

## Operational Trustworthiness Enabling Technologies



### D2.4 – Socio-economic evaluation of trust and trustworthiness

Stefanie Wiegand et al.

<b>Document Number</b>	D2.4
<b>Document Title</b>	Socio-economic evaluation of trust and trustworthiness
<b>Version</b>	2.0
<b>Status</b>	Final
<b>Work Package</b>	WP 2
<b>Deliverable Type</b>	Report
<b>Contractual Date of Delivery</b>	31/10/15
<b>Actual Date of Delivery</b>	31/10/15
<b>Responsible Unit</b>	IT Innovation
<b>Contributors</b>	Laura German, Costas Kalogiros, Michalis Kanakakis, Bassem Nasser, Sophie Stalla-Bourdillon, Shenja van der Graaf, Wim Vanobberghen, Stefanie Wiegand
<b>Keyword List</b>	Trust, Trustworthiness, Semantic modelling, User trust
<b>Dissemination level</b>	PU

## Document Review

Review	Date	Ver.	Reviewers	Comments
<b>Outline</b>	16/06/2015	0.1		
<b>Draft</b>	09/10/2015	0.2		merged iLaws/iMinds contributions in
	12/10/2015	0.3		replaced section 3 with Michalis' content
	14/10/2015	0.4		added updated iLaws/iMinds contributions
	20/10/2015	0.5		checked references, updated ToC and ordered contributors by name
	21/10/2015	0.6		extended section 2, added executive summary
	23/10/2015	0.7		corrected (cross-) references and updated introduction and summary
	23/10/2015	0.8		added missing contributions
	27/10/2015	0.9		QA
	28/10/2015	1.0		proof-reading
	30/10/2015	1.1		Integrated feedback and made the requested changes
	30/10/2015	2.0		Finalised document
<b>QA</b>			Karin Bernsmed, Vasilis Tountopoulos	
<b>PCC</b>				

## Glossary, acronyms & abbreviations

Item	Description
<b>AAL</b>	Ambient Assisted Living
<b>DADV</b>	Distributed Attack Detection and Visualisation
<b>E2E</b>	End to End
<b>GE</b>	Generic Enabler
<b>OPTET</b>	Operational Trustworthiness Enabling Technologies
<b>OWL</b>	Web Ontology Language
<b>PCC</b>	Project Coordination Committee
<b>RDF</b>	Resource Description Framework
<b>SMC</b>	System Model Compiler
<b>SMQ</b>	System Model Querier
<b>SPARQL</b>	SPARQL Protocol and RDF Query Language (recursive acronym)
<b>SPIN</b>	SPARQL Inferencing Notation
<b>SSD</b>	Secure System Designer
<b>SWC</b>	Secure Web Chat
<b>TME</b>	Trust Metric Estimator
<b>TW</b>	Trustworthiness
<b>TWME</b>	Trustworthiness Model Editor
<b>WP</b>	Work Package

## Executive Summary

In this deliverable, we present the work done on Trust and Trustworthiness models after the D2.3 milestone. The work focused on extending the models, enhancing their performance as well as accuracy when used across the socio-technical system lifecycle. This deliverable also presents the details of the validation and evaluation of these models, and their integration into the WP8 use cases (DADV, AAL and SWC).

The Trustworthiness model was enhanced with new asset types, threats and controls restructured in a modular way to allow easier future extension and performance optimisation as systems complexity grows. The GE presented in the 2<sup>nd</sup> year review was made more robust and finally used for the evaluation requested by the reviewers to show how both the model as well as the GE support the system design and provide additional value compared to the traditional modelling process.

The evaluation on the Trust model was done by conducting a large-scale experiment on users of a fictional search engine and questioning them about their perception of trust into the system depending on various factors. Furthermore, we analysed the effect of user trust on the legal framework.

The evaluation results and identified software bugs have already been taken into consideration in the final release of the OPTET GE's.

## Table of Contents

1. Introduction .....	7
1.1. Document Organisation.....	8
2. Trustworthiness Model Implementation and Evaluation .....	9
2.1. Introduction.....	9
2.2. OPTET core model.....	9
2.2.1. Roles .....	11
2.2.2. Patterns .....	11
2.2.3. Threats .....	12
2.2.4. Misbehaviours.....	12
2.2.5. Controls .....	12
2.3. OPTET generic model .....	13
2.3.1. Assets .....	13
2.3.2. Patterns .....	13
2.3.3. Threats .....	14
2.3.4. Controls, Control Sets and Control Strategies .....	15
2.4. The Compilation Process .....	17
2.4.1. Inputs.....	17
2.4.2. Algorithm .....	17
2.4.3. Output .....	19
2.4.4. Run-time Model Instantiation .....	19
2.5. Software components .....	19
2.6. Trustworthiness Model Validation & Evaluation.....	20
2.6.1. Evaluation plan and execution.....	20
2.6.2. Evaluation results and discussion.....	20
3. Trust Model Implementation and Evaluation.....	25
3.1. The Experiment.....	26
3.1.1. The experiment research-context .....	26
3.1.2. The experiment description.....	27
3.2. Users' Segmentation .....	34
3.2.1. Overview of the research approach .....	34
3.2.2. Derived Segments: Characteristics and validation .....	35
3.2.3. Third year research results .....	36

3.2.4. Fundamental expected properties.....	38
3.3. High level Evaluation of the results .....	38
3.3.1. The performance metric .....	38
3.3.2. The privacy metric .....	40
3.4. TME Revisited .....	43
3.4.1. The theoretical framework supporting TME .....	43
3.4.2. Comparison between the three approaches.....	47
3.4.3. An approach for finding the optimal time-fading TME parameters .....	48
3.4.4. Validation of TME.....	51
3.5. Post-questionnaire analysis .....	54
3.6. Post-questionnaire findings .....	56
4. Signalling (un)trustworthiness to end users – legal signposts.....	66
5. Summary and Future Work .....	76
6. References .....	77
Appendix .....	82

# 1. Introduction

---

This deliverable is an update of D2.3 [1], focusing on the scenarios used by WP8 for the final evaluation case studies in OPTET, namely AAL and DADV. As with D2.3, the report is accompanied by an updated version of the socio-economic model and threat model as well as the metrics, and describes the application and evaluation of models for trust and trustworthiness, and how these models are being used in WP3-WP6.

The OPTET threat model is the backbone of the threat identification process which leads to developing trustworthy systems and maintaining this trustworthiness during runtime using the threat diagnosis tools. The semantic models stack we developed in OPTET addresses these different phases of the system lifecycle. The trustworthiness expertise within a particular domain (e.g. healthcare) are encoded in the generic model based on an abstract core model defining high level concepts and terminology as asset, threat, misbehaviour, etc.

During the design phase, a system-designer models their system and generates a design-time trustworthiness model by applying the aforementioned trustworthiness knowledgebase onto their specific system. This is done automatically using semantic rules within the generic model that map the trustworthiness knowledgebase threats to the specific system based on its architectural patterns. In the deployment phase, the deployed assets of the system can be represented as instances of the asset types specified in the design time trustworthiness model. This marks the start of the runtime phase. During this phase, the dynamic system is monitored as it evolves using a runtime model. The runtime model is used for threat diagnosis by executing similar reasoning as in the design time but this time applied on the asset instances to detect threats based on their misbehaviours. The identified threats are then highlighted to the system operator alongside the potential controls for a faster mitigation.

In this deliverable, we present the updates to the models including the core and generic model as well as threats and controls. The updates were necessary for usability, performance enhancement and manageability of the knowledge base. The updates also include an extended asset model and the possibility of composing controls to form control strategies. An evaluation of the modelling approach was carried out via the Trustworthiness System Model Editor tool (product name: Secure System Designer). System designers went through the modelling exercise of the AAL system. The results of this evaluation highlighted the value of the automation allowing the design and analysis of the AAL system within couple of hours instead of days. The evaluation also highlighted the need for more restrictions and guidance during the modelling phase using the software. However, it did not point out any issue with the actual modelling approach or methodology. While encountered bugs were fixed and incorporated in the final tool release, some other enhancements will be addressed after OPTET and before getting the product to the market (for instance adding new asset types to cover wearable sensors and devices).

On the trust side, we extend in this deliverable our work on the socio-technical and legal factors that affect the subjective nature of trust and drive individuals' decisions in online environments. Our objective, which relates to these two former factors, is both to validate our user segmentation approach into clusters of similar trust-related behaviour, but also to further investigate trust shaping towards different metrics that characterize the performance of the system of interest.

The findings are utilized to improve the theoretical framework that supports the TME (Trust Metric Estimator) [2], and to conclude on the computational models that best estimate the actual trust values. We incorporated the legal aspects, by identifying the impact of legal information and

guarantees, e.g. signalling (un-)trustworthiness through signposting/cues, on individuals' trust responses. We emphasize here once again, that it is our methodology to discriminate between a user's trust level and their decision to pay the relevant price and engage (or not) with a system. This approach captures the real market conditions, where a rational potential user would weigh up their trust level by taking into account the monetary risks, costs and benefits.

Thus, the derived knowledge of the TME may be utilized on the provider's side as a powerful tool to compute the optimal price of the offered system, targeting their profit maximization. In D2.3 [1], section 4.1, we presented the provider's optimization problem at the design-time and quantified the additional gains achieved when the TME was applied, compared to the case of its absence, i.e., where all users are supposed to accurately assess the actual trustworthiness. In this deliverable we go farther with the TME incorporation into the provider's optimization problem during run-time, covering the whole life-cycle of a socio-technical system.

## **1.1. Document Organisation**

Section 2 is about the updated version of the trustworthiness model and how it was evaluated.

In section 3, we discuss the trust model and how socio-technical and legal factors affect user trust in online environments.

Section 4 presents the analysis of the impact of legal information and guarantees, e.g. signalling (un-)trustworthiness through signposting/cues, on individuals' trust responses.

Finally section 5 provides a summary with suggestions on how this work can be continued in the future.



## 2. Trustworthiness Model Implementation and Evaluation

---

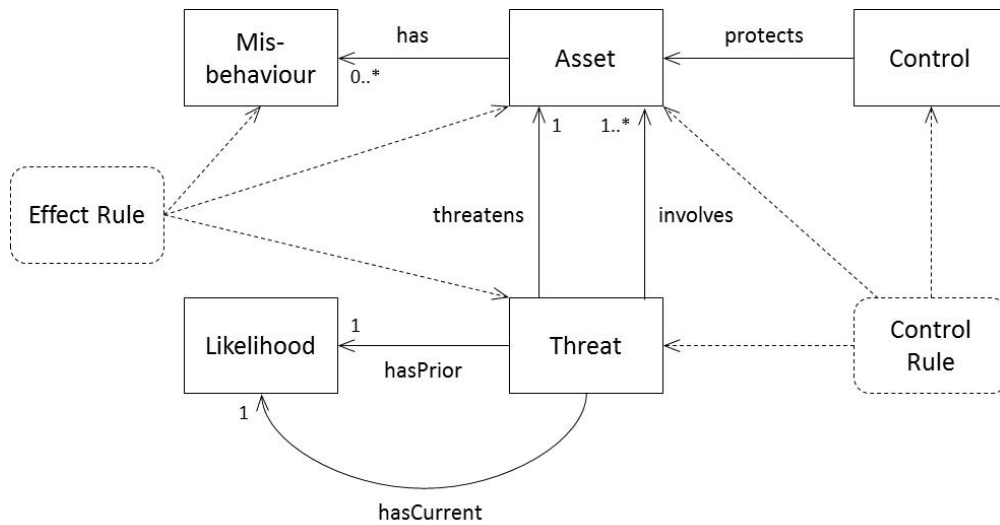
### 2.1. Introduction

The OPTET threat model is the backbone of the threat detection. Its application assesses the trustworthiness of a system and helps increase it by providing control strategies to mitigate threats during all phases of the OPTET lifecycle. During the design phase, it helps a system-designer to create a trustworthy abstract system model by analysing a system and identifying potential threats so the system can be redesigned before being deployed. In the deployment phase, the trustworthiness assessment of a concrete system takes place, applying the same reasoning as during the design phase but this time to OWL instances rather than classes. Then, during the runtime phase, the dynamic system can be monitored by periodically executing the reasoning and detect threats based on observed asset misbehaviours. The threats are highlighted to the system operator alongside the potential controls for a faster mitigation.

This section covers the lifecycle of the OPTET threat model through design-time and run-time. We describe the OPTET model stack, consisting of the core model (see section 2.2) and the generic model (see section 2.3) and how they are combined with user input obtained through the usage of the Secure System Designer GE (formerly TWME but renamed for marketing reasons) [1] to be compiled (see section 2.4) into a full design-time model, which adds threats found by pattern matching to the asset model defined by the system designer. Finally we will evaluate this approach in section 2.6 in a small-scale modelling experiment using the SSD.

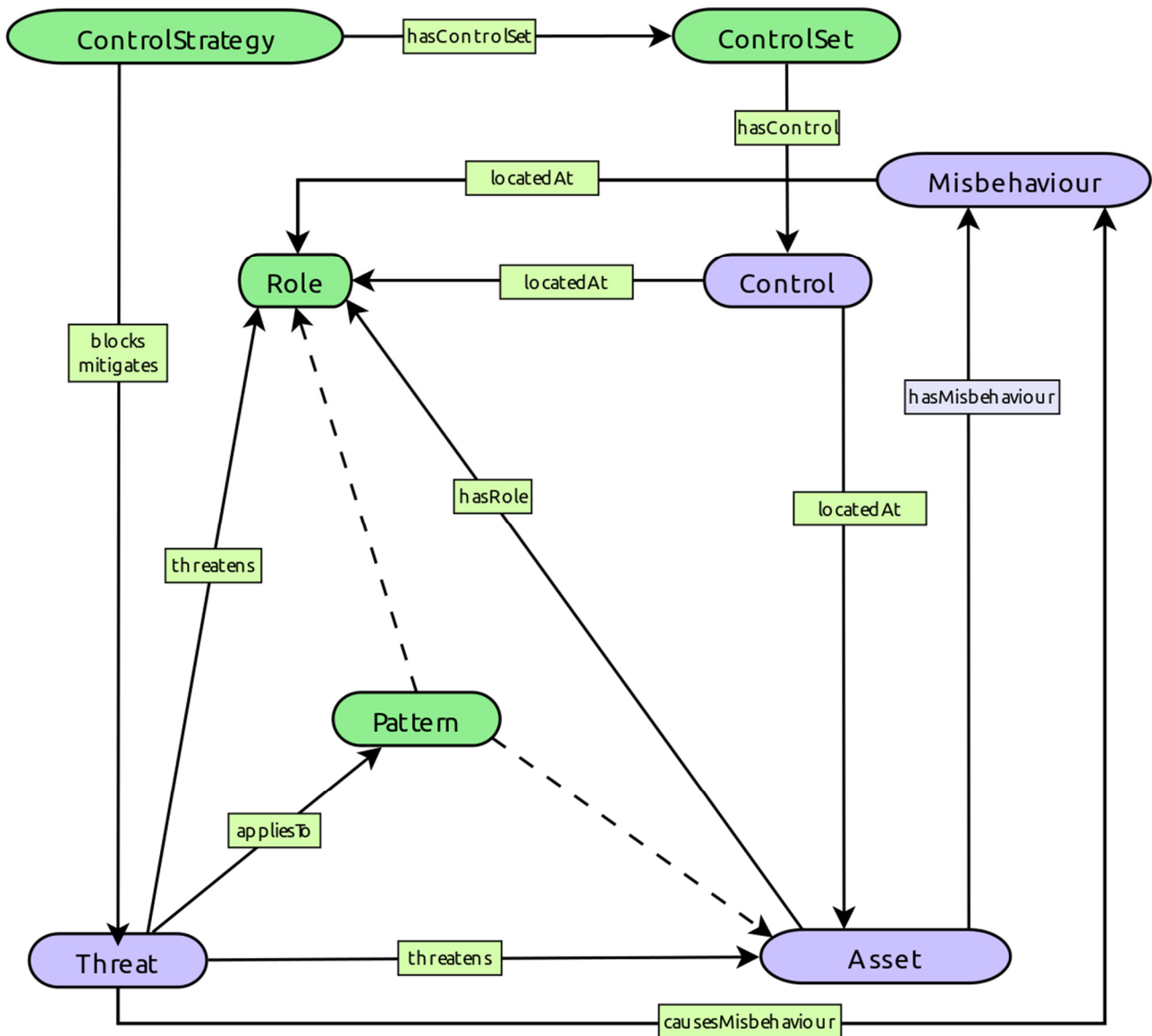
### 2.2. OPTET core model

Since D2.3 [1], the underlying core model has changed significantly to reduce the amount of redundant information included in the ontology and to make it easier to add extensions, for example new (generic) misbehaviours, controls, assets or threats. Figure 1 shows the previously used core model as of D2.3. This has some shortcomings, which this new model aims to address, like the inability to retain the connection between assets (and their roles) in specific patterns only and the lack of an efficient way to maintain and enhance threats. This section will highlight the key differences between the two versions of the model.



**Figure 1 – The old core model [1]**

The new core model as shown in Figure 2 looks much more complex as it has a lot more classes and properties. However, this makes defining generic assets, patterns and threats much easier. The purple objects stem from the original core model while the green ones have been added in this version. The dashed lines represent indirect connections that have been simplified in this figure for the sake of clarity.



## Figure 2 – The new core model

### **2.2.1. Roles**

The old model classified system-specific assets by attaching **Asset** subclasses to them based on their relationships. This was basically mixing asset classes ("What is this asset?") with roles ("What does this asset do?") that assets can have within a certain pattern. To better distinguish between an asset's class and its role in the system, a new **Role** class has been introduced.

### 2.2.2. Patterns

While in the old model a threat already applied to a pattern, we never explicitly defined patterns. As a consequence, repeating patterns (like Client-Service) had to be re-matched every time a threat is applied to the system topology. The new model now introduces explicitly defined patterns using the

new **Pattern** class. A pattern is a representation of a directed graph and contains a number of nodes (at least one) and a number of links (can be 0). Each **Node** has an **Asset** and a **Role** it represents. A **Link** links nodes and has a link type which represents an object property.

These patterns are used to create subclasses based on them to represent actual patterns found in the abstract system model. These subclassed patterns will contain all the system-specific asset subclasses involved in the pattern and for each of the involved system-specific asset subclasses which role they have in this particular pattern subclass.

### **2.2.3. Threats**

Where threats mainly consisted of SPIN templates, threats are now defined semantically. They no longer have to match a pattern but are linked to a generic pattern. Once the pattern subclasses have been created, a threat subclass will be generated for each of the patterns it applies to.

The "involves" object property has been removed as it is now redundant: a threat applies to a pattern and implicitly involves all the assets within this pattern.

Threats can have **SecondaryEffectConditions**, which is a means of expressing conditions, under which they would be considered to be secondary effects (knock-on consequences) rather than primary effects. This means if the secondary effect conditions are met, the threat is caused by the misbehaviours given in the conditions, otherwise it is just a normal threat. A secondary effect condition describes a misbehaviour located at a role from the pattern to which this threat applies. If all the conditions are satisfied (i.e. the secondary effect's defined misbehaviours are present on the assets in the pattern this threat applies to), a threat can be classified as a secondary effect.

The remaining properties (hasAction, hasConsequence, hasCurrentLikelihood and hasPriorLikelihood) have been left untouched and work like they used to in older versions of the ontology model.

### **2.2.4. Misbehaviours**

Misbehaviours are malfunctions that assets can exhibit and that are ideally measurable when monitoring the asset. A threat which is active can cause misbehaviours in an asset taking a certain role within the pattern the threat applies to. However, it also works the other way round: SecondaryEffects are a way of describing that a Misbehaviour could have caused a threat. This ability is defined using the causesMisbehaviour object property in the threat class definition.

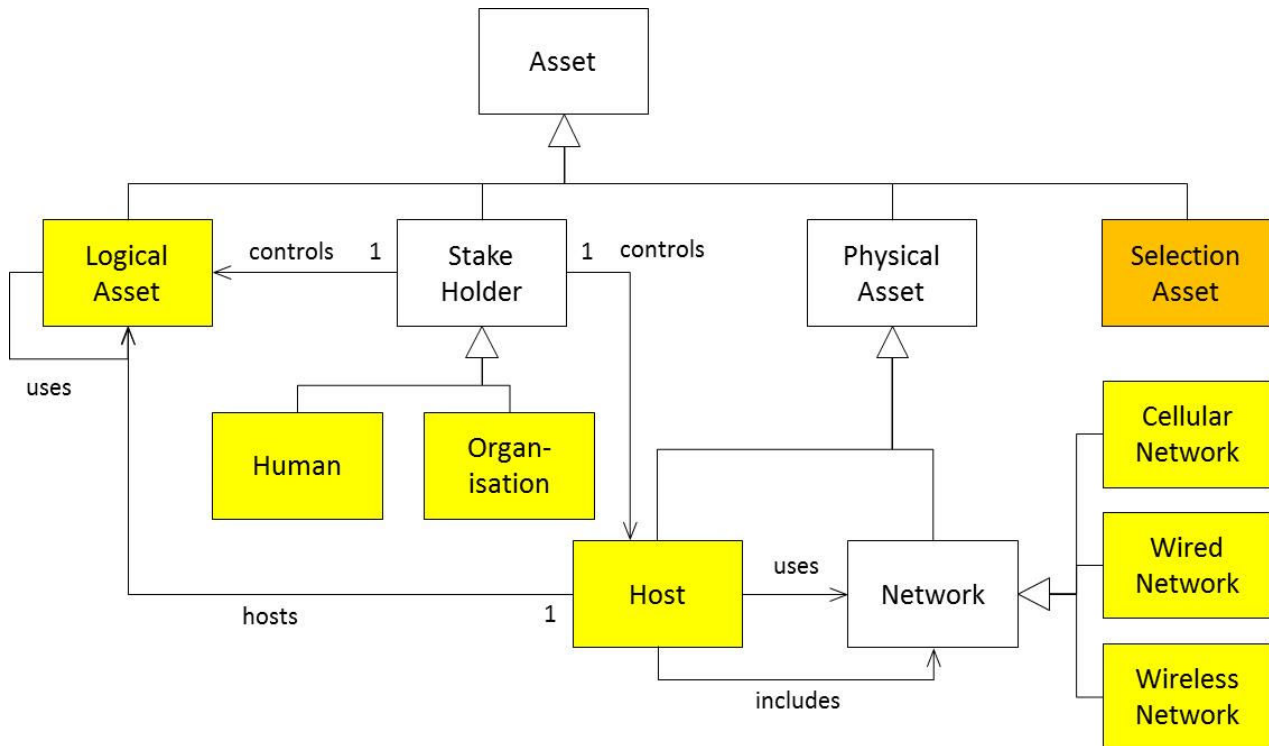
### **2.2.5. Controls**

Each threat can have one or more control strategies to block or mitigate it. Each **ControlStrategy** has one or more **ControlSets**, which consist of one **Control** and one **Asset** at which the control can be located.

A threat is blocked/mitigated when one of its control strategies is implemented, i.e. if the control(s) contained in the control strategy are implemented on the assets which have the roles specified in the control strategy within the pattern the threat applies to.

## 2.3. OPTET generic model

The following figure shows the previous generic asset model, which is still valid. However, all the patterns have been implemented as described above in 2.2.2.



**Figure 3 – Generic asset classes**

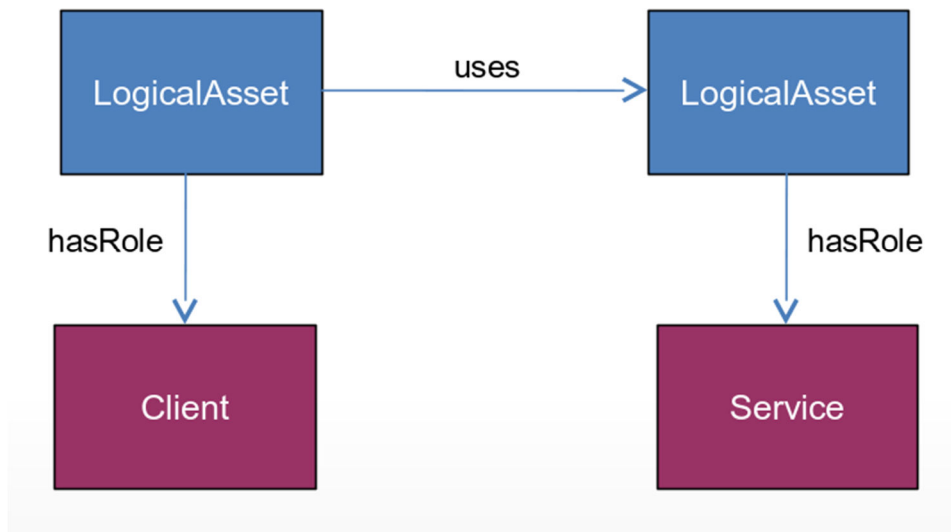
### 2.3.1. Assets

Introducing roles slimmed down the generic asset subclass tree significantly. All the assets present in the generic model now equal the assets classes that can be used when creating a new system model, with exception of ServicePool and Interface – those can be inferred automatically.

### 2.3.2. Patterns

In the previous version of the model, there was a SPIN template for every pattern which encoded the assets and relations to be matched. Each of these would be run to assign the asset classes to the assets which encoded what is [1] now modelled as roles. In the new model, there are no individual threat rules; the patterns are encoded in OWL/RDF. It means that when specifying patterns, the security expert no longer needs in-depth semantic knowledge; instead all that is required is an understanding of graphs as a means to express patterns.

The simplified graph shown in Figure 4 illustrates the knowledge encoded in a pattern. It is a representation of the pattern as a graph, recording all its nodes (assets and roles) and relationships.



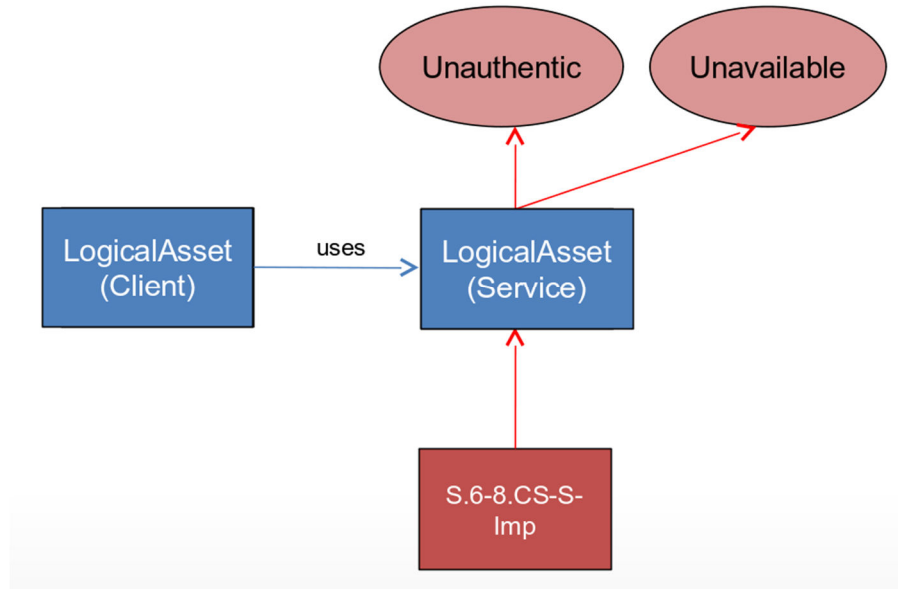
**Figure 4 – The Client-Service pattern**

### ***2.3.3. Threats***

As described above, threats no longer model the whole pattern they apply to. Instead, they are now quite simple.

Each threat still threatens one asset only, except in the generic threat classes - as opposed to the system-specific class - it threatens the role rather than the asset itself because the system-specific asset subclass which will be threatened is unknown at this time. However, by referring to a role which is unique within a pattern and giving the pattern, the system-specific asset subclass can be queried during compilation and linked directly in the system-specific threat subclass. The "involves" relationship is no longer explicitly asserted; instead all assets within a pattern are considered to be involved in a threat which applies to the pattern.

Each threat can cause 0..n misbehaviours (as shown in Figure 5) and have 0..n control strategies.



**Figure 5 – An example threat**

Also it is possible for a threat to specify 0..n secondary effect conditions. If all of these are satisfied, the threat can be a secondary effect – though this won't be classified until the system operation phase where we encode the observed conditions within the runtime model and identify secondary threats.

#### **2.3.4. Controls, Control Sets and Control Strategies**

This whole concept of the new threat model is based on the definition of controls. Each control can be located at a number of roles. These ontology rules prevent any potential errors a designer might make while trying to put a control on an asset which is not compatible with that control. The following table shows the link between the controls and the possible assets where they may be located at.

Control	Selection Asset	Stake holder	Host	Inter face	Net work	Logical Asset	Comments
Firewall classes				X			
Secure Configuration			X				
Software Patching			X			X	

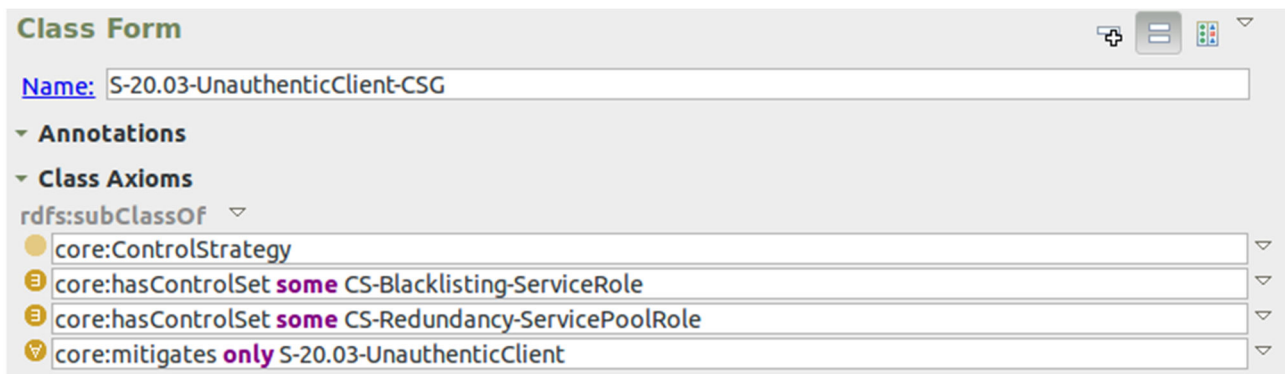
Software Testing			X			X	
Secure Transport					X		
AntiMalware			X				
MailScanning						X	Only MailAgent, MailStore and associated clients
Sandboxing						X	
UserTraining		X					Only Humans
Identification		X	X			X	
Strong Identification		X	X			X	
Delegation						X	
Client Authentication			X		X	X	Better with physical networks
Service Authentication			X			X	
AccessControl			X		X	X	Better with physical networks
Trust Management			X		X	X	Better with physical networks
Blacklisting		X	X			X	
Blacklisted		X	X			X	
Redundancy	X						
Service Switching	X						Only ServicePool
Input Checking			X			X	
Scalability			X			X	

**Table 1 - Controls**

A control set is any combination of control and asset according to the table. Because system-specific control sets cannot exist before all the system-specific asset subclasses are known, a control set definition can also be given as a combination of a control and a role. As soon as it becomes known which asset subclass has the role, the system-specific control set can be generated.

Once all the control sets have been generated, the security expert can bundle them in control strategies, as shown in Figure 6, indicating all the specified control sets that should be in place in order to block – or in this case mitigate – a specific threat. Currently this is done using an ontology editor (such as Protégé [3] or TopBraid Composer [4]) however we plan to provide a GUI to facilitate this task (not within OPTET).





**Figure 6 – A control strategy**

Each threat can have multiple control strategies which might require different controls and might be easier, cheaper or have another advantage. For a given system model, it can then be queried which control strategies can be implemented e.g. using the fewest or cheapest controls.

## 2.4. The Compilation Process

This step has become more complex but at the same time much faster. The previous compilation process required to first run an OWL reasoner (e.g. Hermit [5] or Pellet [6]) to assign Asset classes (now roles) to system-specific asset subclasses via the `rdfs:subClassOf` property. This step is now redundant and the OWL reasoner has been eliminated altogether from the process making it a good deal faster. However, the compilation is now more of an incremental process with lots of smaller (thus faster) rules building on top of each other. This section describes how it all works.

### 2.4.1. Inputs

Like before, the inputs will be the generic model, containing generated control sets, and a system-specific asset model (produced by the SSD).

### 2.4.2. Algorithm

#### 2.4.2.1 Generic model compilation

**Prerequisites:** a generic model containing control definitions and misbehaviours.

This step does several different things:

- It generates the generic control sets. To do this, it runs a template which finds all the roles on which a control can be deployed. Then it creates control sets for these combinations.
- It creates one instance per misbehaviour
- It creates one instance per control

The inferred triples are then added to a separate file and imported into the generic model.

### 2.4.2.2 Implicit system-specific asset class generation

Prerequisites: 2.4.2.1

This step runs a template to generate implicitly defined asset classes such as interfaces and network groups, whose existence can be inferred completely.

### 2.4.2.3 System-specific pattern subclass generation

Prerequisites: 2.4.2.2

This step subclasses the generic patterns, replacing the generic assets in the pattern definition with their system-specific counterparts. Instead of one single template, it is written in Java and generates SPARQL queries for the generation of subclasses for each pattern using the information from the generic pattern class definition.

### 2.4.2.4 System-specific control set generation

**Prerequisites:** 2.4.2.1, 2.4.2.3

This step reads all the generic control sets and then identifies all the system-specific asset subclasses that take the role specified in the generic control set within one of the generated system-specific patterns. For each match, it creates a system-specific control set putting the generic control on the system-specific asset subclass.

### 2.4.2.5 System-specific threat subclass generation

**Prerequisites:** 2.4.2.2, 2.4.2.3

This step has three different parts:

- First, a template is run to create system-specific threat subclasses that apply to a system-specific pattern subclass and threaten a system-specific asset subclass.
- Next it runs a template to attach all the possible misbehaviours to the threat that can be caused by it.
- Finally it runs a template to attach all the secondary effect conditions to the newly created threat subclass.

The results of this step are complete system-specific threat subclasses, containing all the necessary information (the pattern it applies to, the threatened asset, applicable control strategies, caused misbehaviours and secondary effect conditions).

### 2.4.2.6 System-specific control strategy generation

**Prerequisites:** 2.4.2.4, 2.4.2.5

This step does two different things:

- First it runs a template to generate system-specific control strategies and link them to the system-specific threat subclasses.
- Then it runs another template to attach the matching system-specific control sets to the newly generated control strategies.

### **2.4.3. Output**

All of the information (i.e. the inputs as well as all the inferred triples) is saved to a new file, commonly referred to as the compiled or full system model.

### **2.4.4. Run-time Model Instantiation**

All of the above steps happen during design-time and are only preparations for using the model at run-time. To use it, instances need to be created based on monitoring information.

Whenever a new asset instance is detected, it is added to the model.

Then, the pattern instance generation template has to be run again to detect if the instances form a new pattern.

Following this, the threat instance generation template is run to create new threat instances that might affect any newly created pattern instance.

At any time during run-time, the operator can blacklist asset instances to exclude them from any potential threats (this is a means of manually overriding the system). Also control instances can be assigned to the asset instances to protect them.

The monitoring also gives information about misbehaviours that can be detected on the asset instances. As soon as this happens, the threat instances have to be reassessed to see whether they are vulnerabilities, secondary effects, blocked or mitigated.

## **2.5. Software components**

This is a brief overview of the different software components developed in WP2 and how they contribute to the OPTET lifecycle.

- **System Model Compiler (SMC)**

This component coordinates the compilation process as explained in section 2.4. It can compile generic, design-time and run-time models. In order to do this, it provides a high level API which is called by other components, such as the SSD, SMQ or the System Analyser.

- **System Model Querier (SMQ)**

This component represents the query interface to a system model. It contains a number of preconfigured, parameterised methods to retrieve various parts of the model which are needed to answer questions such as "How many threats does this generic model contain?", "How many threats would be mitigated in this design-time model if this control was deployed on all the assets of this class?", "Which of the threats in this run-time model are currently active?". Like the SMC, the SMQ provides an API for other components to use.

- **System Analyser**

This component wraps the SMQ functionality and provides a Restful service that allows other WP components (e.g. WP3 E2E TW calculator) to query the knowledge in the models.

- **Secure System Designer GE (SSD, formerly TWME)**

The SSD is the GUI for designing abstract design-time models. It uses the generic model, the SMC and SMQ to compile a design-time model based on user input. After compiling, the user can navigate the model and view potential threats to the system and control strategies to block/mitigate them.

Since it is still a design-time model, the design can be changed. Finally, a report is generated, containing all assets, relations and threats within the modelled system.

For more detailed documentation of this component, see D7.2 [7] and D7.3 [8].

## **2.6. Trustworthiness Model Validation & Evaluation**

This subsection covers the evaluation done on the trust model using the SSD. It contains two parts and aims to validate the OPTET model itself in terms of functionality as well as the SSD and its capabilities and usability.

### ***2.6.1. Evaluation plan and execution***

To get an evaluation for the above parts, the best way was to run an experiment on a small group of system designers. We were able to recruit 5 IT Security Students from the University of Duisburg Essen. They were provided with the SSD documentation and installation instructions a couple of days prior to the experiment and one A4 page description of the system to be modelled on the day (the AAL scenario [9]). We held a presentation covering the introduction to threat modelling prior to the actual experiment to introduce OPTET and the scope of the experiment. They then had time to familiarise themselves with a test version of the SSD running on their own hardware. The modelling of the system (including threat generation) took about 2 hours and we obtained the resulting model for analysis. After the modelling, each participant filled out a questionnaire [10] about the user experience related to using the SSD.

### ***2.6.2. Evaluation results and discussion***

#### **2.6.2.1 SSD questionnaire**

The general reception of the SSD was positive. The participants reported being faster in threat modelling due to the SSD and likely to use it again in the future. There were several comments on bugs/glitches in the software as well as some feature requests which were very useful for us and will be considered for future versions of the SSD.

Some of the positive comments we received include

- It's fast compared to the manual process
- It's easy to use and not overly complex

The criticism included

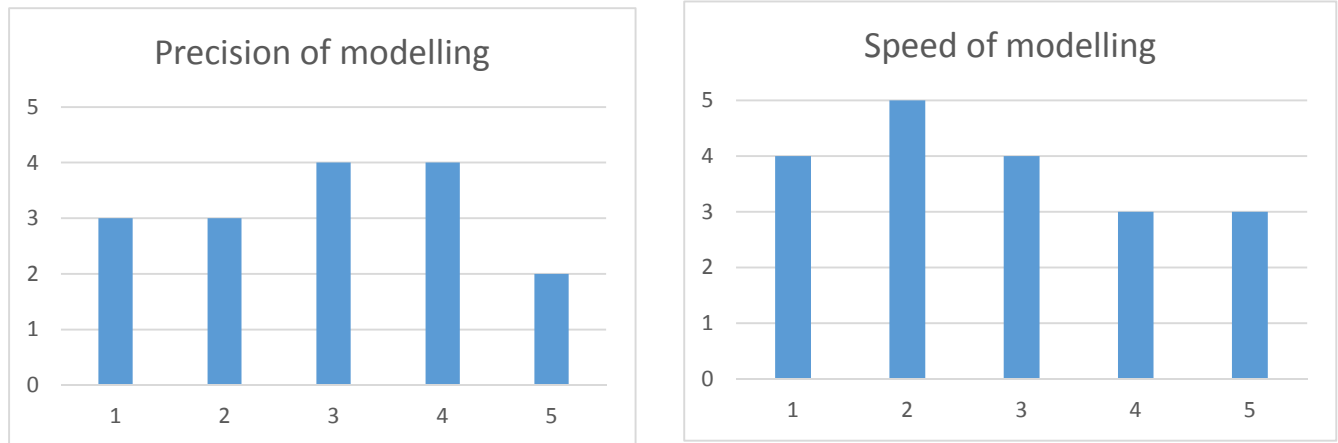
- The software needs an introductory tutorial to be used effectively
- The performance (of the compilation) can be improved
- It is restrictive in terms of assets available for modelling (e.g. wearable devices)

Users asked for the following features and enhancements to be available in future versions:

- More asset- and relationship types to choose from
- Usability enhancements such as multi-select, resizable canvas and sorting options
- Insertion of patterns (e.g. logical asset with a host) directly via drag and drop

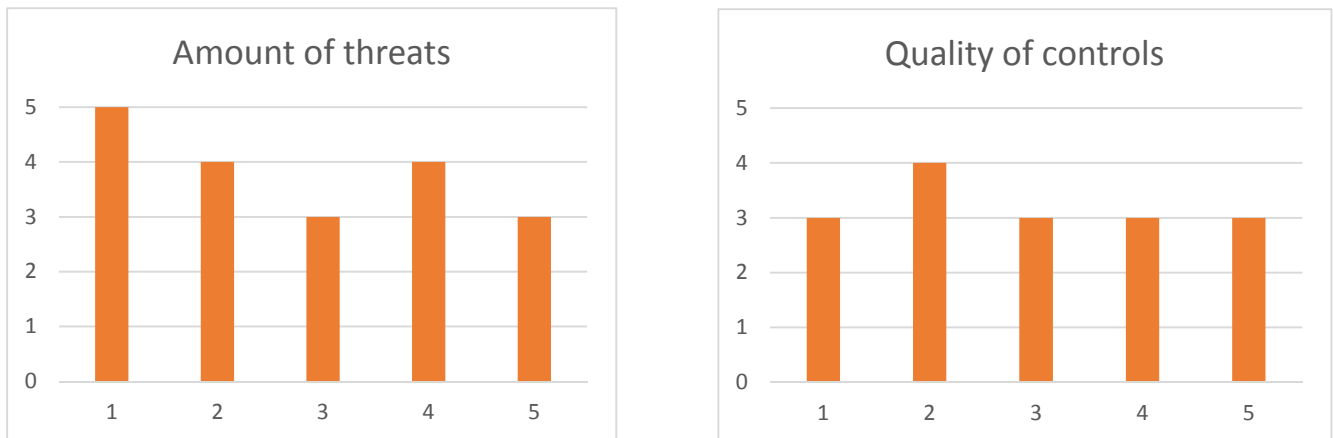
Generally, the most important aspect of threat modelling to the users was the ability to precisely define the assets and their relations and model the system as quickly as possible. For this, the system achieved a satisfactory score.

The scale in the following figures is always 0-5 where 0 means disagree strongly/very bad/very few and 5 means agree strongly/very good/many. The x-axis shows the individual participants.



**Figure 7 – Modelling experience**

Figure 7 shows the feedback on the modelling experience in general. While we achieve good scores for speed of modelling, the participants found it too restrictive in the amount of assets offered for the modelling. This was intentional from our side to reduce potential mistakes by matching patterns in the modelled systems which can only be achieved in a manageable way by restricting the amount of available, generic assets the user can use (subclass).



**Figure 8 – Quality of the OPTET threat model**

The quality of the model itself is highlighted in the responses shown in Figure 8. While the participants seemed happy with the number of threats found in their system, they welcome more options when it comes to the choice of controls. This highlighted the need to include more details about the current controls which should be:

1. communicated to the user more clearly to show that the presented controls would in fact be sufficient to block and/or mitigate the threats
2. allow users to include more controls and build custom strategies based on their system knowledge.



**Figure 9 – Software usability aspects**

Finally, Figure 9 shows that the SSD as a threat modelling tool was perceived to be well presented and easy to use while still covering the threat modelling task in the design stage of system design.

### 2.6.2.2 AAL system models

The following table provides a summary of the system models created and compiled by the participants for the AAL scenario.

Creator	Assets	Relations	Patterns	Threats
<b>OPTET</b>	<b>16</b>	<b>20</b>	<b>119</b>	<b>230</b>
Participant1	21	24	132	217
Participant2	18	21	125	192
Participant3	23	25	153	225
Participant4	19	21	174	205
Participant5	13	13	148	235

**Table 2 - Evaluation model stats**

**Assets:** The scenario itself was descriptive but abstract enough to allow different ways to model it. This is reflected to a certain extent in the number of assets chosen by the modellers. The number of assets did not fall below our minimal reference model (in the first row) except in one case (participant 5). This is mainly because the actual constructed model was restricted to cover some but not all the interactions required.

**Relations:** Given that the number of assets was in general higher than our reference model, the number of relations in certain cases did not reflect this. The produced models lacked in some places relations amongst logical assets or between logical assets and their hosts. Though it is possible to have no relations amongst the logical assets, it is counter-intuitive to have a logical asset without a host. More checks on the model need to be introduced in order to help the modeller in producing a realistic model.

**Patterns and threats:** the number of patterns found in the evaluators models was generally higher than our minimal model. This is expected given the higher number of assets and relations used in the evaluation models. Each pattern corresponds to at least one threat. However, Service pools, required in considerable number of threat patterns, was not easy to understand and thus to specify. This meant that the number of threats identified varied amongst the models in relation with our reference model.

However, from a practical perspective, the number of threats identified in each model shows the advantage of our automated approach. Identifying such number of threats manually would require much longer time and security expertise without guaranteeing consistency.

### 2.6.2.3 Future enhancements recommendations

The results of this evaluation highlighted the need for more restrictions and guidance during the modelling phase using the software. However, it did not point out fundamental issues with the actual modelling approach or methodology. Enhancements can be made in the future for a better user experience and output models:

1. The user constructs their model by dragging and dropping assets from the side bar. In order to avoid errors like having a logical asset without a host, we can make sure that the

tool inserts such assets automatically. Overall checks of the system can also be introduced to warn the user about logical assets that are not interacting with any other logical assets (it is possible to have a standalone logical asset, and thus a warning is only needed)

2. The service pool notion did not prove easy to use in the SSD which caused some related threats to be missed. This can be avoided easily by having default service pools configurations avoiding user confusions. The default configurations still allow advanced users to tailor the model to their scenarios by editing them when needed.
3. The scenario included wearable assets which were not explicitly supported by the SSD. However it was possible to model them using generic subcomponents (i.e. logical asset and host). In order to facilitate the modelling task for the user, assets will be:
  - a. accompanied with detailed description so that the user is clear on their semantics
  - b. organized in clusters for clear display, easy search and retrieval
  - c. more asset types will be added following the current information system trends ( e.g. wearable sensors, smartwatch, etc.)
4. Allow more customization when it comes to the selection of control strategies. The current version of the tool does not provide a way for users to choose/add controls in order to build custom strategies based on their intimate knowledge of their information system and organizational culture.



### **3. Trust Model Implementation and Evaluation**

---

In this section, we extend our research on the socio-technical and legal factors that affect the subjective nature of trust and drive individuals' decisions in online environments. Our objective, which relates to these two former factors, is both to validate our approach that clusters users into segments of similar expected trust-related behaviour, but also to further investigate trust shaping towards different metrics that characterize the performance of the system under interest. Our findings are utilized to improve the theoretical framework that supports the TME (Trust Metric Estimator) [2], and to conclude on the computational models that best approximate the actual trust values. Concerning the involvement of legal issues, we intend to identify the impact of legal information and guarantees, e.g. signalling (un-)trustworthiness through signposting/cues, on individuals' trust responses.

The process followed is aligned with the one employed during the previous two years of the OPTET project: we designed and performed an experiment where participants engaged with a fictitious on-line service, observed its functionality and reported their trust values concerning two metrics namely "performance" and "privacy". Additionally they answered a post-questionnaire, containing two sets of questions: the first related to a user's perception of trust within the context of privacy and personal data; and the second examined the impact of legal cues on trust formulation.

Section 3 explains our research activities and how we met our research objectives – it is organised as follows: in section 3.1, we describe in detail how the experiment was conducted, including the steps that allowed us to derive actual trust values. In section 3.2, we present our findings related to the user's segmentation and compare them with those of the previous years. In section 3.3, we depict the actual user responses and provide our insights on the major attributes that cause trust differentiations among them. In section 3.4, we describe mathematically a variation of the TME, aiming to better capture the user's trust evolution. We evaluate its accuracy by means of comparative analysis, juxtaposing their results against the actual trust measurements. Finally, in section 3.5 we analyse both the link between the attributes of each segment with the sensitivity of the type of personal data revealed, and the way in which legal guarantees affect their trust.

We emphasize here once again, that it is our methodology to discriminate between a user's trust level and their decision to pay the relevant price and engage (or not) with a system. This approach captures the real market conditions, where a rational potential user would weigh up their trust level by taking into account the monetary risks, costs and benefits. Thus, the derived knowledge of the TME may be utilized on the provider's side as a powerful tool to compute the optimal price and trustworthiness of the offered system, targeting their profit maximization. In D2.3 [1], section 4.1, we presented the provider's optimization problem at the design-time and quantified the additional gains achieved when the TME was applied, compared to the case of its absence, i.e., where all users are supposed to accurately assess the actual trustworthiness. In section 3.4, we present the TME incorporation into the provider's optimization problem during run-time, covering the whole life-cycle of a socio-technical system.

## 3.1. The Experiment

### *3.1.1. The experiment research-context*

In this section, we describe our prerequisites that the chosen application should satisfy, so as to meet our research objectives and provide reliable results. Firstly, we agreed that users should be familiar with the application that they are asked to engage with. This would allow us to setup a realistic scenario, aligned with real-life experience and avoid the time-consuming explanatory phase. Secondly our aim was to extend the experimental scale (compared to the second year) with respect to three factors.

1. The number of metrics under investigation.
2. The number of participants.
3. The number of trials (sequence of outcomes) that each participant observed.

The first point, above, captures the core direction of the OPTET-project, which perceives trust as a multidimensional magnitude with respect to the metrics characterizing the system. We investigate two metrics, "performance" and "privacy". The interpretation of the former is equivalent with the previous year and refers to its ability to provide the anticipated results. We focus again on results of binary form, meaning that each one may be characterized as a success or failure according to an objective criterion. The latter refers to the user's personal data processing and usage by the application, aiming to serve its own interests. Notice that the experimental context should allow for users to detect that such a breach has occurred, which drastically limits our options.

We chose the search engine as the most suitable application that satisfies the aforementioned requirements. Recall that such applications return two sets of outcomes after each search, i.e. the proposed webpages and the advertisements of relevant products or services. We associate the former with the performance metric and the latter with the privacy metric. A detailed explanation of our approach is presented in the next subsection.

For the implementation and execution of the experiment, we utilized the "Amazon Mechanical Turk" platform [11], which is a crowdsourcing Internet marketplace that brings together registered individuals with businesses and/or academic institutes. A business or academic institute posts a task, namely a HIT (Human Intelligence Task) and the individuals are incentivized to participate by means of a monetary reward. This pool allowed us to reach participants beyond the OPTET consortium; hence we have a representative statistical sample not limited to the IT field that would bias the results. Furthermore, in contrast to the second-year experiment, the current one was performed independently of any other OPTET generic enablers (GEs), because here we only focus on the investigation of trust-related behaviours. Additionally, our experience indicates that the combination of GEs in a single experiment requires long time execution periods. Moreover, participants may lose interest and/or concentration. This context allowed us to meet our requirements at points 2, 3 above, as the number of participants increased by almost 7.5 times (204 vs 27) in comparison to the second-year experiment. 100 of the subjects that participated in the experiment were recruited from Amazon Mechanical Turk, and the rest were invited by OPTET partners following an online open call.

In comparison to the second-year experiment, the number of observed trials were increased slightly (12 vs 10). We were aware of the need to keep the overall time of the experiment as succinct as possible.

### **3.1.2. The experiment description**

In this section, we describe the sequence of actions within the experiment that allowed us to both cluster the participants into segments of common trust-related behaviour, and collect their actual trust evolution towards the implemented application via their responses. As we have already mentioned, this knowledge is utilized to design and evaluate the Trust Metric Estimator.

#### **1) Introduction: Instructions and the experimental scenario**

The participants were first given instructions and a brief overview about the purpose of the experiment and were introduced to its underlying fictional scenario. They were asked to imagine that they were organising a "surprise birthday party", and therefore needed to find a set of ten (10) items e.g. a "bespoke birthday cake". They were then requested to use the fictional ACME search engine in order to attempt to locate these items and perform a couple of more queries for a side-project, resulting in a total of twelve (12) trials. The participants were explicitly informed that this experiment focused on analysing their trust evolution towards the socio-technical system in question (i.e. the fictional ACME Search Engine) with respect to the two metrics under investigation (i.e. performance and privacy). Furthermore, as we wanted to guarantee that the participants had fully understood the criteria applied to determine an outcome as successful or not in terms of the performance metric (for more information see the next subsection), we implemented a number of interactive instruction steps where the users were unable to proceed to the experiment until they had provided the right response to the practice trial.

#### **2) Main Body: The batch of trials**

After the introduction phase, each participant performed the batch of 12 trials, which formed the main body of the experiment. At each trial, the search phrase was predefined and statically provided in the relative textbox. In order to ensure consistency and the reliability of the derived results, the participants observed the same sequence of trials. This helped us to guarantee that any potential deviations in the participants' trust levels were isolated on their personal attributes, and were unaffected by a different level or sequence of system performance.

The participants interacted with the application by clicking on the "Search" button, which caused the search engine results to appear (as depicted in Figure 10 and Figure 11). Their positioning is aligned with the design of an actual relative application. Note that the search results appeared on the left hand side of the screen, and the right hand side displayed the advertisement message.

You are currently at trial: 1 / 12

**ACME**  
*Search engine*

Our history – Contact us – Careers

purple sparkly birthday balloons






Click the Search Button below to see the results!

Search

To-do list:

1. Purple sparkly birthday balloons
2. Bakery – to make a bespoke, diabetes-friendly chocolate cake
3. Catering – Spanish cuisine
4. Music – live jazz band
5. Purple sparkly invitations
6. A diabetic chocolate fountain!
7. Party hats – mixed colours
8. 30 Cake Candles – pink
9. Karaoke equipment
10. Buy Digital Camera
11. Air-tickets to Brussels
12. Learn Russian

ACME search results – page 1

ACME featured advertising

Need office stationery supplies for your business - such as pens and toner cartridges - then look no further!

To what extent do you have the confidence that ACME Search Engine will deliver at least one useful result on the first page during your next search?

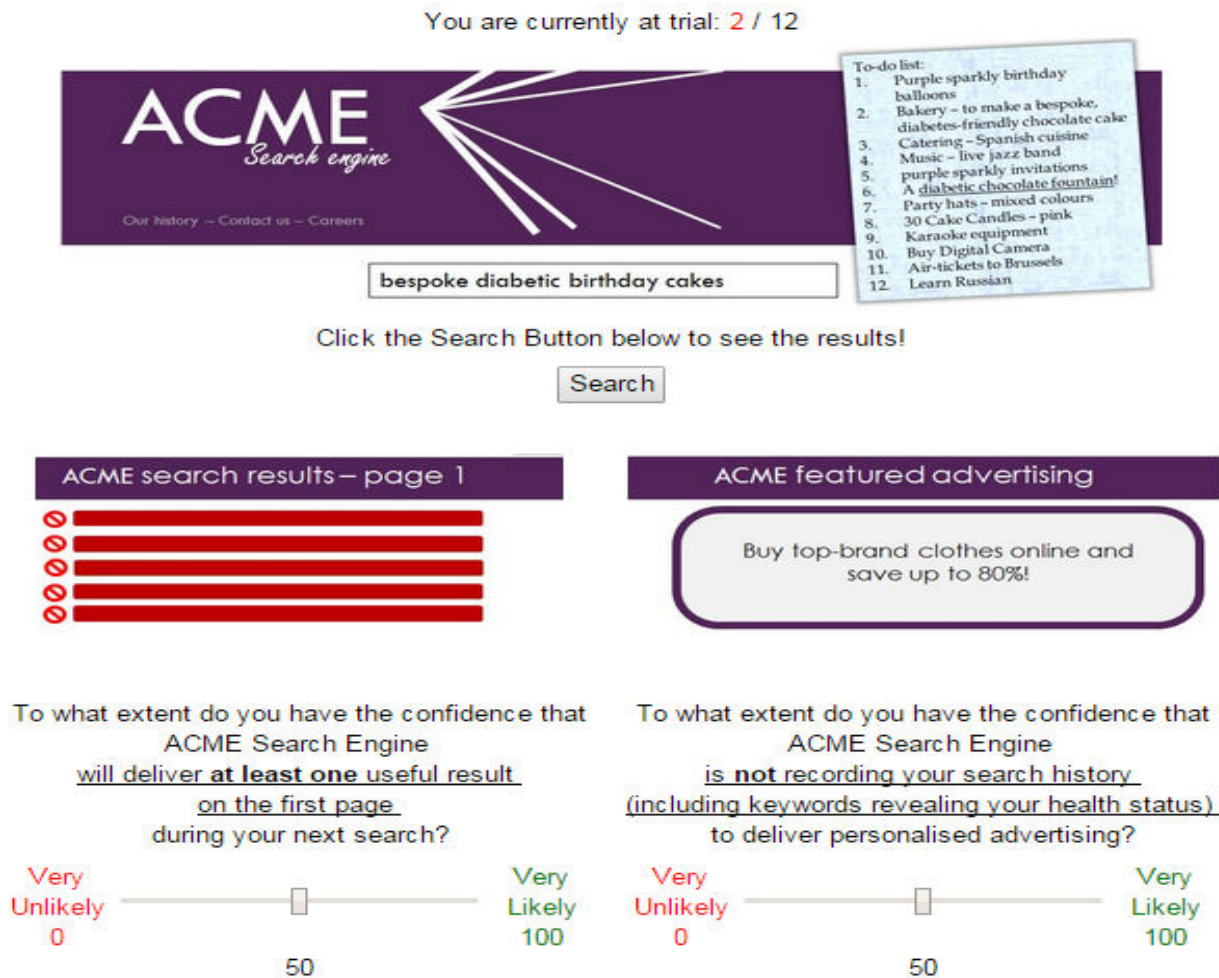
Very Unlikely 0 50 Very Likely 100

To what extent do you have the confidence that ACME Search Engine is not recording your search history (including keywords revealing your health status) to deliver personalised advertising?

Very Unlikely 0 50 Very Likely 100

When you are ready, please press Enter to continue to the next trial

**Figure 10: A snapshot of the experiment, where the search engine provides three useful results (trial-success) on the first page and the advertisement does not relate to the search history.**



**Figure 11: A snapshot of the experiment, where the search engine provides no useful results (trial-failure) on the first page and the advertisement does not relate to the search history.**

The search results appeared in the form of a list complete with figures indicating a successful (depicted by a "smiley face" sign) or non-successful outcome (represented by a "no" sign). They did not include any websites links or descriptions. This is because we wanted to avoid any unnecessary information that could potentially divert attention away from the main task. During each trial, the participants responded to the search engine results displayed by recording their perceived level of trust via the associated slide bar (0-100) to answer the following performance metric question:

*"To what extend do you have the confidence that the ACME Search Engine, will deliver at least one useful result on the first page during your next search?"*

Notice, that according to the applied criterion, a trial is characterized as a success if the search engine provided at least one useful result. Thus, our decision was to use the same figure for all



successful trials i.e. three useful (green) and two irrelevant (red) results (as depicted in Figure 10). Here, our aim was to ensure that trust evolution would not be affected by a varying number of successful links. The absence of any useful result, indicating a trial-failure, is depicted in Figure 11. The sequence of the search engine successes and failures with respect to the performance metric is presented in Figure 13. This figure is illustrative of the user's reactions and their trust evolution towards this metric, which is presented and analysed in section 3.3, below.

Concerning the advertisement message, it was either totally irrelevant to the search-phrase, or related to search-phrases (indicating that the user's search history had been used by the search-engine to provide personalized advertising). The latter case, was explicitly revealed by an advertisement related with search-phrase at the current trial, or implicitly if it referred to a keyword during a previously performed search. Aware of the importance of offering a user-friendly environment, we continuously provided the full list of keywords throughout the whole experiment, thus the participants were able to overview all the previous search activity with ease.

The participants responded to the advert displayed during each trial by answering the following question and using the associated slide-bar (0-100) to indicate their perceived level of trust:

*"To what extend do you have the confidence that the ACME Search Engine, is not recording your search history (including keywords revealing your health status) to deliver personalized advertising?"*


In Table 3 below, we document the search keywords and the advertisements that appeared during each trial. Furthermore this table will be utilized in order to explain trust evolution relating to the privacy metric.

For the sake of completeness, we mention the further measures we took in order to ensure that users were able to navigate the experimental environment effectively. First, the slide-bar during each trial (apart from trial 1) remained at the value that the user had selected during the previous trial. This allowed us to overcome the "lack of memory" problem we observed during the experiment conducted in the previous year i.e. where in some cases users reacted with a trust decrease even after a success. We reasonably assume that such a reaction resulted from a lack of ability for participants to inspect their previous responses. Second, the "Enter" button (that when pressed by the user would cause the application to proceed to the next trial) was activated only after the participant had interacted at least once with each slid-bar; even if its value was left the same as in the previous trial. This precaution guaranteed that the users could not (un-)intentionally skip a trial without recording a specific trust level relating to that particular trial.

In Figure 12 we present the first implemented trial, where no search-keyword or results appeared and the users were asked to report their trust before observing any evidence of its performance. A major difference, compared to the second-year experiment, is the absence of the "about pages" that provide information concerning the actual trustworthiness of the application (D2.3 [1], section 3.3.1, point 2). Before proceeding, recall that in D2.3 we presented two theoretical models, the former trying to estimate initial trust while the latter ("machine-learning") utilized the trust values as input at this first moment. Their comparison (D2.3, sections 3.3.3 and 3.3.4) showed that the second provides more accurate results (over the whole moment), thus in this deliverable we only focus on

this approach. This is the reason why we omit the "about pages" and consequently, the users rely on their previous experience with such systems to provide their initial trust level.

You are currently at trial: 0 / 12



Our history – Contact us – Careers

ACME Search engine

1. Purple sparkly birthday balloons

2. Bakery – to make a bespoke, diabetes-friendly chocolate cake

3. Catering – Spanish cuisine

4. Music – live jazz band

5. purple sparkly invitations

6. A diabetic chocolate fountain?

7. Party hats – mixed colours

8. 50 Cake Candles – pink

9. Karaoke equipment

10. Buy Digital Camera

11. Air-tickets to Brussels

12. Learn Russian

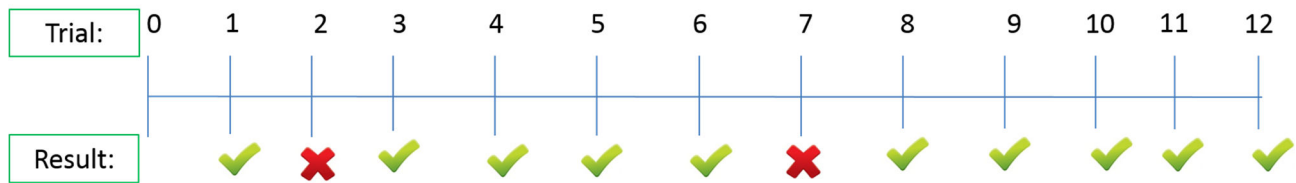
To what extent do you have the confidence that ACME Search Engine will deliver at least one useful result on the first page during your next search?

Very Unlikely 0 ————— 50 ————— 100 Very Likely

To what extent do you have the confidence that ACME Search Engine is not recording your search history (including keywords revealing your health status) to deliver personalised advertising?

Very Unlikely 0 ————— 50 ————— 100 Very Likely

**Figure 12: The first trial, where participants were asked to initialize their trust level for both metrics, based on their previous experience with similar applications.**



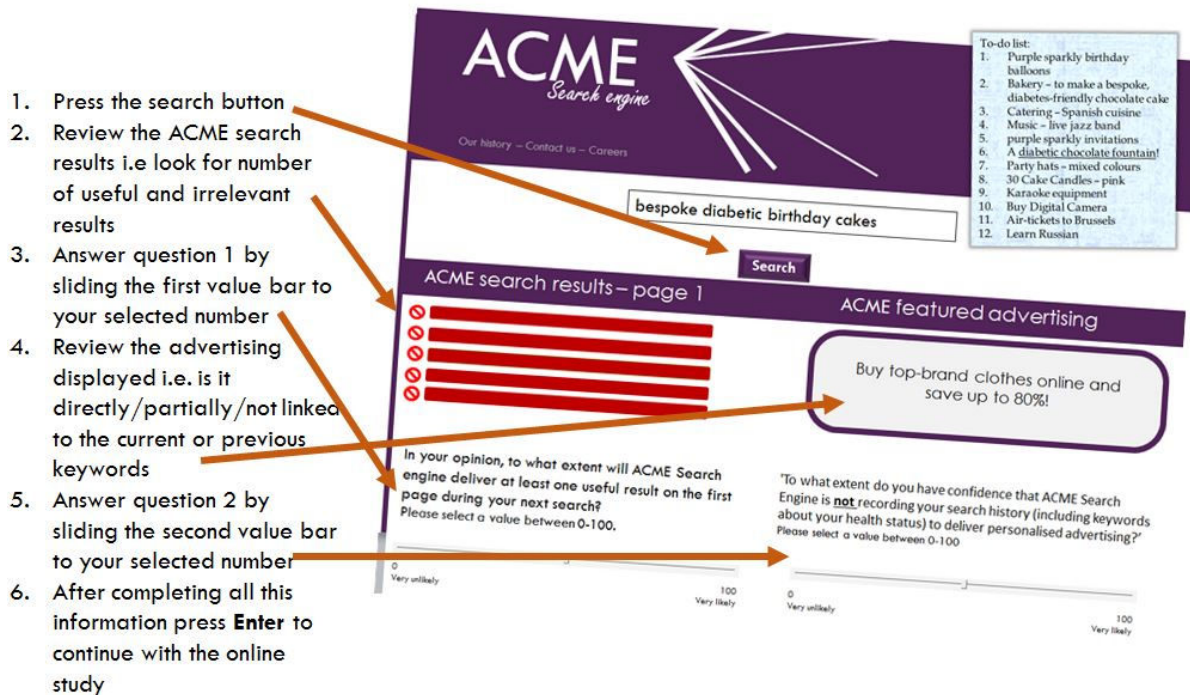
**Figure 13: The sequence of trials that resulted to success or failure, according to the criterion applied for the performance metric.**

**Table 3: The search-phrases and the advertisement that appeared during each trial.**

Trial	Search phrase	Advertisement message
1	Purple sparkly birthday balloons	Need office stationery supplies for your business – such as pens and toner cartridge? – then look no further
2	Bespoke diabetic birthday cake	Buy top-brand clothes online and save up to 80%.
3	Catering companies Spanish cuisine	Buy our low calorie diabetic cookbook for hundreds of tasty recipes
4	Live jazz band	Host an unforgettable party by having a famous chef to cook low calories dishes for your guests. Learn how by clicking here.
5	Bespoke purple sparkly invitations	Ever dreamed of living like a celebrity? Find a personal trainer that will help you shape the perfect body.
6	Diabetic chocolate fountain	Need help losing weight? Get slimming aids from your licensed local pharmacy.
7	Party hats colourful	Love chocolate but worried about your health? Then visit our website. 20% discount on confectionary end Tuesday.
8	Cake candles pink	Finest Belgian Dark chocolate. Expedited shipping.
9	Karaoke equipment	Your special occasion deserves a special treat! We will create a customized dessert table for you.
10	Digital Camera	Keep track of your daily calories with our mobile app! Use OFFER coupon for 10% discount.
11	Air Tickets Brussels	Enter our competition to win a luxury bag and one year admission to your local gym.
12	Russian language course	Visit our store in Brussels to buy 3 boxes of the finest low calorie truffles for 2.



In Figure 14, we present the whole process that each participant followed during each trial, via a single snapshot. This figure sums up the whole set of actions performed and was provided at the instructions phase of the experiment to the users for a clear understanding of their role.



**Figure 14: The whole sequence of actions performed at each trial.**

### 3) The post-questionnaire

After the participants had completed the twelve trials, they answered a questionnaire. The first part of the questionnaire was identical to the questionnaire conducted during the previous years of the project. Its aim was to identify the personal attributes that play a dominant role in trust decisions and results to the users' segmentation. A detailed overview of these questions is available in D2.2 [12] and D2.3 [1]. The second part investigated the importance that users place on the protection of their personal data, depending on the type of revealed information. Additionally, it included questions which aimed to investigate the trust responses in the presence of legal information and guarantees. In Figure 15, we depict the set related with demographic issues, while an extensive analysis of this research work is presented in section 3.5, below.

Next, we ask you a few questions about how you feel about disclosing personal information online.

Q6: What is your gender? 1. Male ▼

Q7: What is your year of birth?

Q8: What is your nationality?

Q9: What is your highest educational degree? Please choose one answer from the drop box. -- select an option -- ▼

Q10: What is your living situation? Please choose one answer from the drop box. -- select an option -- ▼

Q11: In what city do you live?

Q12: In what city do you work?

Q13: Where did you spend your last holiday?

Q14: How sensitive do you find the information you had to reveal about your location (Q6 to Q13)? (1 = Not Sensitive at all; 7 = Very Sensitive) -- select an option -- ▼

Q15: Do you think this information is available online about you by performing a search on, for example, Google? -- select an option -- ▼

Q16: How important is it for you that this type of information about you is not publicly available? (1 = Not Important; 7 = Very Important) -- select an option -- ▼

**Figure 15: A set of questions included in the post-questionnaire, aiming to identify the sensitivity of users on their personal data protection, depending on the type of information revealed.**

## 3.2. Users' Segmentation

### 3.2.1. Overview of the research approach

Throughout our work we have focused on examining socio-technical and economic factors affecting the subjective nature of trust associated with the stakeholder's decisions in online environments. More specifically, we have sought to build a theoretical framework that captures these aspects and reflects trust differentiations among users. This has enabled us to address the increasing complexity of trust in the digital realm and the conditions that affect it in systems development, especially those presented in other OPTET Work Packages. Since the beginning of the OPTET project, it has been our role to explore the socio-economic and legal drivers in such environments so as to develop a trust computational model assessing a user's trust level regarding the performance of a particular system.

Our journey started out by focusing on several studies that recognized the need for models of trust and credibility in technology-mediated interactions, particularly, those that aimed to be domain agnostic and technology-independent. These models have been found to offer guidance for researchers across disciplines that study various technologies and contexts (see D2.1 [13]), focusing, among others, on: antecedents (i.e. preconditions of trust), processes of trust building (e.g., interdependence), the context of shaping trust-building (e.g., social relations, regulation), decision-making processes in trust (e.g., rational choice, routine, habitual), implications and uses of trust

(e.g., interpersonal entrepreneurial relations, moralistic trust), and lack of trust, distrust, mistrust and repair (e.g., risks, over-trust, trust violations).

In order to elaborate on existing insights, we examined how different trust-related user experiences seem to be guided by different sets of trustor's attributes. Guided by our first task (2.1) the linkage of socio-economic and legal components was examined. The generic and exploratory outcomes of survey (and interview) research were presented in D2.1 [13] (section 6) yielding insights – via factor analysis, reliability testing and regression – into so-called 'trust levels' for end users. This was followed-up in D2.2 [12] (section 3) by a 'segment-specific' approach so as to learn about different types of subjective trust-related user experiences in this context. More specifically, a survey was conducted (1) to provide data that allows to identify key attributes impacting the subjective trust experience; and, (2) to develop ways to adapt a computational trust model parameterization based on these key attributes.

For this purpose, we deployed several statistical methods, in particular, regression (to come to scales), reliability (of the scales), cluster analysis (K-means to come to the segmentation and analysis of proximities), one-way ANOVA, comparing means and post hoc tests. Based on findings conducted over several empirical cycles – presented in D2.2 and D2.3 (n= 232) – linkages between different sets of trustor attributes were detected, corresponding to trust-related concepts of (1) Trust stance: the tendency of people to trust other people across a wide range of situations and persons; (2) Trust beliefs in general professionals; (3) Institution-based trust; (4) General trust sense levels in online applications and services; (5) ICT-domain specific sense of trust levels; (6) Trust-related seeking behaviour; (7) Trust-related competences; and, (8) Perceived importance of trustworthiness design elements. These concepts guided the development of the segmentation study of trust-related user experiences on trustor attributes of the OPTET project.

Each of the aforementioned items was tested to see whether statistical significance differences could be retrieved between the uncovered trust-related user experience segments. Iterative clustering and testing resulted in a four segment-solution that could best explain differences in trust-related user experiences. Consequently, the segments were labelled by, 'High trust' (HT), 'Ambivalent trust' (A), 'Highly active trust seeking' (HATS) and 'Medium active trust seeking' (MATS). We found that they seem to differ on a number of aspects. However, based on our analyses, three concepts are sufficient to explain these main differences. These underpinning concepts are 'trust stance' (e.g., 'I usually trust a person until there is a reason not to'), 'motivation to engage in trust-related seeking behaviour' (e.g., 'I look for guarantees regarding confidentiality of the information that I provide') and 'trust-related competences' (e.g., 'I'm able to understand my rights and duties as described by the terms of the application provider'). They could be measured on 3, 7 and 4 item-scale with a reliability coefficient of, respectively, .69, .89 and .87 (see section 3.1 in D2.2 [12]).

### ***3.2.2. Derived Segments: Characteristics and validation***

To recap our initial findings, the user experience for the *High Trust* (HT) segment could be characterized by a so-called high level trust stance. This means an overall high trust level for the various online applications, such as social networks and online banking, accompanied by only few trust seeking behaviours, such as checking for trust marks, even though the competences are present to cognitively assess the trustworthiness of online applications and services.

For the *Highly active trust seeking* (HATS) segment, the user experience highlighted a high level of trust seeking behaviour beyond the mere scanning of trustworthiness cues. It also showed that individuals seem to be informed about procedures in case of harms and misuse, pointing to the capacity of certain competence level that facilitate the assessment of trustworthiness and to possess,

at least, a minimal understanding of the rules and procedures to look for in case of complaints and misuse. Varied trust stance and trust levels were observed including medium to low trust stance/trust levels.

The user experience for those clustered as *Medium active trust seeking* (MATS), was relatively similar to the highly active one. However, trust seeking behaviour was less apparent. In other words, while drivers for trust seeking behaviour, such as a relatively low trust stance, could be detected as well as competences to assess trustworthiness, the motivation to look for trustworthiness cues was less apparent or even absent.

The *Ambivalent* (A) group showed an obvious perceived inability to assess the trustworthiness of online applications and services. This could partially be explained by one's personal competence level – only a few active trust seeking behaviours could be observed, however, those do not equal low(er) trust levels per se. Trust seemed to be derived from either the general trust stance or 'basic heuristics', such as 'public organizations are more trustworthy than commercial companies'. It seemed that the 'ambivalent' nature of user experience could be explained by a failure to cognitively assess the trustworthiness and a certain need to trust in order to avoid, or to lower the omnipresence of cautious and other negative feelings (so-called 'forced trust'. Thus, pointing to understand trustworthiness indicators based on the experience of others ('referrals'), as the main source of 'trustworthiness information' that is accessible and underpinning the outcome of assessing trustworthiness.

These findings were further investigated and elaborated in the context of the OPTET DADV experiment (see D2.3 [1], section 3.1, and D8.5 [14]). Here, we learned that the HT segment possessed again the highest trust stance of all. This time showing somewhat higher competence levels, hence, the trust seeking behaviours decreased somewhat than before. The HATS segment could be characterized better by their competencies, while they were also quite motivated to look for trust marks and so forth. MATS are again somewhat similar to HATS in their trust seeking behaviours and showed a decrease in motivations vis-à-vis their competences. The A segment showed the lowest competence levels as well as trust stance, suggesting users are likely to be more motivated to look for trust cues.

Despite the minor variations between these exploratory analyses, the dominant drivers describing the users in each of the four segments seemed relatively constant. Accepting this, we could differentiate among trustors based on these drivers and infer some expected trust-related behavioural properties for each segment. This linkage of dominant drivers to certain expected properties were validated by means of comparison with the actual trust measurements as reported by the participants in the Cyber Crisis Management experiment (see D2.3, section 3.1). We could conclude that our analysis was valid and resulted in the capacity to steadily detect dominant drivers affecting the subjective nature of trust. Based on these findings we sought to derive the expected users' behaviour, considering also the technical factors that seemed to determine system performance. To this end, trust was explicitly formulated as a function of both aforementioned aspects, while shaping the expected behaviours of each segment.

### **3.2.3. Third year research results**

In order to further validate the user segmentation we have sought to elaborate the trust-related behaviours in the context of the hypothetical search-engine as well as posing contextual questions so as to decrease likelihood of users not being sufficiently truthful or competent enough to understand the questions (see D2.3 [1]).

We followed the same procedure as for conducting the previous segmentation analyses. The respondents who were interested to participate in our experiment (see section 3.1 above) were first asked to conduct the trial of using a hypothetical search engine, and after the last trial they had to answer several questions. First, they had to fill in the online segmentation-related questionnaire. The answers were used to test and validate the four segment solution again, and after some reliability testing, each individual could be clustered according to the most suitable segment. Furthermore, the mean values of the three trust-related concepts<sup>1</sup> could be computed, as documented in Table 4:

	<b>Total (n=204)</b>	<b>HT (n=51)</b>	<b>HATS (n=50)</b>	<b>MATS (n=31)</b>	<b>A (n=72)</b>	<b>Anova</b>	
	<b>Mean</b>	<b>Mean</b>	<b>Mean</b>	<b>Mean</b>	<b>Mean</b>	<b>F</b>	<b>Sig.</b>
<b>Trust stance</b>	3,15	3,37	2,84	3,26	3,15	5,021	<b>,000</b>
<b>Trust related seeking behaviour</b>	2,61	3,19	2,99	2,06	2,19	3,228	<b>,000</b>
<b>Trust related competences</b>	2,63	2,50	3,48	3,37	1,82	2,314	<b>,000</b>

**Table 4**

\* HT = High Trust; HATS = Highly active Trust seeking; MATS = Medium active Trust seeking; A = Ambivalent

If we look at these results we come to similar findings as for the previous conducted segmentation analyses, presented in D2.2 and D2.3. More specifically, "HT" segment appears here also with the highest trust stance, while the combined factor of "seeking-behaviour" and "competences" is higher for "HATS" among all segments. Furthermore, the F-scores have improved, and again significance results are confirming that our overtime trust-related segmentation model seems relevant. Recall, that according to our methodology, the aforementioned dominant drivers were correlated with three fundamental expected properties that characterize the trust shaping of segments. In section 3.3 we validate them by means of the derived actual trust values. Before proceeding, let us briefly remind them in the following section.

<sup>1</sup> Our segmentation solution is based on what trust-related concepts are predictive towards our explored trust levels (see D2.1, D2.2 and D2.3). Based on various internal validity testing and factor-analyses conducted for each of the trust concepts (measured on a scale level), the following trust constructs were developed: Trust stance (2 item scale,  $\alpha=.79$ ); Trust related seeking behaviour (5 item scale,  $\alpha=.86$ ); and Trust related competences (4 item scale,  $\alpha=.88$ ). These scales show sufficiently good to excellent reliability coefficients. Regression analyses were performed to study what trust concepts are predictive and to what degree are they predictive towards trust in a particular set of web-based sites.



#### ***3.2.4. Fundamental expected properties***

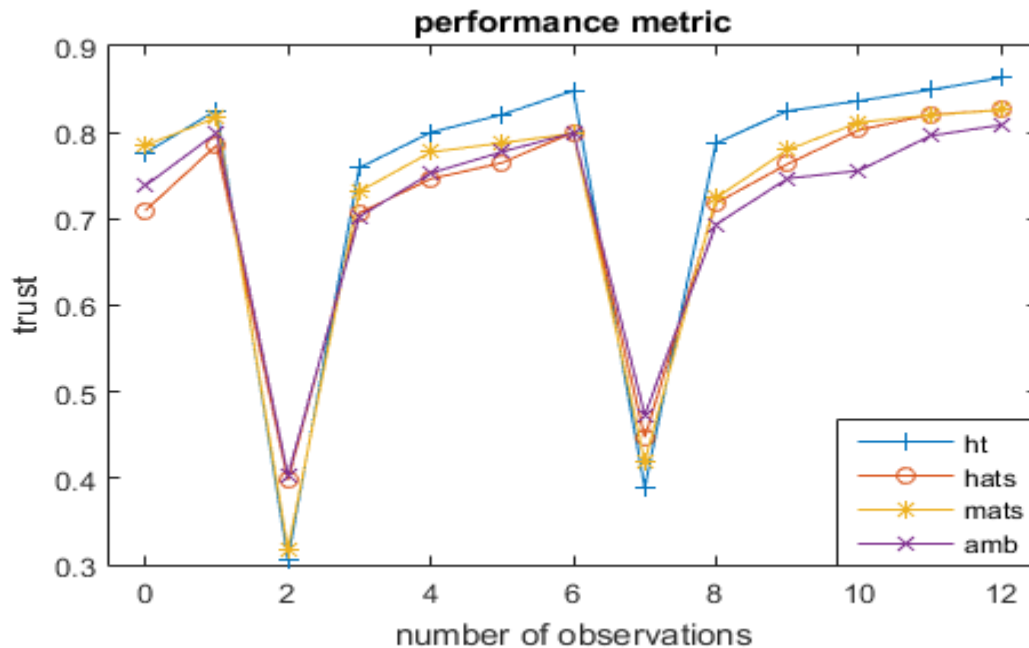
- 1) The combined impact of "competences" and "motivation" concepts implies that the miss-estimation (absolute difference between trust and trustworthiness) in "HATS" segment must be the lowest among all.
- 2) The impact of "trust stance" concept implies that the trust level in "HT" segment must be higher than the corresponding trustworthiness of the system and the highest among all segments.
- 3) All segments have the minimum level of competences and motivation so as to distinguish between two differently performing system, i.e., with different levels of trustworthiness.

### **3.3. High level Evaluation of the results**

In this section we present the users' trust evolution towards the two metrics and provide a high-level analysis of our findings. We use the term "high level" because we mainly focus on trust shaping and fluctuations and not on specific values, contrary to section 3.4.4 below where we evaluate the TME by comparing its estimations with the actual trust measurements. Thus, our target here is to validate the fundamental expected properties and also to identify if the trust responses are consistent with the three underpinning concepts that characterize trust-behaviour within each segment.

#### ***3.3.1. The performance metric***

In Figure 16, we present the actual trust measurements towards the performance-metric, as reported by the participants after observing the outcome of each trial. Notice that according to the applied criterion which characterizes an outcome as successful, the ACME search engine achieved to return ten (10) successful results out of totally twelve performed trials, meaning that its trustworthiness over this set equals  $10/12=0.83$ .



**Figure 16: The trust evolution for the four segments towards the "performance" metric.**

We can make the following general observations: Firstly, notice that the reactions of all segments are consistent with the outcome, i.e., trust deteriorates after a failure while it increases after a success. Once again (as it also happened in the second-year experiment) this fact justifies our approach to capture trust updates by means of the increment coefficients in the Beta probability density function. Additionally, the trust change is most intense after an outcome interchange (from success to failure and vice-versa) while it slightly increases after two consequent successes. This fact motivates our extension to the theoretical framework that supports the TME and will be presented next.

Contrary to D2.3, where we evaluated the expected properties both at the initial and final phase, here we only focus on the latter. As we have already mentioned, we skip the former because we did not provided the "about pages" at the beginning of the experiment that consist of the evidence for the trustworthiness assessment. Thus initial trust is only based on previous experience with similar systems. However, after the final outcome, the participants have access to all the necessary information to assess the aforementioned magnitude (due to the specific batch of outcomes as mentioned earlier). Properties (1) and (2) are valid: the trust of "HATS" is closer to trustworthiness compared with all segments ("HT": 0.863, "HATS": 0.8261, "MATS": 0.8260, "A": 0.808 vs tw: 0.83). Additionally, "HT" users overestimate it (0.863 vs 0.83) and more specifically they appear with the highest values compared to any other type of users. Finally, note that during the latest seven trials, the trust evolution of "HATS" and "MATS" users is almost identical. This shaping is aligned with their characteristics in section 3.2.2, where the two segments appear with relatively similar properties and almost common expected trust behaviour.

Despite the fact that the two first properties are met, the responses are very close between each other and their deviations are not that distinct compared to the second-year results. The reason is most probably the nature of the chosen application and the fact that participants are very familiar with its functionality. Indeed, the every-day experience with such systems suggests a-priori a high

level of trustworthiness with respect to their performance. Thus, even these minor (in absolute terms) differentiations among users should be considered as relatively strong validation results in this, arguably, unfavourable experimental context, for our segmentation approach [15], [16].

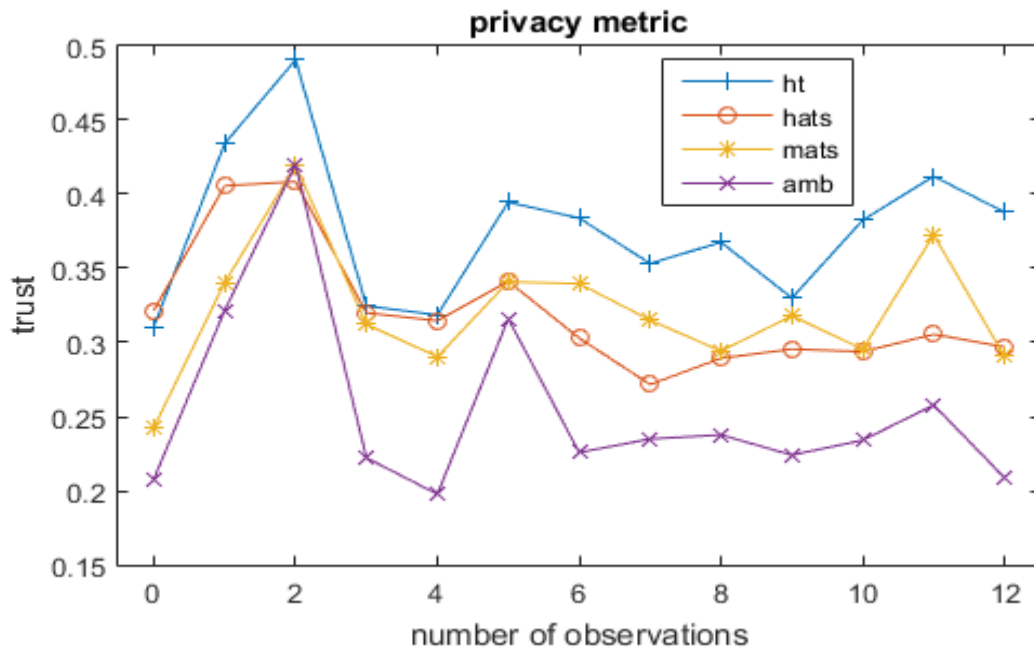
As one may easily conclude, the third property above may not be easily evaluated as the experimental setup considered only one system. Recall that in the second-year experiment, participants had observed the performance of both vanilla and OPTET-enabled DADV and all segments reported higher trust values for the latter, i.e. the system with the higher trustworthiness (see D2.3, section 3.3.2). But, the validation of the third property may now be based on the comparison of the trust values within the same segment towards the two different metrics. We present our finding in the next section, after introducing the trust evolution towards privacy.

### ***3.3.2. The privacy metric***

The analysis of trust in relation to the privacy metric is more demanding compared to the performance metric. Notice that in this case, we did not provide an explicit interpretation of the outcome (as we did for performance), because we aimed to identify the way in which users perceive each message. Before proceeding, let us first describe the general form of the chosen advertisements: they could be totally irrelevant to the search words or reveal an explicit or implicit linkage. More specifically, in the latter case they propose goods or services that are related with the health-status or the location of the participants, which are commonly considered as sensitive personal data. In other words, here also we challenge our approach in an unfavourable context, as our choice makes it more difficult for any differentiations to appear and validate the segmentation.

In Figure 17, we present the actual trust measurements towards the privacy-metric, as reported after each trial. It becomes apparent that apart from the unequal levels of trust, the segments also react differently after a subset of intermediate steps. Thus, we observe an absence of "global consensus" on the search-history recording activity at the engine side. We utilize this fact to further investigate the impact of personal attributes on trust shaping:





**Figure 17: The trust evolution for the four segments towards the "privacy" metric.**

We present our major findings as follows:

- 1) The advertisements that appeared in the first two trials (1, 2) were completely unrelated to the search-words and the general theme of the scenario. As a result, all segments increase their trust after both these observations. Our objective here was to confirm that users pay the attention needed to distinguish between relevant and irrelevant messages, but also to deliberately cause a trust increase that allows for distinct fluctuations at the following trials.
- 2) We placed an advertisement containing the word "**diabetic**" in the following trial (3), which was explicitly related to health issues and was also a search-term in trial 2. We observe a clear and steep trust deterioration for all segments. This fact indicates that all segments consider the personal data related to their health-status as very sensitive and have the adequate competence to understand this direct correlation.
- 3) In trial 4, the advertisement proposes "a famous chef to cook **low calories** dishes". Our aim here was to include a term that refers indirectly to the health-status and implicitly reveals a search-history recording. Interestingly, all segments decrease again their trust values, with the change being smoother compared to the transition between trials 2 and 3. This smoother reaction indicates an expected behaviour that users are mostly affected by a straight correlation between the search-phrase and the advertisement compared to an implicit one.
- 4) Our analysis above is justified also by the reaction of the users at trial 5. Recall that the advertisement proposes a "personal trainer", but does not include a buzzword such as "diabetic" and "low calories" as above. Notice that the target group of this message is not limited only to diabetics but refers to the whole population. Our aim here was to identify if users perceive it as a "hidden" correlation with search history or an irrelevant one and investigate the impact of biased-terms on their responses. From their reactions, we observe that they interpreted the advertisement as irrelevant with their search history, a fact that triggered the trust increase for all segments. Interestingly, this change is less intense for

"HATS" and "MATS" compared to "HT" and "A" segments, a fact which is aligned with their higher trust-related competence.

- 5) The advertisements at the following two trials (6, 7) contain the phrases "lose weight" and "worry about your health". For both of them we observe a trust decrease for all segments, apart from the ambivalent users for trial 7. The analysis for this reaction assembles the one for trial 4 (point 3 above). Notice that the behaviour of ambivalent users at trial 7 is aligned with their personal attributes of trust-related competence, which is the lowest among all segments.
- 6) At trials 8, 9 we placed two advertisements containing the phrases "dark chocolate" and "customized dessert" respectively. Notice that even though they do not directly refer to health issues, they are strongly related with the nutrition habits and precautions of a diabetic person respectively. It seems that the interpretation at the user's side is ambiguous and we cannot identify a clear tendency from their reactions. Thus, here once again (as for trial 5 – point 4) the comparison with cases containing buzzwords (trials 3, 4, 6, 7) proves their negative impact on trust shaping.
- 7) One of the most interesting findings appears at trial 10. The advert displayed contains the word "calories"; however, the segments reacted differently. While the perceived levels of trust increased for "HT" and "A", it decreased for "HATS" and "MATS". The trust levels of the "HT" and "A" segments may have increased as the search terms used during trials 7-10 were all unrelated to health. Given that trial 6 was the last occasion when a health-related search term was used ("diabetic"), this may indicate that these two segments are less focused on remembering and/or taking into account health-related terms from the distant past. In contrast, as "HATS" and "MATS" seem to be more focused, a fact which is aligned with their higher competence and consequently it is unsurprising that their perceived level of trust decreased.
- 8) Finally, we provide a combined analysis for the last two trials (11, 12). Note that during the former, the user searches for "Air Tickets to Brussels", and the advertisement promotes a competition to "win a luxury bag and 1 year admission to your local gym". All segments increase their trust and the analysis is the same as for trial 5 (point 4 above). At trial 12, the user is supposed to search for "Russian language course", while the engine proposed a store in Brussels, directly related with the previous search. All segments decreased their trust, and more specifically the slope of this reaction assembles the one when health-related advertisements appeared (for three out of four segments – apart from "HATS"). This fact indicates that users consider the personal data related to their location as very sensitive and of similar importance to data related to their health-status.

The absence of common reactions makes it impossible to provide a mathematical formula for the trustworthiness value. However, despite the aforementioned differentiations in trust fluctuations, common behaviours appear for the majority of trials (8/12) a fact that allows us to provide a loose computation: we could consider as irrelevant advertisements, those that trust increased for all segments (1, 2, 5, 11), i.e., the trials where there was a consensus for the outcome. In this case, the trustworthiness equals  $4/12=0.33$  and the first two expected properties are validated again: "HT" overestimate it and "HATS" have the most accurate assessment.

Another computation would consider the consensus on the relevant advertisements (the cases when all trust values decrease – trials 3, 4, 6, 12). In this case, the trustworthiness equals  $1-4/12=0.67$

and our assumptions on expected properties do not hold. Despite this fact, we observe that trust of "HT" is the highest among all segments. Also, notice that over both computations above, the trustworthiness for this metric is lower compared to the respective one for performance. Additionally, trust within the same segment is lower towards privacy compared to performance metric, due to the lower trustworthiness of the system with respect to this factor. Thus, the third property also holds. Finally, once again the trust values and reactions of "HATS" and "MATS" resemble each other most, compared to any other segments (their reactions differ only at trial 8). Thus, for this metric also we have strong indications that our approach of users clustering is aligned with actual behaviours.

### 3.4. TME Revisited

In this section, we investigate the performance of the TME in the context of the fictitious search-engine experiment. Notice that in D2.3 we provided two proposed theoretical models, i.e., the one based only on assumptions about trust differentiations among segments and the second which utilized a subset of actual trust responses, aiming to calculate the trust update coefficients after each observation. Their comparison in terms of accuracy concluded that the latter, namely "machine-learning", outperforms the former. Thus here, we only focus on this approach and provide a variation aiming to better capture the trust shaping. Our target is to demonstrate that the proposed trust computational model is a flexible tool which may be easily adjusted to reflect actual trust reactions.

#### 3.4.1. The theoretical framework supporting TME

##### 3.4.1.1 Overview of the "machine-learning" formulation

In this section, we provide a brief description of the "machine-learning" approach, which will be used as a theoretical basis for the extensions of TME. Before proceeding, recall that throughout our research activity, we quantified trust as the mean of a Beta probability density function, characterized by two parameters ( $\alpha$  and  $\beta$ ). Thus, trust is a function with respect to the aforementioned parameters according to the formula below:

$$\tau_l^j(t) = \frac{\alpha_l^j(t)}{\alpha_l^j(t) + \beta_l^j(t)}$$

Notice, that we use three indicators, where "l", "j" and "t" refer to the segment, the metric and the time respectively. The two former reveal our approach to differentiate among segments and capture the multi-dimensional nature of trust in terms of the various metrics that characterize the performance of a system. The time indicator reflects the fact that trust evolves over time, based on the evidence that users observe for the system functionality.

According to our methodology, trust evolution is captured by means of the coefficients that are used to update the values of the Beta parameters after each trust-related event, such as a system outcome. More specifically, their update after each intermediate outcome is as follows:

$$\alpha_l^j(t) = \alpha_l^j(t-1) + A_l^j * r^j(t) \quad \text{and} \quad \beta_l^j(t) = \beta_l^j(t-1) + B_l^j * (1 - r^j(t))$$

, where "r" is the vector of the results that the participants observed. Each value  $r(t)$ ,  $t > 0$  may take a binary value, with each "zero" (0) and "one" (1) representing the case of a failure and a success respectively. If the system outcome is not binary by default, then a threshold could be used to characterise it accordingly. For example, the "performance" metric could be: [1,0,1,1,1,0,1,1,1,1]. In words, the trust update process suggests that parameter "a" is updated only after a success, while "b" only after a failure, resulting to the trust increase and decrease respectively.

It becomes apparent that the target of the TME is to provide values for the update coefficients  $A_l^j$  and  $B_l^j$ , that best approximate this evolution. This is easily feasible by utilizing a subset of actual trust measurements (as reported by the participants in the experiment or in a real-world scenario by users during a trial period) to compute the average actual trust level  $m_l^j(t)$  of each segment and setting the TME trust estimation equal to that value as follows:

$$\tau_l^j(t) = \frac{\alpha_l^j(0) + s(t)A_l^j}{\alpha_l^j(0) + s(t)A_l^j + b_l^j(0) + f(t)B_l^j} = m_l^j(t)$$

, where  $s(t)$  and  $f(t)$  stand for the number of successes and failures observed until time  $t$ . Note that if we know the initial trust value (at  $t = 0$ ) and apply this rule for two different moments in time ( $t_1 \neq t_2$ )<sup>2</sup>, then we derive a unique pair of increment coefficients, assuming that they remain constant for all observations.

### 3.4.1.2 Related Work

In our basic methodology described above, we place equal weight on each outcome independently of the time moment that it occurred. Despite the fact that this approach provided satisfying estimation during the second-year experiment, it may not always achieve to accurately fit the actual trust reactions. More specifically, the equal importance of each outcome is motivated by a trust evolution where fluctuations tend to fade out as the number of observations increase. In other words, it best fits the situation where trust tends to converge after a large number of evidences and any further outcomes affect it only slightly thereafter.

Related work (RW) ([17]) has investigated the case where users place greater importance on more recent outcomes compared to those in the distant past. This is a reasonable behaviour, because recent evidences reveal the current state of the system under interest and as a result they should strongly affect trust despite the number of outcomes already observed. In technical terms this effect can be formulated by applying a time-fading coefficient in the Beta parameters' update, as follows:

$$a_l^j(t) = u_l^j * a_l^j(t-1) + (1 - u_l^j) * A_l^j * r^j(t) \quad , \quad \beta_l^j(t) = u_l^j * \beta_l^j(t-1) + (1 - u_l^j) * B_l^j * (1 - r^j(t))$$

where  $0 \leq u_l^j \leq 1$ . In other words, these are quantified as a weighted average of the past and current outcomes. More specifically, higher values of  $u_l^j$  place a greater importance on past

---

<sup>2</sup> Thus 3 actual trust values are used in total.

observations and vice-versa. Note that in this case, contrary to our basic approach, both parameters are updated after each outcome. More specifically, both " $\alpha$ " increases and " $\beta$ " decreases after a success, while the opposite occurs in the case of a failure.

Notice that for this case also, we may compute the values of Beta parameters at time  $t$ , using the following non-recursive set of equations:

$$\alpha_l^j(t) = \alpha_l^j(0) * (u_l^j)^t + (1 - u) * A_l^j * \sum_{k=0}^{t-1} ((u_l^j)^k * r^j(t - k))$$

and

$$\beta_l^j(t) = \beta_l^j(0) * (u_l^j)^t + (1 - u) * B_l^j * \sum_{k=0}^{t-1} ((u_l^j)^k * (1 - r^j(t - k)))$$

Thus, despite the fact of the different update process (compared to our basic method), here also we may apply the "machine-learning" method at two time moments ( $t_1 \neq t_2$ ) and derive the respective update coefficients. Notice that this approach is necessary for a fair comparison between our proposed models and related work.

A major difference compared to the basic scenario, which appears also in our extension that follows, is that here the derived values are not unique but a function of parameter  $u_l^j$ . In section 3.4.3, we introduce an optimization problem to overcome this issue, and calculate unique values for  $u_l^j$ ,  $A_l^j$  and  $B_l^j$ , for a given set of actual trust values.

### 3.4.1.3 Our Extension

In this section, we present our extension of the theoretical framework that supports the TME, aiming to better capture the trust evolution towards the two different metrics. Our design objective is to keep the fluctuations property of Related Work (RW), but only after an outcome interchange (from success to failure and vice-versa), while the trust evolution should follow a sub-linear form when similar outcomes are observed.

Let us first describe an update approach which is introduced as an Intermediate Step (IS) for what follows. More specifically, the two parameters are updated exactly as in the related work but not over all outcomes, i.e., " $\alpha$ " is updated only after a success, while " $\beta$ " only after a failure. In a mathematical formulation:

$$\alpha_l^j(t) = \begin{cases} u_l^j * \alpha_l^j(t-1) + (1 - u_l^j) * A_l^j * r^j(t), & \text{if } r^j(t) = 1 \\ \alpha_l^j(t-1) & , \quad \text{if } r^j(t) = 0 \end{cases}$$

, or equivalently:  $\alpha_l^j(t) = (1 - r^j(t) + r^j(t) * u_l^j) * \alpha_l^j(t-1) + (1 - u_l^j) * A_l^j * r^j(t)$

While

$$\beta_l^j(t) = \begin{cases} u_l^j * \beta_l^j(t-1) + (1 - u_l^j) * B_l^j * (1 - r^j(t)), & \text{if } r^j(t) = 0 \\ \beta_l^j(t-1) & , \quad \text{if } r^j(t) = 1 \end{cases}$$

, or equivalently:  $\beta_l^j(t) = (r^j(t) + (1 - r^j(t)) * u_l^j) * \beta_l^j(t-1) + (1 - u_l^j) * B_l^j * (1 - r^j(t))$

For this case also, we may provide a non-recursive close-form of their values at time  $t$ :

$$a_l^j(t) = a_l^j(0) * (u_l^j)^{s(t)} + (1 - u) * A_l^j * \sum_{k=0}^{s(t)-1} (u_l^j)^k$$

and

$$\beta_l^j(t) = \beta_l^j(0) * (u_l^j)^{f(t)} + (1 - u) * B_l^j * \sum_{k=0}^{f(t)-1} (u_l^j)^k$$

, where recall that  $s(t)$  and  $f(t)$  represent the number of successes and failures until time moment  $t$  respectively.

Our proposed extension (EX), is a hybrid of RW and IS: The update is identical to RW after each outcome interchange and to IS after the second same outcome in sequence.

In a mathematical concrete formulation:

$$a_l^j(t) = \begin{cases} u_l^j * a_l^j(t-1) + (1 - u_l^j) * A_l^j * r^j(t), & \text{if } r^j(t-1), r^j(t) = [11, 10, 01] \\ a_l^j(t-1) & , \quad \text{if } r^j(t-1), r^j(t) = [00] \end{cases}$$

or equivalently:

$$a_l^j(t) = (r^j(t) + r^j(t-1) - r^j(t-1) * r^j(t)) [u_l^j * a_l^j(t-1) + (1 - u_l^j) * A_l^j * r^j(t)] + (1 - r^j(t) - r^j(t-1) - r^j(t) * r^j(t-1)) * a_l^j(t-1)$$

While

$$\beta_l^j(t) = \begin{cases} u_l^j * \beta_l^j(t-1) + (1 - u_l^j) * B_l^j * (1 - r^j(t)), & \text{if } r^j(t-1), r^j(t) = [00, 10, 01] \\ \beta_l^j(t-1) & , \quad \text{if } r^j(t-1), r^j(t) = [10] \end{cases}$$

or equivalently:

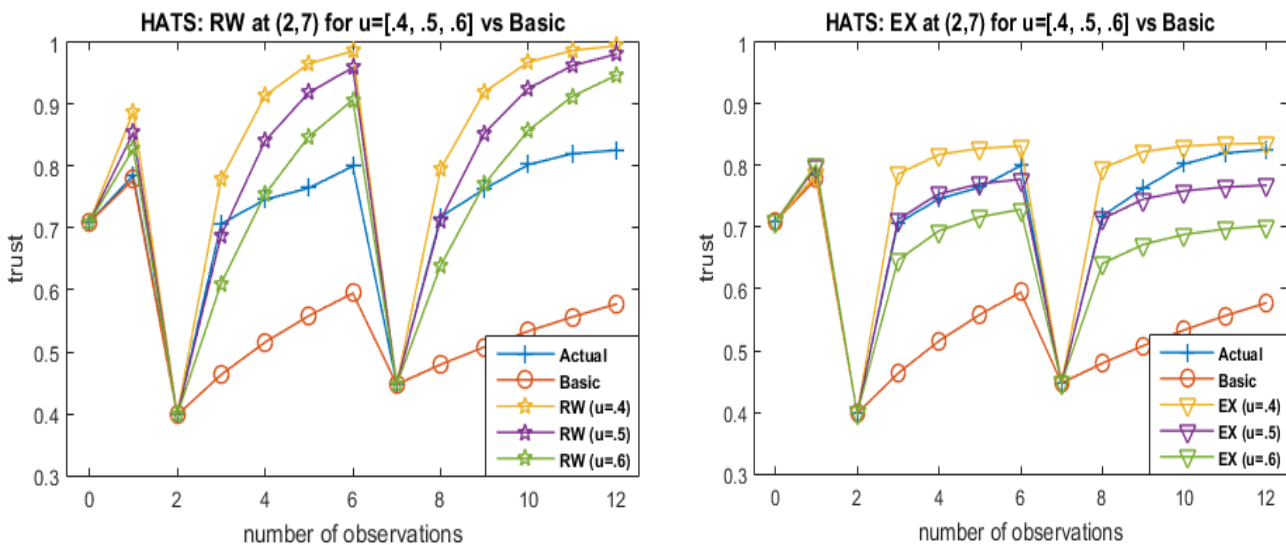
$$\beta_l^j(t) = (1 - r^j(t-1) * r^j(t)) [u_l^j * \beta_l^j(t-1) + (1 - u_l^j) * B_l^j * (1 - r^j(t))] + (r^j(t) * r^j(t-1)) * \beta_l^j(t-1)$$



Notice that the equivalent formulas clearly reflect the hybrid nature of this extension for both parameters, as the first term is the same with RW and the second with IS. For this update process also, we can compute the values of both parameters at any time  $t$ , with respect to their initial values, the update coefficients and the vector of results. We skip the presentation of its close-form for simplicity reasons. Thus we may apply the "machine-learning" methodology at 1+2 moments in time (the initial and two additional ones) and derive a formula of the update coefficients as a function of parameter " $u$ ".

### 3.4.2. Comparison between the three approaches

In this section, we present two illustrative examples of the trust estimations provided by the three described models (basic, RW and EX). Our aim is to depict both their design properties and the impact of the time-fading coefficient " $u$ ". At the left-hand-side of Figure 18, we juxtapose the actual trust responses and the graphs of the basic model with the one from RW, while at the right-hand-side with one from the proposed extension, all towards the performance/accuracy metric. Notice that for both RW and EX we present three graphs as a function of " $u$ ", while the graph of the basic model is unique. Without loss of generality, we include only the "HATS" users, as a detailed analysis of their performance for all segments follows in section 3.4.4.



**Figure 18: The trust estimations of RW (left) , EX (right) and Basic (both), towards the performance metric for different values of coefficient  $u$ .**

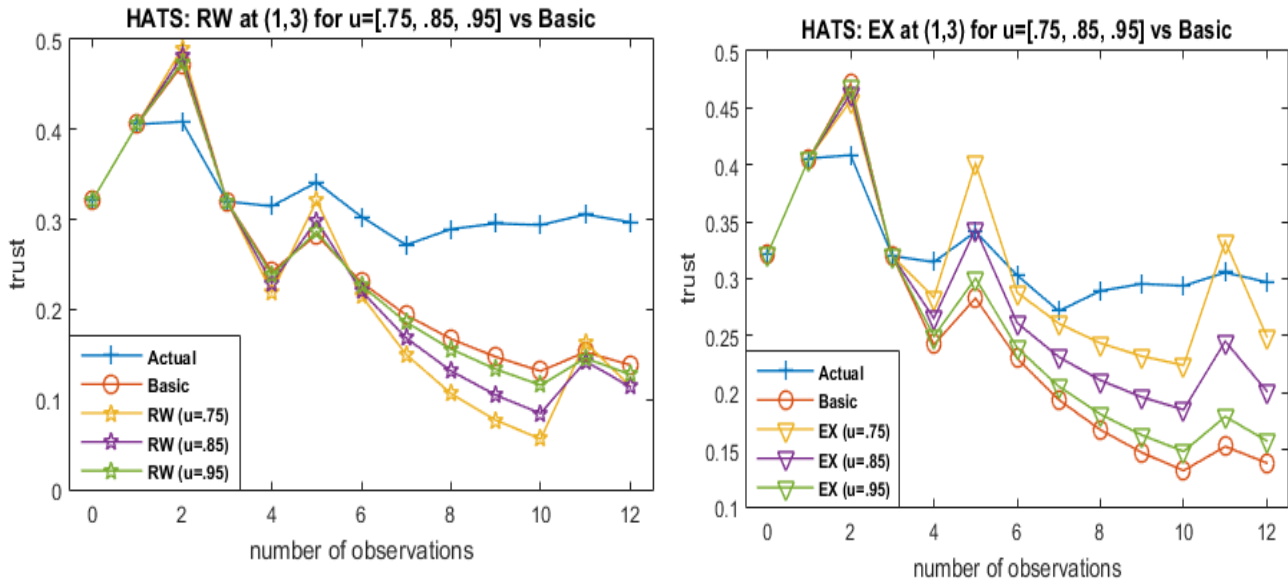
Firstly, notice that all graphs are identical at the 3 provided points, as a result of the applied "machine-learning" methodology. Concerning the basic model, its linear evolution fails to capture the trust recovery at trials (3, 8), resulting to significant miss-estimations. On the other hand, both RW and EX can be fine-tuned to reflect the severe trust fluctuations after each outcome interchange. But, their difference lies to their evolution after each adjacent same outcome (successes in this case). More specifically, note that for the same values of " $u$ ", EX appears with greater concavity at the adjacent successes, a property that fits better the actual trust responses. Finally, we mention the trade-off that parameter " $u$ " causes within the same model. Note that lower values result to



both higher increase (at interchange) and concavity (at adjacent same results), while the reverse effect appears for lower values.

In **Figure 19** we present the relevant graphs for the privacy metric. Recall that for this metric there was not consensus over all trials on how the different segments perceived the outcome. Thus we only consider trials as "successes" when trust increased for all segments, i.e., the vector of results for this metric is  $[1,1,0,0,1,0,0,0,0,0,1,0]$ . Notice that all theoretical models fail to capture the trust evolution shaping mostly due to the misalignment of the reactions: accuracy for all three models tend to decrease at trials (8, 9, 10), while actual trust remains almost constant or increases. Despite this fact, observe that our proposed extension achieves modest miss-estimations. At this case also, its better performance is due to the design properties: the greater convexity (at consecutive failures) keep the graphs close to actual responses at trials 8, 9 and 10, while the more intense reactions (at outcome differentiations) cause for it to recover at trials 5, 11.

Despite the fact that the trust form of our proposed extension resembles most the actual trust evolution, the applied value of coefficient " $u$ " may result to trust estimations that are far from the actual, or even less accurate compared to the other two models. This fact motivates our approach to identify its unique value as the solution of an optimization problem, which is described in the next section.



**Figure 19: The trust estimations of Basic (both), RW (left) and EX (right) towards the privacy metric, for different values of coefficient  $u$ .**

### 3.4.3. An approach for finding the optimal time-fading TME parameters

As we depicted in the previous section, the "machine-learning" approach over the time-fading process returns an infinite set of update coefficients, which are a function with respect to " $u$ ". In this section, we introduce our methodology aiming to identify a unique pair of values that best approximates the trust evolution. To this end, we utilize an additional actual trust measurement (at time moment  $t_3$ ) and derive the unique values by the solution of an optimization problem. More specifically, our objective is to minimize the absolute distance between actual and estimated trust

at  $t_3$  subject to the constraints that the two aforementioned magnitudes remain equal at  $t_1$  and  $t_2$ . The reason why we do not require equality also at  $t_3$ , is because there may not exist a value of " $u$ " within the  $[0,1]$  interval that satisfies all three constraints.

In a mathematical concrete formulation:

$$\min_{u_l^j, A_l^j, B_l^j} |\tau_l^j(t_3) - m_l^j(t_3)|$$

Such that:

$$\tau_l^j(t_1) - m_l^j(t_1) = 0$$

$$\tau_l^j(t_2) - m_l^j(t_2) = 0$$

$$0 \leq u_l^j \leq 1$$

For completeness reasons, we mention that in this form the problem cannot be solved by known solvers (e.g., Matlab), so below we provide the equivalent. In words, we introduce an additional optimization variable ( $d$ ) and correlate it with the distance between actual and estimated trust at the constraints:

$$\min_{u_l^j, A_l^j, B_l^j, d} d$$

Such that:

$$\tau_l^j(t_1) - m_l^j(t_1) = 0$$

$$\tau_l^j(t_2) - m_l^j(t_2) = 0$$

$$\tau_l^j(t_3) - m_l^j(t_3) - d \leq 0$$

$$-\tau_l^j(t_3) + m_l^j(t_3) - d \leq 0$$

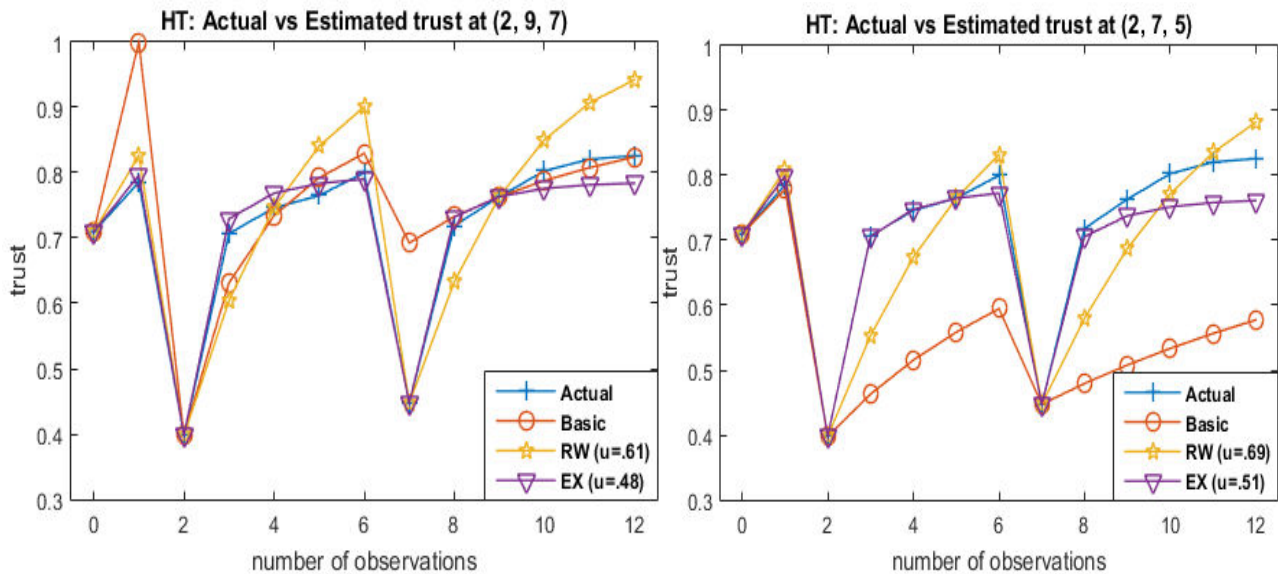
$$0 \leq u_l^j \leq 1$$

In the rest of the section we validate the accuracy of the TME and compare its performance with the other two models when we derive the optimal value of coefficient " $u$ ", as the solution of the optimization problem just described. In **Figure 20**, we juxtapose the actual trust responses with the graphs of all three models for two input sets in the form  $(t_1, t_2, t_3)^3$ . Firstly notice that in this case the values of RW and EX are equal with the given trust level not only at periods  $t_1, t_2$  (as in section 3.4.2), but also at  $t_3$ . This fact indicates that the optimization process identified a value for " $u$ " that achieves to minimize the objective, i.e., results to a zero distance between the estimated and actual trust. At the left-hand side, we apply the equality constraints after a failure (trial 2) and a success (trial 9), while the objective is at trial 7 (a failure). At the right-hand side, we apply the equality constraints after two failures (trials 2, 7), while the objective is at trial 5 (a success). Note

---

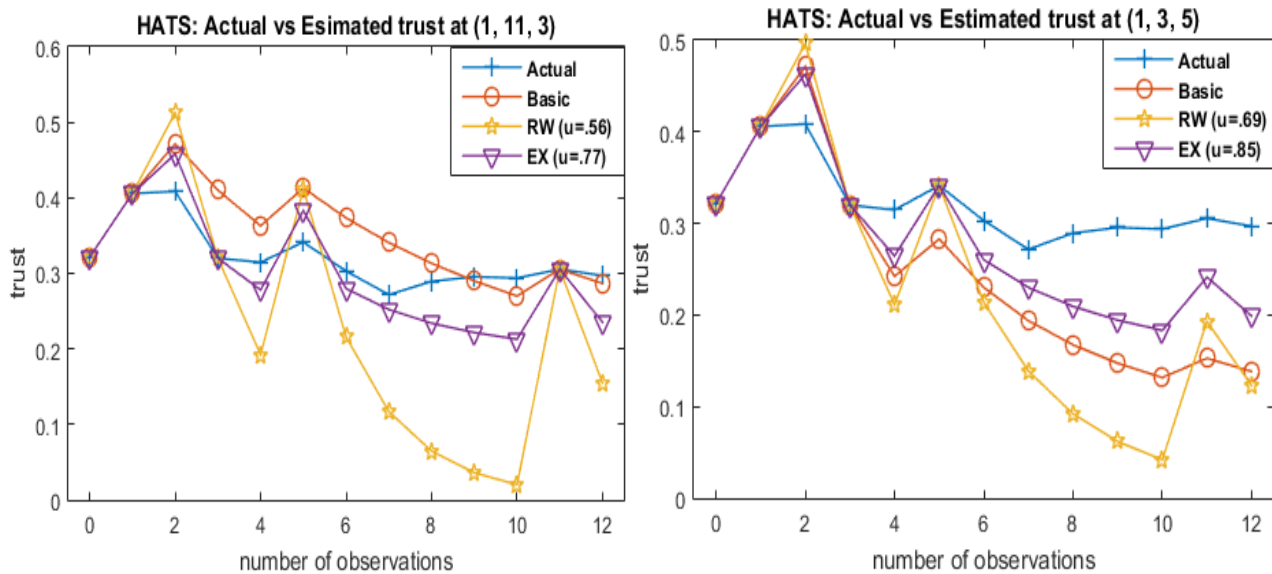
<sup>3</sup>The fourth actual trust value for each segment is the initial one (before any transaction is evidenced) and thus  $t_0$  can be skipped for brevity.

that the different input sets, have a marginal effect on the RW and EX graphs, revealing their design property to adjust the coefficient " $u$ " and capture trust fluctuations with accuracy. On the contrary, the basic graph is highly affected and the performance of the model strongly depends on the input set. Finally, a comparison between the right-hand-side of **Figure 20** with **Figure 18** justifies the core impact of the optimization framework at the computation of " $u$ " and consequently to the performance of the models.



**Figure 20: The trust estimations of Basic, RW and EX towards the performance metric, for two different input sets and the optimal values of coefficients  $u$ .**

In **Figure 21**, we present the relevant graphs towards the privacy metric. Their analysis is the same as above: the three points equation property holds also here and a comparison with **Figure 19** suggest the trust-estimations improvements achieved by the optimal value of coefficient " $u$ ".



**Figure 21: The trust estimations of Basic, RW and EX towards the privacy metric, for two different input sets and the optimal values of coefficients  $u$ .**

#### 3.4.4. Validation of TME

We now quantify the accuracy of the theoretical models by means of the Absolute Average Difference (AAD) between the actual and estimated trust of all trials that are not given as input. For completeness reasons we mention that the average for the basic methodology is over ten ( $13 - t_1 - t_2 - \text{initial} = 10$ ) points, while for RW and EX over nine ( $13 - t_1 - t_2 - t_3 - \text{initial} = 9$ ). Notice that the number of the potential sets that could be fed in the models is exponential to the number of trials. Thus our aim is not to cover all the possible cases, but to provide a guide for the major requirements that the input set should satisfy, for a meaningful comparison between the three approaches. Intuitively, it should reveal the participants' reactions at points of interest for the design purposes of the theoretical models, i.e. when an outcome interchange occurs (from success to failure or vice versa). As it becomes apparent next, this requirement allows for the properties of our proposed extension to appear and consequently achieve significant improvements. In the following tables, the input set is in the form  $(t_1, t_2, t_3)$ , followed by an explanatory parenthesis indicating whether each of the three outcomes resulted to a success or failure. Notice that we chose a representative subset of the potential input sets, providing various combinations of successes and failures for both metrics.

In **Table 5**, we report the AAD measurements and their average (last column) for the performance metric. We clearly observe that our proposed extension outperforms on the average the other two models. More specifically, it provides more accurate estimations for all cases, compared to RW. As we have already described (see also **Figure 18** and **Figure 20**), both methodologies achieve to capture the severe trust fluctuations at any outcome interchange, but EX better fits the actual responses at the adjacent same outcomes. Additionally, it outperforms the basic models for all input sets, apart from the last one (3, 6, 4) which includes trials that the search-engine provided only successful results. This combination does not reveal the user's reaction after a failure, thus the designed properties of EX and RW do not appear with the desired intensity. This is the reason why, both RW and EX, provide the poorest estimations over this set. On the contrary, their higher

improvement compared to the basic model is for the first input pair where the equality constraints are applied at the two failures. As we have already mentioned (see also **Figure 18** or **Figure 20-right**), the linear evolution of the basic model causes the inherent weakness in capturing the trust fluctuations and thus it returns greatly miss-estimated values. Notice that this set alone is enough to cause better average results for RW compared to the basic model, despite the fact that the latter outperforms at all the other cases (where we provide at least one success at equality constraints).

Performance Metric – AAD						
Segment-Method	Outcomes used as Input ( $t_1, t_2, t_3$ )					Average
	(2,7,5)- (f,f,s)	(2,6,8)- (f,s,s)	(2,4,7)- (f,s,f)	(2,9,7)- (f,s,f)	(3,6,4)- (s,s,s)	
HT-BASIC	0.3109	0.0827	0.0720	0.0728	<b>0.0833</b>	0.1243
HT-RW	0.0945	0.0922	0.0925	0.0949	0.1138	0.0976
HT-EX	<b>0.0478</b>	<b>0.0397</b>	<b>0.0523</b>	<b>0.0399</b>	0.0926	<b>0.0545</b>
HATS-BASIC	0.2170	0.0747	0.0663	0.0648	<b>0.0645</b>	0.0975
HATS-RW	0.0665	0.0830	0.0728	0.0729	0.1037	0.0798
HATS-EX	<b>0.0289</b>	<b>0.0436</b>	<b>0.0303</b>	<b>0.0229</b>	0.0717	<b>0.0395</b>
MATS-BASIC	0.2517	0.0884	0.0853	0.0743	<b>0.0855</b>	0.1170
MATS-RW	0.0916	0.0983	0.0957	0.0943	0.1499	0.1060
MATS-EX	<b>0.0708</b>	<b>0.0411</b>	<b>0.0467</b>	<b>0.0473</b>	0.0950	<b>0.0602</b>
A-BASIC	0.1811	0.0610	0.0766	0.0589	<b>0.0609</b>	0.0877
A-RW	0.0630	0.0668	0.0793	0.0720	0.1192	0.0801
A-EX	<b>0.0233</b>	<b>0.0529</b>	<b>0.0249</b>	<b>0.0238</b>	0.0677	<b>0.0385</b>

\* HT = High Trust; HATS = Highly active Trust seeking; MATS = Medium active Trust seeking; A = Ambivalent

\* AAD = Absolute Aggregate Difference; RW = Related Work; EX = Extension

**Table 5: AAD of the three models for the performance metric, over various input sets.**

In **Table 6**, we present the relevant results for the privacy metric. We chose to provide input sets that include only trials where consensus appeared on the reaction among all segments, i.e., either a trust increase or decrease (trials: 1, 2, 3, 4, 5, 6, 11, and 12). We mention beforehand that the

aforementioned misalignment between outcomes and responses at trials (8, 9, 10) does not allow for a robust analysis. Despite this fact we can still make the following observations:

- Here also, the accuracy of our proposed extension is higher on the average compared to the other models, but the improvement is clearly less impressive compared to the basic one.
- Furthermore, the RW model strongly fails to capture trust reactions, especially during the trials that the misalignment appears and consequently it is dominated even from the basic one for input sets.

Privacy Metric – AAD						
Segment-Method	Outcomes used as Input ( $t_1, t_2, t_3$ )					Average
	(1,3,5)- (s,f,s)	(2,3,5)- (s,f,s)	(1,11,3)- (s,s,f)	(2,5,4)- (s,s,f)	(3,4,5)- (f,f,s)	
HT-BASIC	0.1711	0.1647	<b>0.0674</b>	<b>0.0806</b>	0.0788	0.1125
HT-RW	0.2124	0.2027	0.2145	0.1169	0.1391	0.1771
HT-EX	<b>0.1076</b>	<b>0.1041</b>	0.0791	0.0868	<b>0.0575</b>	<b>0.0870</b>
HATS-BASIC	0.1088	0.0867	<b>0.0476</b>	<b>0.0401</b>	<b>0.0267</b>	0.0620
HATS-RW	0.1537	0.1360	0.1599	0.0445	0.1176	0.1223
HATS-E2	<b>0.0714</b>	<b>0.0495</b>	0.0492	0.0445	0.0284	<b>0.0486</b>
MATS-BASIC	0.0979	0.0994	0.0547	<b>0.0602</b>	0.0466	0.0718
MATS-RW	0.1379	0.1304	0.1537	0.0668	0.0510	0.1080
MATS-EX	<b>0.0674</b>	<b>0.0713</b>	<b>0.0251</b>	0.0668	<b>0.0421</b>	<b>0.0545</b>
A-BASIC	0.1098	0.1107	0.0494	<b>0.0432</b>	0.0861	0.0798
A-RW	0.1252	0.1213	0.1256	0.0851	0.1239	0.1162
A-EX	<b>0.0475</b>	<b>0.0537</b>	<b>0.0473</b>	0.0453	<b>0.0458</b>	<b>0.0479</b>

\* HT = High Trust; HATS = Highly active Trust seeking; MATS = Medium active Trust seeking; A = Ambivalent

\* AAD = Absolute Aggregate Difference; RW = Related Work; EX = Extension

**Table 6: AAD of the three models for the privacy metric, over various input sets.**

Concluding, the extended "machine-learning" model best fits the actual trust shaping and achieves to estimate the actual values with greater accuracy compared to the other described approaches. This fact holds both for the performance but also for the challenging case of the privacy metric,



where the unpredictable user's reaction create an unfavourable context for any theoretical framework to be applied.

### 3.5. Post-questionnaire analysis

In addition to the trials and validating the segmentation and models, we also explored how to further elaborate on the four segment solution in the context of perceived legal awareness. For this purpose, we turned to a hugely relevant development of understanding trust in the context of online environments (that is, socio-economic-technical systems) that have recognized the dynamic of trust and sharing of personal information, *warranting our interest to expand our segmentation with insights into the kinds and extents users are providing certain information bits about themselves.*

More specifically, technological progress is said to profoundly change the way connections and interactions between people are made and sustained over time and space, highlighting that increasingly personal information about these users is being collected, accessed and used. So-called digital footprints are left with every click individual users make online, such as on search engines, social network sites, location-based services, and transactional services. All kinds of existing and emerging companies collect and process these personal data streams on an unprecedented scale, often, serving as input for their economic and social activities.<sup>4</sup> Over the last three decades or so, the study of personal data-driven markets has evolved, yet it is currently not an easy task to keep track of what is happening in real markets [18]. For example, already in the 1980s the Chicago school of economics explicated privacy as a cause for information asymmetries, creating inefficiencies in the marketplace [19] [20] [21]. By the mid-1990s voices can be heard that draw attention to a marketplace in which individuals have the right, but not the obligation, to sell their information. Yet, a critical reflection can also be detected on the potential 'privacy divide' between those who can afford paying for privacy and those who do not. At the turn of the century, more attention was given about the dis/advantages of letting people freely negotiate over their personal data. Also, research now differentiates between the economics of privacy at the stage of information disclosure and later stages of information use as well as, increasingly, links personal information markets to the concept of multi-sided markets. Other themes of scrutiny include enhancing insights into empirical-based research and methods, the meanings of individual empowerment in personal information driven-societies, ethics, and the emerging phenomenon of big data. In summary, the research on personal information and markets is vibrant and cuts across disciplines.

If we look at established practices within social science research (in particular, communication studies), we can subscribe to a need to understand changing user practices and expectations concerning online trust in the context of privacy, personal data protection and personal data value definition [22]. This is highly relevant in the changing ICT context of mass self-communication, typically exemplified by the proliferation of socio-technical systems, such as online social networks. These transitions in the online technological world also fit in with new thinking in communication studies. According to Deuze (2012) [23] the key challenge of communication and media studies in the 'media life' of the 21st century is, or will be, the disappearance of media, where people

---

<sup>4</sup> This has drawn attention from policy makers that make a strong case for individual rights and to protect people's privacy and personal data vis-à-vis these commercial practices. Up to this day, however, large quantities of personal information are easily transferred between jurisdictions and are not well-balanced, such as global data flows between the United States and Europe.



increasingly are living 'in' media instead of living 'with' media. This perspective also fits in the notion of 'mediation' stating that 'mediated connection and interconnection' are part of the infrastructure of most people's lives in the Internet age [24] [25]. This is in line to similar schools of thought that have been developed in social sciences by scholars like Bauman (2000) [26] on 'liquid modernity', Wellman (2002) on 'networked individualism' [27], Orgad (2007) on 'online and offline' [28], and Couldry (2011) on 'media practices' [29]. The integration of online and offline social life then also has an impact on trust, privacy and surveillance of users, which is being identified as 'liquid surveillance' (Bauman & Lyon, 2013) [30].

Regarding technological tools for supporting user trust/privacy in online (mediated) environments, in particular, various Privacy Feedback and Awareness (PFA) tools have been developed. The general effectiveness of the feedback approach has been tested in a wide range of empirical and experimental demonstrations of the effects of feedback on human behaviour such as academic learning [31] (where the awareness is constituted by metacognition of one's learning activities) and health-related areas (Lehrer, 1996) [32]. Moreno et al. (2009) demonstrated that even mild feedback on public profile contents in online social networks delivered via email, can be effective in the sense of leading to a more trustworthy and privacy-conscious self-portrayal [33]. Although the general effect of feedback towards users has already been tested, there is not yet sufficient proof of the evaluation of feedback and awareness tools in the context of trust-privacy setting. This kind of evaluation is a complex process. Not only usability criteria are into play, but also the correctness and comprehensiveness of the feedback should be evaluated. Capturing the user context and usage is especially of importance. Different gaps currently exist in this research domain. Lessons learned for end-user programming environments and human-based computation games tend to be taken into account when defining possible implementations based on social requirements capturing. In that way a transfer of knowledge from data controller to data subject yielding insights into the logic of data processing, is aimed for. Furthermore, protection and revelation of personal data flows involve tangible and intangible trade-offs for the data subject as well as the potential data holder underpinned by implicit assumptions of (economic) value.

Research has tended to concentrate on (explicitly or implicitly) measuring the amount of money (or, benefit) an individual is likely deem sufficient in order to give away their personal data (Wathieu and Friedman, 2005 [34]; Huberman et al., 2006 [35], Hui et al., 2007) [36] as well as the investigation of tangible prices or intangible costs that individuals are willing to pay to protect their privacy (Acquisti and Grossklags, 2005 [37], Varian et al., 2005 [38]). Following a 'canonical' economic stance, however, individuals tend to be approached as having stable preferences over privacy that underlie mental trade offs they between costs and benefits of sharing and protecting personal data suggesting individuals make rational decisions about what personal information they reveal and what to protect, and, hence, suggests that there is no need for market (regulatory) intervention.

Yet, contemporary debates focusing on the user-defined values of trust-privacy urge for a behavioural understanding of the economics of privacy rather than approaching privacy decision making as mere 'rational'. For example, user-defined values in privacy valuations have been found to be inconsistent (e.g. people seem to make inconsistent privacy-relevant decisions) and a control paradox can be detected (e.g. providing users with more control over information publication points to increase their willingness to disclose sensitive information). In fact, recent findings have suggested that individuals seem to assign different values to their data privacy depending on, the one hand, on "whether they consider the amount of money they would accept to disclose otherwise private information, or the amount of money they would pay to protect otherwise public information" and, on the other hand, "the order in which they consider different offers for that data" (Acquisti, John, and Loewenstein, 2013: 1 [39]); the 'price' to protect a piece of information differs from the

price people give it when considering sales. It is difficult, therefore, to infer exact evaluations of user-defined value toward their personal privacy, guiding their likelihood of trust and trustworthiness of online environments (and, a gap that is much wider than for ordinary consumer goods).

Against this backdrop, we have sought to elaborate our trust-related segmentation with (emerging) insights based on revealing or sharing personal information in the context of the search engine experiment described in the previous sections. Questions (see Annex) focused on disclosing personal information online about "demographics" (e.g. gender, year of birth, nationality, living situation), "online consumption" (e.g. what kind of social media are used, top 3 sites for online purchases, how online payments are made), and these were followed by asking "How sensitive do you find the information you had to reveal about ..?", "Do you think this information is available online about you by performing a search, for example, Google?", and "How important is it for you that this type of information about you is not publicly available?". What follows are the main findings.

### 3.6. Post-questionnaire findings

In order to elaborate the four segment solution here relating privacy to trust, we identified a list of personal data attributes that have been qualified as sensitive or valuable by the users (e.g. privacy scoring framework such as developed by USEMP<sup>5</sup>).<sup>6</sup> For users, to better perceive the different aspects of their trust-related privacy, identified attributes can increasingly be seen to be organized in a number of high-level categories, so-called privacy dimensions.<sup>7</sup> This organization allows for a

---

<sup>5</sup> The '*User Empowerment for Enhanced Online Management Project*' (USEMP), funded by EUFP7, builds on the Personal Data Economics Paradigm (PDE), which wants to empower data subjects with regard to the sharing of their personal data. PDE refers to an architecture that (1) returns a measure of control to the users of online and offline services that require the sharing of information, and (2) provides a measure of transparency as to who profits how from processing and sharing their data. The most cited format of PDE is that of the Personal Data Store, Locker or Vault. It entails a single point of control or secure dashboard, which allows users, consumers, citizens, data subjects to manage their personal data with something like a secure dashboard. The idea is related to that of Vendor Relationship Management (VRM), which counterbalances the more common idea of Customer Relationship Management (CRM). Though USEMP is not focused on the idea of the Personal Data Vault, it will develop a set of tools allowing users of OSNs greater control over the personal data they share within the network while also providing them with tools to enable the use of their data by entities outside the OSN, e.g. in the form of licensing agreements. In that sense USEMP builds upon the notion of PDE and may in fact assume a Personal Data Vault to provide the secure environment for effective control over the relevant data. It works around two use cases: the development of an USEMP OSN Presence control tool, allowing users to improve their control over information and content they share online, and an USEMP Economic Value Awareness Tool, raising the awareness of users about the economic value of their data which is currently utilised exclusively by the data controller (see [usemp-project.eu](http://usemp-project.eu) for more information about USEMP).

<sup>6</sup> We recognized that we need to qualify this as 'perceived' sensitivity, since when the law qualifies certain data as sensitive, based on art. 8 Data Protection Directive (DPD), this has major legal effect, which, however, does not depend on how a user 'feels' about the data.

<sup>7</sup> Clearly, these dimensions are not exhaustive and they not necessarily match with the legal right to privacy as stipulated in art. 8 of the European Convention of Human Rights, or with the fundamental rights to privacy and data protection of the Charter of Fundamental Rights of the European Union. It is pivotal that perceived

clear and intuitive presentation and handling of the different aspects of a user's personal information, here, vis-a-vis the trust segmentation solution. One of the privacy dimensions to be considered within OPTET is demographics, which includes user attributes such as age, sex, etc., and another is about consumption factors, which includes attributes such as payment means, purchasing – another example, is health and which can include smoking and drinking, etc.<sup>8</sup> Such a grouping has multiple benefits for the user. First, it enables him/her to form a succinct, easy to grasp mental model of his/her private information and to prioritize its different parts. Second, it enables the use of different compact visualization methods that will further augment the user's awareness with respect to his/her private information and trust perception.

Demographics and consumption profile were selected for users conceiving them of their private nature (perceived privacy), while also encompass information that is considered sensitive from a legal perspective (legally sensitive data). In addition, based on current business practices (mainly stemming from the marketing industry), the identified dimensions are associated with certain value levels, i.e. they carry a certain level of utility for (marketing) companies that are interested in targeting consumers, and which underpins our experiment of deploying a hypothetical search engine. Table 6 summarizes the two selected privacy dimensions, along with the value levels associated with them for the purpose of the empirical investigation.

#	Name	Description	Trust-threats sensitivity	Value (search engine/advertising)
1	Demographics	Personal data, such as Gender, Age, Nationality, Living situation, Location, etc.	Discrimination in a variety of settings. The most frequently used type of information.	High: advertisers wish to target users of certain demographic criteria
2	Consumption profile	Preferred products and brands, means of payment	(Ad) targeting and discrimination in online price-setting	High: advertisers wish to target consumers based on their consumer profile attributes like the devices the use to access digital content

**Table 7: Overview trust-privacy dimension for ACME search engine survey**

### Trust-privacy scoring

Starting from the level of values, the primary findings are presented next.

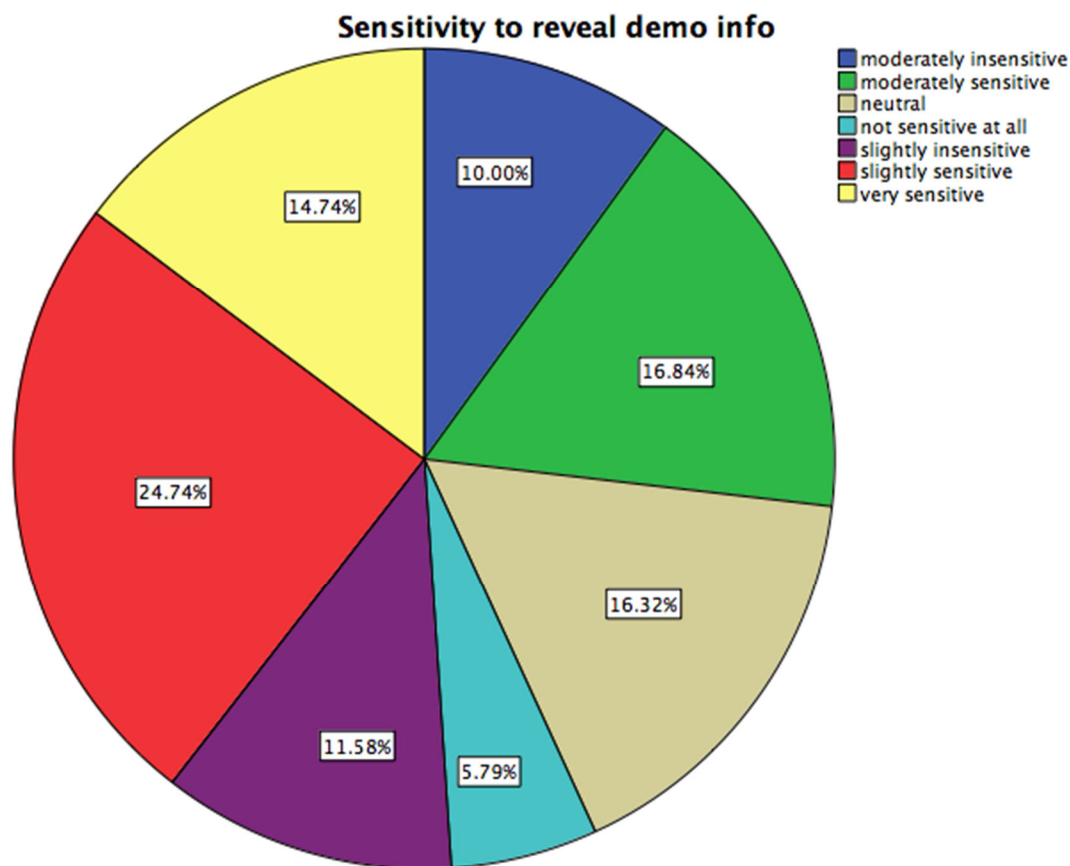
privacy and the right to privacy are understood on their own merits, taking note that the latter aims to provide the level playing field for users to develop their own privacy preferences.

<sup>8</sup> Key categories of personal attributes associated with privacy dimensions can include, among others, Demographics, Psychological Traits, Sexual Profile, Political Attitudes, Religious Beliefs, Health Factors & Condition, Location and Consumer Profile.

### General demographics

Based on sex indicators, 60% of the respondents reported to be male and 40% to be female (n=211); everybody disclosed this information. The respondents were born between 1946 and 1997 (n=205), and with about 70% indicating to be older than 27. A relatively large group of respondents came from Belgium, Greece, UK, and the USA. Asked about their highest level of education, 22% reported to have finished high school, 34% reported to possess a Bachelor's degree, 25% a Master's degree, and 14% an advanced/PhD degree (n=211). More than half reported to live with a partner, of which equally reported to live with/without kids (23%). Some 21% reported to live alone, and of which 4% alone but with children, and 15% said to live with their parents.

When asked about how they perceive the level of sensitivity to reveal this information, on a Likert-scale 1 to 7, the following Figure shows the results (M=4.54, SD=1.735, N=190).



**Figure 16: Sensitivity to reveal information about one's demographics**

Furthermore, 65% vs 27% (n=194) thinks this information is available/not-available about them online when a search is performed on a search engine, such as Google. Of the respondents nearly 60% finds it import that this information is *not* publicly available about them when a search is performed (M=5, SD=1.941).

The respondents were asked about their general trust stance, by how they tend to trust other people in their daily live environment. Some 53% reported their tendency to trust another person 'until

there is a reason not to' ( $M=3.67$ ,  $SD=1.045$ ,  $N=210$ ). Even when the stakes are high 45% said to believe that most people tend to be honest in their dealings with others ( $M=3.19$ ,  $SD=1.07$ ,  $N=211$ ). Consistently, a little less than half reported to believe that people may not care about the well-being of others ( $M=2.58$ ,  $SD=1.061$ ,  $N=210$ ). Furthermore, when online applications are considered, 57% of the respondents claimed that they would not easily trust an online provider ( $M=2.78$ ,  $SD=1.13$ ,  $N=211$ ). More than half of the respondents do not find that the higher the price for an online application makes it more/less trustworthy, however, about 80% ( $n=211$ ) perceives an application provided by a public organization as more trustworthy than those provided by a commercial organization.

### General consumer profile

Several questions were asked about frequently consuming social network sites, sites for online purchases, and means of payment used in order to learn about perceived sensitivity of the respondents towards experiencing perceived levels of trust-privacy in these settings.<sup>9</sup> Table 7 shows how the respondents ( $n=180$ ) ranked their preferred social networks (based on providing us with their Top 3).

What social media do you use?	% ( $N=180$ )
Facebook	66
LinkedIn	8
Instagram	6
Google+	2
Twitter/WhatsApp	1
Reddit	0.5
Other	17

Source: OPTET ACME Search Engine Survey

<sup>a</sup> Ranking social media sites

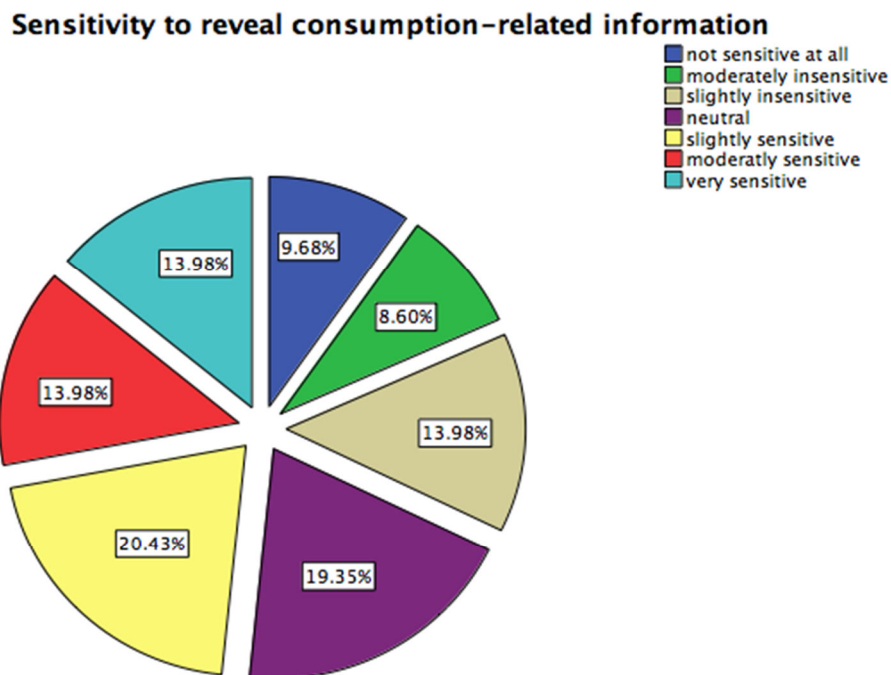
<sup>9</sup> In D2.1 we found that significant predictors for trust in online stores included structural assurance ( $\beta = .294$ ), the design element 'works well technically' ( $\beta = .245$ ), monetary advantage as motivation to engage with online stores. ( $\beta = -.214$ ) and trust related seeking behavior ( $\beta = -.189$ ). The regression model for online stores showed a proportion of explained variance of 30,8%. These results indicated that for users, in order to display trust in online stores, it is important that the site is working well technically and that users can belief that structures such as regulations and safeguards exists to assure safe and trustworthy interaction experiences. Trust related seeking behavior in this context, could be seen as an indication of low trust levels and as a way to eventually increase one's trust level, hence a high amount of seeking behavior does correlate with low trust levels. Concerning "social networks", significant predictors included structural assurance ( $\beta = .211$ ) and display of seals of approval ( $\beta = .211$ ). This regression model has an explained variance of 17%. These results stressed the need for third party trust certificates and the belief that regulations should be in place to allow trustworthy interactions helps in eliciting trust in social networks sites. These trust levels should be considered as general and a priori trust levels towards a particular set of technologies, hence, not towards any specific application or on the basis of a concrete experience. For instance we asked in general about trust in online stores and not specifically about trust in individual stores such as Amazon. Users without first hand experience could leave the question unanswered; hence they were not forced to express their opinion.



**Table 8: Social media use**

When asked about which sites they tend to use to make online purchases (Top 3), Amazon was with 67% (n=178) by far the most frequently mentioned site, followed by eBay (18%) and iTunes (6%) and remaining sites were deemed as other. We also asked respondents about how they tend to pay for online purchases. Here, the results (n=191) returned almost a half split between using a credit or debit card, with a small number of people who said to prefer to use services such as PayPal, a phone app or to receiving an invoice.

When asked about how they perceive the level of sensitivity to reveal this kind of information, on a Likert-scale 1 to 7, the following Figure shows the results (M=4.30, SD=1.818, N=186).



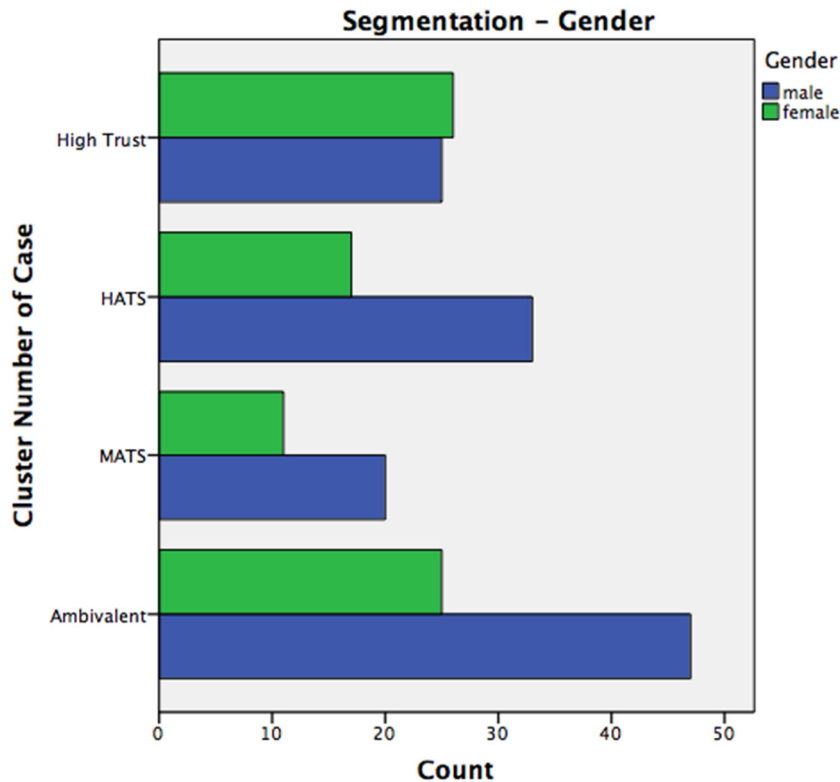
**Table 9: Consumption-related information**

Furthermore, 35% vs 65% (n=191) thinks this kind of information is not/available about them online when a search is performed on a search engine, such as Google. Of the respondents nearly 70% find it important that this information is *not* publicly available about them when a search is performed (M=4.93, SD=1.910, N=203). In addition, 70% of the respondents (M=1.32, SD=.567, N=195) said that it would impact their online behaviour, if they would know that this kind of information about them would be publicly available.

Against this backdrop, we will now elaborate these results in the context of the four segment solution.

*Elaborating trust-privacy indicators in 4 segment solution*

In order to expand our insights into more contextual social-economic personal information, the analysis based on male/females across the segmentation, shows that for A (n=72) consists of 23% males vs 12% females; MATS (n=31) 10% vs 5%; HATS (n=50) 16% vs 8%; and, HT (n=51) 12% vs 13%. This is pictured next:



In terms of education (n=204), the advanced graduate/PhD degrees can be found among the A segment (48%), followed by HATS (31%) and HT (17%); master's degree also A (43%) and HT (29%); bachelor's degree HT (29%), A (27%) and MATS show that this is the highest degree among this segment (45%) and has relatively few advanced degrees (3%). HATS have with 34% the largest group with (only) a high school diploma. When considering the respondents in terms of their living situation (n=204), the MATS segment shows relatively the fewest respondents having/living with children, followed by the A segment. The HATS group shows relatively the most people living with children, while HT seems to have most people living alone living with children.

To better understand how people across the segments perceive of their online control and safety, that is, in terms of being 'protected' associated with information availability, the survey asked respondents about how they feel about the following statements. The first column provides information of the total group of respondents (n=211) while the others present the mean distribution per segment. In order to check for validity/reliability between and within groups ANOVA tests were conducted, all indicating significance, respectively, 1 (F=6.523, Sig.= .000), 2 (F=4.551, Sig.= .004), 3 (F=2.341, Sig.= .075), 4 (F=4.542, Sig.= .004), and 5 (F=2.937, Sig.= .034).



#	TOTAL Mean <sup>a</sup>	HT (n=51)	HATS (n=50)	MATS (n=31)	A (n=72)
1. I have lost control over how my personal information is collected and used by digital applications	3.45	3.45	3.18	2.84	3.93
2. Organizations handle the personal information they collect about me in a proper way	2.59	2.75	2.64	2.84	2.25
3. Existing laws and organizational practices adequately protect me from online problems or risks today	2.29	2.24	2.52	2.42	2.07
4. I think that (future) legal measures are necessary to increase my control	4.11	4.22	4.04	3.71	4.35
5. I feel confident that encryption and other technical advances online make it safe for me to use	3.14	3.14	3.14	3.52	2.90

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-5 (Statements, 1=strongly disagree; 2=disagree; 3=neither dis/agree; 4=agree; 5=strongly agree).

**Table 10: Perceived control and safety indicators**

When it comes down to feelings of "loss of control over one's information" the MATS seem the least to perceive that as such, while the A segment seem to indicate to agree with this perceived experience. In conjunction with this, they also seem to be the least trustful with how organizations may handle personal information as well as that current laws and practices are adequate enough to offer protection and hence, new ones need to be put into place and they are least feeling that technical solutions are an answer to increasing one's perceived control and feelings of safety. The HT segment feels that various new legal measures can be beneficial, while MATS show that they may tend to agree but seem to show more trust in technical solutions to increase perceived feelings of control and safety.

In order to learn about how people across the segments perceive of providing personal information online questions were asked about their gender, year of birth, nationality, highest educational degree, living situation, city in which they live, city in which they work, and where they spent their last holiday. In addition, to answering these questions they were asked about how they feel about the following statements to reveal this sort of information. The first column provides information of the total group of respondents (n=211) while the others present the mean distribution per segment. In order to check for validity/reliability between and within groups ANOVA tests were conducted, all indicating significance, respectively, 1 (F=4.085, Sig.= .008), 2 (F=1.643, Sig.= .081), and 3 (F=3.801, Sig.= .011).

#	TOTAL Mean	HT (n=51)	HATS (n=50)	MATS (n=31)	A (n=72)
1. How sensitive do you find the information you had to reveal about your demographics	4.54 <sup>a</sup>	4.82	4.27	3.58	4.85
2. Do you think this information is available online about you by performing a search on, for example, Google?	1.34 <sup>b</sup>	1.29	1.45	1.46	1.24
3. How important is it for you that this type of information about you is not publicly available?	4.79 <sup>c</sup>	5.22	4.73	3.75	4.94

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-7 (Statements, 1=not sensitive at all; 7=very sensitive).

<sup>b</sup> Values range: yes/no.

<sup>c</sup> Values range from 1-7 (Statements, 1=not important; 7=very important).

**Table 11: Perceived sensitivity personal information – demographics**

From the findings, we can see that HT and A segments indicated that providing demographic-related information is a quite sensitive act, while MATS seem to still find it sensitive but less so (also, in comparison to the overall non-segmented respondent group). Awareness of whether this kind of personal information may be available online MATS and HATS show somewhat 'doubt' that this may or may not be the case, while HT and A segments are inclined to think that this is likely. Looking at the scores indicating how important it is for this kind of information to *not* be publicly available, we can see that that for the HT segment this is quite important, followed by the A and HATS segments. For MATS is seems important but less so, especially compared to the HT segment.

Also, questions were asked to yield insights into how people across the segments perceive of providing personal information about their Top 3 social media sites they use, where they make online

purchases, and their preferred means of paying for purchases. Again, they were asked to not only answer these questions but also how they felt about revealing this kind of information, and which findings are presented below. The first column provides information of the total group of respondents (n=211) while the others present the mean distribution per segment. In order to check for validity/reliability between and within groups ANOVA tests were conducted, all indicating significance, respectively, 1 (F=4.499, Sig.= .005), 2 (F=.073, Sig.= .097), and 3 (F=3.558, Sig.= .015).

#	TOTAL Mean	HT (n=51)	HATS (n=50)	MATS (n=31)	A (n=72)
1. How sensitive do you find the information you had to reveal about your online consumption	4.33 <sup>a</sup>	4.88	4.61	3.41	4.03
2. Do you think this information is available online about you by performing a search on, for example, Google?	1.65 <sup>b</sup>	1.66	1.66	1.67	1.63
3. How important is it for you that this type of information about you is not publicly available?	4.96 <sup>c</sup>	5.54	5.06	4.14	4.81

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-7 (Statements, 1=not sensitive at all; 7=very sensitive).

<sup>b</sup> Values range: yes/no.

<sup>c</sup> Values range from 1-7 (Statements, 1=not important; 7=very important).

**Table 12: Perceived sensitivity personal information – consumption**

Quite similar as the results for demographic-related information are the results in terms of perceived sensitivity for personal information being available about one's consumption online. Here, we see that the MATS segment is again relatively less preoccupied with providing sensitive information, or that this kind of information is publicly available. Overall, the segments seem to move towards thinking that this information is less likely to be found online and also find it quite important that it indeed is not publicly available. Across the segments, the HT segment scores relatively lower in the willingness that this kind of personal information is disclosed or available.

Lastly, we asked if being aware of the kinds of information un/available may impact their attitude towards their online behaviour; in order to get a feel for a "sample" awareness the question was asked if they ever read Google's privacy policy. Next, we present the results. The first column provides information of the total group of respondents (n=211) while the others present the mean distribution per segment. In order to check for validity/reliability between and within groups ANOVA

tests were conducted, all indicating significance, respectively, 1 ( $F=3.260$ ,  $\text{Sig.}=.023$ ) and 2 ( $F=3.418$ ,  $\text{Sig.}=.018$ ).

#	TOTAL Mean <sup>a</sup>	HT (n=51)	HATS (n=50)	MATS (n=31)	A (n=72)
1. Does your awareness of the kind of data that is processed change your attitude towards your online behaviour?	1.31	1.24	1.18	1.57	1.35
2. Have you ever read the privacy policy of a search engine, such as Google?	1.75	1.60	1.59	1.87	1.90

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range: yes/no.

**Table 13: Perceived sensitivity personal information**

The findings presented in this Table suggest that MATS seem to be relatively less inclined to possible change their online behaviour in conjunction to their level of awareness of which personal information may or may not be available online, while HATS are most likely to do so. When looking at the results whether people have read Google's privacy policy (as an indicator of "awareness") the A and MATS segment seem to have less or not done so relatively to HATS and HT, though all segments seem not so inclined to do so.

When we take a look at the legal aspects, we build further on the D2.1 contribution [13] (focused on the interplay between the law and trust) by examining the pragmatic effect of legal awareness on an individual's perceived level of trust. The legal objective here is to explore individuals' responses to legal (sometimes personal) information and guarantees i.e. signalling (un)trustworthiness through legal signposting/cues. This objective is examined by analysing the responses to relevant questions asked during the segmentation questionnaire. As the OPTET project centres on optimising online trust and trustworthiness primarily within the EU, the analysis also compares the mean responses between EU respondents ( $n = 105$ ) and non-EU respondents ( $n = 106$ ) in order to highlight any regional differences and similarities to the socio-legal questions posed.<sup>10</sup> Overall, this legal objective is concerned with investigating individuals' perceived awareness of the law in practice, and whether this plays a role in forming their perceived level of trust.

<sup>10</sup> EU respondents are defined as individuals who were resident in the EU at the time they responded to the segmentation questionnaire ( $n = 105$ ). Non-EU respondents are defined as individuals who were not resident in the EU at the time at the time they responded to the segmentation questionnaire ( $n = 106$ ).

## 4. Signalling (un)trustworthiness to end users – legal signposts

From the segmentation questionnaire responses, a significant majority of respondents either strongly disagree (21.8%) or disagree (43.1%) that existing laws and organisational practices adequately protect them from online problems or risks today; only a small minority of respondents either strongly agree (1.4%) or agree (12.8%).<sup>11</sup> Furthermore, the vast majority of respondents either strongly agree (35.5%) or agree (48.3%) that legal measures such as 'the right to be forgotten' (i.e. the ability to erase your personal information) are necessary to increase control.<sup>12</sup> In addition to this, there is only minor variance between the mean responses for EU and non-EU respondents:

COMPARING EU AND NON-EU RESPONSES		
Segmentation questions	EU mean response	Non-EU mean response
Q10: Existing laws and organizational practices adequately protect you from online problems or risks today? <sup>a</sup>	<b>2.21</b>	<b>2.36</b>
Q11: Do you think that legal measures such as 'the right to be forgotten' (i.e. the ability to erase your personal information) are necessary to increase control? <sup>a</sup>	<b>4.20</b>	<b>4.03</b>

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-5 (Statements, 1=strongly disagree; 2=disagree; 3=neither dis/agree; 4=agree; 5=strongly agree).

In consequence, while it is perceived that (on the whole) existing laws are not fit for purpose both within and outside EU, the majority of respondents would appear to support a robust legal framework.

ICT platform providers may imply trustworthiness through their (past) behaviours, brand, reputation and goodwill. From the segmentation questionnaire responses, reputation appears to be an important consideration for end users, as the vast majority of respondents always (35.1%) or sometimes (48.3%) look for information about the reputation of an organisation.<sup>13</sup> Moreover, there is only minor variance between the mean responses for EU and non-EU respondents:

<sup>11</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q10: Existing laws and organizational practices adequately protect you from online problems or risks today? 21.8% strongly disagree; 43.1% disagree; 20.4% neither dis/agree; 12.8% agree; 1.4% strongly agree; and, 0.5% no usable response (n = 211).

<sup>12</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q11: Do you think that legal measures such as 'the right to be forgotten' (i.e. the ability to erase your personal information) are necessary to increase control? 2.4% strongly disagree; 3.3% disagree; 10.4% neither dis/agree; 48.3% agree; and, 35.5% strongly agree (n = 211).

<sup>13</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q13: In general, do you look for information about the reputation of the organization? 4.7% never; 11.8% hardly; 48.3% sometimes; and 35.1% always (n = 211).

COMPARING EU AND NON-EU RESPONSES		
Segmentation question	EU mean response	Non-EU mean response
Q13: In general, do you look for information about the reputation of the organization? <sup>a</sup>	3.09	3.19

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-4 (Statements, 1=never; 2=hardly; 3=sometimes; 4=always).

However, where an ICT platform provider lacks distinguishable characteristics (e.g. a unique selling point for trustworthiness), is relatively unfamiliar (e.g. a new entrant to market) and/or there is no explicit trustworthiness information given (e.g. codes of best practice openly published on their website), end users are likely to face greater difficulties when attempting to verify the trustworthiness attributes of a specific ICT platform. An information asymmetry arises, as end users are without access to the relevant underlying socio-economic, technical and legal information, and the necessary interdisciplinary knowledge bases required to make a robust trustworthiness assessment.

ICT platform providers can opt for various signalling methods to better-communicate trustworthiness attributes to end users, e.g. by openly releasing notices, terms and conditions, codes of conduct, membership of authoritative organisations, (industry) awards, certificates, policies, informed consent, click-wrap licences, and certification/trust marks/seals.<sup>14</sup> From the segmentation questionnaire responses, these legal cues seem to be useful. For example, most of the respondents either always (26.1%) or sometimes (44.1%) look for any guarantees regarding the confidentiality of the information they provide; a very small minority (9%) never look for such guarantees.<sup>15</sup> Again, there is only a minor difference between the mean responses by EU and non-EU respondents to this question:

<sup>14</sup> For further information see [43] which examines cue-based trust for small online retailers, including seals of approval, [46] which investigates how online privacy policy influences consumer trust, and [57] which investigates: "*four common trust indices (i.e. (1) third party privacy seals, (2) privacy statements, (3) third party security seals, and (4) security features).*" Signalling certain attributes to consumers is not only confined to trustworthiness, e.g. see [60] which focuses on signalling the environmentally-friendly features and attributes of products, and [59] which focuses on signalling the quality of agricultural seeds through certification. For a more general overview of signalling theory see, e.g., [45] and [54].

<sup>15</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q16: In general, do you look for any guarantees regarding confidentiality of the information that you provide? 9.0% never; 20.4% hardly; 44.1% sometimes; 26.1% always; and, 0.5% no usable response (n = 211).

COMPARING EU AND NON-EU RESPONSES		
Segmentation question	EU mean response	Non-EU mean response
Q16: In general, do you look for any guarantees regarding confidentiality of the information that you provide? <sup>a</sup>	2.83	2.92

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-4 (Statements, 1=never; 2=hardly; 3=sometimes; 4=always).

While the segmentation questionnaire was not able to cover the entire range of legal cues, the 2015 version was modified to include two additional questions about trust marks (Q.20 and Q.21).<sup>16</sup> These questions were added in order to take into consideration: (a) the increased attention given to trust marks by European policy makers,<sup>17</sup> e.g. Article 39 of the proposed European General Data Protection Regulation (GDPR)<sup>18</sup> [40] is set to encourage the use of certification and seals for personal data processing activities; and (2) the various – current and proposed – European trust marks endorsed by public regulators e.g. the European Trustmark (EMOTA)<sup>19</sup>, EuroPriSe<sup>20</sup>, the ICO Privacy Seal<sup>21</sup>, CNIL labels,<sup>22</sup> and ULD Gütesiegel.<sup>23</sup> Trust marks appear to have an important role in signalling trustworthiness, as the majority of the respondents either always (28.9%) or sometimes

<sup>16</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q20: I tend to be more cautious about visiting websites or using online services, systems or applications which have no trust marks? Q21: I tend to find a trust mark to be more authoritative if it has been accredited or endorsed by a public regulator? 'Trust mark' definition: a symbol used to represent that a website, system, application and/or service has passed a particular set of best practice criteria i.e. for quality, privacy or security. Trust marks are also known by (but not limited to) the following terms: certification marks, authentication marks, quality assurance labels and seals of approval.

<sup>17</sup> For instance, there are already a number of trust mark and certification programmes available for ICT platforms outside the EU, e.g. TRUSTe (USA) and Privacy Mark (Japan) [55].

<sup>18</sup> Refer to the OPTET D2.5 Report, Section 4 for more information about the proposed European General Data Protection Regulation (GDPR).

<sup>19</sup> On 1 July 2015, the European eCommerce Association (EMOTA) launched the European EMOTA trust mark for online shopping [65]. Its release was welcomed by the European Commission as a means to facilitate and increase online purchasing across member states: "[t]oday's launch will help build consumers' trust in the digital world. Currently, only 15% of European consumers buy online from other Member States" [65].

<sup>20</sup> EuroPriSe [61] is the European privacy seal for IT products and IT based services; enabling companies to display their privacy compliance. On 13 July 2008, Ixquick – a meta-search engine based in the Netherlands – became the first recipient of the EuroPriSe privacy seal [67]. According to EuroPriSe's Register of Awarded Seals, there are currently twenty active privacy seals [64] (correct on 19 August 2015).

<sup>21</sup> Proposals for a privacy seal scheme endorsed by the Information Commissioner's Office (ICO) in the UK are set to be launched in early 2016 [41], [53].

<sup>22</sup> CNIL labels [66] are administered by La Commission Nationale de l'Informatique et des Libertés (CNIL) – National Commission on Informatics and Liberties – in France.

<sup>23</sup> The ULD Gütesiegel [63] (ULD Seal of Approval) is managed by the Unabhängiges Landeszentrum für Datenschutz (ULD) – Independent Centre for Privacy – in Germany.



(36%) look for trust marks or seals of approval.<sup>24</sup> Furthermore, the majority of the respondents are either always (34.6%) or sometimes (34.6%) cautious about visiting websites or using online services, systems or applications which have no trust marks.<sup>25</sup> Finally, the majority of the respondents either always (32.2%) or sometimes (43.6%) tend to find a trust mark to be more authoritative where it has been accredited or endorsed by a public regulator.<sup>26</sup> Only 11.4% of respondents stated that they would never find such trust marks to be more authoritative. This finding seems to correspond with an earlier Information Commissioner's Office (ICO) survey [41], which held that 80% of its respondents would approve the introduction of an ICO endorsed privacy seal.<sup>27</sup> However, the mean responses for Q20 and Q21 show a small degree of variance between EU and non-EU respondents:

COMPARING EU AND NON-EU RESPONSES		
Segmentation questions	EU mean response	Non-EU mean response
Q. 19 I look for trust marks or seals of approval when visiting a website or using an online service, system or application. <sup>a</sup>	2.67	2.96
Q.20 I tend to be more cautious about visiting websites or using online services, systems or applications which have no trust marks. <sup>a</sup>	2.75	3.13
Q21 I tend to find a trust mark to be more authoritative if it has been accredited or endorsed by a public regulator. <sup>a</sup>	2.85	3.15

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-4 (Statements, 1=never; 2=hardly; 3=sometimes; 4=always).

This small mean response increase for non-EU respondents could be due to a more mature trust mark ecosystem operating outside of Europe; particularly in the USA, who accounted for the majority (around 84%) of non-EU respondents (n = 89). An inference could be that websites, online services, systems or applications which have no trust marks may be more of a concern where trust marks are in wider use. Furthermore, in a digital environment that is more saturated with trust marks, it may be useful for those end users to distinguish via levels of authority e.g. trust marks which are

<sup>24</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q19: I look for trust marks or seals of approval? 12.8% never; 21.8% hardly, 36.0% sometimes; 28.9% always; and, 0.5% no usable response (n = 211).

<sup>25</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q20: I tend to be more cautious about visiting websites or using online services, systems or applications which have no trust marks? 11.4% never; 17.5% hardly; 34.6% sometimes; 34.6% always; and, 1.9% no usable response (n = 211).

<sup>26</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q21: I tend to find a trust mark to be more authoritative if it has been accredited or endorsed by a public regulator? 8.5% never; 14.7% hardly; 43.6% sometimes; 32.2% always; and, 0.9% no usable response (n = 211).

<sup>27</sup> For an extensive overview concerning the role of trust marks and trust mark organisations in e-commerce refer to [58]. For examples of studies examining the use of web assurance seals see [49], [51], [52], [50] and [47].

accredited or endorsed by a public regulator. This would be a useful area to take forward as further research.

However, to what extent end users fully appraise (e.g. read the small-print), acknowledge and understand legal signposts is unclear.<sup>28</sup> Despite the fact this is only one very specific legal cue, the majority (65.9%) of the segmentation questionnaire respondents have not read Google's privacy policy.<sup>29</sup> As Google is widely used, this lack of engagement is intriguing. Furthermore, there is hardly any difference between the mean responses for EU and non-EU respondents:

COMPARING EU AND NON-EU RESPONSES		
Segmentation question	EU mean response	Non-EU mean response
Q49: Have you ever read Google's privacy policy? <sup>a</sup>	1.72	1.70

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-2 (Statements, 1= yes; 2=no).

In addition to this, while a small majority (47.9%) of respondents feel able to understand their rights and duties as described by the terms of the application provider, a sizeable minority (32.7%) of respondents feel unable to understand these rights and duties.<sup>30</sup> However, a distinction also needs to be drawn between: (a) those individuals who both perceive themselves to be legally-aware and have actual legal-awareness i.e. sufficient legal understanding; and, (b) those individuals who perceive themselves to be legally-aware, but in fact have insufficient legal understanding. It is unclear what number of segmentation questionnaire respondents would fall in these two categories. Furthermore, there is a degree of variance between the mean responses by EU and non-EU respondents to Q22:

COMPARING EU AND NON-EU RESPONSES		
Segmentation question	EU mean response	Non-EU mean response
Q22: When using an online application, do you feel you are able to understand your rights and duties as described by the Terms of the application provider? <sup>b</sup>	2.69	3.52

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-5 (Statements, 1=strongly disagree; 2=disagree; 3=neither dis/agree; 4=agree; 5=strongly agree).

<sup>28</sup> For example, see [48] which focuses on the reasons why people do (not) read online privacy notices.

<sup>29</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q49: Have you ever read Google's privacy policy? 30.3% yes; 65.9% no; and, 3.8% no usable response (n = 211).

<sup>30</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q22: When using an online application, do you feel you are able to understand your rights and duties as described by the Terms of the application provider? 12.3% strongly disagreed; 20.4% disagreed; 19.4% neither dis/agreed; 40.3% agreed; and, 7.6% strongly agreed (n = 211).

An inference could be that websites outside the EU, particularly in the USA have more legal cues (i.e. trust marks) and therefore these respondents feel they are able to understand their rights better as they are (potentially) more informed.<sup>31</sup> Once again, more information is required, and this could be a useful area to take forward as further research.

Furthermore, it appears that the respondents look for certain types of legal information. While most respondents sometimes (40.3%) look for any information about complaint procedures in the case of problems,<sup>32</sup> the majority of respondents either never (32.7%) or hardly (32.2%) look for information about laws that are applicable with regard to their interaction with an organisation.<sup>33</sup> There was a mixed response to Q18: in general, do you look for any information about who is liable in case of problems? The most popular answer was sometimes (35.1%); however, a small majority of respondents hardly (25.1%) or never (26.1%) look for who is liable.<sup>34</sup> Therefore, it would be interesting to further evaluate what types of legal signposts people find most useful; in order for ICT platform providers to tailor more "popular" legal cues to include other relevant legal information, which an end user might not initially consider. There is little difference between the MDVs for Q17 and Q18. However, there is a slight variance between the EU and non-EU mean responses for Q15:

COMPARING EU AND NON-EU RESPONSES		
Segmentation questions	EU mean response	Non-EU mean response
Q15: In general, do you look for information about laws that are applicable with regard to your interaction with the organization? <sup>a</sup>	1.84	2.39
Q17: In general, do you look for any information about complaint procedures in case of problems? <sup>a</sup>	2.50	2.58
Q18: In general, do you look for any information about who is liable in case of problems? <sup>a</sup>	2.26	2.47

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-4 (Statements, 1=never; 2=hardly; 3=sometimes; 4=always).

<sup>31</sup> ICT platforms providers need to consider the clarity and content of their legal information and guarantees. For instance see [44] where a number of privacy policies – including those published by LinkedIn and Twitter – were ranked based on the level of plain language used and their overall presentation.

<sup>32</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q17: In general, do you look for any information about complaint procedures in case of problems? 19.9% never; 21.8% hardly; 40.3% sometimes; 16.6% always; and, 1.4% no usable response (n = 211).

<sup>33</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q15: In general, do you look for information about laws that are applicable with regard to your interaction with the organization? 32.7% never; 32.2% hardly; 26.1% sometimes; and, 9% always (n = 211).

<sup>34</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q18: In general, do you look for any information about who is liable in case of problems? 26.1% never; 25.1% hardly; 35.1% sometimes; and, 13.7% always (n = 211).

Legal signposts are not only able to indicate trustworthiness to end users, but also untrustworthiness e.g. by notifying end users of a data breach incident. Two such untrustworthiness scenarios are highlighted in the segmentation questionnaire: (a) data misuse, and (b) unauthorised access. However, from the segmentation questionnaire results, it appears that the methods used to signal untrustworthiness require greater strengthening. First, the majority (62.1%) of respondents feel unable to detect when their personal data has been misused; in comparison to a minority (20.4%) of respondents who feel able to detect misuse.<sup>35</sup> Second, the majority (62.5%) of respondents feel unable to detect when a third party has gained access to an application without authorisation; in contrast to a minority (22.6%) of respondents who feel able to detect this unauthorised access.<sup>36</sup> One inference is that while the majority of end users appear to be aware of and utilise legal trustworthiness cues, there is more limited signalling of untrustworthiness (i.e. data misuse and unauthorised access). Moreover, there is only very minor differences between the mean responses by EU and non-EU respondents to Q23 and Q24:

COMPARING EU AND NON-EU RESPONSES		
Segmentation questions	EU mean response	Non-EU mean response
Q23: When using an online application, do you feel you are able to detect when your personal information has been misused? <sup>a</sup>	2.23	2.62
Q24: When using an online application, do you feel you are able to detect when a third party has gained access to the application without authorisation? <sup>a</sup>	2.22	2.58

Source: OPTET ACME Search Engine Survey, N=211

<sup>a</sup> Values range from 1-5 (Statements, 1=strongly disagree; 2=disagree; 3=neither dis/agree; 4=agree; 5=strongly agree).

In some instances, mandatory notification of untrustworthiness is a legal requirement; although this is currently limited. For instance, under EU law, the revised e-Privacy Directive (2009/136/EC) [42] imposes a personal data breach notification requirement on the electronics communication sector; this legal obligation is further clarified by the European Commission Regulation (EU) No. 611/2013.<sup>37</sup> Furthermore, Article 31 of the proposed European General Data Protection Regulation (GDPR) [40]

<sup>35</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q23: When using an online application, do you feel you are able to detect when your personal information has been misused? 20.9% strongly disagreed; 41.2% disagreed; 17.5% neither dis/agreed; 15.2% agreed; and, 5.2% strongly agreed (n = 211).

<sup>36</sup> OPTET WP2 Segmentation Questionnaire 2015 – Q24: When using an online application, do you feel you are able to detect when a third party has gained access to the application without authorisation? 27.0% strongly disagreed; 35.5% disagreed; 14.7% neither dis/agreed; 16.1% agreed; and, 6.6% strongly agreed (n = 211).

<sup>37</sup> For more information about European legal notification requirements for personal data breaches see [56].

is set to establish a legal obligation where a controller<sup>38</sup> must notify the supervisory authority<sup>39</sup> in respect of personal data breaches.<sup>40</sup>

Trustworthiness indicators are not enough for end users to make fully informed decisions. Moreover, it is not possible to completely eradicate all untrustworthy practices. Therefore, it appears that there needs to be greater transparency over untrustworthiness. This could potentially enhance the reputation of an organisation by demonstrating that it takes data breaches seriously through showcasing how it has confronted these challenges (e.g. via more robust organisational and/or technical measures). Furthermore, it must be noted that just because there has been a data breach, it does not mean people will no longer use that platform. In consequence, a key grey area for ICT providers is the extent in which they should be more transparent with regard to untrustworthiness. The overarching question is: how should legal signalling methods be further strengthened in order to better-notify end users about untrustworthiness?

In summary, the interplay between the law and trust is yet again unclear.<sup>41</sup> From the responses to the segmentation questionnaire, there is a perception that existing laws are not fit for purpose and end users seem largely unable to uncover untrustworthy events. Furthermore, it is uncertain to what extent the respondents are legally-aware (i.e. fully understand legal requirements) even where they believe the law is important for trust. Rather than sole reliance on subjective knowledge or overlooking legal issues entirely, a better approach would be to: (1) objectively identify relevant trustworthiness attributes; (2) force designers and developers of ICT platforms to embed these attributes at the earliest point in time; (3) make sure data controllers using these platforms adopt adequate technical and organisational measures to ensure these trustworthiness attributes are effective; and, (4) finally signal them to end-users at a later stage through legal cues such as trust marks (which appear to be well-received by the majority of respondents). This high-level (un)trustworthiness signalling could therefore reduce information asymmetry between end users and ICT platform providers by strengthening legal awareness and ultimately facilitating more informed decision-making. In consequence, it seems that trustworthiness should be at the core of a legal strategy in an overall approach to trust optimisation; the interplay between the law and trustworthiness is therefore explored in OPTET deliverable D2.5.

As a final point, there is only the smallest difference between the majority of EU and non-EU mean responses. However, there are four notable exceptions. First, the mean for non-EU responses is higher than EU responses to Q20: I tend to be more cautious about visiting websites or using online services, systems or applications which have no trust marks. Second, the mean for non-EU responses

---

<sup>38</sup> Article 4(5) of the GDPR (European Commission draft text [40]) defines 'controller': "[...] means the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes, conditions and means of the processing of personal data; where the purposes, conditions and means of processing are determined by Union law or Member State law, the controller or the specific criteria for his nomination may be designated by Union law or by Member State law".

<sup>39</sup> Article 4(19) of the GDPR (European Commission draft text [40]) defines 'supervisory authority': "[...] means a public authority which is established by a Member State in accordance with Article 46".

<sup>40</sup> Article 31(1) of the GDPR states (European Commission draft text [40]): "In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 24 hours after having become aware of it, notify the personal data breach to the supervisory authority. The notification to the supervisory authority shall be accompanied by a reasoned justification in cases where it is not made within 24 hours." Moreover, mandatory notification is not only a European legal requirement; for instance, forty-seven US states compel "businesses to notify individuals when data security breaches compromise their personal information. [62, p. 8]"

<sup>41</sup> For more information about the interplay between law and trust refer to the OPTET D2.1 Report [13].



is again higher than EU responses to Q21: I tend to find a trust mark to be more authoritative if it has been accredited or endorsed by a public regulator. This may be explained by the more matured trust mark ecosystem in place outside of Europe; in particular respondents from the USA (who accounted for 89/106 of the overall non-EU respondents) should be more familiar with trust marks as they are in wider use. Third, the mean for non-EU responses is higher than EU responses to Q15: In general, do you look for information about laws that are applicable with regard to your interaction with the organization? Fourth, the mean for non-EU responses is again higher than EU responses to Q22: when using an online application, do you feel you are able to understand your rights and duties as described by the Terms of the application provider? As an area of future research interest, it would be useful to undertake a cross-comparative study to examine the different types and levels of legal information and guarantees available within EU and non-EU countries and how this impacts on perceived legal-awareness and trust.

If we then take another look at the four segment solution in the context of our trust-privacy indicators explored above in conjunction to our trust scales of trust stance, trust seeking behaviour and trust competences, we can elaborate on the four segment solution. Generally speaking, the evolution of the trust level as regards the privacy metric appears to be coherent with our segmentation approach.

The user experience for the *High Trust* (HT) segment can be characterized by a so-called high level trust stance. This refers to a tendency of an overall high trust level vis-a-vis (socio-)technical systems accompanied by a relative trust in future (legal) measures and other technical safe-guarding. Also, several trust seeking behaviours can be seen, while a relatively high "aversion" can be distilled in providing certain bits of personal information and having a rather high interest in that such information is not (publicly) available online underpinned by their available competences to cognitively assess the trustworthiness of online applications and services. Considering the privacy metric independently, The HT segment tends to have the highest level of trust over time.

The *Highly active trust seeking* (HATS) segment has been described by a high level of trust seeking behaviour beyond the mere scanning of trustworthiness cues. Also, individuals seem to have a higher interest to inform themselves about procedures in case of personal data provision and availability, which may impact their online behaviour to a greater extent than for others. This confirms and strengthens the capacity of possessing a certain competence level that facilitates the assessment of trustworthiness and cues. And, increases the likelihood to address or act upon such events.

Consistent with earlier findings, the user experience for the *Medium active trust seeking* (MATS) segment remains relatively similar to the HATS. Trust seeking behaviour, however, is less apparent as they seem less preoccupied with (possible risks of) providing sensitive information, or whether this kind of information is publicly un/available. In conjunction with this, they are also relatively less inclined to possibly change their online behaviour. While the drivers for trust seeking behaviour, such as a relatively low trust stance, can still be detected as well as competences to assess trustworthiness, the motivation to look for trustworthiness cues is still relatively low. Considering the privacy independently, the levels of trust of the HATS and MATS segments are very similar.

Lastly, the *Ambivalent* (A) segment shows a perceived "ambivalence" to assess the trustworthiness of online applications and services. The 'ambivalent' nature of user experience can be explained by a difficulty to cognitively assess trustworthiness and a certain need to trust (according to 'basic heuristics'), such as perceived feelings of loss of control over one's personal information, how personal information is handled and whether current laws and practices are adequate enough to offer protection. Unsurprisingly, as regards the privacy metric, the A segment presents the lowest level of trust over time.

In order to avoid, or to lower the omnipresence of cautious and other negative feelings (so-called 'forced trust') we are led to consider trustworthiness indicators based on the experience of others ('referrals'), as the main source of 'trustworthiness information' that is accessible and underpinning the outcome of assessing trustworthiness. The A segment showed the lowest competence levels as well as trust stance, suggesting users are likely to be more motivated to look for trust cues, while, in fact, they actually may not particularly do so.

Based on our work conducted throughout the OPTET project and the experiment presented in this deliverable, we can conclude that the four segment solution is relevant and valid, providing and strengthening (existing) research into social, legal and economic drivers of trust. Moreover, the primary socio-legal findings are:

### Key socio-legal highlights from the questionnaire

- (1) A distinction needs to be made between: (a) those **individuals who have actual legal-awareness** i.e. sufficient legal understanding; and, (b) those **individuals who perceive themselves to be legally-aware**, but in fact have insufficient legal understanding;
- (2) The **reputation** of an ICT platform provider appears to be an **important implied signpost** for trustworthiness;
- (3) **Trust marks** appear to be **important explicit cues** for trustworthiness, and seem to be **more authoritative when accredited or endorsed** by public regulators;
- (4) To what extent end users fully **appraise, acknowledge and understand** legal cues is unclear;
- (5) The extent in which more "**popular**" **legal cues can be tailored** to include important, but less searched for, legal information needs further examination;
- (6) More attention needs to be given to **signalling untrustworthiness** to end users; and **Trustworthiness** should be at the **core of a legal strategy** in an overall approach to trust optimisation.



## 5. Summary and Future Work

---

We have presented a new, enhanced OPTET model stack to facilitate the design of trustworthy systems from the design stage and for use during runtime to support automated threat detection based on monitored assets' misbehaviours. This work can be extended by for instance modelling more patterns, threats, misbehaviours and controls. Also the SSD GE can be optimised in various ways (like decreasing compilation time or implementing usability enhancements) to promote its usage among system designers. Our next target beyond OPTET is to apply the SSD in the context of 5G networks where the dynamics of such systems cause security challenges that can be addressed using a risk based approach based on our semantic models. A new generic model for 5G networks needs to be developed capturing the domain expertise including an asset model focusing on the network assets, accompanying threats and control strategies.

The experiment testing the validity of the trust model presented in this deliverable has highlighted its relevance to the social, legal and economic domains. The results are cues for the creation of legal strategies based on the trust of users and their perception thereof. Using this information, it can improve applications such as targeted advertising and help raise awareness of trustworthiness and the legal implications that come with it.

## 6. References

---

- [1] O. C. «D2.3: Socio-economic evaluation of trust and trustworthiness,» 2014.
- [2] O. C. «OPTET Wiki GE Catalogue: TME,» [Online]. Available: <http://eu-sites.atc.gr/optetwiki/index.php/TrustMetricEstimatorCatalogue>. [Consultato il giorno 30 10 2015].
- [3] S. C. f. B. I. Research. [Online]. Available: <http://protege.stanford.edu/>. [Consultato il giorno 13 30 2015].
- [4] TopQuadrant. [Online]. Available: <http://www.topquadrant.com/downloads/topbraid-composer-install/>. [Consultato il giorno 30 10 2015].
- [5] D. o. C. S. U. o. O. Information Systems Group. [Online]. Available: <http://www.hermit-reasoner.com/>. [Consultato il giorno 30 10 2015].
- [6] C. Inc. [Online]. Available: <https://github.com/complexible/pellet>. [Consultato il giorno 30 10 2015].
- [7] O. C. «D7.2,» 2015.
- [8] O. C. «D7.3,» 2015.
- [9] O. C. «D8.4.1: AAL and CCM use case implementations (first release),» 2014.
- [10] O. C. «System and Threat Modelling Questionnaire,» 2015. [Online]. Available: <http://tinyurl.com/modelQuestions>. [Consultato il giorno 30 10 2015].
- [11] Amazon. [Online]. Available: <https://www.mturk.com/mturk/welcome>. [Consultato il giorno 30 10 2015].
- [12] O. C. «D2.2: Socio-economic models for trust and trustworthiness evaluation,» 2014.
- [13] OPTET Consortium, «D2.1: Socio-economic requirements for trust and trustworthiness,» OPTET – 317631, FP7-ICT-2011-8, 2014.
- [14] O. C. «D8.5: Results of the Intermediate UCA Evaluations,» 2015.
- [15] B. Jansen e A. Spink, «How We Are Searching The World Wide Web? A Comparison of nine search engine transaction logs,» *Information Processing & Management*, vol. 42, n. 1, pp. 248-263 , 2006.
- [16] A. Langville e C. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton: Princeton University Press. .

- [17] S. Buchegger e J.-Y. L. Boudec, «A Robust Reputation System for P2P and Mobile Ad hoc Networks,» 2004.
- [18] S. van der Graaf, «Imaginaires of Ownership; the logic of participation in the moral economy of 3D software design,» *Telematics and Informatics*, Vol. 32, number 2, n. special issue Ethics in the information Society, p. 400–408, 2015.
- [19] R. POSNER, «The right of privacy,» *Georgia Law Review*, vol. 3, n. 12, pp. 393-422, 1978.
- [20] S. (G.J.), «An introduction to privacy in economics and politics,» *The Journal of Legal Studies*, vol. 4, n. 9, pp. 623-644, 1980.
- [21] R. POSNER, «The economics of privacy,» *The American Economic Review*, vol. 2, n. 71, pp. 405-409, 1981.
- [22] J. Pierson, «Online privacy in social media: a conceptual exploration of empowerment and vulnerability,» *Communications & Strategies (Digiworld Economic Journal)*, n. 4thQ (88), pp. 99-120, 2012.
- [23] M. Deuze, «Media life,» *Cambridge: Polity*, 2012.
- [24] R. Mansell, *Imagining the Internet: Communication, Innovation and Governance*, Oxford University Press, 2012.
- [25] R. Silverstone, *Media and morality: on the rise of the mediapolis*, Cambridge: Polity press, 2006.
- [26] Z. Bauman, «Liquid modernity,» *Cambridge*, 2000.
- [27] B. & H. C. Wellman, «The Internet in everyday life,» *Oxford: Blackwell*, 2002.
- [28] S. Orgad, «The interrelations between online and offline: questions, issues, and implications,» in *R. Mansell & C. Avgerou & D. Quah & R. Silverstone (Eds.) The Oxford handbook of information and communication technologies*, Oxford: OUP, 2007.
- [29] N. Couldry, «The necessary future of the audience ... and how to research it,» in *V. Nightingale (Ed.) The handbook of media audiences*, Malden: Wiley-Blackwell, 2011.
- [30] Z. & L. D. Bauman, *Liquid surveillance*, Cambridge: Polity Press, 2013.
- [31] B. Berendt, «Learning Paths and Metacognition,» Berlin, Raabe Fachverlag für Wissenschaftsinformation, 2006, pp. 1-34.
- [32] P. Lehrer, «Biofeedback: A practitioner's guide,» in *Applied Psychophysiology and Biofeedback 21*, 1996, pp. 199-202.

- [33] M. A. e. a. Moreno, «Display of health risk behaviors on myspace by adolescents: prevalence and associations,» in *Archives of pediatrics adolescent medicine* 163(1), 2009, pp. 27- 34.
- [34] L. a. A. F. Wathieu, «An Empirical Approach to Understanding Privacy,» *Proceedings of the Fourth Workshop on the Economics of Information Security (WEIS '05)*, 2005.
- [35] E. A. a. L. F. B. Huberman, «Valuating Privacy,» *Proceedings of the Workshop on the Economics of Information Security (WEIS '06)*, 2006.
- [36] H.-H. T. S.-Y. L. K.-L. Hui, «The Value of Privacy Assurance: An Exploratory Field,» *MIS Quarterly*, n. 31(1), pp. 19-33, 2007.
- [37] A. a. J. G. Acquisti, «Privacy and Rationality in Decision Making,» *IEEE Security and Privacy*, n. 3(1), pp. 26-33, 2005.
- [38] H. F. W. a. G. W. Varian, «The demographics of the do-not-call list,» *IEEE Security & Privacy*, n. 3(1), pp. 34-39, 2005.
- [39] A. J. L. & L. G. Acquisti, «What is Privacy Worth?,» *The Journal of Legal Studies*, n. 42(2), pp. 249-274, 2013.
- [40] «Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) /\* COM/2012/011 final - 2012,» 25 January 2015. [Online]. Available: <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52012PC0011>. [Consultato il giorno 17 August 2015].
- [41] G. Farmer, «Information Commissioner's Office (ICO) Blog: ICO Privacy Seal,» 28 January 2015. [Online]. Available: <https://iconewsblog.wordpress.com/2015/01/28/what-you-need-to-know-about-ico-privacy-seals/>. [Consultato il giorno 21 September 2015].
- [42] European Parliament and of the Council, «Revised e-Privacy Directive 2009/136/EC,» 25 November 2009. [Online]. Available: <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1443018436127&uri=CELEX:32009L0136>. [Consultato il giorno 23 September 2015].
- [43] S. Wang, S. E. Beatty e W. Foxx, «Signaling the trustworthiness of small online retailers,» *Journal of Interactive Marketing*, vol. 18, n. 1, p. 53–69, 2004.
- [44] K. Steinmetz, «These Companies Have the Best (And Worst) Privacy Policies,» *TIME*, 6 August 2015.
- [45] M. Spence, «Signaling in Retrospect and the Informational Structure of Markets,» *The American Economic Review*, vol. 92, n. 3, pp. 434-459, 2002.
- [46] Y. Pan e G. M. Zinkhan , «Exploring the impact of online privacy disclosures on consumer trust,» *Journal of Retailing*, vol. 82, n. 4, p. 331–338, 2006.

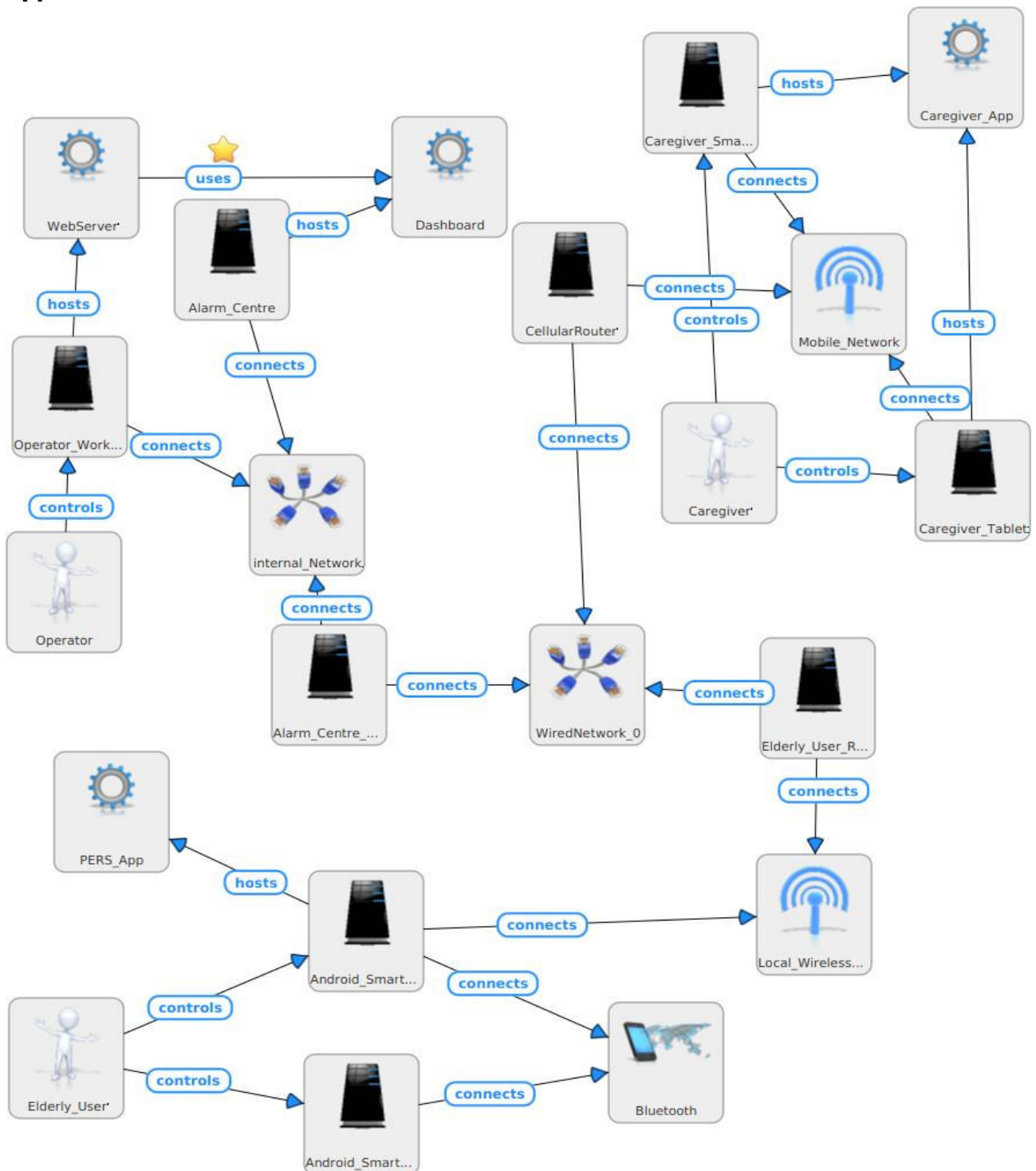
- [47] K. Özpolat e W. Jank, «Koray Özpolat and Wolfgang Jank, Getting the most out of third party trust seals: An empirical analysis,» *Decision Support Systems*, vol. 73, pp. 47-56, 2015.
- [48] G. R. Milne e M. J. Culnan , «Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices,» *Journal of Interactive Marketing*, vol. 18, n. 3, p. 15–29, 2004.
- [49] V. Lala, V. Arnold, S. G. Sutton e L. Guan, «The impact of relative information quality of e-commerce assurance seals on Internet purchasing behavior,» *International Journal of Accounting Information Systems*, vol. 3, n. 4, p. 237–253, 2002.
- [50] D. J. Kim, C. Steinfield e Y.-J. Lai, «Revisiting the role of web assurance seals in business-to-consumer electronic commerce,» *Decision Support Systems*, vol. 44, n. 4, p. 1000–1015, 2008.
- [51] S. E. Kaplan e R. J. Nieschwietz , «A Web assurance services model of trust for B2C e-commerce,» *International Journal of Accounting Information Systems*, vol. 4, n. 2 , p. 95–114, 2003.
- [52] X. Hu, G. Wu, Y. Wu e H. Zhang, «The effects of Web assurance seals on consumers' initial trust in an online vendor: A functional perspective,» *Decision Support Systems*, vol. 48, n. 2, p. 407–418, 2010.
- [53] G. Farmer, «Information Commissioner's Blog: What's the latest on the ICO privacy seals?,» 28 August 2015. [Online]. Available: <https://iconewsblog.wordpress.com/2015/08/28/whats-the-latest-on-the-ico-privacy-seals/>. [Consultato il giorno 23 September 2015].
- [54] B. L. Connelly, S. T. Certo, R. D. Ireland e C. R. Reutzel, «Signaling Theory: A Review and Assessment,» *Journal of Management*, vol. 37, n. 1, pp. 39-67, 2011.
- [55] J. Cline, «Will the EU Privacy Reform Boost Privacy Seal Adoption?,» *International Association of Privacy Professionals (IAPP)*, 13 April 2012.
- [56] D. Brennan, «New rules on breach notification by telecoms and ISPs - clarity at last?,» *Privacy & Data Protection*, vol. 14, n. 1, pp. 4-6, 2013.
- [57] F. Belanger, J. S. Hiller e W. J. Smith, «Trustworthiness in electronic commerce: the role of privacy, security, and site attributes,» *The Journal of Strategic Information Systems*, vol. 11, n. 3–4 , p. 245–270, 2002.
- [58] P. Balboni, *Trustmarks in E-Commerce: The Value of Web Seals and the Liability of their Providers*, The Hague: T.M.C Asser Press, 2009.
- [59] E. Auriol e S. Schilizzi, «Quality signaling through certification in developing countries,» *Journal of Development Economics*, vol. 116, p. 105–121, 2015.
- [60] L. Atkinson e S. Rosenthal, «Signaling the Green Sell: The Influence of Eco-Label Source, Argument Specificity, and Product Involvement on Consumer Trust,» *Journal of Advertising*, vol. 43, n. 1, pp. 33-45, 2014 .

- [61] European Privacy Seal: EuroPriSe, «Welcome!», [Online]. Available: <https://www.european-privacy-seal.eu/EPS-en/Home>. [Consultato il giorno 19 August 2015].
- [62] «United States student data breach laws», *Criminal Lawyer*, vol. 222, pp. 8-9, 2014.
- [63] EuroPriSe, «ULD-Gütesiegel», [Online]. Available: <https://www.european-privacy-seal.eu/EPS-en/ULD-Guetesiegel>. [Consultato il giorno 21 September 2015].
- [64] European Privacy Seal: EuroPriSe, «Register of Awarded Seals», [Online]. Available: <https://www.european-privacy-seal.eu/EPS-en/Awarded-seals>. [Consultato il giorno 19 August 2015].
- [65] European eCommerce Association (EMOTA), «Press release: European Trust Mark for online shopping launched today Commissioner Jourova welcomes initiative to provide confidence to European consumer», 1 July 2015. [Online]. Available: <http://www.emota.eu/#!publications/c1351>. [Consultato il giorno 19 August 2015].
- [66] La Commission Nationale de l'Informatique et des Libertés (CNIL), «Label CNIL procédures d'audit de traitements», [Online]. Available: <http://www.cnil.fr/linstitution/labels-cnil/procedures-daudit/>. [Consultato il giorno 21 September 2015].
- [67] European Privacy Seal: EuroPriSe, «European Privacy Seal for ixquick.com - de-080001p», 13 July 2008. [Online]. Available: <https://www.european-privacy-seal.eu/EPS-en/ixquick>. [Consultato il giorno 19 August 2015].
- [68] S. A. JansenB., «How We Are Searching The World Wide Web? A Comparison of nine search engine transaction logs», *Information Processing & Management*, vol. Vol. 42(1), pp. 248-263, 2006.
- [69] A. & M. C. Langville, «Google's PageRank and Beyond:», *Princeton: Princeton University Press*, 2012.
- [70] O. C. «OPTET Wiki GE Catalogue: SSD», [Online]. Available: <http://eu-sites.atc.gr/optetwiki/index.php/SecureSystemDesignerCatalogue>. [Consultato il giorno 30 10 2015].



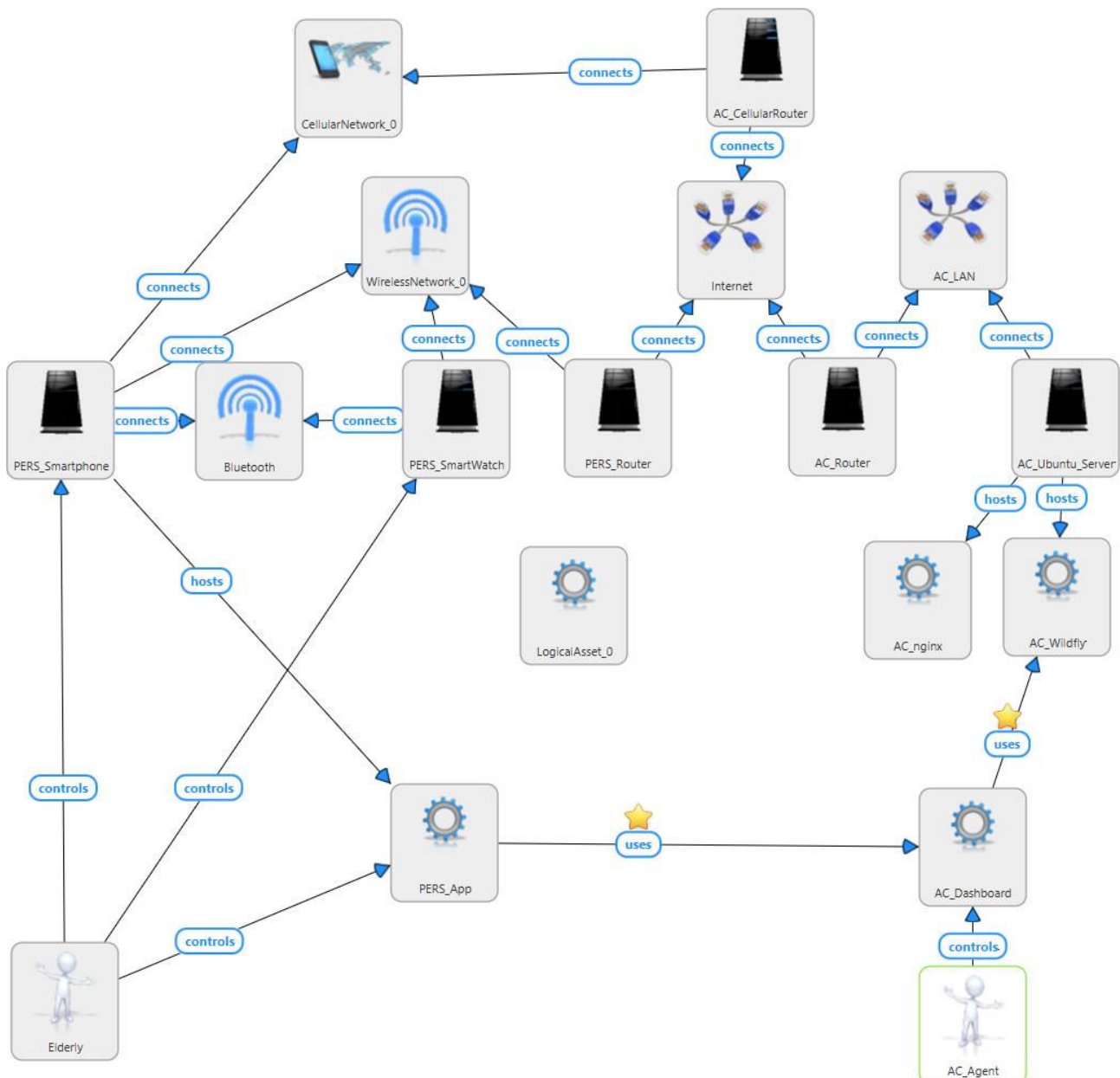
## Appendix

### Appendix I – Evaluation models

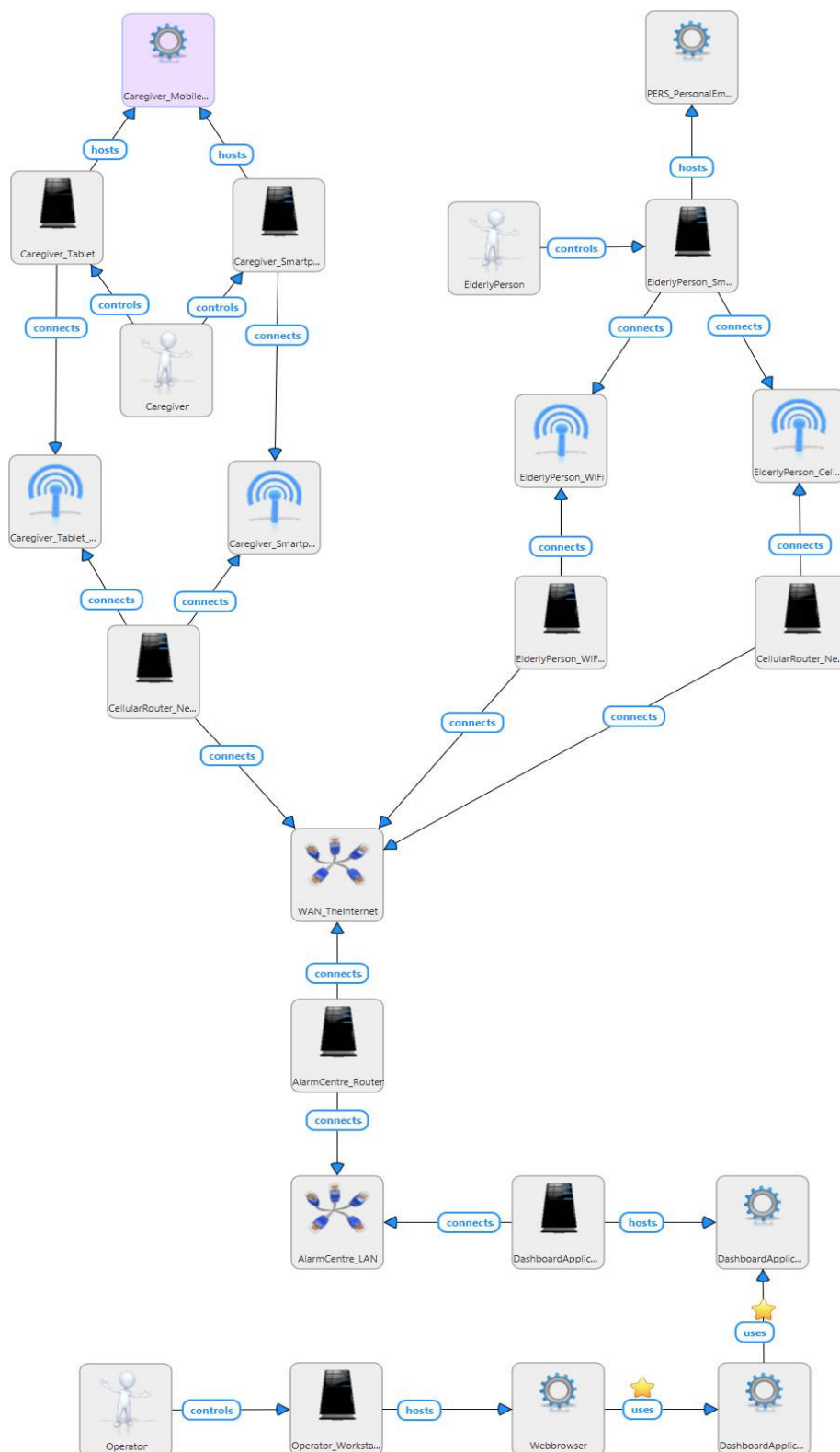


**Figure 22 – Participant 1's model**

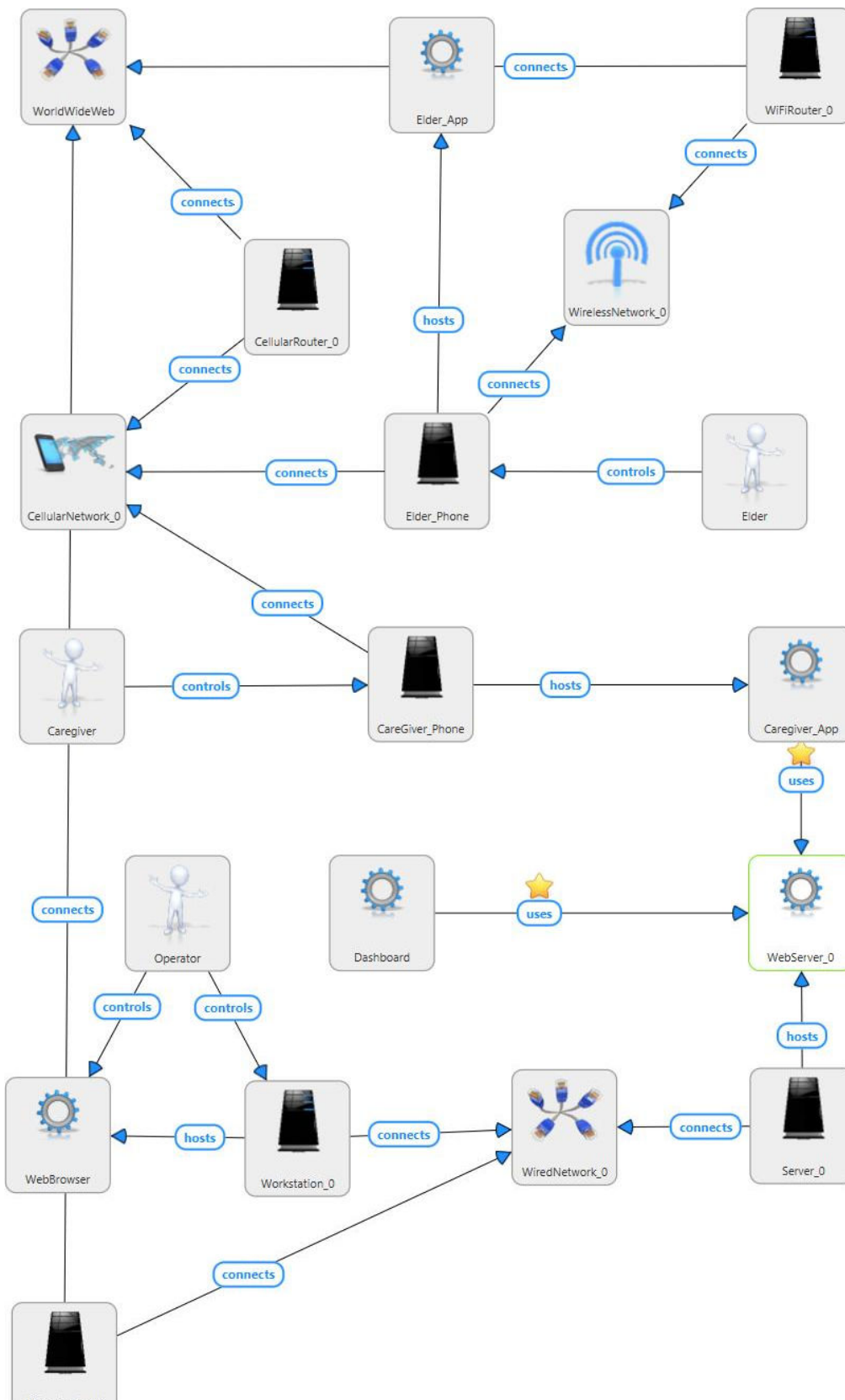




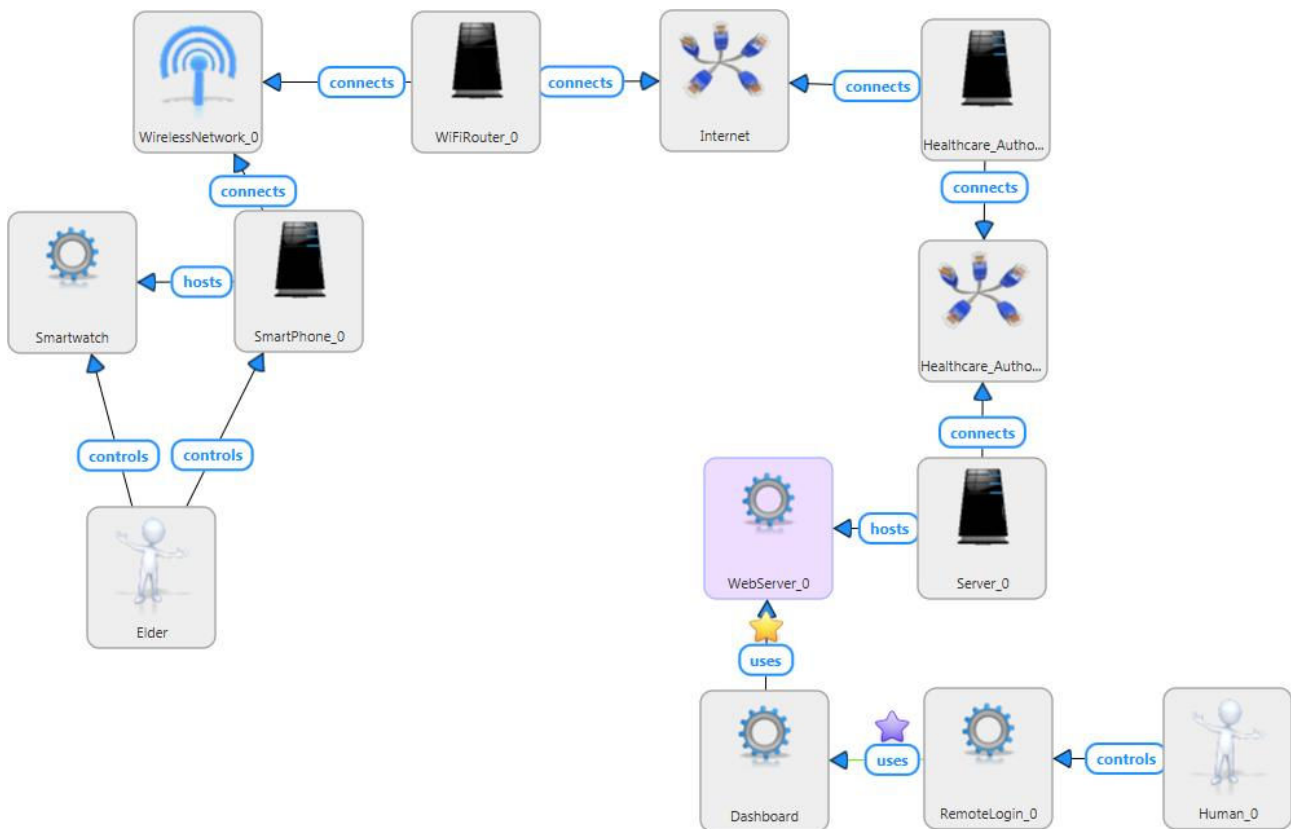
**Figure 23 - Participant 2's model**



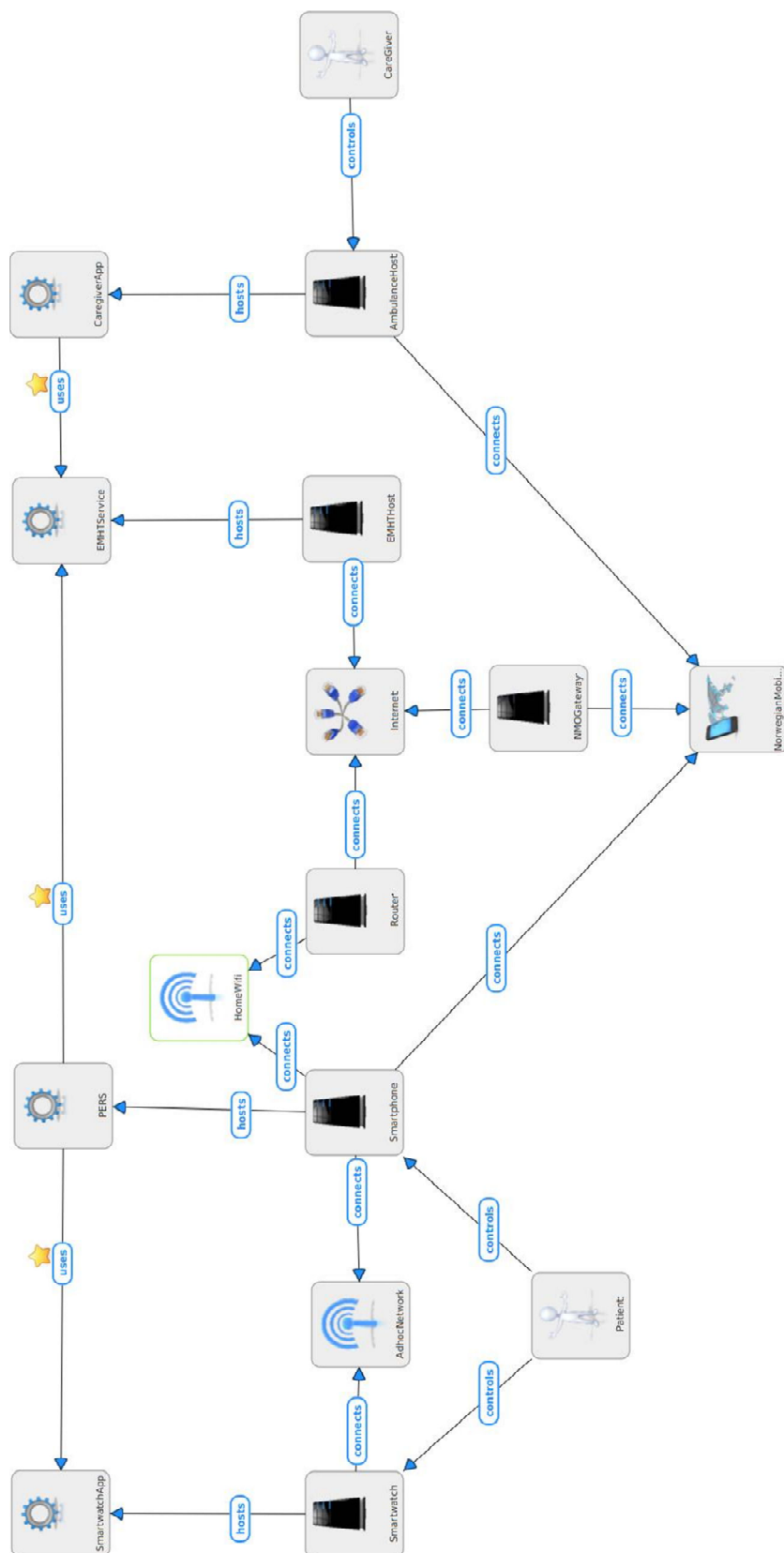
**Figure 24 - Participant 3's model**



**Figure 25 - Participant 4's model**



**Figure 26 – Participant 5's model**



**Figure 27 – OPTET reference model**

## **Appendix II - Post-questionnaire**

Q1 In your daily life, how do you tend to trust another person?

Please indicate on a scale of 1 to 5 to what extent you agree with the following statements. (1 = Strongly Disagree; 2 = Disagree; 3 = Neither Dis/Agree; 4 = Agree; 5 = Strongly Agree)

1. I usually trust a person until there is a reason not to
2. Even when the stakes are high, I still think that most people are honest in their dealings with others
3. In general, people do not really care about the well-being of others

Q2 In general, to what extent do you trust online applications and services (such as those accessible through your PC, laptop, smartphone or tablet)?

Please indicate on a scale of 1 to 5 to what extent you agree with the following statements. (1 = Strongly Disagree; 2 = Disagree; 3 = Neither Dis/Agree; 4 = Agree; 5 = Strongly Agree)

1. I easily trust online applications
2. The higher the price, the more I tend to trust the online application or service
3. I am very suspicious towards free online applications
4. I tend to trust online applications offered by public organizations more than those offered by commercial organizations

Q3 In general, how do you perceive your level of protection when using the Internet?

Please indicate on a scale of 1 to 5 to what extent you agree with the following statements. (1 = Strongly Disagree; 2 = Disagree; 3 = Neither Dis/Agree; 4 = Agree; 5 = Strongly Agree)

1. I have lost control over how my personal information is collected and used by digital applications and services
2. Organizations handle the personal information they collect about me in a proper way
3. Existing laws and organizational practices adequately protect me from online problems or risks today
4. I think that legal measures such as 'the right to be forgotten' (i.e. the ability to erase your personal information) are necessary to increase control

5. I feel confident that encryption and other technical advances online make it safe for me to use

Q4 When using an online application, do you feel you are able to...?

Please indicate on a scale of 1 to 5 to what extent you agree with the following statements. (1 = Strongly Disagree; 2 = Disagree; 3 = Neither Dis/Agree; 4 = Agree; 5 = Strongly Agree)

1. Understand your rights and duties as described by the Terms of the application provider
2. Detect when your personal information has been misused
3. Detect when a third party has gained access to the application without authorisation
4. Assess the effectiveness of available redress mechanisms to remedy any problems or harms

Q5 In general, do you look for information or (legal) guarantees when you decide to use an application or service?

Please indicate on a scale of 1 to 4 to what extent you agree with the following statements. (1 = Never; 2 = Hardly; 3 = Sometimes; 4 = Always)

1. I look for information about the reputation of the organization
2. I look for information about the (physical) location of the organization
3. I look for information about laws that are applicable with regard to my interaction with the organization
4. I look for any guarantees regarding confidentiality of the information that I provide
5. I look for any information about complaint procedures in case of problems
6. I look for any information about who is liable in case of problems
7. I look for trust marks\*\* or seals of approval when visiting a website or using an online service, system or application
8. I tend to be more cautious about visiting websites or using online services, systems or applications which have no trust marks
9. I tend to find a trust mark to be more authoritative if it has been accredited or endorsed by a public regulator

\*\* trust mark = A symbol used to represent that a website, system, application and/or service has passed a particular set of best practice criteria i.e. for quality, privacy or security.

Trust marks are also known by (but not limited to) the following terms: certification marks, authentication marks, quality assurance labels and seals of approval.



Next, we ask you a few questions about how you feel about disclosing personal information online.

Q6 What is your gender?

1. Male
2. Female

Q7 What is your year of birth? Please choose one answer from the drop box.

Q8 What is your nationality? Please choose one answer from the drop box.

Q9 What is your highest educational degree? Please choose one answer from the drop box.

1. Nursery school
2. High school
3. Bachelor's degree
4. Master's degree
5. Advanced graduate or PhD
6. Not sure

Q10 What is your living situation? Please choose one answer from the drop box.

1. Living alone
2. Living alone with children
3. Living with my partner, without children
4. Living with my partner, with children
5. Living with parents
6. Living with friend(s)
7. Other

Q11 In what city do you live?

Q12 In what city do you work?

Q13 Where did you spend your last holiday?

Q14 How sensitive do you find the information you had to reveal about your demographics (Q6 to 13)? (1 = Not Sensitive at all; 7 = Very Sensitive)

1 - 7

Q15 Do you think this information is available online about you by performing a search on, for example, Google?

1. Yes
2. No

Q16 How important is it for you that this type of information about you is not publicly available? (1 = Not Important; 7 = Very Important)

1 – 7

THE FINAL SET OF QUESTIONS CONCERN YOUR ONLINE CONSUMPTION INTERESTS.

Q17 What social media (such as Facebook, LinkedIn, Instagram, Twitter) do you use? Name your top 5.

1.

Q18 What sites (such as iTunes, Amazon, Cheaptickets, H&M, Nike) do you use to make online purchases? Name your top 5.

Q19 How do you tend to make online payments (such as credit card, debit card, PayPal, smartphone app, invoice)? Name your top 3.

Q20 How sensitive do you find the information you had to reveal about yourself (Q17-19)? (1 = Not Sensitive at all; 7 = Very Sensitive)

1 - 7

Q21 Do you think this information is available online about you by performing a search on, for example, Google?

1. Yes
2. No

Q22 How important is it for you that this type of information about you is not publicly available? (1 = Not Important; 7 = Very Important)

1 – 7

Q23 Does your awareness of the kind of data that is processed change your attitude towards your online behaviour?

1. Yes
2. No

Q24 Have you ever read the privacy policy of a search engine, such as Google?

1. Yes
2. No