Information Acquisition for Bayesian Network-based Threat Analysis

Ronnie Johansson[a,b], and Christian Mårtenson[b]
[a] Informatics Research Centre, University of Skövde
[b] Swedish Defence Research Agency (FOI)

Abstract— In this work, we describe and evaluate a proof of concept implementation of a resource allocation mechanism for a threat analysis support system. The system uses a Bayesian network to structure information requests, and the goal is to minimize the uncertainty of a non-observable variable of interest.

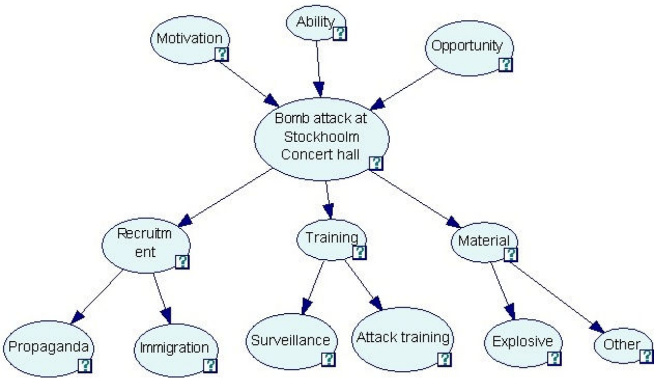
I. Introduction

Determining how to utilize information acquisition resources optimally is a difficult task in the security domain. Nevertheless, a security analyst can expect little or no support for this from software tools today. At the Swedish Defence Research Agency, we are developing and situation and threat analysis support tool, Impactorium, which helps an analyst to model, structure, fuse and visualize information. In the tool, an incoming request for information can be turned into a detailed problem decomposition of sub-problems, which we represent using a Bayesian network (BN). However, even though the sub-problems (i.e., nodes in a BN) can be analyzed separately, the available information acquisition resources, used for observing nodes, are limited and shared. In addition all observations are uncertain due to sensing imperfection.

The goal of information acquisition is here to improve the belief estimate of a non-observable variable of interest, i.e., the *hypothesis* variable. This is accomplished by assigning resources to other (observable) variables called *indicators*. If multiple resources are at a system's disposal, more than one resource may be used for one indicator (hence resulting in multiple observations on the same indicator). Thus, the information acquisition mechanism we propose optimizes the expected belief estimate improvements while considering the results of possible resource assignments.

To concretize, consider the following example. The police are interested in knowing if there are any criminal groups that are planning a bomb attack against the Nobel Prize ceremony in the Stockholm Concert Hall. To investigate the case, an analyst makes a decomposition of the information request (hypothesis) into sub-problems with sub-hypotheses. The decomposition ends when the sub-hypotheses are concrete enough to enable real world observations to confirm or reject them. Figure X shows the decomposition of the main hypothesis "Bomb attack…" into a Bayesian Network. If the analyst finds information about e.g. increased activity on a related forum/website, the evidence for the "Propaganda" indicator is strengthened, which through Bayesian inference will affect the belief of the main hypothesis.

Now, consider the case where the analyst has a limited number of investigative resources (policemen) at hand to search for indicator evidence. Which is the optimal allocation strategy to minimize the uncertainty of the main hypothesis?



II. Results

We propose a resource assignment strategy (denoted Hypent) which assigns the resources in such a way that the expected decrease in entropy of the hypothesis variable variable according to information theory is maximized. We compare Hypent to a few different strategies to select a resource assignment: Rand, Ex-Greedy and Nx-Greedy. Rand assigns resources randomly to the indicators. The latter two strategies assign all resources to the single indicator with the highest entropy and the set of indicator*s* with the highest entropy, respectively.

In our initial experiments [1], we used a prototype BN consisting of a hypothesis variable and 4 indicator variables. When varying the observation uncertainty of the resources from 50 – 100 %, the results show that the different strategies have similar performance from 55% to 70%. From 75% the Hypent strategy dominates the others. As Hypent optimizes its performance with respect to the hypothesis variable, this could be expected, but the actual degree of observation certainty where this dominance starts is harder to guess. The disadvantage of the Hypent strategy is its running time, see [1].

The results indicate potential benefits of the proposed approach and further work will study the implications of these for the type of security application described above.

Acknowledgement

This work was funded by the R&D programme of the Swedish Armed Forces.

References

[1] R. Johansson, C. Mårtenson , "Information Acquisition Strategies for Bayesian Network-based Decision Support " in *Proceedings of the 13th Int Conf on Information Fusion,* 2010.