# Validating the Raster Risk Assessment Method in Practice

**Eelco Vriezekolk**
Radiocommunications Agency
Netherlands & University of Twente
e.vriezekolk@utwente.nl

**Sandro Etalle**
Eindhoven University of
Technology
s.etalle@tue.nl

**Roel Wieringa**
University of Twente
r.j.wieringa@utwente.nl

## ABSTRACT

Telecommunication services are essential to modern information systems, especially so for crisis management. Telecoms systems are complex and difficult to analyse. Current risk assessment methods are either not used because of their complexity, or lack rigorous argumentation to justify their results because they are oversimplified. Our challenge has been to develop a risk assessment method that is both usable in practice and delivers understandable arguments to explain and justify its risk evaluations. After experiments to validate the method in laboratory environments, we now present the first results from successful application with practitioners in a regional crisis organization that provides evidence about the practical usability of the method.

### Keywords

Risk assessment, telecommunications, validation, field test.

## INTRODUCTION

Information systems are nowadays inseparable from telecommunication. This is especially true for information systems for crisis management, where there is a trend towards more, and more connected systems, information sharing, and online interaction. Unavailability of telecom services therefore poses a large risk to effective crisis management.

Analysis of availability risks for telecom services is difficult. These systems are highly complex, information on infrastructures and their vulnerabilities is incomplete, and uncertainties are high. Risk assessments therefore rely heavily on expert judgments and may lack objectivity.

To overcome these difficulties we have been developing a risk assessment method called Raster (“Risk Assessment by Stepwise Refinement”). We have conducted several lab experiments, which showed that the method produces correct results in a reliable way when used in controlled circumstances (Vriezekolk, Etalle, Wieringa, 2012, 2015). We have not yet validated the method in practice, under uncontrolled circumstances. The remaining validation step is to show that, in the field, Raster yields correct results against an acceptable cost, and that experts perceive the method as usable and useful. In this paper we describe a field test to that purpose, performed with a regional crisis organization in the Netherlands.

A practical validation is essential before we can conclude that Raster offers definite advantages over existing risk analysis methods in this domain. Our field test provides evidence about the practical usability and usefulness of Raster.

We first describe the Raster method, past validation in lab settings, and one particular traditional risk analysis method for comparison. We then describe our field test that is to answer practical questions on the reliability of telecom services while at the same time collecting information to validate our method. We present and discuss our results, and conclude with an outlook and invitation for further research.

## BACKGROUND

### Previous Research

Raster is a risk assessment method for availability risks to telecommunication services used by crisis organizations (Raster, 2015). Its target users are telecom experts together with domain experts from crisis organizations.

Development of a new risk assessment method requires a number of steps, from initial design through validation to implementation (that is, transfer to practice). Validation requires a demonstration that the method can be performed and yields results that are reliable and correct. Implementation requires at minimum that prospective users perceive the method to be useful and easy to use.

Raster uses diagrams to model telecom services (Vriezekolk et al., 2011). Missing knowledge about the technical infrastructure is represented in diagrams by cloud-shaped subsystems, so that in all cases an end-to-end model can be created. The model is refined iteratively, driven by the outcome of prior risk assessment steps. Availability risks are estimated qualitatively for each diagram component, in a group and consensus based process. A special feature of Raster is the analysis of common cause failures (CCFs). A CCF occurs when multiple components fail because of a single incident, such as power failure affecting several components, or a flawed software update to multiple routers. CCFs are analysed by clustering components based on shared properties. For example, for power failures components are clustered based on geographical proximity; two components cannot be affected by the same power failure if they are sufficiently distant. Raster also includes in its risk evaluation potential public sentiments to incidents and risk reduction measures, as these sentiments are relevant to decision makers acting in the public domain.

Based on these ideas we continued the development of the method. The first validation showed that the method is executable, and produces meaningful results in a lab setting (Vriezekolk et al., 2012). One observation was that tool support is essential in doing iterative risk assessment on networks of components. Our next step in validation was to show that Raster is reliable. By reliable we mean that the method can be repeated with the same results. Other terms for this concept are repeatability, stability, consistency, and reproducibility. Our second validation showed that Raster has a relatively low reliability, but that this is to be expected of any method that depends on expert judgment with scarce data (Vriezekolk et al., 2015). In our view, experts retain a large responsibility for their results. To make this explicit, a Raster assessment not only yields risk evaluations, but also the justification for these evaluations. There justifications make clear to the decision makers what the limitations of the evaluations are.

Both previous validation experiments took place in a lab setting. Based on their outcomes we improved the method and its documentation, and the Raster tool.

### Current Risk Assessment in Safety Regions

We conducted our field test in a Safety Region. A Safety Region in the Dutch system of crisis management is an organization responsible for fire services, emergency medical care and crisis management and response. Safety Regions serve areas with an average population of 650 thousand; there are 25 Safety Regions in the Netherlands. Information and communication technology are essential in all activities of a Safety Region (Boersma, Wolbers and Wagenaar, 2010).

Safety Regions have a legal obligation to plan for unavailability of electrical power, information systems and telecommunications. A study commissioned by the Ministry of Safety and Justice revealed that in 2012 most Safety Regions did not have such a plan (Dorssen, Holzmann, Franx and Rens, 2012). The ministry organized drafting sessions, based on a model crisis plan to stimulate compliance and improve the quality of plans.

The model plan for the Safety Regions starts with an inventory of critical processes that must at all times be continued. Then, supporting applications (defined as information systems and networks) are identified, and their use in critical processes is recorded. Applications are then specified in more detail, on the level of major hardware components. Each component receives a three-level score to indicate its overall reliability. For example, the middle level is defined as *"Somewhat resistant to failure of parts. Redundant parts and supply lines. Not always resistant to maintenance (probably available during power failures, not during maintenance)"*. The reliability of an application is the lowest reliability figure of any of its hardware components. The reliability of a process is the lowest of that of its supporting applications. The plan then continues with a description of existing countermeasures, and concludes with a recommendation for further improvement and risk mitigation measures.

The model crisis plan is useful, as it makes clear which applications and hardware components are important, based on a pre-existing list of critical processes. The plan is therefore closely tied to organizational priorities. However, we believe that the analysis part of the model is weak. Analysts do not justify their selection of applications or hardware components. If crucial hardware components are overlooked, this will have a significant effect on the overall risk evaluation. Recall that reliability of an application is defined as the minimum reliability of its hardware components. Omission of even a single low-reliability hardware component will inflate the reliability of the application.

## METHOD

Our object of study is the Safety Region Groningen (SRG). SRG serves a relatively sparsely populated province of about 2,300 km$^2$ and a population of 575 thousand people. The region has a single large city, and is characterized by installations for the winning of natural gas. As from 2014, SRG has an information management department that is responsible for information and communication systems deployed by the Safety Region. This department was the *sponsor* of the study. Their objective was to receive an independent assessment of whether the level of reliability of their services was suitable, given the needs of its internal customers. The SRG adopted and completed the model crisis plan. The main author of this plan did not participate in the study.

Given the sensitivity of SRG's operations, all participants agreed to confidentiality. Publishing details of the risk analysis or its results could jeopardize the effective operation of emergency services, or could make them a target for malicious actions. In this paper we can therefore only describe the results in general terms and cannot mention specific risks found.

For the design of our validation we followed the checklist provided by Wieringa (2014). In Wieringa's terminology, the field test described in this paper is Technical Action Research, an application of a still-experimental technique in the real world to help a client. In the field test described here, the first author of the paper (the *experimenter*) participated in a project within SRG in a real operational environment. The experimenter acted as project leader, facilitating the application of the Raster method. The goal of the field test was therefore twofold: the sponsor (as the client) had an actual question that needed answering, and the experimenter wanted to validate the Raster method.

For validation we had three research questions:

Q1. Does Raster produce, in practice, risk evaluations that are correct? That is, are all relevant risks included (completeness), are low risks excluded (conciseness), and are risks presented with the right priorities?
Q2. What are the costs and effort involved in execution of Raster?
Q3. Are the target users willing to use the Raster method?

We describe each question in turn. Correctness of risk evaluations is a difficult concept. First, it cannot be determined objectively, because there is no available known good standard to compare against. Secondly, information is incomplete and the amount of expert judgment is necessarily large. Lastly, risk evaluations are uncertain predictions, meaning that they state that some future impact is likely or less likely. Crises are infrequent, but even when the impact materializes the fact that the event did happen in no way validates or invalidates the risk assessment. It is only with a large set of predictions and their outcomes that we can make statistical statements about the accuracy of predictions. In that respect risk evaluations for crises are very different from other uncertain predictions, such as weather forecasts. Because of these difficulties, we must determine correctness in a subjective way. Firstly, we asked participants about their personal belief in correctness of their risk evaluations; secondly, we will ask participants from independent field tests to assess the correctness of each other's evaluations. Note that so far we only completed a single field test; a second field test is planned and will be followed by a cross-validation.

We now turn to the research question Q2 on costs and effort. Since the parties agreed to participate without charge, only effort was relevant. We recorded the time spent during project meetings, and the participants reported the amount of time they spent on project matters between meetings.

To see whether target users would want to use Raster themselves (Q3), we used the Technology Acceptance Model TAM (Davis, 1989). TAM has been studied extensively and, although it has received criticism and several extensions have been proposed, its usefulness is well established in research. We use TAM because it employs a list of 12 standardized questions that accurately predict whether users would adopt new information technology. The list could be integrated easily into our own post-test questionnaire. The Raster method falls within the scope of TAM, because telecommunication services are information technology, and because the use of the Raster software tool forms an essential part of the method.

TAM measures perceived usefulness and perceived ease of use as predictors for future behaviour; TAM does not (nor does it intend to) measure objective benefits or efficiency gains. The standard TAM questions were translated from English into Dutch.

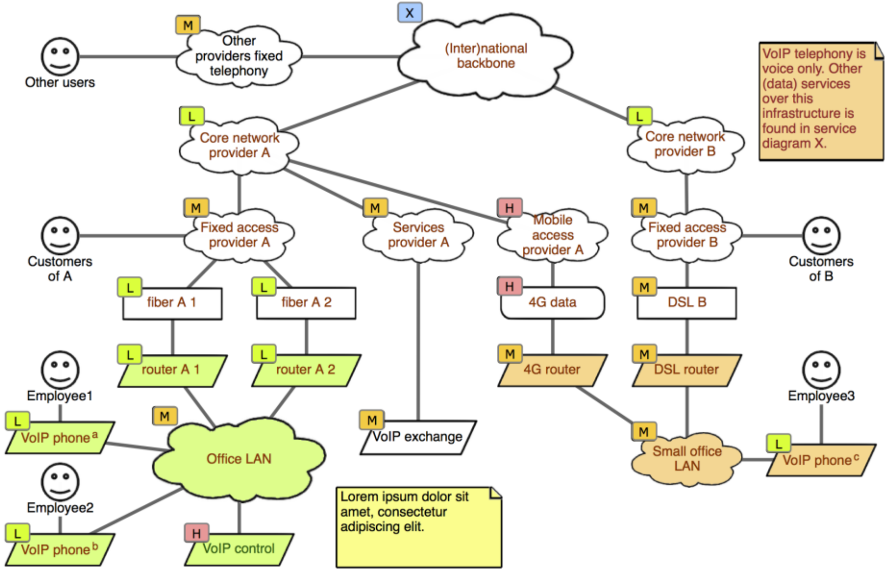
Before we started the project, a project plan was presented to the sponsor for approval. The plan presupposed the participation of experts from diverse backgrounds. These experts would perform the Raster method, with the experimenter acting only as facilitator and chair.

At the end of the project, each participant completed a questionnaire, and a guided group discussion was held. The purpose of the questionnaire and discussion was to provide further information to answer our three research questions. The questionnaire contained 38 statements with five possible answers from “Strongly disagree” to “Strongly agree”.

## RESULTS

After receiving approval on the project plan a kick-off meeting was held, at which participants from the various disciplines were present. The managing director of SRG was present to express his support. We had representatives from emergency medical care, police, civil support, both water boards present in the province, fire services (both volunteer and professional), and members of the SRG’s information management department.

All project members were present (with a few apologies) during each of the five half-day project meetings. The purposes of plenary meetings were to introduce and explain the steps of the Raster method, to collect ideas from all project members, and to collectively discuss the main approach to conducting the risk analyses. The experimenter and members from the information management department convened between plenary meetings to complete tasks that were left unfinished during the plenary meetings. The full results were then presented at the next plenary meeting for the group’s approval. This way of working allowed for efficient use of the expert’s time.



VoIP telephony is voice only. Other (data) services over this infrastructure is found in service diagram X.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

**Figure 1: The project’s diagram for fixed telephony (anonymized). Components have a shape, colour and emblem to indicating type, location and risk level respectively.**

The group inspected 10 telecommunication services: civil warning sirens, personal mobile radios, paging, VPN connections, satellite telephony, mobile telephony, fixed telephony (including Voice over IP, see Figure 1 for an anonymized version), the national emergency telephony service, data services, and vehicle automation. Together these 10 services contained 205 telecom infrastructure components. In all, the project recorded over 1,800 estimates of likelihood and impact of vulnerabilities. The final risk list contained 19 important availability risks.

At the project's end seven participants completed the exit questionnaire, and a one-hour guided discussion was held. Copies can be made available on request.

## DISCUSSION

We now return to our three research questions. For completeness, almost all participants agreed that Raster helps to find all large risks quickly (86% agreed or strongly agreed), but were less certain when asked whether there are large risks that have not been found (14% agreed, 57% had no clear opinion). For conciseness, participants almost all agreed that Raster helps to ignore small risks quickly (86% (strongly) agreed), but were slightly less certain whether the final risk list contains risks that are actually not that important (71% believed this not to be the case). All participants agreed that the priorities on the risk list were right. We also asked whether participants trusted the results of the project (86% agreed strongly), and whether they were willing to take responsibility for the results towards their colleagues (all agreed strongly). Overall, we conclude that participants believe the results to be correct, but also that they recognize that uncertainties are still large.

The second research question concerns the effort required in execution of Raster. Five project members spent on average 16 hours on the entire project; the two core members who actively worked on the project between meetings spent on average 29 hours. Their combined total of 138 staff-hours excludes time spent by the experimenter. According to the participants, the effort was sufficient. Only one participant believed that the results would have been better if more time had been available (all others scored neutral). Also, it would not have been impossible for them to participate if the required time had been higher.

The questionnaire included the 12 TAM questions in random order and position. The 6 questions on perceived ease of use were answered positively: (strongly disagree, disagree, neutral, agree, strongly agree) = (0%, 7%, 27%, 63%, 2%). The answers for perceived usefulness were (0%, 7%, 17%, 60%, 17%). These are positive results, especially since perceived usefulness is an even stronger predictor of future use than perceived ease of use (Davis, 1989).

All participants were provided with a booklet explaining, in English, the Raster method and tool. It surprised us that experts found the lack of a Dutch translation a stumbling block. We have since localized the tool, and are completing a Dutch translation of the Raster manual.

The project again confirmed that tool support is essential. Participants found the use of colour in diagrams very useful to convey relevant information. We have refined the tool based on experiences from the project, and will continue to do so.

Overall, the participants found the project interesting, stimulating, and very useful, not just because of the final risk list but also because it gave them an understanding of the workings and complexities of telecom services that otherwise remain hidden behind the scenes. We hope that this translates to better preparedness and effectiveness; the results do not allow us to draw conclusions about that.

Finally, an interesting question is how the risk list from this project compares to the results from the model crisis plan. Comparing the two is difficult, because the information management department of SRG had made significant changes to the infrastructure after the model crisis plan was written, and before we did the Raster assessment. For example, all servers had been outsourced and desktop equipment currently consists of thin clients; fixed telephony had been replaced by a hosted VoIP solution. Nevertheless, we see that some of the telecom services that contained the most risks in the Raster project were assessed as 'reliable' in the model crisis plan. We believe that Raster gave the more accurate description, because its risk evaluations are grounded in a detailed analysis of the technical architecture of the telecom services, whereas the model crisis plan did not offer any argumentation for its assessments.

## CONCLUSIONS

A single field test is not sufficient to draw conclusions. We therefore cannot state with certainty whether Raster produces correct results with acceptable effort, nor whether target users are likely to adopt the method. However, we do find these first results very encouraging. The project ran smoothly and easily, project members found the risk analysis to be interesting and participated in lively discussions. Using the coloured Raster diagrams as a common ground, responders and planners were able to collaborate with telecom experts. Together they drafted and justified risk evaluations that were approved by all participants as well as the sponsor.

We also tentatively conclude that Raster's results are complete, concise and prioritized correctly. The effort required, about 140 staff-hours over a period of six weeks for a medium-sized crisis organization, is significant but appears to be acceptable. Perceived ease of use and perceived helpfulness suggest that the Raster method would be adopted by its target users, provided that localized versions of the manual are available. Successful implementation will probably require localized training materials, such as examples and tutorials, as well.

These results suggest to us that Raster is feasible and useful in practice.

We need to perform further field tests before we can conclude that Raster offers an improvement over current approaches to analysis of telecom service availability risks. It would be especially interesting to execute Raster with a team that does not include the original researchers. We welcome suggestions from the research community for further field tests.

## ACKNOWLEDGEMENTS

We thank all participants from SRG for their support.

9.

## REFERENCES

1. Boersma, K., Wolbers, J., and Wagenaar, P. (2010) Organizing Emergent Safety Organizations: The travelling of the concept 'Netcentric Work' in the Dutch Safety sector, *Proceedings of the 7th International ISCRAM Conference*, Seattle.
2. Davis, F.D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS quarterly*, 319–340.
3. Dorssen, M. van, Holzmann, M., Franx, K., and Rens, L. (2012) Continuïteitsplannen ICT/elektriciteit: inventarisatie van continuïteitsplannen in geval van grootschalige uitval van ICT en/of elektriciteit bij gemeenten, provincies, veiligheidsregio's, politieregio's en waterschappen [in Dutch, with English summary], *Berenschot Groep/I&O Research.*
4. Raster documentation website (2015), http://wwwhome.ewi.utwente.nl /~vriezekolk/Raster/ .
5. Vriezekolk, Etalle, and Wieringa (2011) A new method to assess telecom service availability risks, *Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management ISCRAM2011.*
6. Vriezekolk, Etalle, and Wieringa (2012) Design and initial validation of the Raster method for telecom service availability risk assessment, *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM2012.*
7. Vriezekolk, Etalle, and Wieringa (2015) Experimental validation of a risk assessment method, *Proceedings of the 21st Intl. Working Conference on Requirements Engineering: Foundation for Software Quality*, LNCS, vol 9013, Springer (accepted for publication).
8. Wieringa, R. J. (2014) Design Science Methodology for Information Systems and Software Engineering, Springer.