predict
prioritise
prevent

TRESPASS

Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security

Deliverable D2.5.1

TRESPASS Information Testing and Degradation Tools

Project:	TRESPASS
Project Number:	ICT-318003
Deliverable:	D2.5.1
Title:	TRESPASS Information Testing and Degradation Tools
Version:	1.0
Confidentiality:	Public
Editor:	Sven Uebelacker
Cont. Authors:	M. Ford, D. Gollmann, C. Heath, M. Stoelinga, S. Uebelacker, A. S. Yesuf
Date:	2016-10-31

Part of the Seventh Framework Programme
Funded by the EC-DG CONNECT

Members of the TRESPASS Consortium

1.	University of Twente	UT	The Netherlands
2.	Technical University of Denmark	DTU	Denmark
3.	Cybernetica	CYB	Estonia
4.	GMV Portugal	GMVP	Portugal
5.	GMV Spain	GMVS	Spain
6.	Royal Holloway University of London	RHUL	United Kingdom
7.	itrust consulting	ITR	Luxembourg
8.	Goethe University Frankfurt	GUF	Germany
9.	IBM Research	IBM	Switzerland
10.	Delft University of Technology	TUD	The Netherlands
11.	Hamburg University of Technology	TUHH	Germany
12.	University of Luxembourg	UL	Luxembourg
13.	Aalborg University	AAU	Denmark
14.	Consult Hyperion	CHYP	United Kingdom
15.	BizzDesign	BD	The Netherlands
16.	Deloitte	DELO	The Netherlands
17.	Lust	LUST	The Netherlands

Document History

Authors		
Partner	Name	Chapters
CHYP	Margaret Ford	2.1
GUF	Ahmed Seid Yesuf	2.1
RHUL	Claude Heath	1, 2.2, 3
TUHH	Dieter Gollmann	1, 5, ALL
TUHH	Sven Uebelacker	1, 2, 4, ALL
UT	Mariëlle Stoelinga	4

Quality assurance		
Role	Name	Date
Editor	Sven Uebelacker	2016-10-31
Reviewer	Jan-Willem Bullée	2016-10-12
Reviewer	Ahmed Seid Yesuf	2016-10-11
Task leader	Sven Uebelacker	2016-10-31
WP leader	Michael Osbourne	2016-10-31
Coordinator	Pieter Hartel	2016-10-31

Circulation	
Recipient	Date of submission
Project Partners	2016-09-30
European Commission	2016-10-31

Acknowledgement: *The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318003 (TRESPASS). This publication reflects only the authors' views and the Union is not liable for any use that may be made of the information contained herein.*

Contents

List of Figures

Management Summary

Key takeaways:

- The TRE$_S$PASS tools equally work with empirical data and expert judgements; users of the tools need strategies for dealing with uncertainty in both cases.
- The TRE$_S$PASS process includes models and tools for capturing uncertainties during data collection.
- There exist inherent conflicts between protecting the security of data and collecting data for security.

Deliverable D2.5.1 is the final deliverable of task T2.5 that covered “Information Testing and Degradation”. The TRE$_S$PASS project proposal had been written in the spirit that empirical data, when available, constitute more reliable input to model-driven risk analysis than expert judgements that would inherently suffer from some degree of uncertainty. The term “degradation” reflects this expectation which, however, has not been confirmed by the experiences gained from the TRE$_S$PASS case studies. Experts from our Technical Advisory Board have affirmed our observation that quantitative empirical data are difficult to obtain in security related risk analysis.

Uncertainty, moreover, can also apply to collections of empirical data, e.g. to the ratio of relevant events that has actually been recorded. Similar considerations thus apply to risk calculations based on empirical data and risk calculations based on expert judgements. The challenge then lies not so much in comparing the effectiveness of risk analysis methods in those two settings, but in capturing, quantifying, and visualising uncertainty about inputs to risk calculations.

This deliverable delineates the challenges in data collection faced in the TRE$_S$PASS case studies and summarises how methods and tools developed in The TRE$_S$PASS Project can capture and process uncertainties in data.

1 Introduction

The TRESPASS proposal had assumed that in the best case risk analysis would be based on suitable collections of empirical data. Possible data sources had been identified when preparing the project proposal. In the suboptimal situation where empirical data is lacking risk analysis would have to take recourse to the 'second best', e.g. to expert judgements afflicted by a higher degree of uncertainty. Degradation, the subject of this deliverable, reflects this assumption.

The expectations on the availability of empirical data were, by and large, disappointed. Empirical data on security incidents are not always existent. If they exist, they may be too sensitive to be fully made available for public research. For example, the European Statistical System does provide public information for decision-making purposes, research, and debate (Batini & Scannapieca, 2006), and such data has been used in The TRESPASS Project. However, when data sources like survey results cannot be re-evaluated due to privacy issues it becomes difficult to challenge potentially subjective selections of data that could present a specific preconceived view on an issue. As another example for the difficulty of obtaining relevant data, TUHH had attempted to get access to information about cybercrime events recorded in the criminal databases of the Hamburg State Office of Criminal Investigation (LKA) for research purposes, but did not succeed.

The question of data collection for security remains an ongoing issue in general, as stated in the recently published Royal Society report on *Progress and Research in Cybersecurity* (The Royal Society, 2016):

> *The sensitive nature of the material protected by cybersecurity can affect how much information is shared about protective measures, vulnerabilities and breaches. This knowledge is an important collective resource for improving cybersecurity defences, but its use is often limited by lack of transparency. This lack of transparency is often based on justified concerns about the risks of releasing information, but sometimes risk-aversion or organisational culture drives greater secrecy than is warranted.*

As a further limiting factor, empirical data by its very nature captures the past and thus, by default, can contribute to risk prediction only partially: there is no empirical data on new types of attacks. Quoting a member of our Technical Advisory Board, "in security the past is a poor predictor of the future". It has moreover been observed that too strong a focus on known threats can turn out to be counterproductive when new types of threats get too little attention (Kang, 2013).

Feedback from other researchers on risk analysis confirms that quantitative empirical data are rarely available in practice. The experiments sketched in the project proposal could thus, due to a lack of available data, not be conducted in the way anticipated.

Risk analysis is not only performed for existing systems, where data might have been recorded, but also for systems and services under development. The problem of uncertainty is well recognised in the field of design innovation. Uncertainty is of necessity commonly embraced wherever innovation within organisations is taking place, and wherever new service designs are being considered. Frequently in these contexts groups of interested parties gather in order to co-design solutions. This vague state of affairs is the starting point for a narrowing of the field down to relevant and manageable targets, as the following quote describes:

> *Innovation management literature describes the front end of innovation as the stage where the generation of an idea and concept design takes place. The stage roughly ends around the concept's approval for development or its termination as a viable project to continue. It is considered to be "fuzzy" because information is usually scarce, costs associated with change low, and ability to influence results high.* (Botero, 2013, p.47).

The *design space* in these cases is actually constituted by the range of 'possible solutions' (Westerlund, 2009). This range of potentially advantageous solutions may also be

> *a way to map trajectories, histories, and resources around and beyond the fuzzy front end. The design space could be a performative tool to communicate where we have been and where we are now, or even "where we could be", to all those with a stake in the process.* (Botero, 2013, p.58)

1.1 Goals

This deliverable will illustrate how the TRE$_S$PASS process can deal with capturing uncertainties during data collection using primarily the example of the IPTV case study with the TRE$_S$PASS LEGO^{TM} approach.

1.2 Choices Made

We focus on the IPTV case study and its LEGO^{TM} modelling approach.

1.3 Foreground and Background

The tools and concepts presented in this deliverable are 95% foreground, 5% is background.

1.4 Document Structure

Chapter 2 discusses data collection in the TRE$_S$PASS process. Section 2.1 focuses on the challenges that arose during data collection in the TRE$_S$PASS case studies. Section 2.2 describes two TRE$_S$PASS data collection tools that can handle uncertainty during data collection, *InterActor* developed in many LEGO^{TM} modelling session, and a graphical interface for entering fuzzy inputs to the TRE$_S$PASS Attack Navigator Map. Section 3 presents relevant modelling insights from the IPTV case study and further risk assessment workshops. Section 4 summarises WP3 results on data uncertainty in quantitative risk analysis. We finish with an outlook in Chapter 5.

2 Data Collection – Challenges and Tools

The purpose of task T2.5 was to investigate in which way a lack of data would impede the effectiveness of the TRE$_S$PASS process. 'Data' here stands for 'empirical data' but expert opinion can also be 'data' that serves as input to the TRE$_S$PASS process. The quality of the results of applying the TRE$_S$PASS process depends on the quality of the inputs available, be they empirical data or expert opinions.

To set the context, we recollect our experience from modelling an attack tree for the IPTV case study at an early stage of The TRE$_S$PASS Project. This was a novel scenario so no pertinent historic data was available. Risk analysis had to rely on expert knowledge derived from experience in similar cases. We observed on some occasions difficulties in agreeing on a single value per node and domain; we faced the challenge of converting qualitative values like "less likely" (expert "feeling") into quantities, e.g., 0.1. It turned out that a group session enables discussion on discrepancies, enhancing the overall confidence in the finally agreed values.

For a good reason, many risk analysis methods incorporate expert groups (external consultants or internal experts of various fields/departments) in a data quality cycle (cf. model-driven risk analysis with CORAS (Lund, Solhaug, & Stølen, 2010)). Modelling in groups like the TRE$_S$PASS LEGO^{TM} approach supports such activities (Chapter 3). The brainstorming tool ArgueSecure (Ionita, Bullee, & Wieringa, 2014) and its online derivative – both developed by University of Twente in TRE$_S$PASS – implement a qualitative, argumentation-based approach based on expert knowledge.

2.1 Case Studies

TRE$_S$PASS has from the outset been working with case studies in order to test and develop the methods being created by the project. The three case studies encompass a wide range of complex and sensitive data. They cover a diverse range of environments, providing realistic examples of the challenges faced by information security risk professionals in practice:

- Cloud infrastructure;
- Fraud or misuse of a telecommunications company's products and services;
- Customer Privacy Protection — initially exploring home payments made using an IPTV set-top box, subsequently incorporating the ATM case study and ÉpStan (Épreuves standardisées, https://www.epstan.lu/), in order to evaluate the TRE$_S$PASS approach in different contexts, each involving sensitive assets of customers.

Each case study has different features and therefore presents different priorities in the handling of data, as well as a variety of challenges in dealing with issues of data quality and availability:

- Three inputs are considered sensitive for the cloud provider: security policies, infrastructure configuration and topology, and work-flows and processes.
- Telcos continuously work against the ambitions of so-called *knowledge insiders* (cf. notion of insiderness in Probst and Hansen (2013)) by evaluating their products with respect to their inherent fraud potential. Different types of sensitive data accumulate in the evaluation of new and existing products, and ultimately when monitoring fraudulent user behaviour as it happens.
- The Customer Privacy Protection case study encompasses a wide and diverse range of data relating to different aspects of service delivery: social, commercial, technical, physical, as well as potentially legal and political.

2.1.1 Cloud

In the cloud case study (The TRE$_S$PASS Project, D7.2.1, 2014; The TRE$_S$PASS Project, D7.2.2, 2016), a cloud provider needs to deliver a secure computing infrastructure to its customers. For the security analysis it is crucial to obtain snapshots and real-time data of the infrastructure configuration and topology, and represent this in the form of a model of the cloud infrastructure. For many cloud providers, the configuration and aspects of the topology are considered confidential and, especially for public cloud providers, a competitive advantage to keep it secret. The data may expose information such as utilisation and size of the business (i.e., number of virtual machines as well as protective infrastructure).

The workflows and processes describe on the technical level operations such as backup, patch and key management, as well as access control. On the social/business level, they also describe aspects such as the education and certification of cloud administrators, as well as sign-offs of infrastructure configuration changes.

These data may be protected by a non-disclosure agreement (NDA) between the service provider and the customer, or otherwise it might be possible to access data after suitable de-sensitisation.

2.1.2 Telco

When evaluating a telco product, whether it is new or an existing product in the market, various types of data are required to understand the strategic importance of the product to the company, e.g. in terms of expected volume of sales or the degree of technological innovation. The types of data required to achieve this evaluation include:

- Confidential data about new products. This type of data is highly sensitive and must not leave the boundaries of the company due to the strategic importance with regard to competitors in the market;
- Customer data from trials. When conducting a trial, user data is inevitably involved, e.g. personal information such as numbers and addresses, account data and passwords on the devices and the software;
- Call detail records (CDRs): connection data from trials. When conducting a trial, users create usage data, including metadata: Information about calling and receiving parties, start time and duration, IP addresses, location information that can be extracted from IP addresses or cell information, and IMEI numbers used. As a result, call detail records contain private, personal information. Hence, processing and storage of CDRs is subject to legal restrictions.
- Confidential data about internal operations. Most features and characteristics of a product depend heavily on the functionalities of the telco's communication network, hardware, software and suppliers as well as successful development and completion of IT projects.

Additionally, the telco will monitor customer behaviour such as usage patterns and the extent to which the product is used or misused, in order to evaluate the effectiveness of measures to detect and defend against fraud.

From the company's perspective, such data should not leave the boundaries of the company unless requested by legal orders. It is therefore considered as uncertain that such data will be made available.

Rather, the focus was to get telco service/product descriptions which are helpful for the TRE$_S$PASS fraud assessment. We have achieved this by having an NDA agreement with a telco company.

2.1.3 Customer Privacy Protection

The initial scenario in this case study is based around the delivery of payments services to individuals via an IPTV set-top box in their own home. This scenario presents all the usual risks associated with a payments environment, including contactless technologies. It also presents a range of risks related to the relationship between the account holder, their carer(s) and service providers.

The IPTV case study has particular features in relation to data handling:

Social the case study focuses on systems designed to deliver vital services to people who are in many respects vulnerable. This raises not only privacy issues regarding personal data, but also potentially safeguarding issues, depending on the personal situation of the system user. The UK TRE$_S$PASS partners have experience of working with people in these circumstances and has taken the necessary steps to protect the individual's data, both within and beyond the project.

Commercial the organisation responsible for the IPTV system, which is a social enterprise, still has commercial reasons for protecting its interests in what is proposed to be a commercially viable provision of services in association with other organisations.

Technical certain technical details may need to be protected, or considered in a more generalised context, in order to avoid unnecessary exposure of future live systems.

Physical gathering of physical data is likely to be both demanding and potentially intrusive. Relevant existing data is available from the Human-Computer Interaction (HCI) literature, as well as the MIT PlaceLab (http://web.mit.edu/cron/group/house_n/placelab.html).

Legal and Political this type of installation is likely to form part of a sizeable initiative in co-operation with social housing providers (either trusts or local authorities). This relationship has the potential to have significant legal and political implications for the organisations involved.

Our case study partner, London Rebuilding Society (LRS), were very helpful in sharing sensitive data under NDA with specific project partners and providing ongoing feedback through the course of the case study.

In addition to the IPTV case study and within Customer Privacy Protection, a further ATM (Automated Teller Machine) case study, which complements the IPTV case study, has been developed in the latter years of the project. The ATM case study has explored a range of attacks to ATM machines, including:

- Software attacks consisting of infecting the machines with malware that allows the attacker to take control of the devices, including the ability to open the machine money vault, and record data from the cardholders
- Physical attacks consisting of stealing the machines to open them in order to access the money vault.

Data issues, in the ATM case study as in the IPTV case study, are related to the confidentiality of the data from the customers' and service provider's point of view. From the provider's perspective, publishing information about successful attacks is highly sensitive since it can negatively impact their reputation. From the authorities' perspective successful attacks also motivate other attackers to start performing these attacks and therefore must be handled carefully. While project partners GMV have access to a range of industry reports, the data included in these is restricted and so it was necessary to develop the case study based on derived data rather than quoting directly from those reports.

Some of the data used to develop geophysical aspects of the ATM case study was drawn from public sources, which enabled us to identify particular features of an ATM location such as distance from the nearest police station, or distance from the nearest motorway exit.

In the final year of the project, a further case study, ÉpStan (Épreuves standardisées, https://www.epstan.lu/), about privacy and Personally Identifiable Information (PII) has

been added. The ÉpStan scenario deals with knowledge tests of Luxembourg students, aged between ten and eighteen years old, in order to assess the quality of the education system. In this scenario, a high level of data protection is required by law, and addressed by pseudonymisation. The ÉpStan Trusted Third Party (TTP) is in charge of creating pseudonyms for each student.

Data uncertainty in relation to the Customer Privacy Protection case study is explored in more detail later in this document. Chapter 3 explains how uncertainty and data provenance can be built into the LEGO™ modelling process in the IPTV case study.

2.2 Data Collection Tools

This section presents two tools data collection tools, the *InterActor* tool that emerged from RHUL's experience with the LEGO™ modelling approach, and a graphical interface for entering fuzzy inputs to the TRE$_S$PASS Attack Navigator Map, that document how uncertain inputs can be collected in the TRE$_S$PASS process.

2.2.1 *InterActor*

From extensive contact with security practitioners during qualitative LEGO™ evaluation and engagement sessions, a need was identified for a method where data can be captured during the co-creation process, in a 'brainstorming' setting. This needs to be deployed during and after engagements as a means of extending the modelling process, before insights that are produced are lost.

Furthermore, a parallel requirement was identified, for practitioners to be able to collate, manage and visualise the complex social interactions across any given scenario, and across any given business (see The TRE$_S$PASS Project, D4.3.3 (2016)). This includes modelling their own roles within this business, as a way of managing and accounting for how an issue is being tracked within its workforce. There was a clear need to manage tasks by establishing a clear and comprehensible narrative that can be understood and developed by all team members and stakeholders involved in a risk scenario.

The 'unique selling point' of any such tool would be to package the analytical processes into a single place, making a potentially complicated task of understanding 'messy' social practices more manageable. The prototype described below has been conceived as a way of extending and facilitating the co-construction process, and is intended to be used on-site during and after workshops, in face-to-face sessions, and that if desired, work can be shared remotely via its web-based architecture. In addition, starting points will be provided so that a user may craft their own forms of input (using a spreadsheet view of their data) that will be relevant to their practice.

This tool is designed as a simple web-based desktop application (Figure 2.1). It has three central technical aims, that it should be **responsive, scalable, and shareable**. It produces a map that is interactive and that grows with each piece of data added to it by the

users (Figure 2.2), linking to a synopsis page for each actor or node, which can in turn be called up from within the Attack Navigator Map (ANM). This linkage on its own is designed to offer practitioners the opportunity to query the technical models constructed within the ANM, probing the models more deeply by being able to examine the social dimensions on more detail than the ANM allows on its own terms.

Its components are *node.js* and *d3.js*. It is a flexible *Document-based* schema, stored in JSON format. Much of the logic in querying/manipulating the data is therefore carried out beforehand, written in the node-based middleware for the tool, or 'app' as this type of application has become known. In terms of how the graphing of elements is presented to the user, there is a web server instance with a d3 and HTML focused front-end. D3 uses JSON to hold the data which it applies into its visualisation code, in a simple network/directed graph that shows a collection of nodes and a collection of relationships. The data generated by the app is held in a document-based database as JSON files.

Queries upon the spreadsheet view of the data result in smaller subsets of the data held by *InterActor*, and these can in turn be exported as `.xml` or `.csv` files (Figure 2.3). By combining the results of several searches on a larger data set of social data, it is possible to assemble a more focused view of where the greatest areas of uncertainty lie. For example, during user feedback sessions on early versions of this tool, it was stated by a government analyst that the tool could help to determine which areas contain the highest threat densities. In this way the 'messy' and vague problems are given increased detail and depth, and while the tool is not designed to address these issues head on it can clearly be of use to help practitioners make decisions based on a greater understanding of why and where certain data may be unavailable and/or uncertain.

2.2.2 Fuzzy Inputs for the Attack Navigator

The TRE$_S$PASS Navigator Map interface requires the user to use fuzzy search fields that are designed to work with the underlying tree-based conceptual models. An input field will appear for the user to enter a label. It functions as a fuzzy search field for the interface user that looks up labels from an internal library and displays matching entries. Each newly entered label will be added to the library, so that it can be easily re-used (Figure 2.4).

For a fuller discussion and examples of visualisations of uncertainty please see The TRE$_S$-PASS Project, D4.2.1 (2014), Section 4.3.1.

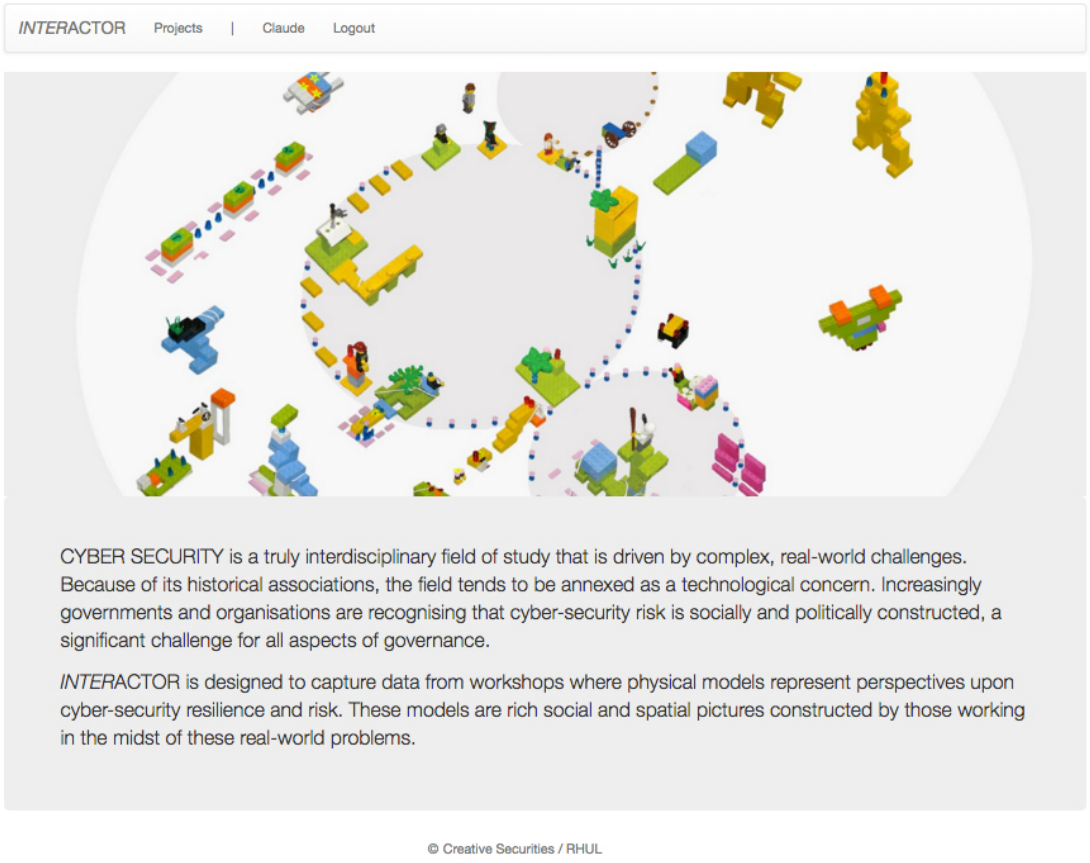

Figure 2.1: *InterActor*: The prototype *InterActor* 'app' as it first appears to the user.

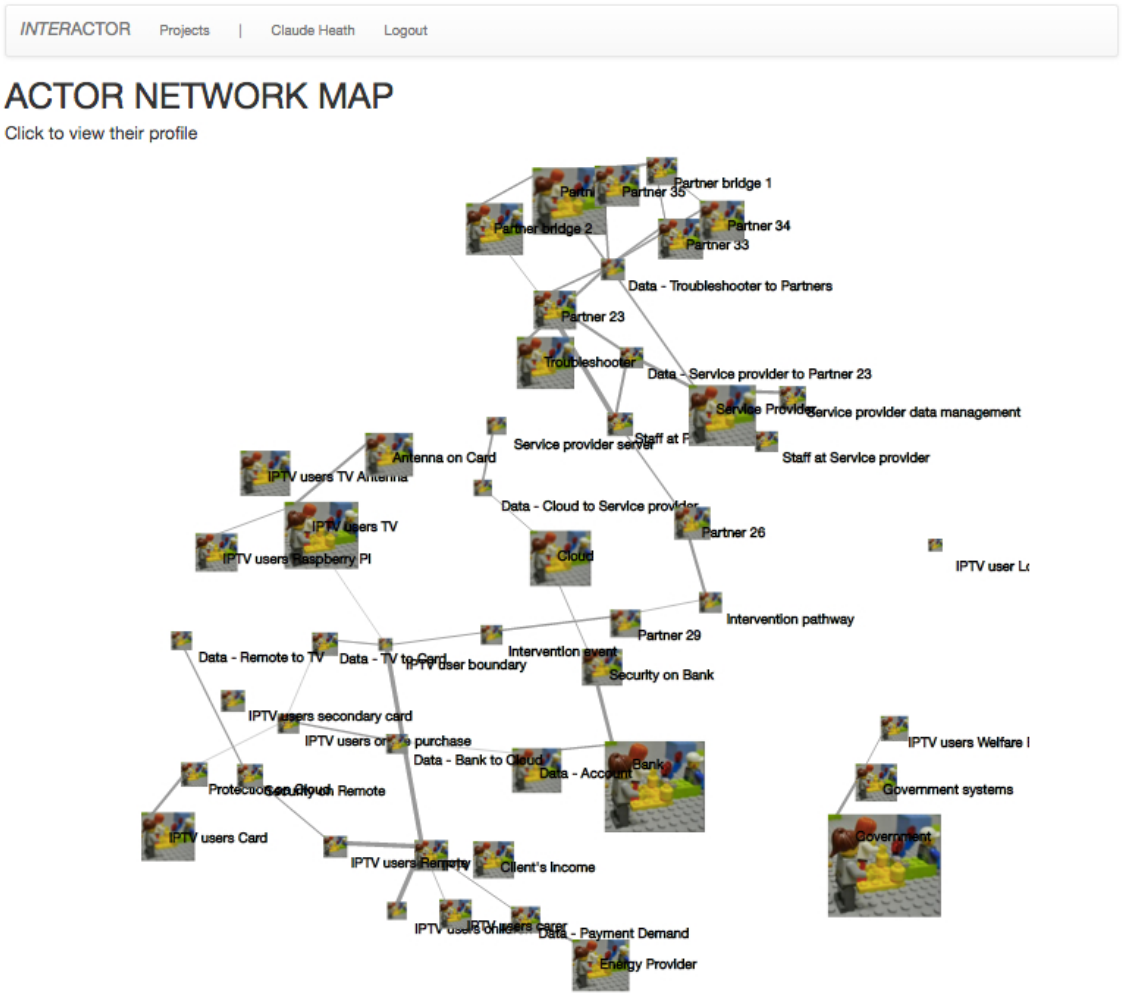

Figure 2.2: *InterActor*: the user can examine the actor graph in more detail, and weighted lines connecting them reflect the data entered (under positive and negative keywords in this case). The graph is also force-directed and can be reconfigured by the user.

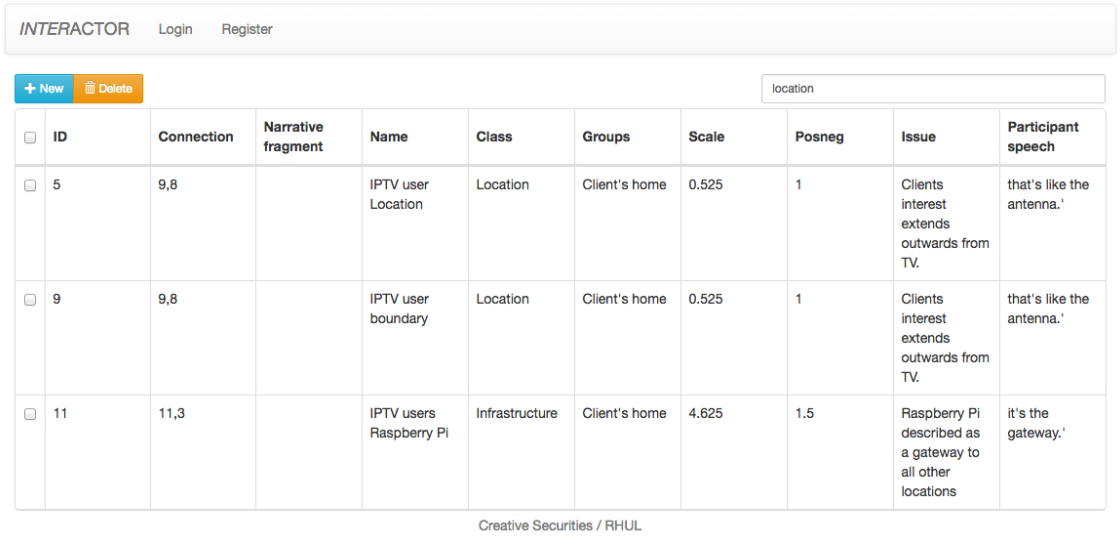

Figure 2.3: *InterActor*: the user can search the data spreadsheet and the page will display the results. The results are exportable.

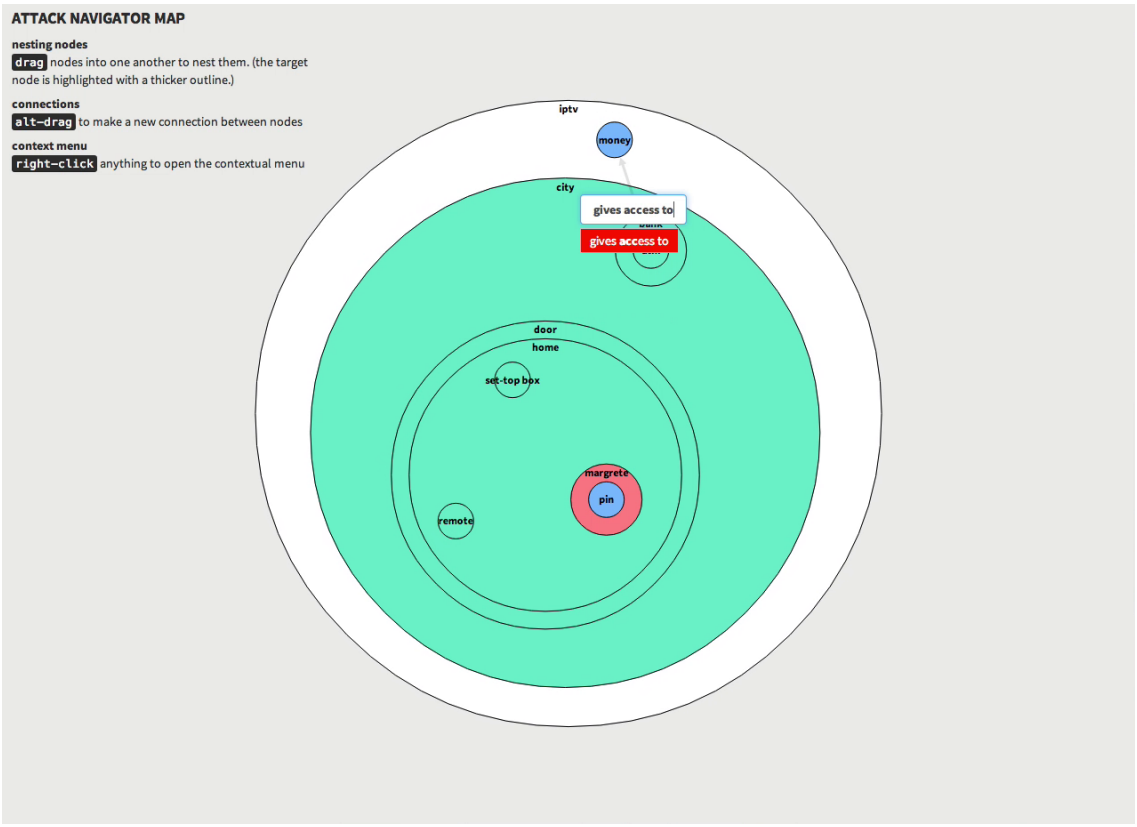

Figure 2.4: A fuzzy search field: an input field will appear for the user to enter a label. Here we define that an ATM gives access to money.

3 Data Quality and Modelling

Building teams: 'n-version modelling' The approach advocated here deals with uncertainty in modelling endeavours, first by having different teams start independently, and then by combining the results that they arrive at, to produce a new level of understanding of the problem of uncertainty. 'Brain-storming' and other forms of engagement offers a unique bottom-up approach, where participants of different disciplines can come together, communicate complicated virtual concepts to one another, and highlight any gaps in their understanding that might exist. This can be described as a qualitative approach to risk assessment, of which there are many different kinds. Visual and other forms of 'rich pictures' that emerge from such an approach, as discussed below, can be the subject of further more rigorous modes of enquiry.

Uncertainty and data provenance in the LEGO^{TM} modelling process The analogue modelling process for qualitative risk assessment (described in more detail in The TRE$_S$-PASS Project, D4.3.1 (2014)) is an alternative means of assessing the uncertainty provenance of data. Modelling the service design of IPTV with LEGO^{TM} has revealed important differences between technically accurate modelling and modelling that accurately portrays the ground-level working environment where appropriate context-specific controls are (or can be) practised. These practices are difficult to capture and analyse using traditional tools, but participatory workshops with practitioners using analogue methods have succeeded in initiating discussion of the risks associated with the future use of the IPTV service.

Testing the unavailability of data requires attention to the following three areas:

- Uncertainty
- Unavailability
- Unreliability

One simple way in which these issues may be addressed is to ask participants who have physically modelled a scenario to enact behaviours that extend their scenario in unusual ways, testing areas of uncertainty within their representations. This approach was employed at a LEGO^{TM} qualitative risk assessment workshop in London, where a small group of security professionals modelled scenarios that were familiar to them in their various professional experiences. The theme proposed to the group was bring-your-own-device (BYOD) in relation to cloud-based organisational practices. In pairs, they made individual models, and later followed this by linking them together into a single larger one, finding ways to connect what at first sight seemed to be disparate scenarios at specific points

that they had in common. During this process they created an oversized avatar to stand for “attackers”. This was not fixed into one position on the model but was allowed to wander and test the model at several points, and a time-lapse movie was made after the workshop, showing how the attacker roamed the scene (Figure 3.1). The unfixed nature of the attacker actor was a two-fold representation of generalised uncertainty, and of the uncertainty about which points on the model might be being exploited by the attackers.

This highly creative strategy devised by the practitioners, to move an attacker avatar around their model in this way, tested the assumptions about strengths and weaknesses of different points, and was designed to determine whether an attacker might be able to find and then open potential breach points, perhaps by establishing unforeseen links between different parts of the model, looking for gaps, moving into places and spaces in such a way as to open new avenues of attack. In this sense, the avatar was of non-fixed (uncertain) position. It was placed temporarily into different spaces, and the stated implication behind this strategy was that there is *little or no information* about where in fact attackers are probing for weaknesses.

Subjective probability: upper and lower limits of uncertainty. The method described above elicits interval-based estimations of certainty versus uncertainty at different points on the tangible models (upper bound, to lower bound). In such a case, sensitivity and uncertainty can be graded by the group: first by positioning avatars and other uncertain elements, and then by establishing whether there is a range that can be attached to spaces that are effected by these uncertain elements.

Practitioners who have tested the above-mentioned prototype (Section 2.2.1) have stated that having a tangible model is a good way of establishing the first points of a new risk register and then refining them (see The TRESPASS Project, D4.3.3 (2016)). Highlighting fields around items that form groups can be achieved in the prototype. This functionality can be used to represent the areas of social practice that are susceptible to the highest degrees of change and uncertainty, for example. In cases where this is indeed the research question that motivates the prototype users, this process of jointly establishing which areas require the most attention is likely to produce decisions based on an informed view of uncertainty and unavailability of data.

Indeed, the app can be used as a vehicle for recording these variable dimensions of data. Meta-information may be attached to parts of the digital model, annotating the “fragility” of organisational structures, as for example it was described by one participant of the London workshop, or describing the porosity of information boundaries at certain places in the representation. ‘Fragility’ could therefore be just one of many criteria that can be created by practitioners to inform judgements about the uncertainty of data.

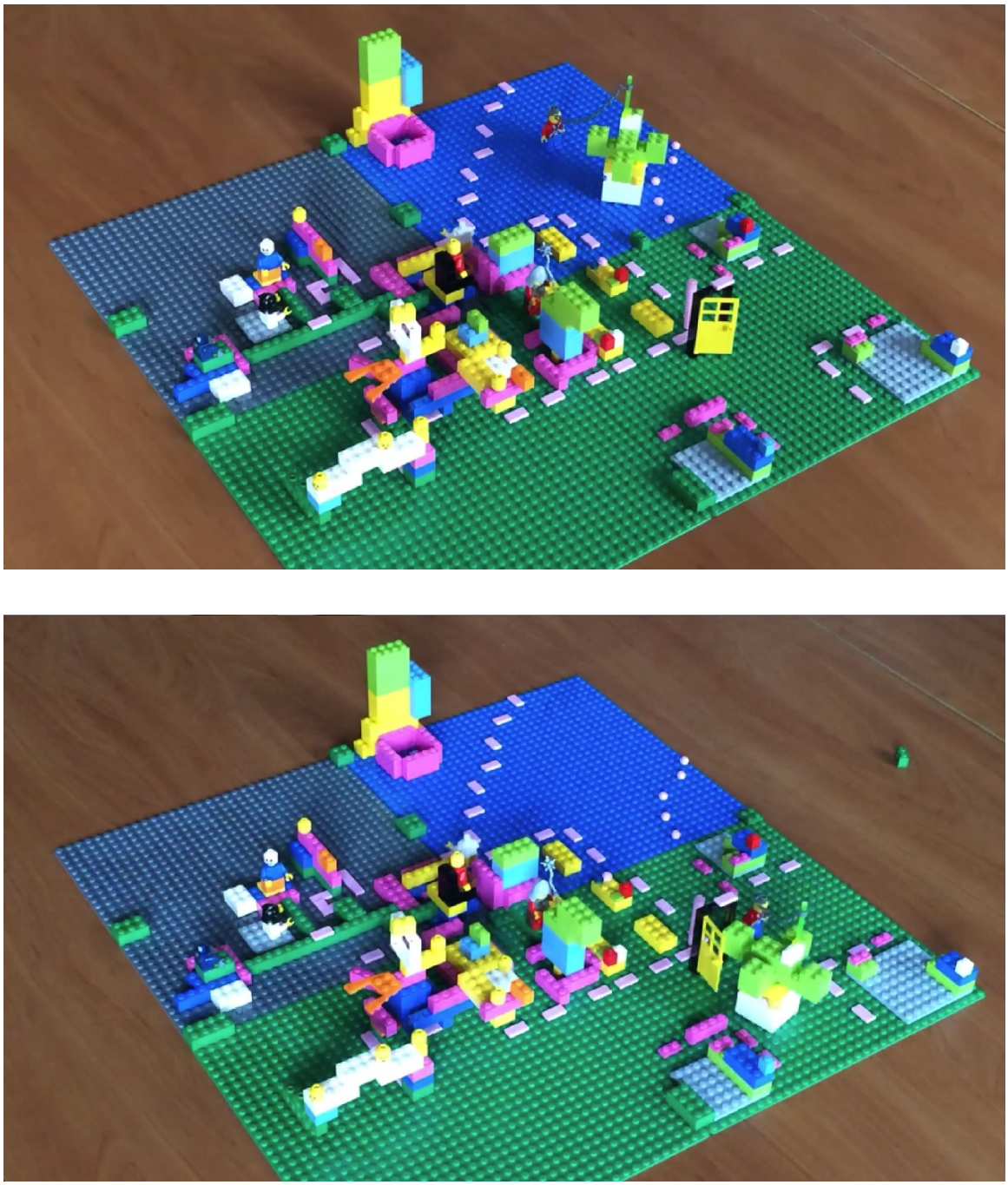

Figure 3.1: Two stills from a short time-lapse movie showing the attacker avatar, a tall structure moving from the upper right panel to the lower right panel. This had been given several heads with menacing facial features facing in all directions. This expressed the qualitative impression that it was not known where the attackers are directing their attentions.

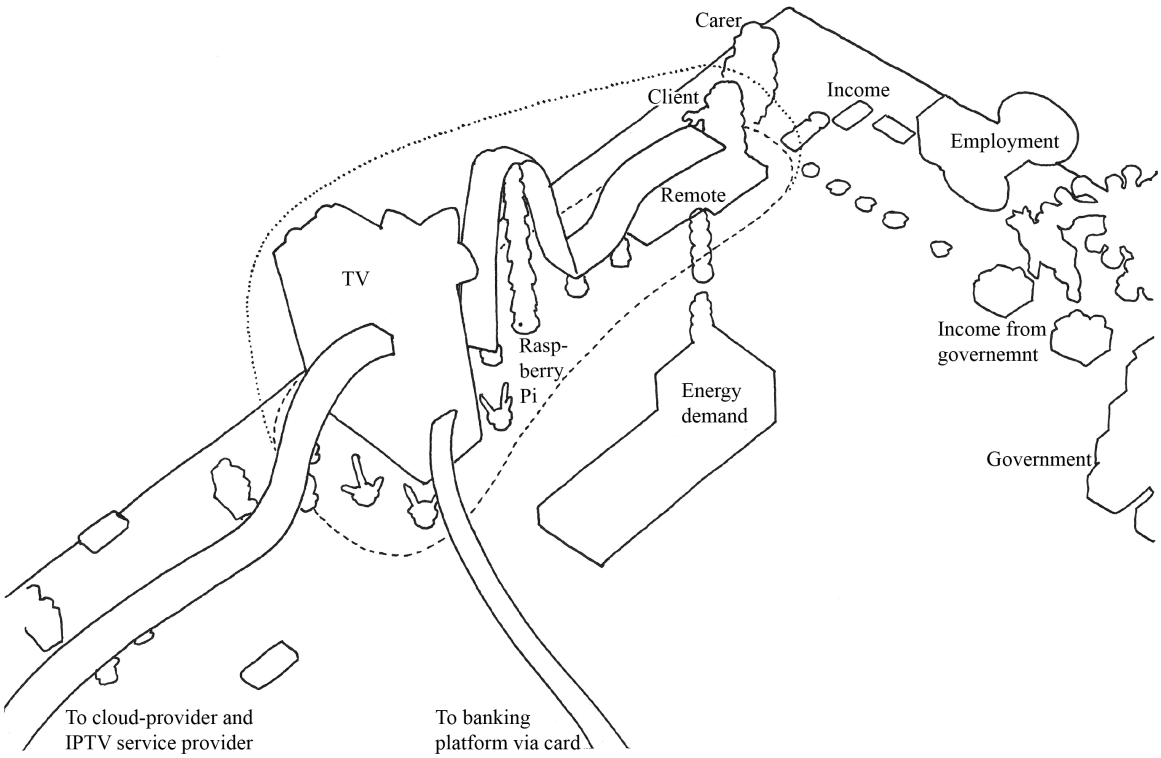

Figure 3.2: A drawing of part of the IPTV LEGO™ model, co-constructed by the service designers, LRS. The dotted line represents the space of the users' practice, around the use of the TV and remote control to access their account. The shape of the field shown here would change if the user's carer or relative was to be given or steal access to the account details. The dotted line refers to a known practice, but its shape is liable to change in response to new and uncertain developments.

4 Quantitative Analysis Tools

Deliverable The TRESPASS Project, D3.4.1 (2014, ch. 6) discusses data uncertainty in quantitative analysis tools. It lists three challenges for data integration in the attack generation process:

- "The data must be presented in an unambiguous form, suitable for analysis;
- The data must be linked to the appropriate Basic Attack Steps;
- Where data is uncertain (as it very often is), the uncertainty itself must be represented unambiguously, and the analysis methods must be able to cope with it."

According to The TRESPASS Project, D3.4.1 (2014) two approaches exist to gather quantitative data for fitting: historical/empirical data and expert opinion (Arnold, Hermanns, Pulungan, & Stoelinga, 2014). In our case the type of data source is irrelevant as far as data are obtained. In case of missing data in attack trees, parameter synthesis can help, i.e., if some probability value is unknown, take it as a parameter p, and see how properties and metrics depend on p.

The final WP3 deliverable The TRESPASS Project, D3.5.1 (2016) focuses on dynamics of stochastic models where sensitivity analysis covers one chapter. Sensitivity analysis measures how changes in input data impact the outcome. Hence, if one variable is very sensitive, then one should re-evaluate results for small changes in the value of this variable. Sensitivity analysis can highlight such variables for which it is most important to get an exact measurement. For variables that are less sensitive, re-evaluation is only needed for large changes in input values. For more details we refer to The TRESPASS Project, D3.5.1 (2016).

5 Conclusions

The TRE$_S$PASS case studies confirm existing knowledge when they list the challenges faced when collecting empirical data for risk analysis. In general, collections of empirical data may not be complete, be it that not all events have been recorded or that aspects relevant for risk analysis are missing from the records. This is particularly true for socio-technical systems where information about the human actors, their background, their social interactions, etc. would hardly be logged by default. Needless to say, such data collections would raise serious privacy concerns, as noted in the Royal Society report on cybersecurity (The Royal Society, 2016).

Data collection in security risk analysis is a dialogue between problem owners and security experts (Kang, 2013). The TRE$_S$PASS LEGO^{TM} modelling process provides means for structuring such a dialogue. Specifically, it supports a strategy where modelling is first performed in several sub-teams and the individual results are then compared and combined. Experimental evaluations have shown that this divide-and-conquer approach helps in the identification of uncertainties in the modelling phase.

Uncertainty about the precision of data collected also applies to empirical data, not only to expert opinion. Risk analysis methods have to accommodate this uncertainty, so the difference between *analysis based on data* and *analysis based on expert opinions* becomes less important to the extent that it does not really matter whether (incomplete) recorded data or data solicited from experts are entered into an analysis tool. The TRE$_S$PASS toolset handles both types of inputs. It has graphical interfaces for entering uncertain inputs and quantitative analysis methods for processing uncertain data.

The TRE$_S$PASS process can thus by design be applied in situations where empirical data is lacking and risk analysis is based on expert judgements. There is no need for separate data degradation tools but for processes and tool support for capturing uncertainty. The TRE$_S$PASS Project has made innovative contributions in this area. The IPTV case study was particularly useful in this respect as it allowed us to accompany the design of a new service from very early stages onwards.

References

Arnold, F., Hermanns, H., Pulungan, R., & Stoelinga, M. I. A. (2014). Time-dependent analysis of attacks. In *Proceedings of the third international conference on principles and security of trust, post 2014, grenoble, france* (Vol. 8414, pp. 285–305). Springer Verlag.

Batini, C., & Scannapieca, M. (2006). *Data Quality. Concepts, Methodologies and Techniques*. Springer. doi: 10.1007/3-540-33173-5

Botero, A. (2013). *Expanding design space (s) design: Design in communal endeavours.* Aalto arts Books, Helsinki.

Ionita, D., Bullee, J.-W., & Wieringa, R. (2014, Aug). Argumentation-based security requirements elicitation: The next round. In *Evolving security and privacy requirements engineering (espre), 2014 ieee 1st workshop on* (p. 7-12). doi: 10.1109/ESPRE.2014.6890521

Kang, M.-C. (2013). *Responsive Security: Be Ready to be Secure*. CRC Press.

Lund, M. S., Solhaug, B., & Stølen, K. (2010). *Model-driven risk analysis: the coras approach*. Springer Science & Business Media.

Probst, C. W., & Hansen, R. R. (2013, December). Reachability-based Impact as a Measure for Insiderness. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *4*(4), 38–48. Retrieved from http://eprints.eemcs.utwente.nl/24198/01/jowua-v4n4-3.pdf

The Royal Society. (2016, July). *Progress and Research in Cybersecurity. Supporting a Resilient and Trustworthy System for the UK* (Tech. Rep.). The Royal Society. Retrieved from https://royalsociety.org/~/media/policy/projects/cybersecurity-research/cybersecurity-research-report.pdf

The TRE$_S$PASS Project, D3.4.1. (2014). *Attack generation from socio-technical security models.* (Deliverable D3.4.1)

The TRE$_S$PASS Project, D3.5.1. (2016). *Dynamics of stochastic models.* (Deliverable D3.5.1)

The TRE$_S$PASS Project, D4.2.1. (2014). *Initial report on visualizations of information security risks.* (Deliverable D4.2.1)

The TRE$_S$PASS Project, D4.3.1. (2014). *Initial visualisations of socio-technical dimensions of information-security risks.* (Deliverable D4.3.1)

The TRE$_S$PASS Project, D4.3.3. (2016). *Visualizations of socio-technical dimensions of information-security risks.* (Deliverable D4.3.3)

The TRE$_S$PASS Project, D7.2.1. (2014). *Results from case study a.* (Deliverable D7.2.1)

The TRE$_S$PASS Project, D7.2.2. (2016). *Final report case study a.* (Deliverable D7.2.2)

Westerlund, B. (2009). *Design space exploration*. Unpublished doctoral dissertation, Phd Dissertation). Stockholm, Sweden: KTH.