An Argumentative-Narrative Risk Assessment Model

Floris Bex, Bas Hovestad
Department of Information and Computing Sciences
Utrecht University
Utrecht, The Netherlands

Abstract— **In this paper, we propose an argumentative-narrative risk assessment model (ANRAM), which takes a formal logical theory of scenarios and arguments, and presents them in a structured, visual language that captures the basic concepts of risk assessment: risk factors, controls, evidence. We also present a set of specific risk scenarios from the domain of football fan violence, and show how ANRAM has been implemented in a collaborative decision-support system for the Dutch Police.**

Keywords— Risk Assessment, Argumentation, Scenarios

I. INTRODUCTION

Assessing and acting on risks is one of the core tasks of any police or intelligence organization. To this end, constant risk assessments are performed in a variety of areas that are part of the organization's main tasks, such as public order, crime and terrorism. The primary purpose of these risk assessments is to identify, describe, estimate and evaluate risks, so as to provide information and insight to decision makers.

To support the risk assessment process, a plethora of languages, techniques and tools to describe and evaluate risks exist [7]. One major problem, however, is that many of these existing techniques are too simple, too complex, or highly domain-specific. Simple techniques such as risk matrices [7] and cause-and-effect diagrams lack the expressive power to capture more complex risk scenarios, in which a series of seemingly unimportant events together cause something that poses a significant risk. Advanced techniques that allow for a precise and structured representation of risk scenarios, such as Bayesian networks [2], require extensive knowledge of mathematical and formal models. Furthermore, less complex structured models are often tailored for specific domains, such as the CORAS method for cyber-risk assessment [5].

In sum, there is no existing language for describing and evaluating risks in intelligence and law-enforcement domains, which can be used by domain experts that might lack the necessary formal or mathematical training. The lack of such a language, and associated tools, hampers risk assessment. Often, there is a proliferation of informal and bespoke analysis methods across departments and domains. Theoretically, many of these methods are quite basic [3], relying instead on the relevant expert's knowledge and experience to evaluate risks and controls in a more ad-hoc way.

What is hence needed is a structured, domain-independent model that can be used to describe risk scenarios and evaluate these risk scenarios based on the evidence provided and the controls proposed. The model should be formal enough to be implemented in decision-support tools, but natural enough to be understood by experts who are not mathematically inclined. Furthermore, while the model itself should be domain-independent, it should be able to also capture more domain-specific knowledge elicitated from experts and documents.

The solution we propose in this paper is the argumentative-narrative risk assessment model (ANRAM), in which *risk narratives or scenarios* can be described and evaluated using *arguments based on evidence*. The model is based on the hybrid theory of evidential reasoning for criminal cases [1]. This hybrid theory combines ideas from computational argumentation and logical model-based reasoning into one logical theory for evidential reasoning. This formal basis makes ANRAM amenable to implementation, whilst the combination of scenarios and arguments is natural enough to be used by domain experts from intelligence and police .

ANRAM was developed in conjunction with the Dutch National Police in the RISK project[1], in which tools were developed to make sense of the risks surrounding football fan violence. In addition to the model, we elicited a number of scenario schemes, templates for typical risk scenarios that often occur in the context of football fan violence, which help in the construction of new scenarios and can be used to analyze existing scenarios. Finally, a risk assessment tool on the basis of a simplified version of ANRAM was developed for the iTable, a 40" multi-touch table that allows users to collaboratively construct and evaluate risk scenarios.

Section II explains ANRAM by means of a small example. Section III briefly discusses the scenario schemes for football fan violence that were developed in conjunction with the Dutch National Police. Section IV then discusses the iTable for collaborative risk analysis, and section V concludes the paper.

II. THE AN-RAM MODEL

The Argumentative Narrative Risk Assessment Model is based on the hybrid theory of stories and arguments [1], which provides a means to make sense of evidence and facts in legal

[1] http://www.florisbex.com/RISK.html.

cases. The core components of the hybrid theory are stories – also called scenarios – and arguments.

Scenarios are sequences of events about what happned in the past or may happen in the future. In the domain of football fan violence, a common scenario is that on a match day, two rival clubs are supposed to play as the away team in different parts of the country. The supporters of these clubs are taken to the two different stadiums of the home clubs by bus. However, the routes of the different buses cross and, to make matters worse, the buses stop at the same petrol station, so the two supporter groups clash violently.

Arguments are used to reason from evidence, such police data and expert reports, to claims. For example, the claim that "the supporters of Feyenoord and Ajax often clash" can be supported by an expert who states that indeed these two clubs have a history of violent clashes. As in logic, the inferences in arguments are based on (defeasible) inference rules. For example, in general we believe an expert who states a claim about something which they are in a position to know about – more specifically, we believe an expert on Ajax and Feyenoord when he states that these clubs often clash.

Arguments can be used to support scenarios: if an expert states that it is very likely two groups of supporters will clash, a scenario in which this takes place also becomes more likely. Arguments can also be used to attack scenarios or other arguments. For example, if another expert says that a clash is highly unlikely, this attacks the scenario in which the supporters fight. Also, if we have another counterargument stating that the first expert – who claimed that a clash is likely –is acting on obsolete information, we can say that his claims no longer support the scenario.

The hybrid theory is a logical theory that integrates logical model-based diagnosis with a computational account of argumentation [1]. Scenarios, in the form of causal sequences, can be compared and analyzed using arguments, inference trees built on evidence from a knowledge base. The acceptability of different sets of scenarios and arguments can then be calculated using argumentation semantics from artificial intelligence.

The formal basis of the hybrid theory makes it amenable to implementation, and additionally allows for automated reasoning given possibly conflicting sets of arguments and scenarios. However, a formal logical theory is difficult to explain to lay-people. In [5], we therefore propose a structured conceptual language that captures the core elements of the hybrid theory in a more simple to understand visual format. Furthermore, we add concepts that are specific to risk assessment, such as *risk factor* and *control* to the language.

A. Syntax of the model

Central to AN-RAM are *scenarios* (blue box). A scenario has a plausibility – how likely is the scenario? – and an impact – if the scenario takes place, how severe are the effects of the scenario? A scenario is made up of at least one *risk factor* (orange box), a single state or event that presents a risk. These risk factors have their own prior plausibility and impact, which is summed for all risk factors in a scenario to calculate the plausibility and impact of the scenario. Fig. 1 shows an example of a simple scenario containing 3 risk factors.

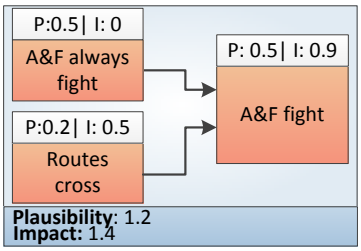

Fig. 1. A risk scenario with three risk factors.

Risk factors can be supported by evidential arguments (purple boxes, Fig. 2). Here, the type of evidence, e.g., expert testimony or data, is mentioned, as well as plausibility. Plausibility is propagated to the risk factor supported by the evidence. For example, the prior plausibility of the risk factor 'A&F often fight' was 0.5 (Fig. 1), but the expert evidence increases this to 0.8 (Fig. 2). Support is indicated by a normal arrow: between risk factors in a scenario, this arrow can be interpreted as a 'causes' relation, while between evidence and a risk factor the arrow denotes an 'is evidence for' relation.

Risk factors can also be attacked (square-headed arrow). One way to attack a scenario is by providing evidence, such as an expert opinion, to the contrary (Fig. 3). Another way is to provide a *control* that mitigates a risk factor (green box) – one way to control fighting hooligans is to deploy riot squads (Fig. 3). Note that a control may itself cause a new risk, further increasing the plausibility and impact of the risk scenario (see Fig. 3, where the control 'deploy riot squads' leads to a new risk factor 'fight with police').

Attacking evidence or controls defeat a risk factor if their plausibility is higher than the risk factor. In Fig. 3, the expert evidence arguing against 'A&F often fight' does not defeat the evidence arguing for this risk factor, but the 'deploy riot squad' control defeats the 'A&F fight' risk factor'. A defeated risk factor's plausibility and impact do not count towards the total of the scenario. In the example in Fig. 3, the high impact of the 'A&F fight' risk factor is deducted from the total scenario impact, but the impact of the new risk factor 'fight with the police' is added to the scenario's impact.

It is further possible to attack the supportive link between evidence and risk factor, so that the evidence does not propagate its plausibility and impact to the risk factor (Fig. 4). If, for example, we have evidence that the expert did not base his claims about A&F on relevant information, we can attack this argument, thus setting the plausibility and impact of the risk factor back to its prior values from Fig. 1.

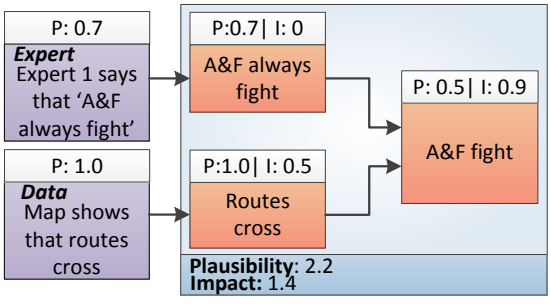

Fig. 2. A risk scenario with three risk factors.

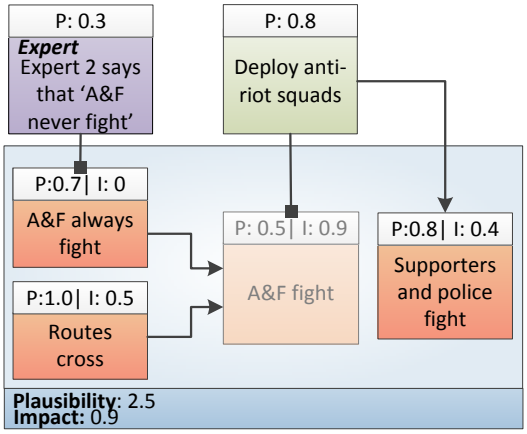

Fig. 3. A risk scenario with three risk factors.

III. Scenario Schemes for Risk Assesment

ANRAM can be used to capture and reason about any type of risk scenario. For risk analysis in a particular domain, however, it is also helpful to have domain-specific scenario schemes. Such schemes, which represent generic types of scenarios that often occur in the domain, can act as a template for constructing and analysing specific risk scenarios. Based on interviews about football fan violence at the Dutch National Police, we have constructed a set of scenario schemes [5], a few of which we will discuss here.

Scenario scheme: risk due to routes

- *Pattern of actions:* crossing routes → fight between group X and group Y
- *Relevant risk factors:* road construction, traffic situation
- *Relevant controls:* change routes
- *Relevant information:* nature of X/ Y, routes of other groups

Scenario scheme: risk due to rivalry

- *Pattern of actions:* group X and group Y always fight due to club rivalry → fight between X and Y
- *Relevant risk factors:* history of the match, position on league table, relationship between supporters/team.
- *Relevant controls:* ban/limit on alcohol, regulation of ticket sales, deploy stewards, contact with individual supporters before match, deploy riot squads
- *Relevant information:* nature of clubs, nature of X and Y

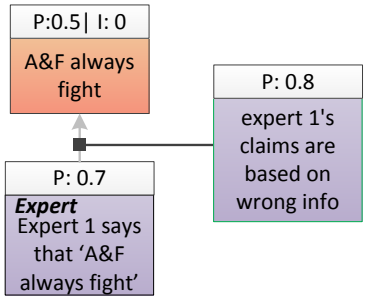

Fig. 4. Attacking the link between evidence and risk factor.

Scenario scheme: risk due to riot squads

- *Pattern of actions:* deployment of riot squads → fight between supporters and police
- *Relevant risk factors:* strictness and enforcement of rules, number of risk supporters
- *Relevant controls:* regulation of ticket sales, deploy stewards, plainclothes police
- *Relevant information:* nature of group X and group Y

Pattern of actions is the central scenario template that can be filled in with specific information to get a concrete scenario. *Relevant risk factors* are other, related risk factors that play a role in the scenario, and *relevant controls* are possible controls that can be used to attack the risk explained or the central action. Finally, *relevant information* contains other information that might be of importance for the scenario.

Scenario schemes enable risk analysts to develop new scenarios and uncover risk factors and controls for existing scenarios more quickly and easily. Take, for example, the risk scenario in Fig. 1. In this scenario, there is a risk due to rivalry and a risk due to crossing routes. Looking at the relevant risk schemes, there are a number of controls that can be applied to mitigate these risks. While 'deploy riot squads' is one of them, we can see that this leads to a new risk, as it is the central action of the 'risk due to riot squads' scheme. The easier solution would be to change the routes of the buses, taking into account the traffic situation, possible road works and the routes of other buses (cf. relevant risk factors and information in the 'routes cross' scenario scheme').

IV. Collaborative Risk Assessment

ANRAM is mainly targeted towards police analysts, who have the time and inclination to work out specific risk scenarios in more detail when performing a risk assessment. Our case-study on football fan violence at the Dutch National Police [5], however, showed that there was also a need for more ad-hoc, collaborative risk assessment, during which the relevant parties should get an overview of the upcoming matches and possible risk scenarios based on the latest information and intelligence data. For this, a software tool called 'Tesseract' was developed by students of the University of Utrecht.

Tesseract visualises a map of the Netherlands to which various types of geo-information layers can be added – e.g. the traffic situation, locations of petrol stations, football stadiums, and so on. Tesseract further allows for simulation of supporter flows across the map, so it can be determined if, when and where these flows of supporters might run into each other. Tesseract is constantly fed by data from the police systems, so new incidents and individuals involved in incidents are added to the system and can subsequently be analysed and viewed.

Tesseract implements ANRAM by allowing users to construct and reason about scenarios based on schemes that have been implemented into Tesseract. Figure 5 shows a scenario in the Tesseract tool. The scenario concerns a possible fight ('Vechtpartij A') near the city of Utrecht, the map of which can be seen in the back. Important risk factors ('Factoren') are the Galgenwaard stadium of FC Utrecht, the bus route ('Combiroute') and a well-known hooligan ('Uli

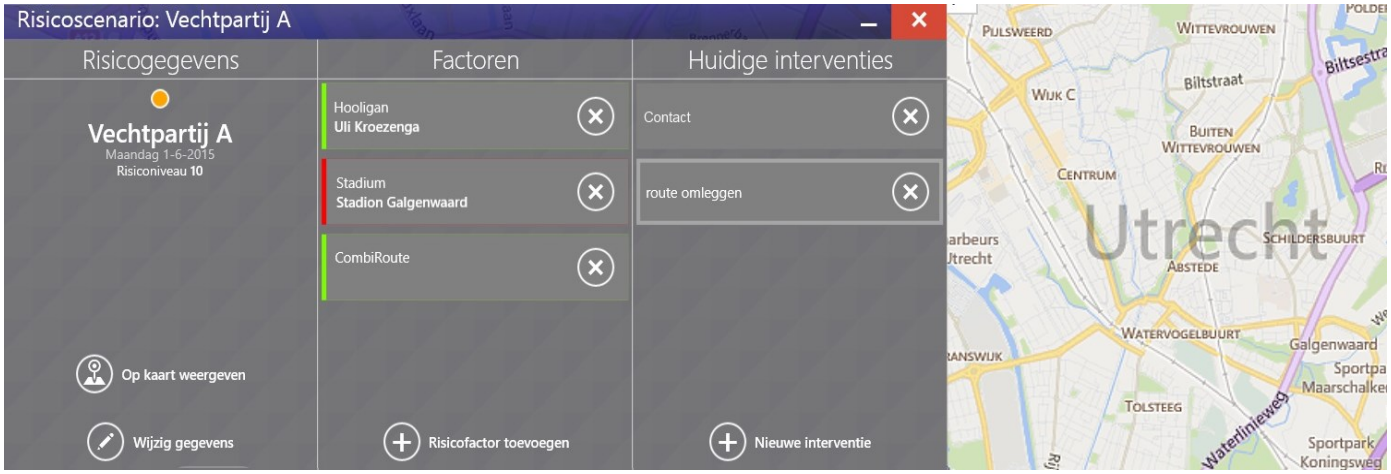

Fig. 5. A risk scenario as rendered on top of a map in Tesseract

Kroezinga'). A number of controls ('Huidige interventies) have been applied: Uli has been contacted ('Contact'), and the route has been changed ('Route omleggen'). These controls attack the relevant risk factors, thus mitigating the risk these factors pose, which is indicated by the green bars next to these two risk factors. The risk factor related to the situation around the stadium has not been attacked and still poses a risk (indicated by the red bar next to it). The scenario as a whole poses a medium risk, shown by the orange circle on the left – if another control would be applied to the stadium, the risk would be further mitigated and this circle would turn green.

Because the objective of the Tesseract tool is to provide collaborative risk assessment, the software runs on an iTable, a 40" touch table that allows multiple users to work simultaneously (Fig. 6). This promotes an informal working styles, where multiple analysts together work on a risk assessment. The idea is that standing around the iTable will promote a dialogue between the analysts, thus resulting in more than just the sum of the individual analyses because of the collaborative, communicative working style [3].

V. Conclusion

We have proposed ANRAM, a structured, semi-formal language that captures risk scenarios, arguments based on evidence that support or counter these scenarios, and controls.

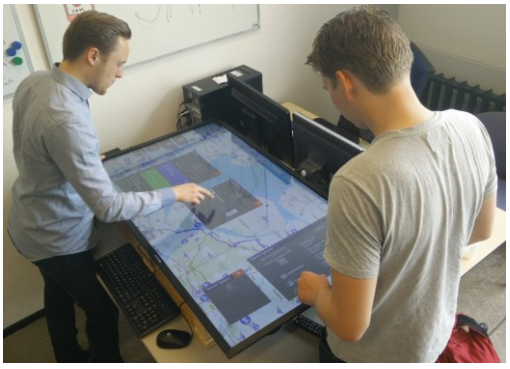

Fig. 6. Multiple people working on the iTable version of Tesseract

The model is broad enough for general risk analysis and adheres to common risk assessment concepts, such as plausibility, impact and cause-and-effect. This makes it easy to use and understand. On the other hand, ANRAM has a formal basis in the form of the hybrid theory [1], which allows for future extensions such as detecting inconsistencies or proposing possible scenarios or controls. With respect to the latter, ANRAM can be easily combined with domain-specific knowledge in the form of scenario schemes.

The version of ANRAM running on the iTable allows for collaborative risk assessment, and shows us that (semi-)formal models such as ANRAM can be used for different purposes; analysts can use the full power of ANRAM, including possibilities for automated reasoning, while at the same time operational teams can use ANRAM at the iTable, which allows for a more informal style of reasoning about risk scenarios. In the future, we will further evaluate the use of ANRAM at different levels of the police organization.

References

[1] Bex, F. J., Van Koppen, P. J., Prakken, H., & Verheij, B. (2010). A hybrid formal theory of arguments, stories and criminal evidence. Artificial Intelligence and Law, 18(2), 123-152.

[2] Fenton, N., and Neil, M. (2012). Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press.

[3] Gutwin, C., Greenberg, S., Blum, R. Dyck, J. Tee, K. and McEwan, G. (2008) Supporting informal collaboration in shared-workspace groupware. Journal of Universal Computer Science, 14:9, 1411-1434.

[4] Hengst, M. den, Rovers, B., and Regterschot, H. (2014). *Intelligence bij evenementen: Een inventarisatie van risicomanagementpraktijken bij de politie.* (EN:*Intelligence at large events: taking stock of risk management practices at the police*). Den Haag: Boom.

[5] Hovestad, B. and Bex, F. (2016). Making Sense of Risks: A Hybrid Argumentative Narrative Approach to Risk Assessment. *Utrecht University Technical report* UU-CS-2016-006.

[6] Lund, M. S., Solhaug, B., & Stølen, K. (2010). *Model-Driven Risk Analysis*. Springer.

[7] Rausand, M. (2011). *Risk assessment: Theory, methods, and applications* (Vol. 115). John Wiley & Sons.

[8] Schank, R.C. (1986) *Explanations Patterns: Understanding Mechanically and Creatively*, Lawrence Erlbaum, Hillsdale (NJ).