Application of URREF Criteria to Assess Knowledge Representation in Cyber Threat Models

Valentina Dragos, Jurgen Ziegler, Johan Pieter de Villiers

Application of URREF Criteria to Assess Knowledge Representation in Cyber Threat Models

1st Valentina Dragos
ONERA - The French Aerospace Lab
Palaiseau, France
valentina.dragos@onere.fr

2nd Jürgen Ziegler
Competence Centres ISR, IABGmbH
Ottobrunn, Germany
jziegler@iabg.de

3rd Johan Pieter de Villiers
University of Pretoria and CSIR
Pretoria, South Africa
pieter.devilliers@up.ac.za

***Abstract*—Systems for threat analysis enable users to understand the nature and behavior of threats and to undertake a deeper analysis for detailed exploration of threat profile and risk estimation. Models for threat analysis require significant resources to be developed and are often relevant to limited application tasks. This paper investigated the implicit and explicit uncertainty assessments to be taken into account for threat analysis systems to be effective for providing a relevant threat characterization. The intent of this paper is twofold. The first is to present and discuss an approach to define a model for cyber threats within a simplified expert model and to translate it into a Bayesian network as a tool for the development of practical scenarios for cyber threats analysis. The second is to address the question of assessing the Bayesian network build and its intrinsic knowledge representation model and to show how modeling decisions impact the outcome of the system. The paper describes the construction of an expert model and the corresponding BN to analyze cyber threats, investigates various types of induced uncertainty with the URREF criteria simplicity and expressiveness and implements an assessment procedure to evaluate the overall approach.**

***Index Terms*—cyber threats, Bayesian inference, knowledge representation, uncertainty, URREF ontology, simplicity, expressiveness**

I. Introduction

Bayesian Networks (BNs) provide a natural and efficient way to represent causal models for decision making under uncertainty. A challenge when using BNs for various applications is the construction and maintenance of the BN - i.e. the conditional probability tables (CPTs) and a-priori distributions. For real life applications, BNs may have hundreds of nodes and complex structures with many nodes having multiple parent nodes.

There are several approaches to perform the parameterization of the CPTs. This can be carried out by building the BNs from large data sets, thanks to machine learning techniques able to process data provided by statistical research or by creating the BNs from expert models, designed to represent expert knowledge.

Whether learned or build by experts, there is a need to continuously asses if the knowledge representation encoded in the BN structure is rich enough to capture data attributes, to represent interactions and causal relations and to offer a reliable support to provide credible results.

In this paper we present a large BN generated from expert models and developed for the holistic assessment for cyber threats and assess the quality of its intrinsic knowledge representation according to URREF criteria. URREF is an ontology developed within the ISIFs Evaluation of Techniques for Uncertainty Representation Working Group (ETURWG) in order to asses different aspects of uncertainty in information fusion systems.

We select a set of URREF criteria that are relevant to analyse the quality of the knowledge representation and implement the associated evaluation process. The discussion is illustrated with an example for cyber threat detection in a system of systems context and we argue that taking into account the uncertainty offers additional output layers that are of interest to end users and analysts. The paper also seek to enrich the ongoing discussion at the ETUR working group on assessment of knowledge representations.

The reminder of the paper is organized as follows : Section II presents the technical and methodological foundations of cyber threat detection with a Bayesian network-based approach and the analysis of uncertainty with URREF criteria. Section III focus on BNs construction for cyber threat detection, discusses knowledge representation and presents the application context. Section IV tackles the analysis of uncertainty with URREF criteria while section V discusses the assessment process and results. Concluding remarks and directions for future work are presented in section VI.

II. Technical and Methodological Foundations

A. Bayesian Networks

BNs represent and depict graphically the cause and effect relationship between various elements and provide a mean to incorporate uncertainty associated to elements and their interactions. [1]. Elements of a domain and their inherent states are represented by nodes of a graph; causality relation is modeled as edges of the graph and uncertainty associated to dependent nodes is quantified or parameterized by CPTs.

A simple example in Fig. 1 illustrates the general principle on BNs.

The probability that a car has a specified color will usually depend on the type of the car (the red Ferrari) and the year of construction (whether a color is fashionable). These dependencies can be modeled by conditional probabilities, e.g. $p(\text{red}|\text{Ferrari}) = 0.9, p(\text{red}|2015) = 0,062$. All combinations of cars, years and regarded colors determine the conditional

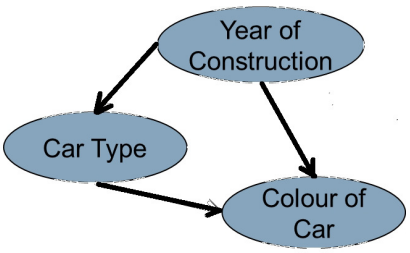

Fig. 1. Simple Bayesian Network

probability tables for this simple model. This BN supports the following reasoning types of *analysis* ("Assumed a car was built in 2011, what is the probability that it is yellow?") and *evidence* ("I observed a yellow Ferrari, what are the probabilities of the possible years of construction?).

The graphical representation of BNs highlights qualitative knowledge about links between variables while quantitative information about the strength of the relationships is captured by local distributions. With this modeling technique it is possible to show (probabilistic) relationships among many causally related variable and thus BNs are used for different types of problems incorporating risk and safety assessment [2]. Examples are in the field of crime risk analysis ([3], [4]), reliability analysis for safety critical environments [2], terrorism risk [5], credit-rating [6] and estimation of default probability for large companies [7].

BNs are also used to investigate some aspects of Air Traffic Management (ATM) applications and for cyber security analysis [8], [9]. Recently, BN were employed to assess cyber threats in the ATM area [10] to identify the existence of the adversary, his intention and the level of competence.

In the context of cyber threat analysis, however, there is limited information available on how BNs can be created and used in practice. This paper contributes to the topic by developing a model that combines observed prior indicators of cyber threats and description of vulnerabilities to predict the probability of a cyber attack at organization level. The model combines both malicious intentions and technical vulnerabilities to detect external or insider threats posed by a group of insiders and relies on both expert knowledge and data (such as journal logs).When changing the observations or the expert knowledge, the probabilities for different scenarios for threat identification can be determined. In this way the best combination of measures maximizing the probability of a cyber attack given certain prior indicators can be identified.

B. Automatic Generation of BN from Expert Models

For the purpose of this work we use a BN which is automatically generated from an expert model [11].

First, the structure of the expert model highlighting domain objects and their dependencies is defined by domain experts. Moreover, dependencies within the expert model are weighted qualitatively according to an ergonomic approach called scale-based distribution retrieval [12]. We used the seven qualitative values: impossible, very unlike, unlike, unclear, probable, very probable and sure.

Then the domain model is translated into a well-defined BN by processing the following steps:

- Analyze whether domain objects comprise mutual exclusive states or not. If yes, the domain object of the expert can be kept as nodes of the BN, otherwise all states of the domain object are represented as binary nodes wherein the original state having true or false values.
- Generate the dependencies within the BN based on the dependencies of the expert model. If the child node in the expert model is divided into n binary nodes, then all the n dependencies to these nodes must be generated.
- Translate qualitative values of the dependencies into numerical values according to the method of scale-based information retrieval. First, the defined weighting values are transformed into scale values. The scale values are predefined values between 0 and 1 preserving the sequence which is given by the meaning of the qualitative values. If the child node has more than one state, the resulting table column values are normalized as demanded by probability theory. In the binary case the column is completed by using 1-translated value. The result of this step are tables representing the qualitative dependencies between the states of the expert model.
- Calculate the final CPTs. If the child node has only one parent node, the already available table is used as CPT. If the child node has multiple parent nodes, it must be determined whether the parent nodes are related with an "OR" relation or with an "AND" relation. In the first case, at the position of the CPT which represents the dependency value of the according states of the parent nodes, the maximum of the according values of the parent node tables must be used . In the second case, the product of these values must be used (see also ([9] for an easy example). If a child node has multiple parent nodes, the last step must be processed iteratively.

It should be noted that the result of the calculation within the BN must be translated back to the qualitative representation. This part is out of the scope of this paper.

C. URREF Ontology for Uncertainty Analysis

The uncertainty representation and reasoning evaluation framework (URREF) ontology [13] is a unified frame developed to provide a set of criteria for uncertainty analysis and evaluation in information fusion systems. The ontology defines criteria to capture different types of uncertainty regarding the sources and data inputs of the system, internal representations of data and knowledge and the associated automated processing and reasoning, and results and outputs of the information fusion process.

The URREF ontology has four main evaluation criteria classes. The first gathers criteria associated with data handling under the general concept DataHandlingCriterion and includes Data interpretation and Traceability as evaluation criteria.

The second class is called RepresentationCriterion and characterizes the quality of domain knowledge representation through five criteria : Knowledge handling, Simplicity, Expressiveness, Adaptability and Compatibility.

The third class named ReasoningCriterion captures how well reasoning procedures performs and includes the following evaluation criteria : Correctness, Consistency, Performance (Throughput and Timeliness), Computational cost and Scalability.

The fourth class is named DataCriterion and consists of criteria relating to quality of input and output data, the reliability of sources and the impact of taking into account specific variables on the results. Criteria of this class are not listed here as they are not considered in the rest of the paper.

The main subjects under evaluation [14] for URREF ontology are uncertainty representation and reasoning components of the fusion systems, but the models define criteria for secondary evaluation subjects such as sources of information, piece of information, fusion methods and mathematical formalisms. URREF criteria have generic definitions and can be instantiated for applications with coarse or finer granularity levels: evaluation metrics can be defined for data analysis [15], or more particularity for data specific types [16] or attributes: reliability and credibility [17], trust and self-confidence [18] or veracity [19]. While allowing a continuous analysis of uncertainty representation, quantification and evaluation [20], URREF criteria are detailed enough to capture model-embedded uncertainties [21], their propagation in the context of the decision loop [22] and offer a basis to compare different fusion methods [23]. URREF criteria served as a basis for uncertainty tracking and investigation for several applications: vessel identification for maritime surveillance [24], activity detection for rhino poaching [25] and imagery for large area protection [26].

III. Modeling Cyber Threats Detection with Bayesian Networks

A. A Holistic Model for Cyber Threats

To make the analysis specific we adopt a scenario related to the cyber security in a large organization, e.g. in air traffic control (ATC) (for example Eurocontrol). The scenario has as central element, namely a system of systems (SoS), which requires cyber security protection, and is composed of a large number of workplaces at different sites. The general approach uses a holistic model as described in [27] to assess the probability of different (future) threats and attack vectors. The system includes computers, software, and the communication links connecting various sites. Elements of the system are continuously under attacks triggered by hackers but also by some more sophisticated actors such as criminal organizations, competitors etc. A temporal development of the scenario is considered, which evolves through a number of stages to generate a situational picture describing the current state of threats and their expected evolution. For the purpose of this paper we adopted a holistic model whose structure is showed in Fig. 2. The holistic model follows the STIX standard, largely adopted within the community [28]. Main components of the STIX model are:

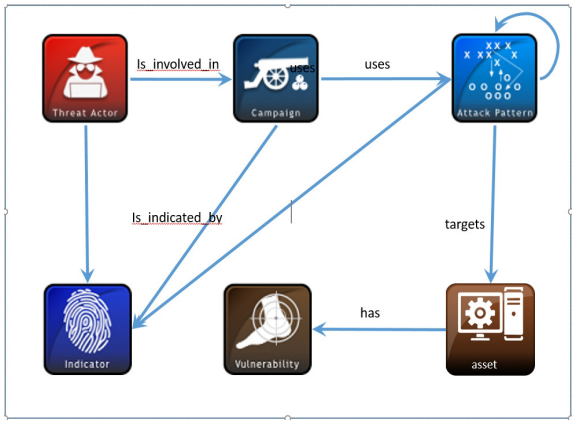

Fig. 2. High Level representation of the holistic model

- threat actor: any individual, group, or organization believed to have malicious intentions.
- threat campaign: adversarial behaviors associated to a set of malicious activities or attacks that occur over a period of time against a specific set of targets.
- attack patterns: type of tactics, techniques, and procedures (TTP) that describes the way threat actors attempt to compromise targets.
- threat indicators: specific pattern that can be used to detect suspicious or malicious cyber-activities.
- system vulnerability: weakness in software that can be directly used by a hacker to gain access to a system or network. In addition, week password protection or leaving a computer turned on or physically accessible to visitors are another class of vulnerabilities not directly related to software.

The model captures the intuition that any system is subject to vulnerability and can become a target to several campaign attacks having specific patterns and perpetrated by different actors. Indicators are needed to detect at early as possible the intent of what an attacker is trying to accomplish.

It should be noted that assets are not defined in the current STIX standard but instead are used to describe SoSs. Arrows describe the set of dependencies of the elements of the expert model. They also loosely follow the STIX standard, and some of them differ from the STIX standard, e.g. we reversed the STIX relation 'Indicator indicates attack pattern' to 'attack pattern is indicated by indicator'. This e.g. allows to analyze which indicators may be relevant for the detection of a specific threat. Every element of this structure is detailed using a set of properties and dependencies.

The next section shows the detailed structure of the holistic model and highlights internal and external relations of attack pattern, vulnerability and asset elements.

B. Identification of Attacks from Expert Knowledge on Vulnerabilities

The assessment of knowledge representation is performed with a specific part of the expert model and the corresponding BN. They describe the estimation of the expected probability of success of an attack against a specific asset of our system of systems. The sub-model of the expert model describing this part of an attack comprises the following elements (see Fig. 3:

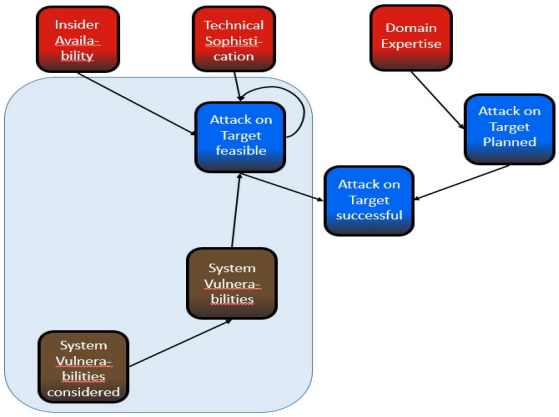

Fig. 3. Detail of holistic model - Vulnerability Model

This part of the expert model comprises the following elements, states and dependencies:

- "Availability of Insiders": States represent insiders with different access to systems. The availability of insiders is not mutually exclusive since a threat actor might have insiders of all types.
- "Technical Sophistication": States are "has ..." and "has not technical sophistication", which are exclusive. A high technical sophistication leads to higher probability of an intrusion attempt.
- "Domain Expertise": States are "has ..." and "has not domain expertise", which are exclusive as well. The domain expertise is required for the appropriate selection of the data element for a domain specific attack (e.g. the selection of an item of a flight plan)
- "System Vulnerabilities considered": This element allows to decide whether a specific component shall be considered within the analysis. We assume as a simplification that a sub-system may comprise an operating system, a security software (e.g. a firewall) and the application software, which is the destination of the attack. We used artificial system components to avoid publishing any real system architecture description. We regarded more than twenty components.
- "System Vulnerabilities": The corresponding BN CPT values of the arrow to the "System Vulnerabilities" encodes the expert guesses about vulnerabilities of these technical components. If System Vulnerabilities considered is "True", the vulnerability values are inherited to these system components which comprise a specific technical part. These states of the expert model are not mutually exclusive.
- "Attack on target feasible": We regarded twelve types of components of the ATM SoS. We assumed that every attack vector is decomposed as : "attack a periphery sub-system" of the architecture, "attack the appropriate protection sub-system" (e.g., a firewall) and after that "attack the internal sub-system". Therefore the periphery system is parent of the protection system, which is parent of the internal system. Altogether we have thirty-six states of this element, which represent whether a specific sub-system can / is attacked or not. They are not exclusive. The internal systems are parents states of "attack on target successful", since the successful intrusion into the appropriate sub-system allows for the malicious actions.
- "Attack on target planned" and "Attack on target successful": An attack is planned if the goals of an threat actor are realizable with the specific malicious action. Malicious actions are different variants of collection of data, denial of service, starting a BOT process, manipulate data, disrupt system processes and manipulate system processes. We regarded more than fifty malicious actions. The attack is successful if it is planned and feasible.

Bayesian networks are constructed to represent the uncertainties in the process of analysis and prediction of the most probable threats as well as for the early detections of dangerous activities of sophisticated attackers (so-called Advanced persistent threats APT). The corresponding BN has the following nodes:

- The elements with mutually exclusive states, i.e. the technical sophistication and the domain expertise are represented by a node with the same states.
- All the other elements of the expert model are represented by binary nodes, e.g. a state representing the SoS component"InuCen 2.0 protection system" coming from the element "attack on target feasible" has the states true or false representing whether the attack against this system part is feasible or not.
- The dependencies of the BN are created according to the dependency values as defined in the expert model using the method as described in II-B. Fig. 4 shows a small part of the BN showing all parents and child of the node InuCen 2.0 internal system, which was a state of attack on target feasible in the expert model.

The InuCen (version 2) internal system consists of hardware components, the Ken Linux Version 6 operating system and the InuCen software vulnerability. The success of an attack depends on the vulnerability values of these components AND on the successful attack against the InuCen protection system. It also depends on the technical sophistication of the threat actor OR whether there is a staff member which can perform the attack. If the InuCen2.0 internal system is attacked successfully the threat actor may e.g. manipulate flight plan data at the internal system. In this case, all parent nodes are binary, resulting in the CPT having 64 values.

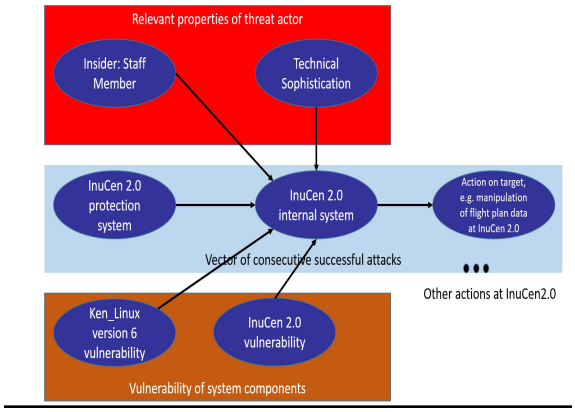

Fig. 4. Small excerpt of the Bayesian Network

IV. Assessment of Knowledge Representation

A. Model Instantiation and Model Based Calculations

The model has two practical applications: analysis of possible threats and their effects and situational analysis (using indicators).

This paper focuses on the first application. The application will be performed in several iterations.

1) A working version of the cyber threat analysis model is created by the domain expert. This step allows the expert to consciously decide how variables, interactions, and causal relations should be modeled.
2) The model is translated into a BN, this BN representation is used in order to analyze different scenarios and to check whether it provides good quality results.
3) The quality of results are assessed and the model is improved until a first acceptable version is created.
4) This accepted version is used e.g. to analyze the danger of various threats orto make recommendations according to system design specifications.

Since the application of the model will lead to improved knowledge of the expert, they will improve the expert model as well and so on. We performed several iterations of the expert model and the BN until we had an initial acceptable model which generated results in accordance with expert expectations.

The design cycle allows for the expert model to be adjusted following e.g. changes of system components (i.e. version update, new patch to be loaded). Hence there is a need to continuously assess the quality of its knowledge representation without carrying out empirical testing. Moreover, the BNs handle uncertainty as a joint probability distribution over a set of uncertain variables. The intrinsic model is parameterized by local probability distributions generated from the assessments of the domain experts.

Creating BNs, especially for end-users not necessarily familiar with probability notions, results in multiple challenges. First, threat indicators and system vulnerabilities related to a specific cyber threat should be searched for, and than tailored to a specific system. Filling the conditional probability tables of the nodes is also quite challenging, since there is a limited amount of data available for this task. In practice, the best way to generate the tables is by using experts assessments. Because elements of the BN are generated in a automatic way and the BNs do not contain detailed variable descriptions it may become difficult to understand the resulting BN structure in detail. Finally, to avoid model complexity, the number of parents of a node and the number of states must be limited. However, the smaller the number of states and connections, the lower the accuracy of the model and the quality of its results.

To investigate the usefulness of the model in practice, we analyze the nature of uncertainty induced by modeling decisions and expert assertions. Tracking uncertainties from BN construction to probability of successful attack estimation is achieved by applying the URREF criteria. The goal is to make explicit the uncertainty arising when the problem to be solved is abstracted by BNs and its intrinsic knowledge structure and data representations are simplified in order to fulfill constraints of its specific formalism.

B. Selection of Relevant URREF Criteria

For the use case considered in this paper, uncertainty enters the BNs in three main forms: uncertainty of variable transformation or accuracy uncertainty, uncertainty of model structure or causality uncertainty, and reasoning uncertainty encompassing uncertainty in the CPTs and its propagation during the estimation of marginal posterior probability distributions.

Those uncertainties can be evaluated according to criteria under two main classes: *Representation* and *Reasoning*.

RepresentationCriterion is a general class for several criteria explaining how uncertainty is characterized, captured and stored and introduces *Simplicity*, *Adaptability* and *Expressiveness* criteria.

ReasoningCriterion is a general class compassing criteria capturing how well the system performs inferences and gathers the following criteria : *ComputationalCost*, *Consistency*, *Correctness*, *Scalability* and *Performance*.

As assertions of experts have an impact of the quality of the model and therefore on the BNs generated, we investigate the uncertainties related to Knowledge representation and more specifically the *Simplicity* and *Expressiveness* criteria. Both criteria are assessed at model level.

C. Metrics Definition and Overall Assessment

We use the *Simplicity* and *Expressiveness* criteria to asses both the expert model and its BN representation. In order to define metrics for both criteria let's consider the following variables :

N the number of nodes in the network and the model
N_s, the number of states, N_c the number of connections,
N_p the number of parameters in the model.
S_g, the average significance of the parameters: If the expert says something is sure it is more expressive than if he says it

is probable. (We used the possible weighting parameter values "impossible", "very unlikely", "unlikely", "unclear", "probable", "very probable" and "sure".) We weigh the parameters' significance with 3 for "sure" and "impossible", 2 for "very probable" and "very unlikely", 1 for "probable and unlikely" and 0 for "unclear" and calculate the average significance by adding the single significance of all parameters and dividing it by the number of parameters.

Currently *Simplicity* is defined in the URREF ontology as: "Simplicity assesses the system's ability to execute common operations without requiring deep knowledge about its inner details"
The introduction of the expert model was triggered by simplicity considerations, since on the one hand the domain experts should be able to encode their knowledge and to work with the system, on the other hand they are usually not able to deal with large Bayesian networks. BNs allow for the segmentation of complex problems into smaller and more manageable subnetworks. As a result BNs have a favorable score when measured according to this qualitative metric of simplicity.

To apply the criterion we must define the operations which are be supported. These are: generation of the expert model by domain experts, calculation of results, understanding the results and update of the model, and finally analysis of the BN by BN experts e.g. for quality assurance.

Intuitively, a first approach to quantitatively measure simplicity is to consider the number of nodes, the number of dependencies, and the number of states and exclusive states in the network.

In order to define a metric for *Simplicity* we define first an additional metric called *ergonomic complexity* EC as:

$$EC = log(N) * [log(N_{ds} - 1) + log(N_c) + log(N_p)] \quad (1)$$

A minimal network with two nodes, two states per node, two connections per node and two parameters per node gets a complexity value 1.0. Theoretically complexity has no upper limit. Therefore the values are within the interval 1.0 up to infinite.

The number of parameters in a BN is dependent on the number of states and the number of connections. It might seem like double counting to include all of these counts in EC, but when it comes to the ease of domain expert knowledge elicitation, the number of nodes and states per node are of interest. It captures whether experts are able to maintain a coherent view of the problem properties during the definition and parametrisation of the models. The same holds for the number of parameters.

Thus we define *ergonomic simplicity* ES as the inverse of *ergonomic complexity*, i.e.

$$ES = (1/EC) \quad (2)$$

Since the criterion *Simplicity* is defined as inverse of *Complexity*, the values of the metric *ergonomic simplicity* are elements of $]0.0, 1.0]$.

This metric captures the intuition that a small network with few nodes, states, dependencies and parameter values is rather simple.

The second URREF criterion considered to assess the knowledge representation if *Expressiveness*. The URREF ontology defines *Expressiveness* as a "Measure of the power of a knowledge representation formalism to convey all relevant aspects of a given fusion problem." Under URREF, *Expressiveness* has two main sub criteria: *Assessment* is a "Measure of the ability of the system to handle the types of uncertainty assessments (e.g., verbal, quantitative, combined) needed for a given problem, and to distinguish them from one another.", while *Dependency* is the "Ability of the uncertainty representation to capture dependency among propositions (e.g., cause and effect, relevance, statistical association)." Considering the notation above, the metric *model expressiveness* ME is defined as:

$$ME = log(N-1) * [log(N_s) + log(N_c) + log(N_p * S_g)] \quad (3)$$

The interval for the values of ME is also limited to $[1.0, \infty]$. According to the URREF definitions the formula is dedicated to the "Dependency" part of the criterion. If $S_g = 1$ (which correlates to the intermediate weightings probable/unlike, ME has the same value as EC. The value for ME is higher than for EC, if the intermediate weight is sharper. This corresponds to the intuitive assumptions, that a model might be complex but not so expressive if the network contains many information sources with low influence on the result.

The intuition behind ME is that the more parameters, connections, states at node level and significant parameters a network has, the more able it is to capture and describe entities and interactions of the model. The assessment criteria can be measured by binary results, i.e. whether a specific knowledge representation different types of uncertainty assessments or not. Both criteria were assessed on iterations of two distinct expert models.

V. Model Iterations and Analysis of Results

A. Iteration Example

Several iterations are needed before building a model with a suitable structure and model parameters. One aspect was the introduction of the domain expertise. Since the original STIX model does not contain this, this model element was introduced within the iteration process. Before introducing this element, the "Action on target planned" depended only on the intended effect of the specific campaign. For illustration we regard the intended effect "Degradation of Service Air Traffic Control" (see Table I). The expert assessment is that this effect can be *probably* realised by the attack "manipulation of Flight Data at InuCen 2.0". The CPT of "Action on Target Planned" contains the translation of all dependencies, the specific weights are translated into Table I.

The *Dependency* between "Domain Expertise" and "manipulation of Flight Data at InuCen2.0" was weighted as if the threat actor "has Domain Expertise". It is "Very Probable" that he knows that and how this attack can be used for the

TABLE I
CPT VALUES WITHOUT DOMAIN EXPERTISE

	Degradation of Service ATC
True	0.667
False	0.333

intended effect. Now the CPT also contains the dependencies of all planned actions from the domain expertise, the resulting values of the CPT are presented in Table II.

TABLE II
CPT VALUES WITH DOMAIN EXPERTISE

	Degradation of Service ATC has Domain Expertise	Degradation of Service ATC has Not Domain Expertise
True	0.91	0.2857
False	0.09	0.7143

Now we regard a threat actor with the type "Market Competitor", who is assured to have "Domain Expertise". If the "Domain Expertise" is not taken into account the probability that this threat actor will "manipulate the flight data at InuCen 2.0" depends only on his "Intended Effect", his (high) "Technical Sophistication" and "Availability of Insiders". The value of "Manipulation of Flight Data Planned" is 0.667, the value of "Manipulation of Flight Data Realized", which takes into account the probability of successful intrusion into InuCen2.0, is 0.468, which is back-translated into the qualitative result value "Unclear". If the "Domain Expertise" is regarded the values change to "Manipulation of Flight Data Planned" = 0.91 and "Manipulation of Flight Data Realized" = 0.586, which is back-translated into the qualitative result value "Probable".

B. Assessment of URREF Uncertainty Criteria

URREF criteria were used to asses the uncertainty representation of two expert models having 16 and 15 elements ('nodes') respectively. Table III shows the values for the input for uncertainty assessment procedure. The first values are for the models without the element "Domain Expertise".

TABLE III
ASSESSMENT INPUT

	Expert Model I	Expert Model II	BN I	BN II
N	15	16	672	673
N_s	780	782	1452	1456
N_c	19	21	1187	1222
N_p	1195	1307	31375	34139

TABLE IV
UNCERTAINTY ESTIMATION: ERGONOMIC COMPLEXITY AND ERGONOMIC SIMPLICITY

	Expert Model I	Expert Model II	BN I	BN II
EC	8.52	8.826	30.339	30.494
ES	0.117	0.113	0.0329	0.0327

Table IV shows the *ergonomic complexity* and *ergonomic simplicity* values for both expert models and BNs. The calculation shows that the expert models are significantly less complex than the BN generated from them. This confirms the experience during the definition and parametrization of these models. Since the definition of the states and the weightings can be performed element by element, it was possible to define the expert model if an appropriate HMI is available to support the task. On the other hand it was sometimes difficult to check whether the generated BN was adequate since there are many connections between the generated nodes, especially if binary nodes were generated from elements of the expert model. Nevertheless the results also imply that the expert models are not so simple either. Since the formulas as defined are not yet tested for different models with different sizes and complexities, it might also be sensible to adopt them after a thorough testing.

The significance S_g of the weightings are calculated as 1.204 for model I and 1.226 for model II respectively. Therefore, the values for *model expressiveness* are a little bit higher than for *ergonomic complexity*. Table V shows the results.

TABLE V
UNCERTAINTY ESTIMATION: MODEL EXPRESSIVENESS

	Expert Model I	Expert Model II	BN I	BN II
ME	9.259	9.674	32.932	33.39

The results show a limitation of the formula as defined. The formula can be used to compare the *model expressiveness* of networks, but the absolute values are not easy to be interpreted. Therefore the results demonstrate the sequence of complexity of our exemplary models as expected, but it is not obvious whether e.g. an absolute ME value of 10.0 is high or not. We close the measurement of this criterion with a remark on the *Assessment* sub-criterion. Both, the expert model and the BN, do only support one type of uncertainty (qualitative and probabilities). If the experts do have statistical results about some of the dependencies, it might become necessary to provide the possibility to define probabilistic weightings within the expert model as well.

C. Implications for the URREF Ontology

The need for a more detailed investigation of uncertainties is exposed when modeling the detection of cyber-threats with Bayesian networks. The URREF ontology offers a unified basis to analyze inaccuracies affecting the intrinsic knowledge representation. The URREF ontology has a good level of granularity and uncertainty decomposition, allowing basic but also more sophisticated uncertainty analysis. However, we see two needs for improvement of the URREF ontology to make it more practicable: Firstly, there is no guidance how to define and use metrics for the criteria to support an overall assessment of an uncertainty representation. It might be helpful to give a guideline for the definition of numerical measures, so that they can be combined for an overall assessment. An initial

attempt is presented for data criteria in [15]. Also in the current version of the URREF ontology criteria do not cover all types of uncertainties induced by practical restrictions; for the presented work the URREF ontology does not allow to estimate the uncertainty induced by mutual transposition of linguistic and numerical values (e.g. *0.55* becomes *probable*)

VI. Conclusions and Future Work

This paper presents a Bayesian network-based approach developed to for cyber-threats analysis and illustrates the use of URREF criteria to characterize the quality of knowledge representation at the core of the approach. A BN for cyber treats is automatically generated from expert knowledge, highlighting vulnerabilities of systems along with threat-specific patterns, actors and indicators.

The BN is used to estimate the probability of detecting successful attacks based on experts' assessment of system vulnerabilities. Uncertainty analysis focuses on characterizing the intrinsic knowledge representation, as the model is dynamic and its structure is continuously adapted to cope with modifications in the real-life system.

The main direction for future work might be to define and apply further metrics for the URREF criteria for knowledge representation, to add an explanatory level to describe the uncertainties induced by experts assertions and modeling choices and how much can be avoided or reduced by taking additional measures. Additionally the criteria shall be applied to several BN with different complexity and size to validate the results and update the criteria definitions if necessary. Another direction might be to analyze, implement and apply links between the criteria for knowledge representation to other criteria of the URREF ontology (e.g. expressiveness should be linked to the accuracy of the results).

References

[1] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.

[2] N. Fenton and M. Neil, *Risk assessment and decision analysis with Bayesian networks*. CRC Press, 2012.

[3] G. C. Oatley and B. W. Ewart, "Crimes analysis software:pins in maps, clustering and bayes net prediction," *Expert Systems with Applications*, vol. 25, no. 4, pp. 569–588, 2003.

[4] R. Boondao, V. Esichaikul, and N. K. Tripathi, "A bayesian network model for analysis of the factors affecting crime risk," *WSEAS Transactions on Circuits and Systems*, vol. 3, no. 9, pp. 1895–1900, 2004.

[5] D. C. Daniels, L. D. Hudson, K. B. Laskey, S. M. Mahoney, B. S. Ware, and E. J. Wright, "Terrorism risk management," *Bayesian Networks: A Practical Guide to Applications*, pp. 239–262, 2008.

[6] P. Wijayatunga, S. Mase, and M. Nakamura, "Appraisal of companies with bayesian networks," *International Journal of Business Intelligence and Data Mining*, vol. 1, no. 3, pp. 329–346, 2006.

[7] E. Ejsing, P. Vastrup, and A. L. Madsen, "Predicting probability of default for large corporates," *Bayesian networks: a practical guide to applications*, pp. 329–344, 2008.

[8] P. Xie, J. H. Li, X. Ou, P. Liu, and R. Levy, "Using bayesian networks for cyber security analysis," in *Dependable Systems and Networks (DSN), 2010 IEEE/IFIP international conference on*. IEEE, 2010, pp. 211–220.

[9] S. Y. K. Mo, P. A. Beling, and K. G. Crowther, "Quantitative assessment of cyber security risk using bayesian network-based model," in *Systems and Information Engineering Design Symposium, 2009. SIEDS'09.* IEEE, 2009, pp. 183–187.

[10] D. Kolev, R. Koelle, R. A. C. Rodriguez, and P. Montefusco, "Security situation management-developing a concept of operations and threat prediction capability," in *Digital Avionics Systems Conference (DASC), 2015 IEEE/AIAA 34th*. IEEE, 2015, pp. 4C2–1.

[11] J. Ziegler and B. Haarmann, "Automatic Generation of Large Causal Bayesian Networks from User Oriented Models," in *Proceedings of the 6th Workshop on Sensor Data Fusion (INFORMATIK 2011: Informatik schafft Communities)*. Citeseer, 2011.

[12] M. Krüger and J. Ziegler, "User Oriented Bayesian Identification and its Configuration," in *Proceedings of the 11th international conference on information fusion, Cologne, Germany*. IEEE, 2008.

[13] P. C. Costa, K. B. Laskey, E. Blasch, and A.-L. Jousselme, "Towards unbiased evaluation of uncertainty reasoning: The urref ontology," in *Information Fusion (FUSION), 2012 15th International Conference on*. IEEE, 2012, pp. 2301–2308.

[14] P. de Villiers, G. Pavlin, P. Costa, A.-L. Jousselme, K. Laskey, V. Dragos, and E. Blasch, "Subjects under evaluation with the urref ontology," in *Information Fusion (Fusion), 2017 20th International Conference on*. IEEE, 2017, pp. 1–8.

[15] J. de Villiers, R. Focke, G. Pavlin, A. Jousselme, V. Dragos, K. Laskey, P. Costa, and E. Blasch, "Evaluation metrics for the practical application of urref ontology: An illustration on data criteria," in *Information Fusion (Fusion), 2017 20th International Conference on*. IEEE, 2017, pp. 1–8.

[16] V. Dragos, "An ontological analysis of uncertainty in soft data," in *Information Fusion (FUSION), 2013 16th International Conference on*. IEEE, 2013, pp. 1566–1573.

[17] E. Blasch, K. B. Laskey, A.-L. Jousselme, V. Dragos, P. C. Costa, and J. Dezert, "Urref reliability versus credibility in information fusion (stanag 2511)," in *Information Fusion (FUSION), 2013 16th International Conference on*. IEEE, 2013, pp. 1600–1607.

[18] E. Blasch, A. Jøsang, J. Dezert, P. C. Costa, and A.-L. Jousselme, "Urref self-confidence in information fusion trust," in *Information Fusion (FUSION), 2014 17th International Conference on*. IEEE, 2014, pp. 1–8.

[19] E. Blasch and A. Aved, "Urref for veracity assessment in query-based information fusion systems," in *Information Fusion (Fusion), 2015 18th International Conference on*. IEEE, 2015, pp. 58–65.

[20] J. P. de Villiers, K. Laskey, A.-L. Jousselme, E. Blasch, A. de Waal, G. Pavlin, and P. Costa, "Uncertainty representation, quantification and evaluation for data and information fusion," in *Information Fusion (Fusion), 2015 18th International Conference on*. IEEE, 2015, pp. 50–57.

[21] A.-L. Jousselme, "Semantic criteria for the assessment of uncertainty handling fusion models," in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 488–495.

[22] J. P. de Villiers, A.-L. Jousselme, A. de Waal, G. Pavlin, K. Laskey, E. Blasch, and P. Costa, "Uncertainty evaluation of data and information fusion within the context of the decision loop," in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 766–773.

[23] V. Dragos, X. Lerouvreur, and S. Gatepaille, "A critical assessment of two methods for heterogeneous information fusion," in *Information Fusion (Fusion), 2015 18th International Conference on*. IEEE, 2015, pp. 42–49.

[24] A.-L. Jousselme and G. Pallotta, "Dissecting uncertainty-based fusion techniques for maritime anomaly detection," in *Information Fusion (Fusion), 2015 18th International Conference on*. IEEE, 2015, pp. 34–41.

[25] H. Koen, J. P. de Villiers, G. Pavlin, A. de Waal, P. de Oude, and F. Mignet, "A framework for inferring predictive distributions of rhino poaching events through causal modelling," in *Information Fusion (FUSION), 2014 17th International Conference on*. IEEE, 2014, pp. 1–7.

[26] E. Blasch, P. C. Costa, K. B. Laskey, H. Ling, and G. Chen, "The urref ontology for semantic wide area motion imagery exploitation," in *Aerospace and Electronics Conference (NAECON), 2012 IEEE National*. IEEE, 2012, pp. 228–235.

[27] T. Kiesling, M. Krempel, J. Niederl, and J. Ziegler, "A model-based approach for aviation cyber security risk assessment," in *Availability, Reliability and Security (ARES), 2016 11th International Conference on*. IEEE, 2016, pp. 517–525.

[28] Introduction to stix, https://oasis-open.github.io/cti-documentation/stix/intro.