2011 Third International Workshop on Security Measurements and Metrics

Experiences from Using Indicators to Validate Expert Judgments in Security Risk Analysis

Olav Skjelkvåle Ligaarden*†, Atle Refsdal*, and Ketil Stølen*†
* Department for Networked Systems and Services, SINTEF ICT, PO Box 124 Blindern, N-0314 Oslo, Norway
E-mail: {olav.ligaarden, atle.refsdal, ketil.stolen}@sintef.no
† Department of Informatics, University of Oslo, PO Box 1080 Blindern, N-0316 Oslo, Norway

***Abstract*—Expert judgments are often used to estimate likelihood values in a security risk analysis. These judgments are subjective and their correctness rely on the competence, training, and experience of the experts. Thus, there is a need to validate the correctness of the estimates obtained from expert judgments. In this paper we report on experiences from a security risk analysis where indicators were used to validate likelihood estimates obtained from expert judgments. The experiences build on data collected during the analysis and on semi-structured interviews with the client experts who participated in the analysis.**

***Keywords*-security risk analysis; expert judgment; indicator**

I. Introduction

Much research report on procedures for eliciting expert judgment in risk analysis, decision support, and in general [1], [2], [3], [4]. There is also research that address the quality of expert judgments [5]. In this paper, however, we focus on the validation of the estimates obtained from expert judgments.

One way to validate likelihood estimates based on expert judgments is to use indicators calculated from historical data. Since we base ourselves on historical data, it is in most cases not possible to define indicators from which likelihood values can be inferred directly. For instance, in the case of the unwanted incident “eavesdropper reading a sensitive e-mail”, an obvious indicator would be the number of times this has occurred in the past. However, as it is normally not feasible to calculate this from historical data, we will have to make do with other indicators that are less to the point. One potential indicator for this unwanted incident could for example be “the number of encrypted sensitive e-mails sent”. Together with knowledge about the total number of sensitive e-mails being sent during a given period of time, this provides relevant input for validating the likelihood estimate.

In this paper we report on experiences from using indicators to validate expert judgments in a security risk analysis conducted in 2010. We build on data collected during the analysis and on semi-structured interviews with the client experts that participated in the analysis.

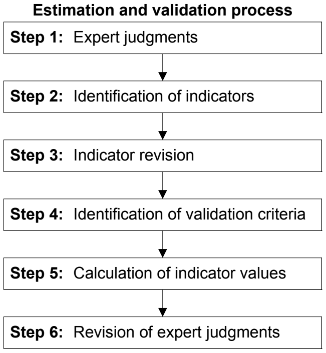
Estimation and validation process
Step 1: Expert judgments
Step 2: Identification of indicators
Step 3: Indicator revision
Step 4: Identification of validation criteria
Step 5: Calculation of indicator values
Step 6: Revision of expert judgments

Figure 1. The estimation and validation process

The rest of the paper is structured as follows: in Section II we present the security risk analysis from 2010. In Section III we present the data collected during the analysis. In Section IV we discuss this data in relation to input from the semi-structured interviews. Finally, in Section V we conclude.

II. Security risk analysis case

Our empirical study was integrated in a commercial security risk analysis based on CORAS [6] conducted in 2010. As the client of this analysis require full confidentiality we can not report on the system assessed, the risk models obtained, the personnel from the client involved, or the name of the client.

Fig. 1 depicts the fragment of the security risk analysis of relevance for this paper as a process of six steps. Step 1 is the likelihood estimation based on expert judgments. Indicators for validation of the likelihood estimates were identified in Step 2. The analysis team proposed a number of indicators, and these indicators were revised during a meeting with the client experts in Step 3. During this meeting some indicators were rejected, some received minor modifications, and new indicators were identified. In Step 4 the analysis team formulated validation criteria for the likelihood estimates in terms of indicators. Each criterion specifies the expected values of the indicators related to the likelihood estimate in question. Here, each criterion makes a prediction about the value of a set of indicators under

978-0-7695-4680-3/12 $26.00 © 2012 IEEE
DOI 10.1109/Metrisec.2011.13

the assumption that the likelihood estimate in question is correct. Indicator values were obtained by the client experts in Step 5. In Step 6 validation criteria were evaluated and some of the initial likelihood estimates were adjusted.

In total, an estimated number of 400 hours were spent on the security risk analysis (not including writing the final report) by the analysis team. Three domain experts (E1, E2, and E3) of the client participated in the analysis. The client experts held degrees equivalent to Master of Science and four to ten years of experience in information security and risk analysis.

III. Data collected from case

For each step of the estimation and validation process, depicted in Fig. 1, we collected data, documented in Table I.

In **Step 1**, we came up with 28 likelihood estimates based on expert judgments.

In **Step 2**, the analysis team proposed at least one indicator for each of the 28 likelihood estimates. In total, 68 indicators were proposed by the analysis team in Step 2.

In **Step 3**, the indicators proposed in Step 2 were revised in a meeting with the client experts. Some of the proposed indicators were rejected during this meeting, because their values were not obtainable within the client's organization. After Step 3, there were 25 out of 28 likelihood estimates with at least one indicator. In total, 57 indicators remained after Step 3 had been conducted.

In **Step 4**, 19 indicators were used by the analysis team to formulate validation criteria for 15 likelihood estimates. For 10 likelihood estimates, validation criteria were not formulated. One of these 10 likelihood estimates was not assigned a criterion because the validation of the estimate was given a low priority by the client experts[1]. For the remaining nine likelihood estimates, the analysis team was not able to come up with good validation criteria, although the indicators were considered to provide relevant information for validating the likelihood estimates.

In **Step 5**, the client experts obtained values for 13 out of the 19 indicators used to formulate the 15 validation criteria. This resulted in that only 10 out of the 15 validation criteria could be evaluated after Step 5.

In **Step 6**, we evaluated 10 validation criteria based on the values obtained by the client experts. The validation criteria were fulfilled for four likelihood estimates, while for two likelihood estimates we could not say whether the criteria were fulfilled or not, because the values of the indicators referred to in the criteria were too uncertain. The criteria were not fulfilled for the remaining four likelihood estimates. For two of these estimates, the client experts decided to adjust the likelihood estimates.

Table I
Relevant data for the different steps of the estimation and validation process

1	Number of likelihood estimates based on expert judgments after Step 1.	28
2	Total number of indicators after Step 2.	68
	Number of likelihood estimates with at least one indicator after Step 2.	28
3	Total number of indicators after Step 3.	57
	Number of likelihood estimates with at least one indicator after Step 3.	25
4	Total number of indicators used to formulate validation criteria after Step 4.	19
	Number of likelihood estimates with a validation criterion after Step 4.	15
5	Total number of indicators used to formulate validation criteria for which the client experts obtained values after Step 5.	13
	Number of likelihood estimates for which validation criteria could be evaluated after Step 5.	10
6	Number of likelihood estimates with a fulfilled validation criterion after Step 6.	4
	Number of likelihood estimates with a not fulfilled validation criterion after Step 6.	4
	Number of likelihood estimates with a validation criterion where it was undecided whether the criterion is fulfilled or not after Step 6.	2
	Number of likelihood estimates with a not fulfilled validation criterion for which the likelihood estimates were adjusted after Step 6.	2

IV. Discussion

With each of the three client experts, we conducted a semi-structured interview, focusing on likelihood estimation based on expert judgments and the use of indicators to validate these. The transcribed interviews were analyzed by the use of a simplified version of thematic analysis [7]. In this section we discuss the data in Table I in relation to the results from the thematic analysis.

A. Step 1

Experts E1 and E2 were quite confident that they their likelihood estimates were correct, while expert E3 did not want to give a clear yes or no answer to this. Even though they believed the estimates to be correct, expert E1 pointed out that validation in terms of indicators still has a purpose: *"... I think there were one or two such cases where we had to adjust the estimates because of their indicator values. So I think the quality was good anyway but it is still an extra quality adjustment when you get acknowledgments or only minor adjustments of the estimates."*

B. Step 2 and 3

It was challenging to identify relevant indicators for which values could actually be obtained within the available time and resources for the analysis. Expert E1 supports this: *"It was a challenge in the least because it is terribly difficult to*

[1] The likelihood estimate was associated with an unwanted incident. For the incident to occur, some technology needed to be used, which was not yet implemented at the time of analysis. Thus, the client considered the likelihood estimate for this incident to be less important.

find good indicators of information security, and there were a number of examples where it was actually not possible to find indicators. Even though we had proposals we discovered later that they were not usable. But there were also areas where we came up with indicators that could be used."

During the revision meeting in Step 3, many indicators were rejected because their values were not obtainable within the client's organization. This resulted in that three likelihood estimates were left without indicators. One might argue that we should have used more time to identify indicators in Step 2, and also that we should have involved the client experts in this step. With respect to the former argument, according to our records we used about 50 hours to identify indicators in Step 2, which is quite a lot. With respect to the latter argument, all three client experts were of the opinion that the analysis team should come up with the initial indicator proposals. Expert E1 even expressed: *"... I also think that when it comes to indicators, it can be a strength that they are proposed by someone else who does not have built-in limitations with respect to ideas."*

On the other hand, we could perhaps have obtained information from the client experts on the kinds of data, in the form of logs and so on, that are available within their company, prior to identifying indicators in Step 2. This would most likely have resulted in fewer indicator proposals being rejected due to their values being not obtainable. On the other hand, proposing relevant indicators where their values are not obtainable at the time of analysis may also prompt the client organization to implement more measurements, as expressed by expert E2: *"It turned out that some of the measurements that had not been carried out should perhaps have been carried out, and that is the experience obtained from what we found here."*

C. Step 4

The analysis team was not able to formulate validation criteria for nine out of 25 likelihood estimates. We do not have the opinions of the client experts on this matter. They were not asked to comment on the formulation of validation criteria in the interviews, since this task was conducted solely by the analysis team.

The indicators of the nine estimates were considered relevant for validating the estimates, but we could not figure out how to link them to the estimates. Common for these indicators is that they are only indirectly linked to the estimates of which they were seen as relevant. An example of such an indicator is "the number of code lines used to produce the web server application" which is indirectly linked with the likelihood estimate of the unwanted incident "hacker takes over the web server by exploiting weaknesses in its code". In many cases it is reasonable to believe that the number of weaknesses will increase with the number of code lines. However, it is not easy to predict how the value of this indicator affects the likelihood estimate since the estimate depends on a lot of other factors as well. On the other hand, the indicator "the number of times the web server was taken over by hackers during the past five years due to weaknesses in its code" is directly linked with the likelihood estimate of the incident. It is not surprising that it is easier to formulate validation criteria based on this kind of more direct indicators than by the use of the more indirect ones. Eight out of the 10 validation criteria evaluated in Step 6 used an indicator that is directly linked to the likelihood estimate. In relation to this it must be pointed out that we would have had seven validation criteria using solely indirect indicators if we had managed to obtained all the indicator values in Step 5 needed for evaluating the 15 validation criteria.

D. Step 5

For five of the validation criteria the client experts did not manage to obtain the indicator values necessary for evaluating the criteria. One reason may be that obtaining all the indicator values required too much effort. The client experts tried to obtain values for 49 out of the 57 indicators remaining after Step 3. Out of the 19 indicators that ended up being used in validation criteria, they managed to obtain values for 13. They may have succeeded for a higher proportion if we had only requested the values for the 19 indicators being used in validation criteria. The reason for requesting all indicator values was that the validation criteria were formulated after the value collection process had been initiated, and before we received the indicator values from the client experts. Thus, we did not know at the time when the value collection process was initiated which indicators we would use to formulate the validation criteria. It would have been better to first identify the indicators needed in the validation criteria, and then ask the client experts to obtain values for those.

Another reason for failing to obtain six of the necessary values may have been that the client experts postponed the task a little too long. This is very likely since we know that many of the indicator values where obtained just before the given deadline. But it can also be the case that the values were not as easily available as first expected. Expert E2 supports the latter: *"... for me the process went pretty smoothly. I got answers if there had been done measurements, but I also got feedback like "we have no idea"."*

All three experts interviewed believe that indicator values of high quality were obtained. It is however a bit uncertain whether this was actually the case. We know, for instance, that some of the values obtained were just new expert judgments by other experts. Expert E2 told us that he obtained indicator values by asking other people working at the company: *"For those indicators where there were numbers it was pretty easy to find the answers, because it is hard to find the right person who has the right competence.*

But in our case it was actually two-three persons who answered all." It is however a bit uncertain how many of the obtained indicator values that were just new expert judgments.

E. Step 6

Two likelihood estimates were changed by the client experts as a result of their validation criteria being falsified. When changing the likelihood estimate of an unwanted incident, its risk level will often change as well, since the risk level depends on the likelihood value and the consequence value of the unwanted incident. A change in the risk level will often have consequences for the type of treatments that are implemented. In our case, however, the risk levels associated with the two unwanted incidents did not change when their likelihood estimates were updated.

In the case of a validation criterion being falsified we can not straightforwardly conclude whether likelihood estimates should be changed or kept as they are. For instance, although we manage to obtain correct indicator values, it may be that the validation criterion does not capture what we believe it does. In the risk analysis we had two cases where the client experts decided not to update the likelihood estimates of two unwanted incidents, even though their validation criteria were falsified. In the case of the first incident, the client experts kept the likelihood estimate because the value of the indicator used in the criterion did not represent a typical value. In the case of the second incident, its likelihood estimate was, according to its validation criterion, equal to zero since some technology required for the incident to occur was not in use at the time of analysis. As a consequence of this the incident should have been removed from the threat model. The client experts wanted the threat model to reflect the situation were the technology was in place, and the threat model was therefore not changed. Also, it was no harm in keeping the incident, since it did not result in any unacceptable risks needing treatments, due to a low likelihood estimate.

V. Conclusion

In this paper we have presented experiences from using indicators to validate likelihood estimates based on expert judgments in a security risk analysis conducted in 2010.

The use of indicators brought forward new information resulting in two out of 28 likelihood estimates being changed.

We also identified some challenges that need to be addressed in order to get the most out of indicator-based validation. First, it is challenging to identify indicators for which it is feasible to obtain values within the available time and resources for the analysis. For a number of the indicators identified, their values were not obtainable within the client's organization. By having some knowledge on the kinds of historical data that are available within the organization and whose responsible for the different kinds of data, it should be easier to both identify indicators and obtaining their values. Unfortunately, it may be difficult to obtain this knowledge since data is often spread across the organization and since few, if none, have a complete overview of the data available. Second, it is challenging to formulate validation criteria for likelihood estimates in terms of indicators. It is especially difficult to predict how indicator values affect a likelihood estimate when the indicators are only indirectly related to the estimate in question. This will typically be a problem when formulating validation criteria for likelihood estimates of incidents that that are not easily observable. Third, the indicator values obtained from an organization may vary when it comes to correctness. In order to get the most out of the validation, the uncertainty of the values should be taken into account. Moreover, one should strive to reduce uncertainty by using several independent indicators to validate the same estimate.

Acknowledgments

The research on which this paper reports has been carried out within the DIGIT project (180052/S10), funded by the Research Council of Norway, and the MASTER and NESSoS projects, both funded from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements FP7-216917 and FP7-256980, respectively.

References

[1] M. A. Meyer and J. M. Booker, *Eliciting and Analyzing Expert Judgment: A Practical Guide*, ser. ASA-SIAM series on statistics and applied probability. SIAM, 2001.

[2] R. M. Cooke and L. H. J. Goossens, "Procedures Guide for Structured Expert Judgment in Accident Consequence Modelling," *Radiation Protection and Dosimetry*, vol. 90, no. 3, pp. 303–311, 2000.

[3] L. H. J. Goossens, R. M. Cooke, A. R. Hale, and L. Rodic-Wiersma, "Fifteen Years of Expert Judgement at TUDelft," *Safety Science*, vol. 46, no. 2, pp. 234–244, 2008.

[4] H. Otway and D. von Winterfeldt, "Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls," *Risk Analysis*, vol. 12, no. 1, pp. 83–93, 1992.

[5] F. Bolger and G. Wright, "Assessing the Quality of Expert Judgment: Issues and Analysis," *Decision Support Systems*, vol. 11, no. 1, pp. 1–24, 1994.

[6] M. S. Lund, B. Solhaug, and K. Stølen, *Model-Driven Risk Analysis: The CORAS Approach*, 1st ed. Springer, 2011.

[7] D. Ezzy, *Qualitative Analysis: Practise and Innovation*, 1st ed. Routledge, 2002.