Here's Johnny: a methodology for developing attacker personas.

ATZENI, A., CAMERONI, C., FAILY, S., LYLE, J. and FLÉCHAIS, I.

2011

Here's Johnny: a Methodology for Developing Attacker Personas

Andrea Atzeni, Cesare Cameroni
Dip. di Automatica e Informatica,
Politecnico di Torino
Torino, Italy
andrea.atzeni, cesare.cameroni@polito.it

Shamal Faily, John Lyle, Ivan Flechais
Department of Computer Science,
University of Oxford
Oxford, UK
shamal.faily, john.lyle, ivan.flechais@comlab.ox.ac.uk

***Abstract*—The adversarial element is an intrinsic part of the design of secure systems, but our assumptions about attackers and threat is often limited or stereotypical. Although there has been previous work on applying User-Centered Design on Persona development to build personas for possible attackers, such work is only speculative and fails to build upon recent research. This paper presents an approach for developing Attacker Personas which is both grounded and validated by structured data about attackers. We describe a case study example where the personas were developed and used to support the development of a Context of Use description for the EU FP7 *webinos* project.**

***Keywords*-Persona, Attacker, Toulmin Model, Attack Tree;**

I. Introduction

A distinguishing feature of secure software engineering is the need for design activities to factor in the adversarial element. Techniques such as Attack Trees [1] and Misuse Cases [2] were devised to encourage such thinking by re-framing system design from the perspective of an attacker. In both cases, the steps carried out by an attacker are modelled and form the basis of threats. Attack Trees model the different paths that an attacker might take when carrying out an attack, while Misuse Cases can be used both to describe how Use Cases can be threatened, and how Use Cases might mitigate particular Misuse Cases.

A hitherto undiscussed weakness of threat modelling approaches is that we do not know, with the certainty that we would like, precisely who these attackers are. To understand why this might be important, we need to consider the process for building such models. Sindre and Opdahl [2] suggest that data contributing to these models can be derived by brainstorming activities guided by a security expert; they argue that this mirrors the way attackers think. However, there is little certainty that the abstractions that we hold about a system are necessarily the same as those held by an attacker. The perceptions held by attackers may be coloured by their own experiences, motivations, and capabilities; these allow attackers to see system vulnerabilities that we, as designers, are unable to see. With this in mind, Schneier acknowledges that, for attack tree modelling to work, attack trees need to be married with knowledge about an attacker [1]. Without grounded information about attackers, we may fall foul of our own stereotypes about who a system's attackers are, and what capabilities they have. For example, attempting to mitigate against attacks precipitated by a *super-hacker* may lead to design decisions which are overly cautious in light of the risks the system does actually face.

In the "Why Johnny can't encrypt" study on usability problems with encryption software, Whitten and Tygar [3] suggest that security needs a usability standard that is different from those applied to 'general consumer software'. Sasse et al. [4] rebut this claim by demonstrating that with proper application of standard HCI methods, usability issues in security can be addressed. To better focus on the goals and characteristics of adversaries, Steele & Jia [5] proposed using User-Centered Design techniques to develop *personas* describing the archetypical behaviour of possible attackers. Two key characteristics of such techniques are an early focus on users and the tasks they carry out, and the empirical measurement of activities carried out by users [6]. Although this proposition is intriguing, it remains speculative; we are unaware of existing work attempting to develop such attacker models based on these principles.

In this paper, we show that in addition to helping with usability issues, HCI techniques can help improve system security. We demonstrate this by presenting a methodology for developing *Attacker Personas*: behavioural specifications of archetypical attackers of a system. In section II we describe the related work that the Attacker Persona methodology builds upon before presenting the methodology in section III. In section IV, we present an illustrative example of how Attacker Personas were developed and applied to develop a Context of Use description for the EU FP7 *webinos* project.

II. Related Work

A. Threat and Attacker Taxonomies

Risk analysis and management plays a central role in designing information security in medium to large systems. A common pre-requisite in risk-based approaches to secure system design, such as [7] is the elicitation and categorisation of *threats*; these cause unwanted incidents leading to harm to a system or organisation [8]. These approaches are somewhat unclear about how to characterise threats with qualitative and quantitative data. Even with

rich taxonomies of possible threats, these need to be customised to the environments where they are to be applied. For example, the Common Attack Pattern Enumeration and Classification (CAPEC) initiative has described *attack patterns* describing methods of exploiting software from an attacker's perspective [9]. One such pattern is a Reflection Attack which describes how an attacker exploits a security protocol weakness to obtain unauthorised access to a server. Assuming that an attacker satisfied the prerequisites for this attack, CAPEC claims that the likelihood of this attack is High, but this rating may be meaningless. For an unskilled attacker, the pre-requisites may be sufficiently daunting that the likelihood of a mounting a successful attack might be low. Similarly, a skilled attacker may not be interested in exploiting the system being designed, thereby making the attack likelihood equally low.

There are two possible sources of data from which detailed models can be derived. The first of these is an appropriate taxonomy of possible attackers. Unfortunately, the lack of data upon which to ground a taxonomy means there has been comparatively little work in this area. One notable exception has been work carried out by the Open Web Application Security Project (OWASP) towards the modelling of threat agents [10]. OWASP defines a Threat Agent as an individual or group that manifests a threat. Several categories of human threat agent have been proposed; these include employees, unintentional and intentional human agents. In particular, OWASP defines specific attacker agents for corporate *intranet* attackers, and external *Internet* attackers such as script kiddies and professional crackers.

The OWASP definition for threat agents includes information about the agent's capabilities, intentions, and past activities, but there is little information about these on the OWASP portal even though this information is essential for grounding attacker models. A more detailed breakdown of a threat agent has been proposed by Jones & Ashenden; they have identified several factors that influence attackers behind malicious threats; these include motivation, the capability of an individual, opportunities for carrying out an attack, catalysts that cause an agent to select a target, and system related factors such as vulnerabilities and high-value assets [11].

Eliciting these factors for different threat agents may also lead to the elicitation of possible attacker characterisation. For this reason, it may be useful, as [5] suggests, to consider the usefulness of User-Centered Design artifacts for embodying these characteristics.

B. Personas and Assumption Personas

Personas are behavioural specifications of archetypical users. These were first introduced by Cooper [12] as a means of dealing with programmer biases arising from the word user. These biases lead to programmers introducing assumptions, bending and stretching the supposed user to meet these needs; Cooper called this phenomena "designing for the elastic user". Cooper's solution was to design for a single user representing the target segment of the system or product being designed. This approach brings two benefits. First, designers only have to focus on those requirements necessary to keep the target persona happy. Second, the idiosyncratic detail associated with personas makes them communicative to a variety of stakeholders. Personas were designed to be data-driven artifacts; they are grounded in empirical data collected about representative users who carry out work within their normal contexts of use. For the purpose of creating attacker personas, it is perhaps unsurprising that we cannot easily elicit empirical data directly from attackers. Secondary sources of publicly accessible *Open Source Intelligence* data on the Internet may, however, act as a suitable proxy given that the purpose of using personas is to articulate attacker viewpoints or explore possible attack ideas.

Assumption Personas are persona sketches created to articulate existing assumptions about a user population [13]. Rather than being based on observed data of prospective users, these are grounded in assumptions that contributors hold about users and the context of investigation. These assumptions may be derived from interpreted or misinterpreted experiences, and coloured by individual and organisational values. Assumption Personas can help people to see the value of personas in design, and how different assumptions can shape these. As a result, when exposed, they guide subsequent analysis or data collection for data-driven personas.

C. Arguing Assumption Personas

Recent work by Faily and Fléchais [14] demonstrated how approaches from Design Rationale can be used to structure the contributions made by assumptions towards personas. They propose associating the narrative structure pertaining to personas with a number of *characteristics* acting as propositions about the persona. As such, the persona is written to satisfy these characteristics, which are analogous to claims that might be made as part of an argument. These characteristics may be backed up as *references*: propositions which act as the grounds of evidence. References may also act as a *warrant* or a *rebuttal* to a characteristic. A warrant is a rule of inference describing how grounds contribute to the characteristic. The origin of a warrants assumption is the backing knowledge for believing the claim. A rebuttal challenges the validity of the claim. Each characteristic is also supplemented by a modal qualifier indicates the degree of certainty about the claim. This persona argumentation model is based on Toulmin's model of argumentation [15]

This argumentation model may be invaluable for grounding attacker-based personas because it not only explicates the origins of assumptions about attackers, but it also presents the claims made by a persona developer argumentatively.

Consequently, this allows attacker personas to be revised in light of alternative perspectives about attackers and their capabilities and motives.

III. Approach

The attacker personas we developed are similar to assumption personas presented in the literature. However instead of being based on the assumptions of stakeholders in a design situation, they are based on various sources of data regarding the types of people who have been known to attack systems. These personas can be used to inform other attack description analysis, notably attack trees as section IV-E will describe. In order to enable the detailed analysis necessary for the creation of attacker personas, we used the CAIRIS [16] tool to support, document and manage the process. The following sections describe the methodology used to build attacker personas.

A. Data Source selection

The first stage involves identifying possible sources of data from which attacker personas can be derived. This data source discovery exercise must be informed by pre-existing knowledge about the problem domain the system will be situated in, together with existing open-source threat taxonomies.

B. Reference elicitation

The data sources must then be analysed on a sentence-by-sentence basis to identify assumptions or claims being made about prospective attackers, and their interest in the operational environment. These assumptions are not elicited verbatim from the text, instead they need to be inferred from the fragments of behaviour that can be reasonably assumed.

C. Affinity diagramming

In the process of culling references from the source material, analysts can then establish some ideas of their own about the kind of attacker they are referring to. To make sure their assumptions do not unduly bias their analysis, sense-making activities are carried out to draw characteristics about the assumption data which, by this stage, is little more than a collection of unrelated references about a user. This activities involve carrying out an affinity diagramming exercise [17], where each reference is written on a post-it note and stuck on a white board. Following this, 2 - 3 people work through the large, un-organised cluster of notes, identifying similar traits in order to form affinity clusters representing different facets of attacker behaviour. This clustering is an interactive process, with participants asking for clarification about what notes mean, discussing the nature of different clusters, and moving notes around. As these clusters appear and became stable, these are then labelled with a concept or phrase describing the affinity cluster, which is represented using a different coloured post-it note.

Not all references may fall under any particular category, and these are kept clustered in an Unknown category. In the next step of developing characteristics for the persona, a possible category for these references sometimes becomes apparent, or, alternatively, references which had fallen under one existing category occasionally become redundant and are reclassified into the Unknown category.

Based on the number and variety of the references, it can be unrealistic to proceed on the assumption that a only single persona will be developed. Consequently, the following steps *D* and *E* can still be followed with the aim of creating multiple personas instead of a single one.

D. Characteristic development

This stage moves from identifying clusters of user behaviour, to developing the characteristics of the persona. Assumptions about personal attributes, such as age and gender, are noted based on indications from the affinity clusters. Following this, individual persona characteristics are developed to correspond to each behavioural cluster. These characteristics are structured according to the argumentation model described in section II-C. These characteristics can be entered into CAIRIS, together with the grounds, warrants, and rebuttals which form the basis of the argumentation model. This allows the visual argumentation models of each persona characteristic to be generated automatically.

E. Author Persona

The final step involves writing a narrative for the personas. This begins by developing a skeleton structure for the persona based on the elicited characteristics. Following this structure, the narrative is then written to describe the persona in a more engaging manner. This step involves some creativity, however, it must always be bounded and informed by the characteristics themselves.

IV. Case Study: *webinos*

The Secure Web Operating System Application Delivery Environment (*webinos*) is defining and delivering an open source platform which will enable web applications and services to be used and shared consistently and securely over a broad spectrum of converged devices, including mobile, PC, home media (TV) and in-car units. From the outset of the project, security and privacy will be designed into *webinos*. The convergence of a large number of different devices (mobile, pc, home media and in-car units) and an increasingly broad and eclectic user base provides new opportunities for attackers to exploit digital assets. These new opportunities for device and application convergence have also led to growing awareness of the importance of users to better control their privacy when interacting with online services.

We developed several attacker personas – summarised in table I – to embody a set of possible threat sources

Name	Synopsys
David	3rd party maintenance operator, with physical device access.
Frankie	Unskilled script kiddie, with access to public domain tools.
Ethan	Semi-professional spammer and botnet herder.
Gary	Irrational (ex) staff member, with un-revoked system access.
Harold	Skilled Grey-Hat cracker.
Irwin	Professional software developer, and Intellectual Property thief.

Table I
webinos attacker personas

which help to explore and understand the security and privacy issues that affect the users, environments and tasks associated with *webinos*. These were developed and applied to supplement a Context of Use description for the *webinos* project. The Context of Use description provides the background necessary to evaluate the usability of a system, and includes characteristics of the intended users, the tasks they perform, the system's goals, and information about the environment in which users are to use the system [18]. More conventional personas and scenario-based task descriptions were developed as part of this context of use description. In particular, the conventional personas were developed using a similar methodology to that described in this paper, i.e. the persona narrative was based on a selection of argued characteristics.

We now describe how the methodology presented in section III was used, and focus our attention mostly on one particular attacker persona: Irwin. Other attacker personas are described in the webinos official deliverable [19].

A. Data Source selection and Reference elicitation

The attacker personas were developed with the following criteria in mind:

- They should be representative of known attacker classes;
- They should be representative of criminals convicted for common online crimes;
- They should be situated within the context of *webinos*.

The attacker personas were chosen to be representative of OWASP human threat agents. To mitigate the risk of developing irrational attacker models, we chose not to model rare but possibly very dangerous attackers, such as government or organised-crime sponsored professional hackers, for which accurate information is not generally available.

We gathered web resources that record judgments, pleas and sentences, clustered similar references, and selected significant examples; these provided information about how the attacker personas might behave. This was supplemented by other OWASP data, and anecdotal data from security domain experts on the *webinos* project team. This data was analysed and summarised into factoids. Between 20 and 50 factoids for each candidate persona were derived. In the case of Irwin, we elicited 23 factoids ranging from personal attitudes to cognitive capabilities. For example, from the excerpt *"He is described by friends as having a quiet sense of humour."* we elicited *"He has a quiet sense of humour"*. Similarly, from the statement *"A Rutgers professor ... described him ... as one of the brightest students ... He was also ambitious and driven and, by the way, an excellent competitive ballroom dancer"* we elicited the reference *"He was a bright student, ambitious and driven"*.

Each elicited factoid was entered into CAIRIS as a reference.

B. Affinity diagramming

Because the process was sensitised by attacker expectations and models, affinity modelling was essential for filtering out unrealistic facts about the attacker. Using a white-board, we transcribed all elicited factoids onto post-it notes, and progressively clustered and de-clustered factoids until a partially disjoint set of 10 clusters was identified. The list below describes some of the criteria that represents each of these emergent categories:

- What kind of unauthorised behaviour might the attacker be inclined to do?
- What kind of assets does the inside attacker manage in an organisation?
- How much does the attacker respect agreements and contracts?
- How much knowledge can the attacker gain of the organization's security posture?
- How much can the attacker exploit his colleague's propensity for insecure behaviour?

C. Characteristic development

After affinity diagramming, the clusters were used to form the basis of more detailed, individual characteristics of the attacker persona. For example, the cluster relating to *How much does the attacker respect agreements and contracts?* formed the basis of the characteristic *Irwin does not respect contracts*.

As figure 1 illustrates, the factoids associated with each cluster formed the grounds and, in one case, the warrant for the claims made by these characteristics. The modal qualifiers, i.e. *Always* and *Presumably* were based on the confidence in the behavioural cluster and the used data. In particular, they reflect how frequent and realistic the relation between the cluster and its respective claim are, as well as the reliability of the data source. Once these characteristics had been drafted, each was associated with a behavioural variable type. These types were based on the model of possible behavioural variable types proposed by [12], i.e. Activities, Attitudes, Aptitudes, Motivation and Skills.

A new persona object was created in CAIRIS, and each respective characteristic was associated with it; the grounds and warrant elements associated with each were associated

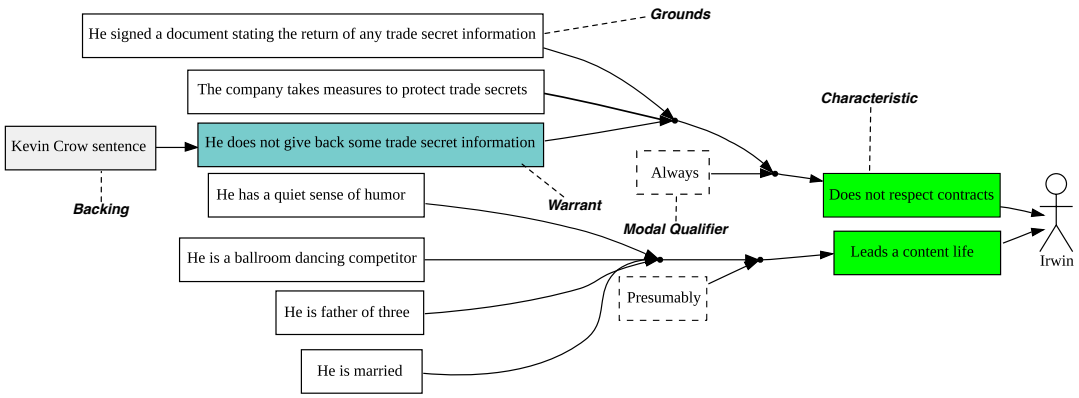

Figure 1. Irwin argumentation model

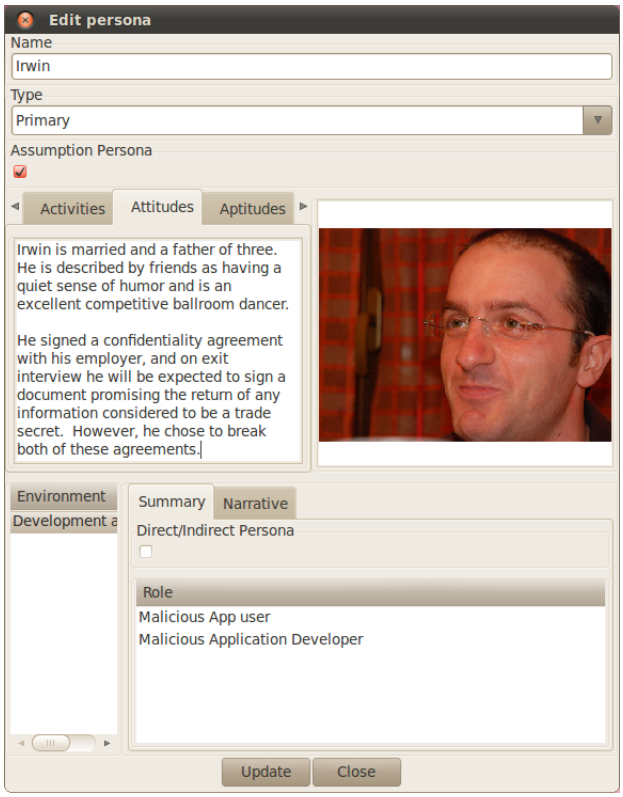

Figure 2. Complete Attitudes narrative for Irwin

with the respective references previously entered. From these, visual argumentation models were automatically generated for different attacker personas and their behavioural variable types. Figure 1 illustrates the argumentation model associated with Irwin's Attitude characteristics.

D. Author Persona

Based on the skeleton provided by the persona characteristic for each attacker persona's behavioural type, a narrative appealing to the characteristics in each section was entered into CAIRIS, and a representative photograph was associated with each persona. Figure 2 illustrates the completed attitudes section associated with Irwin.

E. Applying Attacker Personas

Once an initial version of the *webinos* Context of Use Description had been developed, a two day workshop was held to review the analysis carried out. The workshop participants were mobile application developers, security specialists, and usability experts involved with the *webinos* project. As part of this workshop, a focus group was held to review the conventional personas in order to highlight unrealistic and potentially exploitable user behaviours. Each persona was presented individually and, where queries were raised about individual behaviours, the argumentation structure was used to justify and motivate the persona's characteristics.

On the second day of the workshop, a similar focus group session was carried out to review and discuss the attacker personas. We found that discussing the attacker personas using the same format as the non-attacker personas was useful because session participants had, by this stage, become attuned to the activities associated with validating personas using their argumentation models. Consequently, the participants were able to relate to the attackers with the same ease as they related to the more conventional personas. By the end of this session, all participants were clear about the motives and capabilities of each of the attackers that *webinos* would need to defend against.

Following this attacker persona focus group, one of the security experts facilitated a session where a number of attack trees [1] were developed; these modelled how unauthorised access to user data, application data, and sensitive *webinos* APIs might be obtained, and how unauthorised use of system resources might occur. During this exercise, the attacker personas were frequently used to suggest certain steps that might (or might not) be taken as part of an attack. As well as re-grounding and validating the attacker

personas by using them in practice, this also rationalised the attacker's activities, thereby making the attacks more believable to non-security participants. For example, one of the steps in the attack tree stated "trick the user to use/install Web application". When thinking about how Irwin might carry out such a step, situations were discussed where work based inter-colleague trust relationship might be exploited.

V. Conclusion

This paper presented a grounded approach for developing attacker personas. We have also illustrated this approach using a case study where attacker personas were developed and used to support the development of a Context of Use description for the EU FP7 *webinos* project. Attacker personas are analogous to conventional personas, but differ in the data sources upon which they are grounded. By using open source data about convicted attackers and known attacks, the personas are less biased by individual developer beliefs, and more grounded in reality. Attacker personas helped in focusing on the attackers' characteristics a system realistically has to face.

The grounding of attacker personas is based on three important characteristics : they are representative of known attacker classes; they are representative of criminals convicted for common online crimes; and they are situated within the context of *webinos* by design and workshop discussions. As a result, supplemental threat modelling artifacts appear more realistic, because they are grounded in what a concrete attacker can and is willing to do.

Future work will involve using both the attacker personas and the supplemental analysis to motivate design decisions underpinning the *webinos* security architecture.

Acknowledgment

The research described in this paper was funded by EU FP7 *webinos* Project (FP7-ICT-2009-5 Objective 1.2).

References

[1] B. Schneier, *Secrets and Lies: Digital Security in a Networked World*. John Wiley & Sons, 2000.

[2] G. Sindre and L. Opdahl, "Eliciting security requirements with misuse cases," *Requirements Engineering*, vol. 10, no. 1, pp. 34–44, 2005.

[3] A. Whitten and J. D. Tygar, "Why Johnny can't encrypt: a usability evaluation of PGP 5.0," in *SSYM'99: Proceedings of the 8th conference on USENIX Security Symposium*. Berkeley, CA, USA: USENIX Association, 1999, pp. 169–184.

[4] M. A. Sasse, S. Brostoff, and D. Weirich, "Transforming the 'weakest link': a human-computer interaction approach to usable and effective security," in *BT Technical Journal*, vol. 19, 2001, pp. 122–131.

[5] A. Steele and X. Jie, "Adversary Centered Design: Threat Modeling Using Anti-Scenarios, Anti-Use Cases and Anti-Personas," in *Proceedings of the 2008 International Conference on Information & Knowledge Engineering, IKE 2008*, H. R. Arabnia and R. R. Hashemi, Eds. CSREA Press, 2008, pp. 367–370.

[6] J. D. Gould and C. Lewis, "Designing for usability: key principles and what designers think," *Communications of the ACM*, vol. 28, no. 3, pp. 300–311, 1985.

[7] I. Fléchais, M. A. Sasse, and S. M. V. Hailes, "Bringing security home: a process for developing secure and usable systems," in *NSPW '03: Proceedings of the 2003 workshop on New security paradigms*. New York, NY, USA: ACM, 2003, pp. 49–57.

[8] *ISO/IEC 27002: Information Technology – Security Techniques – Code of Practice for Information Security Management*. ISO/IEC, 2007.

[9] "Common attack pattern enumeration and classification web site," http://capec.mitre.org/. [Online]. Available: http://capec.mitre.org/

[10] O. foundation, "The open web application security project," http://www.owasp.org/index.php/. [Online]. Available: http://www.owasp.org/index.php/

[11] A. Jones and D. Ashenden, *Risk management for computer security : Protecting your network and information assets*. Elsevier Butterworth-Heinemann, 2005.

[12] A. Cooper, *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity (2nd Edition)*. Pearson Higher Education, 1999.

[13] J. Pruitt and T. Adlin, *The persona lifecycle: keeping people in mind throughout product design*. Amsterdam: Elsevier, 2006.

[14] S. Faily and I. Fléchais, "The secret lives of assumptions: Developing and refining assumption personas for secure system design," in *HCSE'2010: Proceedings of the 3rd Conference on Human-Centered Software Engineering*. Springer, 2010, pp. 111–118.

[15] S. Toulmin, *The uses of argument*, updated ed. Cambridge University Press, 2003.

[16] S. Faily and I. Fléchais, "Towards tool-support for Usable Secure Requirements Engineering with CAIRIS," *International Journal of Secure Software Engineering*, vol. 1, no. 3, pp. 56–70, July-September 2010.

[17] H. Beyer and K. Holtzblatt, *Contextual design: defining customer-centered systems*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998.

[18] ISO, "ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on usability," Tech. Rep., 1998.

[19] WebinosConsortium, "User expectations on privacy and security," Tech. Rep., 2011. [Online]. Available: http://webinos.org/archives/559