v31 , v40, v51)	0.588	0.412
P(GS = GSs | v01, v11, v31 , v41, v51)	0.184	0.816

Table 5
Conditional probability distribution of the gateway server (GS) node

AS	AS_0	AS_1
P(AS = ASs | v00, v10, v30 , v40, v50 , v60)	1	0
P(AS = ASs | v01, v10, v31 , v40, v50 , v60)	1	0
P(AS = ASs | v01, v10, v31 , v40, v50 , v61)	0.884	0.116
P(AS = ASs | v01, v10, v30 , v40, v51 , v61)	0.884	0.116
P(AS = ASs | v01, v11, v31 , v40, v51 , v61)	0.748	0.252
P(AS = ASs | v01, v11, v31 , v41, v51 , v61)	0.65	0.35

Table 6
Conditional probability distribution of the admin server (AS) node

This value is calculated by $P_{exploitability}$ (v0) * $P_{exploitability}$ (v3) * $P_{exploitability}$ (v6).

The probability for the SQL server being compromised is P(SS = SS1 | v01, v10, v21) = 0.303. This value is calculated by $P_{exploitability}$ (v0) * $P_{exploitability}$ (v2).

The probability of the top event (TE) node triggering is the combined probability of the admin server being compromised and the SQL server being compromised, which is P(TE = TE1 | v01, v10, v21, v31 , v40, v50 , v61) = P(AS = AS1 | v01, v10, v31 , v40, v50 , v61) * P(SS = SS1 | v01, v10, v21) = 0.116*0.416 = 0.035.

TE	TE_0	TE_1
P(TE = TEs | v00, v10, v20, v30 , v40, v50 , v60)	1	0
P(TE = TEs | v01, v10, v21, v31 , v40, v50 , v61)	0.965	0.035
P(TE = TEs | v01, v10, v21, v31 , v41, v50 , v61)	0.926	0.074
P(TE = TEs | v00, v11, v21, v31 , v41, v50 , v61)	0.926	0.074

Table 7
Conditional probability distribution of node TE (Top event)

4.1. Dynamic Updating of CPTs

as shown above. With time, we collect a series of dataInitially, when t = 0, we use $P_{exploitability,s}$ to define CPTs, points that provide information on threat intel scores and exploit availability. In addition, at each data point, we look for evidence of whether a security node was breached or a top event was triggered. This evidence is used to update the Bayesian probabilities of each security vulnerability and the subsequence event nodes. We have a series of 33 data points with dynamic information available. We will randomly insert evidence of WAF misconfiguration being exploited between these data points to see how the CPTs change over a given period and how the financial asset losses correspondingly change with time.

5. Discussion and Results

This section provides an attack scenario based on the capital breach case to evaluate our model. We utilised 33 datasets where the WAF configuration (top event) was exploited as evidence at t=6. Therefore, we initially started with the probability of exploitability defined by $P_{exploitability,s}$ in the above section. At each time step, the code reads the dataset sheet to look for the dynamic exploitability scores: TI and ES. It also looks for evidence as to whether the top event (WAF misconfiguration) was triggered or not. We synthetically added evidence at a user-defined timestep for the WAF misconfiguration to be

exploited. In the three graphs of Figure 2, this has been done at t=6 units.

At time t<6 units, the code looks through the datasheet and finds no evidence of exploitation, and it hence updates the probability of exploitability of the security vulnerabilities (v0, v1, v2...). After t=6, there is an unexpected jump in the exploitability probability values, the corresponding asset losses, and the probabilities of failure of the different servers. For t>6, the same process is again repeated, where the code looks at the TI and ES scores and identifies whether the WAF misconfiguration was exploited. At t=25, we again see a jump in exploitability values. This again means that the TI and EA scores might have increased at those times. Figure 3 shows another example where WAF misconfiguration was exploited at t=11.

In Figures 4 and 5, we see the dynamic update occurring with a Bayesian approach without considering the TI and EA score update. In this case, all the 3 figures are relatively smooth except a peak that occurs at t=6 and t=11 time units respectively. This is because at these time instances, we find evidence for a WAF misconfiguration as has been shown in Figures 1 and 2. Due to the absence of TI and EA scores,

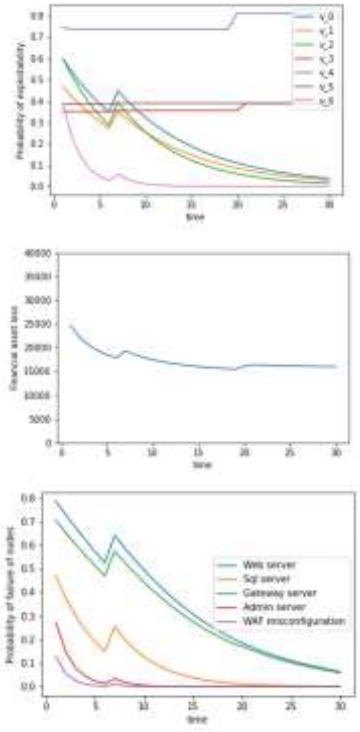

Figure 2: Probability of exploitability, financial asset loss, probability of failure of nodes with time where WAS misconfiguration is exploited as evidence at t=6. Dynamic update includes a Bayesian update plus the inclusion of TI and EA scores.

we do not see the increase in the exploitation probability of exploitability financial asset loss, failure of nodes (at t=21) units.

Another example where WAF misconfiguration was exploited at t=11 is shown in Figure 3.

6. Conclusion

Risk assessment is an integral part of risk management. It is aimed at proactively identifying threats and vulnerabilities that target assets and applying mitigation strategies to reduce the risks to an acceptable level. This paper proposes the Cloud Enterprise Dynamic Risk Assessment (CEDRA) model that uses CVSS, threat intelligence feeds and exploitation availability in the wild using dynamic Bayesian networks to predict vulnerability exploitations and financial losses. The probability of successful exploitation and financial losses is calculated by identifying CVE for each asset and then constructing a bow tie model based on the Capital One breach use case of 2019. The conditional probability distributions are achieved by AND and OR logic gates. We introduced two new parameters into the CVSS 3.1 standard: exploitation availability and threat intelligence. The framework is based on a dynamic Bayesian network that facilitates an underlying process of continuously identifying and assessing risks in the cloud environment. The current

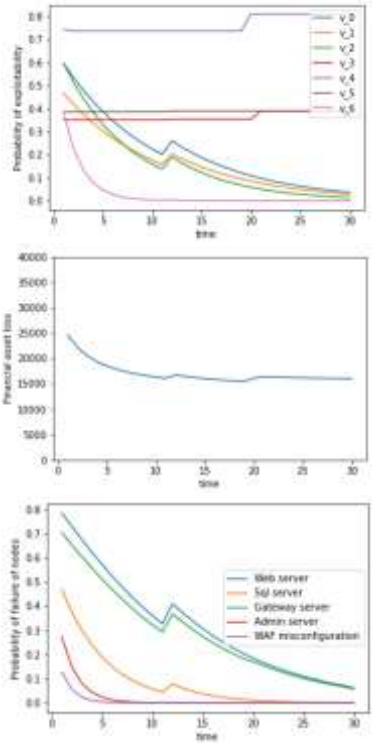

Figure 3: Probability of exploitability, financial asset loss, and probability of failure of nodes with time where WAF misconfiguration is exploited as evidence at t=11. The dynamic update includes a Bayesian update plus the inclusion of TI and EA scores.

study has shown that combination of bow-tie analysis, including dynamic Bayesian network, threat intelligence and and information about exploitation availability in the wild has improved vulnerability and financial losses prediction. However, the work could be further enhanced by introducing data asset value, as it is currently limited to asset purchasing cost and location of the asset.

7. Conflict of interest

The authors have declared no conflict of interest.

8. Declaration

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Dawood Behbehani: Conducted research, wrote the manuscript, and prepared all figures and tables.

Nikos Komninos, Khalid Al-Begain, and Muttukrishnan Rajarajan: Planning and supervision of the work.

All authors reviewed the results and approved the final version of the manuscript

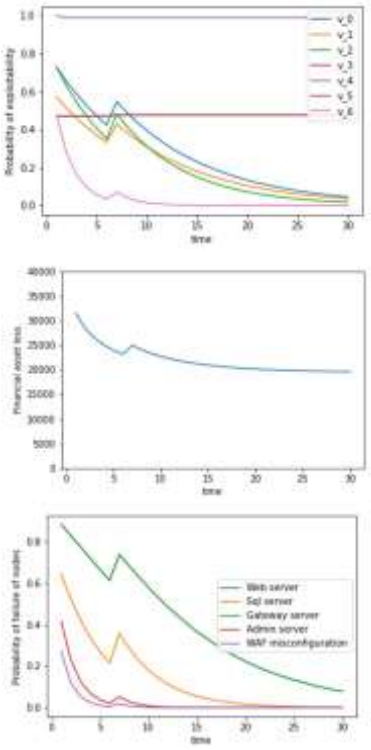

Figure 4: Probability of exploitability, financial asset loss, probability of failure of nodes with time where WAF misconfiguration is exploited as evidence at t=6. Dynamic update includes a Bayesian update without the TI and EA scores.

References

[1] , . NVD - CVE-2019-2828. URL: https://nvd.nist.gov/vuln/detail/CVE-2019-2828. [2] , . What is SSRF (Server-side request forgery)? Tutorial & Examples | Web Security Academy. URL: https://portswigger.net/web-security/ssrf. [3] Ahmadi, O., Mortazavi, S.B., Mahabadi, H.A., Hosseinpouri, M., 2020. Development of a dynamic quantitative risk assessment methodology using fuzzy DEMATEL-BN and leading indicators. Process Safety and Environmental Protection 142, 15–44. doi:10.1016/j.psep.2020.04.038.

[4] Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W., Alyami, H., Ayaz, M., 2021. A Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies. IEEE Access 9, 57792– 57807. doi:10.1109/ACCESS.2021.3073203.

[5] Andrade, R.O., Yoo, S.G., Tello-Oquendo, L., Flores, M., Ortiz, I., 2022. Integration of AI and IoT Approaches for Evaluating Cybersecurity Risk on Smart City , 305–333URL: https://link.springer.com/chapter/10.1007/978-3-030-87059-1_12, doi:10.1007/978-3-030-87059-1{_}12. [6] B, B., George, P., Renjith, V.R., Kurian, A.J., 2020. Application of dynamic risk analysis in offshore drilling processes. Journal of Loss Prevention in the Process Industries 68, 104326. doi:10.1016/j.jlp.2020.104326. [7] Berenjian, S., Shajari, M., Farshid, N., Hatamian, M., 2016. Intelligent Automated Intrusion Response System based on fuzzy decision making and risk assessment, in: 2016 IEEE 8th International Conference on Intelligent Systems, IS 2016 - Proceedings, Institute of Electrical and Electronics Engineers Inc.. pp. 709–714. doi:10.1109/IS.2016.7737389. [8] Huang, K., Zhou, C., Tian, Y.C., Tu, W., Peng, Y., 2017. Application of Bayesian network to data-driven cyber-security risk assessment in SCADA networks, in: 2017 27th International Telecommunication Networks and Applications Conference, ITNAC 2017, Institute of Electrical and Electronics Engineers Inc.. pp. 1–6. doi:10.1109/ATNAC.2017.8215355. [9] Khosravi-Farmad, M., Ghaemi-Bafghi, A., 2020. Bayesian Decision Network-Based Security Risk Management Framework. Journal of Network and Systems Management 28, 1794–1819. URL:

https://www.researchgate.net/publication/343400848_BayesianDecision_Network

-Based_Security_Risk_Management_Framework, _ doi:10.1007/S10922-020-

09558-5.

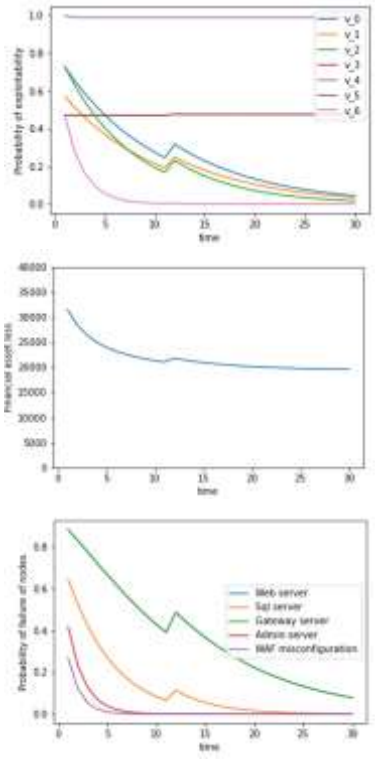

Figure 5: Probability of exploitability, financial asset loss, probability of failure of nodes with time where WAF misconfiguration is exploited as evidence at t=11. Dynamic update includes a Bayesian update without the TI and EA scores.

[10] Kim, J., Shah, A.U.A., Kang, H.G., 2020. Dynamic risk assessment with bayesian network and clustering analysis. Reliability Engineering and System Safety 201, 106959. doi:10.1016/j.ress.2020.106959.

[11] Larriva-Novo, X., Vega-Barbas, M., Villagrá, V.A., Rivera, D., Sanz, M., Álvarez-Campana, M., 2020. Dynamic risk management architecture based on heterogeneous data sources for enhancing the cyber situational awareness in organizations, in: ACM International Conference Proceeding Series, Association for Computing Machinery. doi:10.1145/3407023.3409224.

[12] Li, M., Liu, Z., Li, X., Liu, Y., 2019. Dynamic risk assessment in healthcare based on Bayesian approach. Reliability Engineering and

System Safety 189, 327–334. doi:10.1016/j.ress.2019.04.040.

[13] Liu, Z., Ma, Q., Cai, B., Liu, Y., Zheng, C., 2021. Risk assessment on deepwater drilling well control based on dynamic Bayesian network. Process Safety and Environmental Protection 149, 643–654. doi:10.

1016/j.psep.2021.03.024. [14] Lyu, X., Ding, Y., Yang, S.H., 2020. Bayesian Network Based C2P Risk Assessment for Cyber-Physical Systems. IEEE Access 8, 88506– 88517. doi:10.1109/ACCESS.2020.2993614.

[15] Naumov, S., Kabanov, I., 2016. Dynamic framework for assessing cyber security risks in a changing environment, in: 2016 International Conference on Information Science and Communications Technologies, ICISCT 2016, Institute of Electrical and Electronics Engineers

Inc. doi:10.1109/ICISCT.2016.7777406.

[16] Neto, N.N., Madnick, S., Moraes, A., De Paula, G., Borges, N.M., Novaes, N., Cybersecurity, N., Sloan, M., Madnick Cybersecurity, S., De, G., C6 Bank, P., Malara, N., C6 Bank, B., 2020. A Case Study of the Capital One Data Breach (Revised) A Case Study of the Capital One Data Breach .

[17] Nguyen, S., Chen, P.S.L., Du, Y., Shi, W., 2019. A quantitative risk analysis model with integrated deliberative Delphi platform for container shipping operational risks. Transportation Research Part E:

Logistics and Transportation Review 129, 203–227. doi:10.1016/j.tre.2019.08.002. [18] Oberoi, A., Dave, Y., Patel, B., Anas, M., 2021. Cloud Computing in Banking Sector-A Case Study. International Journal of Scientific Research & Engineering Trends 7, 2395–566.

[19] Poolsappasit, N., Dewri, R., Ray, I., 2012. Dynamic security risk management using Bayesian attack graphs. IEEE Transactions on

Dependable and Secure Computing 9, 61–74. doi:10.1109/TDSC.2011.34.

[20] Qin, Y., Zhang, Q., Zhou, C., Xiong, N., 2020. A Risk-Based Dynamic Decision-Making Approach for Cybersecurity Protection in Industrial Control Systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems 50, 3863–3870. doi:10.1109/TSMC.2018.

2861715. [21] Riesco, R., Villagrá, V.A., 2019. Leveraging cyber threat intelligence for a dynamic risk framework: Automation by using a semantic reasoner and a new combination of standards (STIX™, SWRL and OWL). International Journal of Information Security 18, 715– 739. URL: https://doi.org/10.1007/s10207-019-00433-2, doi:10.

1007/s10207-019-00433-2. [22] Sasubilli, M.K., Venkateswarlu, R., 2021. Cloud Computing Security Challenges, Threats and Vulnerabilities. Proceedings of the 6th International Conference on Inventive Computation Technologies,

ICICT 2021 , 476–480doi:10.1109/ICICT50816.2021.9358709.

[23] Sauve, G., Van Acker, K., . Integrating life cycle assessment (LCA) and quantitative risk assessment (QRA) to address model uncertainties: defining a landfill reference case under varying environmental and engineering conditions. The International Journal of Life Cycle Assessment 1, 3. URL: https://doi.org/10.1007/

s11367-020-01848-z, doi:10.1007/s11367-020-01848-z.

[24] Tam, K., Jones, K., 2019. Factors affecting cyber risk in maritime, in: 2019 International Conference on Cyber Situational Awareness, Data Analytics and Assessment, Cyber SA 2019, Institute of Electrical and

Electronics Engineers Inc. doi:10.1109/CyberSA.2019.8899382.

[25] Tounsi, W., Rais, H., 2018. A survey on technical threat intelligence in the age of sophisticated cyber attacks. Computers & Security 72, 212–233. doi:10.1016/J.COSE.2017.09.001.

[26] Wagner, T.D., Mahbub, K., Palomar, E., Abdallah, A.E., 2019. Cyber threat intelligence sharing: Survey and research directions. Computers & Security 87, 101589. doi:10.1016/J.COSE.2019.101589.

[27] Wang, J., Fan, K., Mo, W., Xu, D., 2016. A method for information security risk assessment based on the dynamic Bayesian network, in: Proceedings

- 2016 International Conference on Networking and Network Applications, NaNA 2016, Institute of Electrical and Electronics

Engineers Inc.. pp. 279–283. doi:10.1109/NaNA.2016.50.

[28] Wu, S., Zhang, L., Zheng, W., Liu, Y., Lunteigen, M.A., 2016. A DBN-based risk assessment model for prediction and diagnosis of offshore drilling incidents. Journal of Natural Gas Science and

Engineering 34, 139–158. doi:10.1016/j.jngse.2016.06.054.

[29] Yu, H., Khan, F., Garaniya, V., 2016. Risk-based process system monitoring using self-organizing map integrated with loss functions. The Canadian Journal of Chemical Engineering 94, 1295–1307. URL:

http://doi.wiley.com/10.1002/cjce.22480, doi:10.1002/cjce.22480.

[30] Zangeneh, V., Shajari, M., 2018. A cost-sensitive move selection strategy for moving target defense. Computers & Security 75, 72–91. doi:10.1016/J.COSE.2017.12.013.

[31] Zhang, L., Wu, S., Zheng, W., Fan, J., 2018a. A dynamic and quantitative risk assessment method with uncertainties for offshore managed pressure drilling phases. Safety Science 104, 39–54. doi:10.

1016/j.ssci.2017.12.033. [32] Zhang, Q., Zhou, C., Tian, Y.C., Xiong, N., Qin, Y., Hu, B., 2018b. A Fuzzy Probability Bayesian Network Approach for Dynamic Cybersecurity Risk Assessment in Industrial Control Systems. IEEE

Transactions on Industrial Informatics 14, 2497–2506. doi:10.1109/

TII.2017.2768998.

Dawood Behbehani is a cybersecurity professional in the financial sector with more than 12 years of experience in IT and information security. He obtained a bachelor's degree in computer information systems from Kingston University, London and completed a master's degree in computer security at the University of De Montfort, UK. Dawood currently works at Kuwait International Bank as the executive manager in the Information Security, Privacy Anti-Fraud department, which is aimed at ensuring customer's privacy and combating information security threats and financial corruption. He has attained numerous professional security certifications, including the Certified Information Security Manager and Certified Data Privacy Solutions Engineer certificates. He is currently pursuing a PhD in information engineering at City, University of London.

Dr Nikos Komninos received his PhD in 2003 from Lancaster University (UK) in Information Security. He is currently a Senior Lecturer (US System: Associate Professor) in Cyber Security in the Department of Computer Science at City University London. Since 2000, he has participated, as a researcher or principal investigator, in a large number of European and National RD projects in the area of information security, systems and network security. He has authored and co-authored more than 80 journal publications, book chapters and conference proceedings publications in his areas of interest. He has been invited to give talks at conferences and Governmental Departments, as well as to train employees in Greek and UK businesses.

Professor Khalid Al-Begain is the founding President of Kuwait College of Science and Technology. He served as the President of the European Council for Modelling and Simulation (ECMS) (2006-2018) and as President of the Federation of European Simulation Societies (EuroSim) (2010-2013). He was the co-founder and chairman of the first National Welsh Industrial Cyber Security Summit, in Newport, UK in 2014 organised jointly with Airbus, General Dynamics and the UK NCSC. He was a consultant for Telekom Malaysia for planning the first 3G network in Malaysia. He won numerous awards including the John von Newman Computer Award (1986) and the Inspire Wales Award for Science and Technology (2013). In 2006, he received Royal Recognition from Her Majesty the Queen for his contributions to the British scientific community. He registered two granted patents, authored/edited 26 books, and more than 200 papers in refereed journals and conferences. He was the general chair of 28 international conferences.

Professor Rajarajan (Raj) is the founding Director of the Institute for Cyber Security at City University of London and the CEO of Citydefend Limited and CTO of TechInspire two successful start-up companies in the area of privacy preserving data sharing online to comply with GDPR requirements. Raj is also an academic advisor to British Telecommunications (BT) security research in Adastral Park, UK and continues to work closely with BT to supervise industry sponsored PhD students in the areas of identity, trust, privacy and security. He was part of the team that worked on the UK Government's Verify UK programme in which he contributed towards the privacy by design aspects of the overall architecture. Raj has extensive experience in working on cutting edge commercial projects in the areas of encrypted search in the cloud, privacy preserving data analytics and cloud security. He has published more than 350 publications, three books and hold two patents in the area of cloud data privacy. He continues to work closely with cyber security start up companies to innovate through novel science of cyber security. He is a Senior Member of Institute of Electrical and Electronic Engineering and a full member of the Charted Institute of Information Security.