Using Risk Analysis to Assess User Trust
– A Net-Bank Scenario –

Gyrd Brændeland[1] and Ketil Stølen[1,2]

[1] Department of Informatics, University of Oslo, Norway
[2] SINTEF ICT, Norway

Abstract. The paper advocates asset-oriented risk analysis as a means to help defend user trust. The paper focuses on a net-bank scenario, and addresses the issue of analysing trust from the perspective of the bank. The proposed approach defines user trust as an asset and makes use of asset-oriented risk analysis to identify treats, vulnerabilities and unwanted incidents that may reduce user trust.

1 Introduction

There is no generally accepted definition of the term "trust". One obvious reason for this is that the meaning of "trust" as the meaning of most other natural language terms depends on the context in which it is used. In this paper we restrict our investigation of trust to a scenario involving a net-bank (online banking), the bank that owns the net-bank, and the net-bank users. We argue that risk analysis is suited to help defend existing user trust. The term "defend" is taken from asset-oriented risk analysis where vulnerabilities of a system are analysed with regard to identified assets. We claim that the user trust is a major asset to the bank. Furthermore, we argue that risk analysis is well-suited to find strategies to defend user trust and prevent unwanted incidents that may reduce user trust.

In order to use risk analysis to asses user trust, we need a way to measure trust in a quantitative or qualitative manner. We argue that it is not the trust itself, but its consequences, such as the number of net-bank users, that is important to the bank. Such observable consequences are often easy to measure and may provide a firm basis for risk analysis.

The paper is divided into six sections. Section 2 introduces a basic terminology with emphasis on factors that affect trust. Section 3 gives a short introduction to asset-oriented risk analysis. Section 4 describes a net-bank scenario on which much of this paper focuses. Section 5 argues the suitability of risk analysis to help defend existing user trust. The evaluation is angled towards the scenario introduced in Section 4. Section 6 summarises our findings, presents the main conclusions and outlines related work.

C.D. Jensen et al. (Eds.): iTrust 2004, LNCS 2995, pp. 146–160, 2004.

2 Trust and Trust Affecting Factors

It is generally accepted that trust is a more general issue than security in particular and dependability in general. Jones et al. [15] argue that "although businesses and consumers may consider underlying systems to be completely dependable in the traditional sense, they may not trust these systems with their business or personal interests unless there exists a suitable legal framework they can fall back on, should problems arise." An analysis of trust will therefore encompass a number of issues like legal, sociological and psychological aspects that are not directly related to security.

2.1 Basic Terminology

Studies of trust distinguish between the trustor, that is, the agent that trusts another agent, and the trustee; the agent being trusted. Trust is a property of the trustor, whereas credibility and trustworthiness are properties of the trustee. Trust can also be seen as a binary relation, from the trustor to the trustee.

Mayer et al. [19] defines trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the agility to monitor or control that other party." Koufaris and Hampton-Sosa [17] use Mayer's definition of trust in their survey of user trust in a web site, which is not so different from our net-bank scenario.

Attributes of the trustee, such as credibility and trustworthiness are considered important factors influencing an agent's trust in another party [19]. In a recent book on the role of computers in influencing peoples attitudes, Fogg [9] is concerned with what constitutes computer credibility. In accordance with existing literature, Fogg defines credibility as "a perceived quality, that has two dimensions: trustworthiness and expertise" (Figure 1).

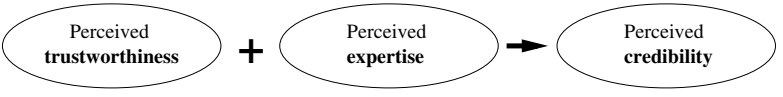

Fig. 1. Fogg – The two key dimensions of credibility

Fogg decomposes the concept of trustworthiness further into the terms *well-intentioned, truthful* and *unbiased*, and expertise into *knowledgeable, experienced* and *competent*. Fogg and Tseng [10] argue that users' evaluation of computer trustworthiness and credibility is a function of both system design features and psychological factors ascribed to the entity behind a system.

2.2 Factors That Affect Trust

Egger [6] has developed a model of trust-relevant factors in e-business that encompasses such features as those discussed by Fogg and Tseng. Egger's model,

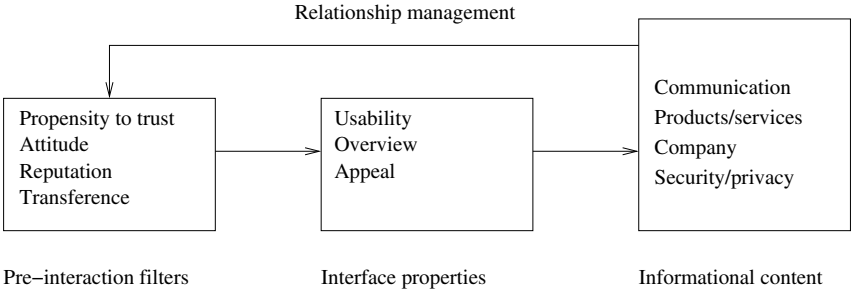

Fig. 2. Egger – Model of trust in e-commerce

shown in Figure 2, identifies factors that affect user trust in a system. *Pre-interaction filters* are factors that may affect a user's trust in a system prior to any interaction. An individual's general *propensity* to trust affects the degree to which she is willing to trust any agent. A user's general *attitude* towards an industry may affect her trust in particular members of that industry. *Reputation* concerns such factors as the strength of a company's brand name and the user's experience with the system through earlier interaction. *Transference* of trust covers the situation where a user trusts a company because a trusted third party has reported that the company is trustworthy.

Interface properties concern the impression that a system gives through its design interface. The significance of such factors are well documented in the literature. An empirical study performed by Stanford Web Credibility Project [8] discovered that users had more trust in a web site with a picture of a nice car in the upper right corner, than the same web site where the picture of the car was replaced by a dollar sign.

Informational content concerns other properties of a system such as security and privacy and how they are conveyed to the user. It is not enough that a system is properly evaluated and certified with regard to security. The user must also be informed that the system has undergone such evaluations. The provider of a web service may for example include information in its web site that the system has undergone a security evaluation that is certified by a licensed certifier.

Egger's model includes factors that are encompassed by both the terms "trust" and "credibility", as introduced in the previous section. In fact, Egger views user trust as perceived trustworthiness. Egger's model is a practical tool for assessing user trust, and has already been tried out in a few cases discussed in Egger's PhD thesis [7].

3 Risk Analysis – The CORAS Approach

The Australian/New Zealand standard for risk management [1] defines risk analysis as the systematic use of available information to determine how often specified risks may occur and the magnitude of their consequences. Furthermore, as illustrated by Figure 3, risk analysis is one of seven risk management processes.

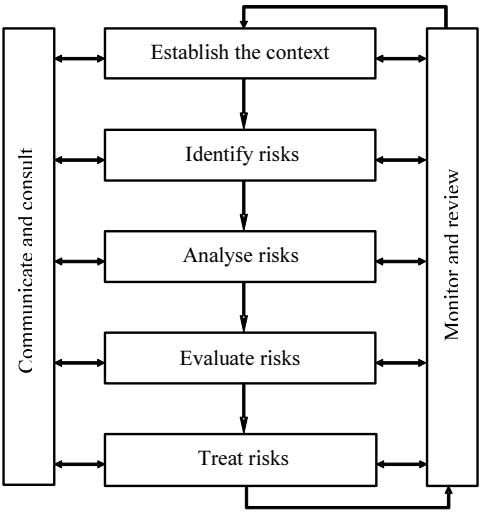

Fig. 3. Risk management overview

In practise, however, the term "risk analysis" normally has a broader meaning covering the five sequentially ordered processes that the Australian/New Zealand standard refers to as: establish the context, identify risk, analyse risk, evaluate risk and treat risk. In this paper we use this broader definition. We refer to what the standard [1] calls "risk analysis" as consequence and frequency analysis.

There are many forms and variations of risk analysis. Asset-oriented risk analysis where system vulnerabilities and unwanted incidents are analysed with regard to identified assets, is a kind of risk analysis often used within the security domain. One such approach to risk analysis is the CORAS [5,2] methodology that will be used in the following.

CORAS is characterised by tight integration of state-of-the-art systems modelling methodology based on UML2.0 with leading methodologies for risk analysis as Hazard and Operability (HazOp) analysis [20], Failure Mode Effect Analysis (FMEA) [4], and Fault Tree Analysis (FTA) [12]. In fact, CORAS comes with its own specialisation of UML, a so-called UML profile for security analysis that has recently become a recommended OMG (Object Management Group) standard integrated in the UML Profile for Modeling Quality of Service and Fault Tolerance [18].

Hence, an important aspect of the CORAS methodology is the practical use of UML to support risk management in general, and risk analysis with respect to security (in the following referred to as security risk analysis) in particular. The CORAS risk analysis methodology makes use of UML models for three different purposes:

– To describe the target of evaluation in a uniform manner at the right level of abstraction.

- To facilitate communication and interaction between different groups of stakeholders, experts and users involved in a risk analysis.
- To document risk analysis results and the assumptions on which these results depend to support reuse and maintenance.

The former two are particularly relevant in the case of trust. To analyse trust, technical system documentation is not sufficient; a clear understanding of system usage and its role in the surrounding organisation, enterprise and society is just as important. UML is well-suited to describe technical aspects as well as human behaviour in the form of work-processes. One major challenge when performing a risk analysis is to establish a common understanding of the target of evaluation, threats, vulnerabilities and risks among the stakeholders, experts and users participating in the analysis. The CORAS UML profile has been designed to facilitate improved communication during risk analysis, by making the UML diagrams easier to understand for non-experts, and at the same time preserving the well-definedness of UML.

4 A Net-Bank Scenario

Figure 4 presents a simple UML class diagram that specifies the overall context of our net-bank scenario. There is a web-service exemplified by the net-bank, the net-bank is owned by a bank, and the net-bank users are account holders performing net-bank transactions via the Internet. The bank, the net-bank, the net-bank users and the Internet exist in an overall context known as Society.

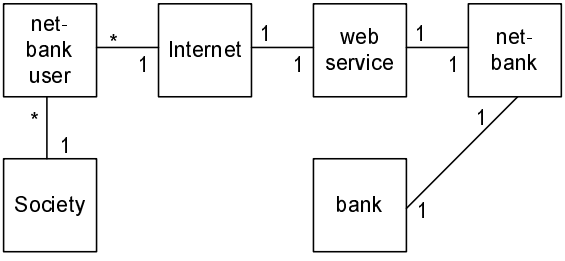

Fig. 4. A net-bank scenario: The main actors

When interacting with a net-bank, a user normally knows very little about the actual design of the bank, but is nevertheless willing to make herself vulnerable, by entrusting her money and personal data with the net-bank. If asked why, the user may answer that she expects the net-bank to perform several actions, such as safe handling of her money, confidential handling of personal data, legislative insurance in the case something goes wrong, and so on.

Figure 5 outlines the relationship between the notions introduced in Section 2.1 and the entities of our net-bank scenario.

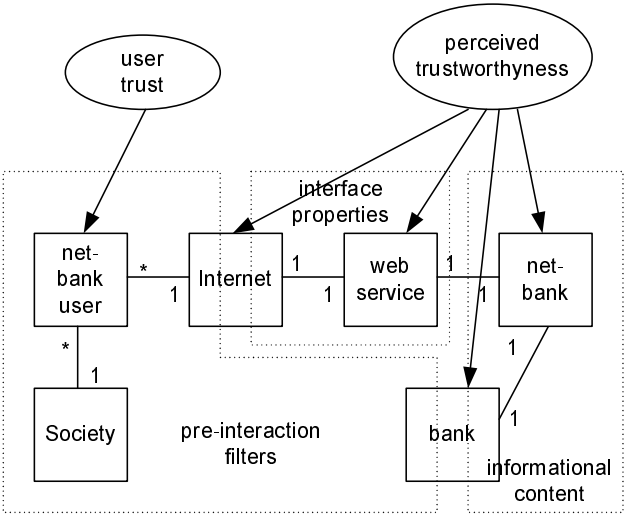

Fig. 5. Trust and trustworthiness in a net-bank scenario

Since this paper restricts its attention to user trust, we consider trust only as a property of the net-bank user. It is in the interest of the net-bank owner that it, as well as the net-bank and the Internet, is viewed as trustworthy by the net-bank users. As explained in Section 2.2, both trust and trustworthiness are influenced by several factors, categorised as pre-interaction filters, interface properties and informational content, respectively. The pre-interaction filter "user's propensity to trust" is clearly a property of the user. The same holds for attitude. The bank's reputation as well as the reputation of the Internet as a medium for interaction has impact on user trust. Transference takes place between actors in society. Properties of the web-service, such as user appeal, are part of the interface properties. Informational content concerns such factors as the net-bank's security and privacy policy.

5 Analysing User Trust in the Net-Bank Scenario

In the following we sketch how the model-based risk analysis approach of CORAS can be used to identify treats, vulnerabilities and unwanted incidents that may reduce user trust. We focus on those aspects where a risk analysis targetting trust differs from an ordinary security risk analysis. An important observation is that there is a difference between perceived security, which is a basis for user trust, and well-founded security, which is the target of conventional security risk analysis. The security of a system may be perceived to be worse than the actual security, if the latter is not properly conveyed to the user, as discussed with regard to informational content in Section 2.2. It may be perceived to be better than the actual security, through for example an appealing interface that gives

a sound impression of the system, regardless of implementation. The presentation is structured into the five main sub processes of risk analysis described in Section 3.

5.1 Subprocess I: Establish the Context

To conduct a risk analysis we need a characterisation of the target of analysis. This target may be a system, a part of a system or a system aspect. The term "system" should be understood in the broadest sense. A system is not just technology, but also the humans interacting with the technology, and all relevant aspects of the surrounding organisation and society. In the following we take the net-bank scenario specified in the previous section as our target of analysis. A risk analysis is always conducted on behalf of one or several customers. In the following we view the bank owning the net-bank as our only customer.

The CORAS risk analysis process is asset-oriented. An *asset* is a part or feature of a system that has a value for one of the stakeholders on behalf of which the risk analysis is conducted, in our case, the bank. A risk analysis makes sense only if there are some assets to be protected. If there are no assets, there is no risk and no need for a risk analysis. Let us therefore identify and value assets from the perspective of the bank. When analysing users' trust in the net-bank, it is not the trust as such, but its direct impact on the market share that the bank is interested in. User trust has impact on the number of users and the amount of time and money they are willing to invest in a system. These are precise factors that are easy to measure. A user's willingness to risk time and money on for example web gambling may also be triggered by other factors than trust, such as addiction to gambling. User trust, however, is clearly one important factor that affects observable customer behaviour, and may therefore be viewed as an asset on its own. Figure 6 makes use of an asset-diagram expressed in the CORAS UML profile to specify that the market share asset depends on other assets like users' trust in the net-bank.

Confidentiality is the property that information is not made available or disclosed to unauthorised individuals, entities or processes [13]. Confidentiality of the customer database is clearly important for the market share since such information could be used by competitors to "steal" customers. That the market share may depend on the confidentiality of the net-bank technology should also be obvious. Furthermore, there are also dependencies between user trust and the confidentiality of the net-bank technology and customer database. We may use Egger's model (see Figure 2) to decompose user trust in more specialised assets. We consider only those factors that the bank can directly influence, leaving reputation as the only asset under pre-interaction filters.

The risk evaluation criteria specify what the customer may tolerate with respect to risk. The risks that do not satisfy the risk evaluation criteria must be treated. Table 1 presents two examples of risk evaluation criteria. In order to assign a quantitative risk acceptance value to user trust, we need a way to measure user trust. We may for example use Jøsang and Knapskog's [16] metric for trusted systems, based on their subjective logic approach. They define trust

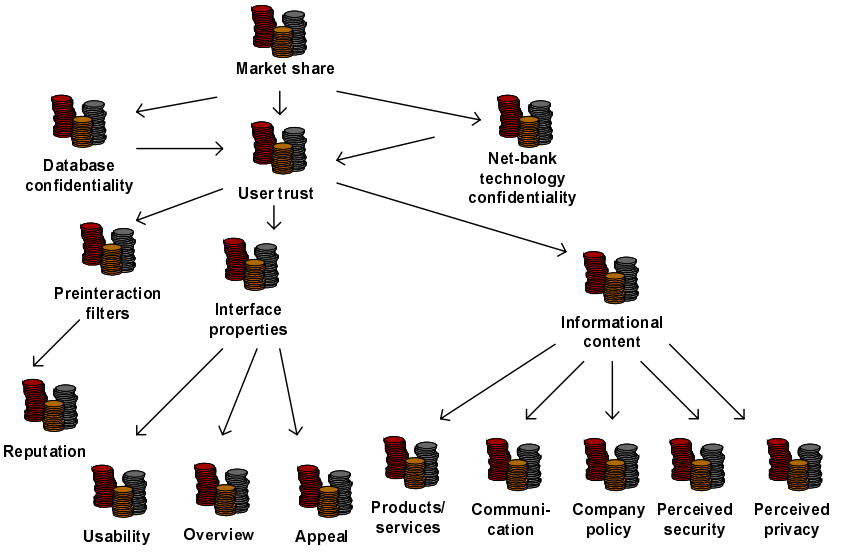

Fig. 6. Hierarchy of trust-related assets

as a subjective belief consisting of three probability values; belief, disbelief and uncertainty that together equal 1. If a person's belief value for a given proposition is 1, she has full trust in the proposition. A decrease in the belief value is a decrease in the trust. Jøsang and Knapskog give guidelines to determine opinions when evidence can only be analysed intuitively, as in the case of trust. Their proposal involves formulating a questionnaire to guide people in expressing their belief as valued opinions. An agent's trust is computed from the value of several beliefs about the trustee, that together constitute the total trust.

Table 1. Risk evaluation criteria

Id	Stakeholder	Asset	Criteria description
1	bank	user trust	if risk impact ¡ "0.15 decrease in user trust within a week" then "accept risk" else "assign priority and treatment"
2	bank	market share	if risk impact ¡ "loss of 500 customers within a week" then "accept risk" else "assign priority and treatment"

For this to make sense, we need an understanding of the effect of reduced user trust on the market share. For example, it seems reasonable to require that any risk that satisfies Criteria 1 also satisfies Criteria 2. Such an understanding may for example be based on statistical data, or be acquired through user surveys.

5.2 Subprocess II: Identify Risks

Identifying risks includes identifying threats to assets, identifying vulnerabilities of assets, and documenting unwanted incidents caused by threats exploiting vulnerabilities. A *threat* is a potential cause of an unwanted incident which may result in harm to a system or organisation and its assets. A *vulnerability* is a weakness with respect to an asset or group of assets which can be exploited by one or more threats. An *unwanted incident* is an undesired event that may reduce the value of an asset. A risk is an unwanted incident that has been assigned consequence and frequency values.

Conventional approaches to risk identification include checklists, judgement based on experience and records, flow charts, brainstorming, systems analysis, scenario analysis and systems engineering techniques. A UML sequence diagram showing normal behaviour of a system in combination with guidewords addressing the various security and trust aspects may be used as a basis for a structured brainstorming to identify possible unwanted incidents. Figure 7 specifies a normal login session.

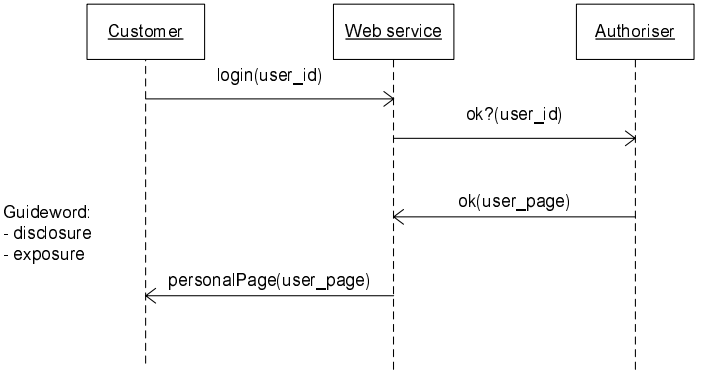

Fig. 7. User login session

Disclosure of personal customer data to outsiders is an example of a security related unwanted incident, during a normal login session. This is what happened in 2000 when it was reported that the net-bank of the Norwegian company Gjensidige NOR made web pages containing confidential customer data accessible to outsiders [3]. This incident was caused by a weakness in the security technology. When analysing trust we may be interested in other types of unwanted incidents than typical security incidents. Whether such an incident is a threat to customer trust depends on public exposure. In our risk assessment we therefore identify exposure of a security incident in media, as the unwanted incident, and disclosure of personal customer data as a threat that may cause the incident, see Table 2.

Table 2. HazOp table: Unwanted incidents

Id	Stakeholder	Asset	Guideword	Threat	Unwanted incident
1	bank	perceived security	disclosure	disclosure of confidential customer data	incident reported to media
2	bank	perceived security	disclosure	security weakness revealed	weakness reported to media
3	bank	user trust	exposure	negative press coverage	loss of user trust
4	bank	market share	exposure	loss of user trust	loss of regular customers

The user trust asset may also be affected by unwanted incidents that are not directly related to security. Lack of support and non-appealing web-sites are two examples.

To model a full scenario corresponding to an unwanted incident we may use the CORAS UML profile as demonstrated in Figure 8. The unwanted incident that a security incident is reported to media, from Table 2, is modelled as a use case including the threat that confidential customer data is disclosed. The threat scenario is caused by a threat agent; in this case an eavesdropper. Each threat scenario and each unwanted incident may be further specified by UML sequence and activity diagrams as in the case of an ordinary UML use case.

5.3 Subprocess III: Determine Consequence and Frequency

A *risk* in the CORAS terminology is an unwanted incident that has been assigned a consequence, in terms of reduced asset value, and frequency values. If the frequency of an incident is not known we can use a fault tree to document the possible routes that can lead to the incident. The objective of a fault tree analysis is to document in a structured way the possible routes that can lead to the violations of security requirements identified by for example HazOp. The unwanted incidents identified in the HazOp table, Table 2, are inserted in a fault tree, see Figure 9, based on abstraction level and the relationship between the incidents. Through the fault tree analysis, the incident is broken down into smaller causes for which we have better estimates of the frequency. The top event of the fault tree in Figure 9, negative press coverage on security leakage, may lead to reduced user trust which may lead to loss of market share. Historical data from similar incidents in the past can be used to estimate the consequences in the form of reduced asset value.

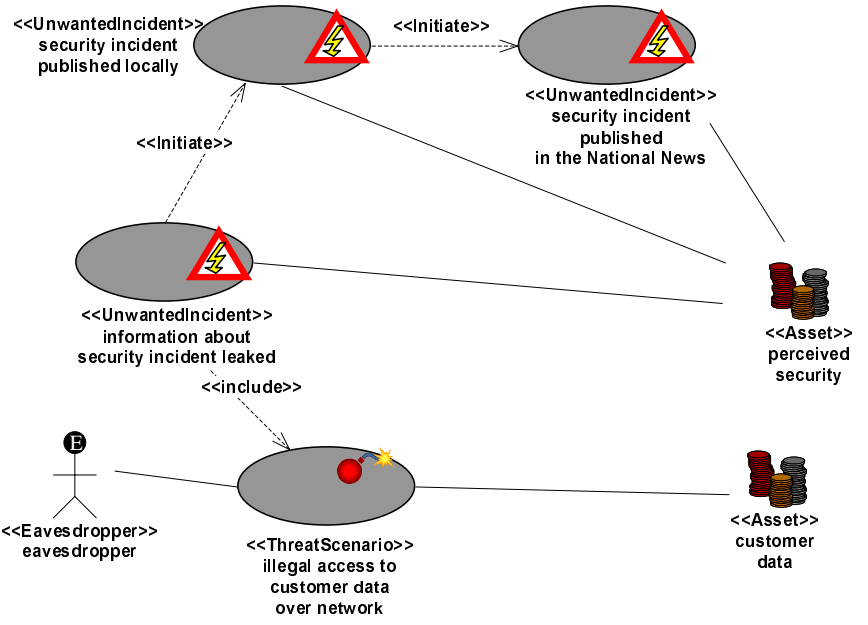

Fig. 8. Specification of threat scenario

5.4 Subprocess IV: Evaluate Risks

Evaluating risks includes determining level of risk, prioritising risks, categorising risks, determining interrelationships among risk themes and prioritising the resulting risk themes and risks. A *risk theme* is a categorisation of similar risks, assigned its own risk value.

When trust relevant risks have been assigned frequency and consequence they can be evaluated in the same manner as any other risk.

5.5 Subprocess V: Treat Risks

Treating risks includes identifying treatment options and assessing alternative treatment approaches. A *treatment* is a way of reducing the risk value of a risk or a risk theme. The Australian/New Zealand standard for risk management [1] identifies five options for treating risks: acceptance, avoidance, reduce likelihood, reduce consequences, and transfer to another party.

Risks having impact on user trust may require other types of treatment than security related risks. In Section 5.2 we identified media coverage of a security incident as an unwanted incident (Table 2) and in Section 5.3 we proposed fault tree analysis to estimate the frequency of such an event. In order to treat this type of risk it may not be enough to fix a programming error. It may also be necessary to do some public relations work, or to prevent information on

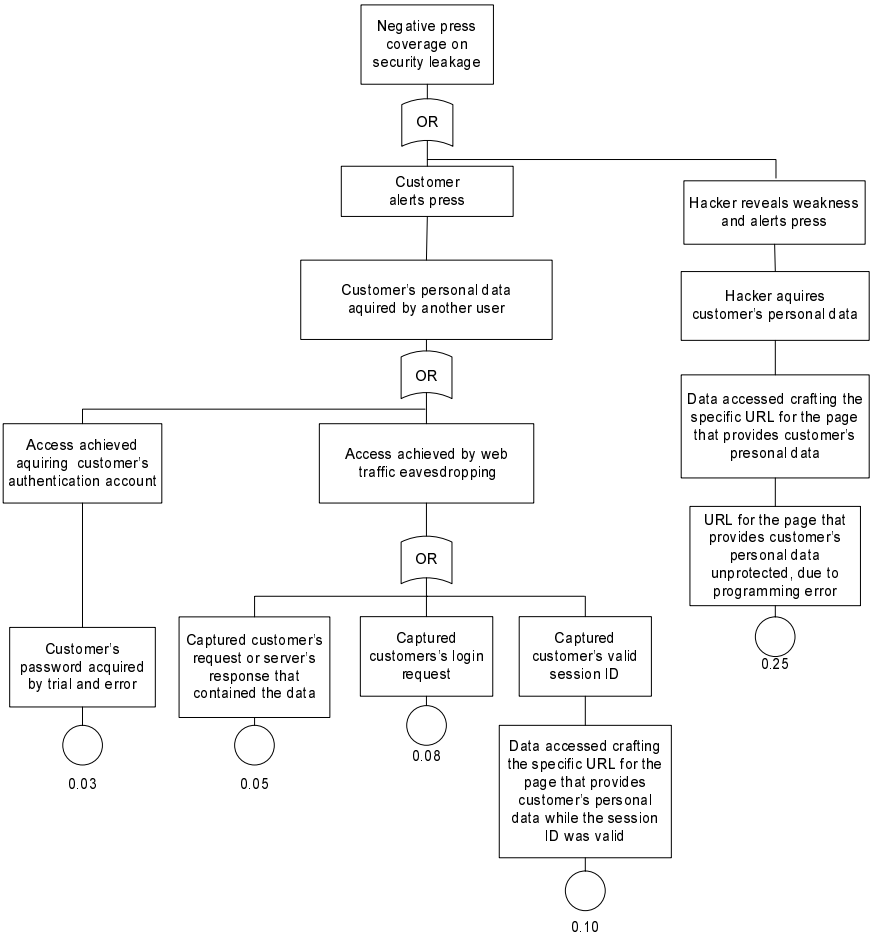

Fig. 9. A fault tree for revealing weak security

the security breach to reach the public. The CORAS UML profile can be used to document security treatments as use case diagrams. As indicated by Figure 10, we may use the CORAS UML profile to document treatments with regard to trust in the same way. We have identified the treatment "public relations work" to reduce the consequences of negative press coverage, and the treatment "authentication" to reduce the likelihood of an intruder obtaining illegal access to customer data.

6 Conclusions

The paper has advocated asset-oriented risk analysis as a means to help defend existing user trust. The proposed approach defines user trust as an asset and makes use of asset-oriented risk analysis to identify threats, vulnerabilities and

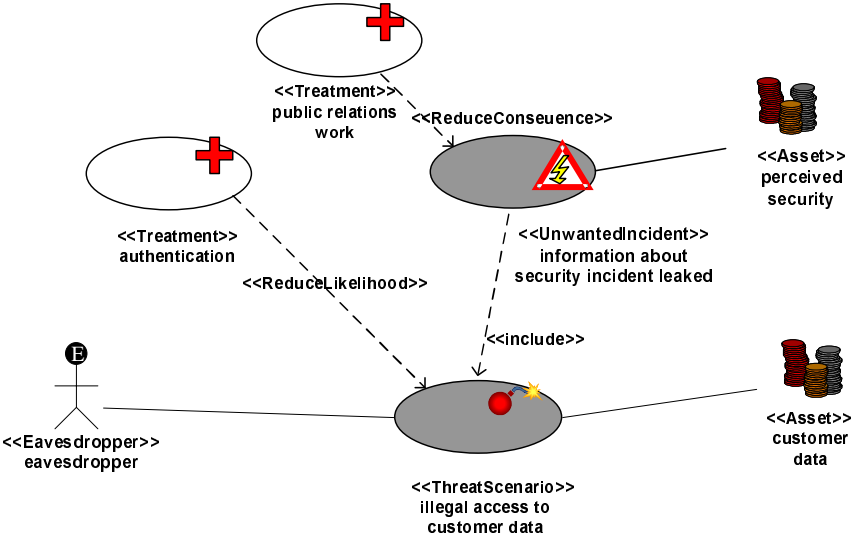

Fig. 10. Documentation of treatments

unwanted incidents that may cause a reduction in user trust. Risk analysis targetting user-trust may be based on the same overall process as risk analysis in the security domain.

Herrmann [11] distinguishes between two principal approaches to integrate trust values into the process of security risk analysis. In the first approach, the auditing process is extended by considering trust values in the risk level computation. Simply said, the higher the trust in the good-naturedness of the involved parties, the lower the likelihood of a successful attack. In the other approach, the auditing process is kept unchanged and the decision about which risk can be run and which not, is made dependent on the trust in the parties. In that case, "an asset owner should be willing to take as greater risks as higher the belief in the benevolent behaviour of the involved parties is."

Herrmann's focus on integrating trust values in the auditing process is clearly different from our use of risk analysis as a means to analyse user trust by interpreting user trust as an asset. Herrmann proposes to express trust relations by so-called trust values, that were first introduced by Jøsang and Knapskog [16].

To accept asset-oriented risk analysis as a means to help defend user-trust entails accepting asset-oriented risk analysis as means to help build new user trust or increase already existing user trust. The same techniques that are used to identify factors that may cause loss of asset value may also be used to identify factors that may increase the value of assets. In that case, a trust incident is wanted and may have positive impact mirroring the negative impact in the "hazard" risk analysis used in this paper.

An interesting issue for further research is the use of modal logic, as for example in [14], in combination with UML to gain the expressiveness that may

be required to describe certain trust relevant scenarios. Furthermore, to estimate frequencies and consequences, we may need a tight integration of methods from decision psychology, or social science. The methods may vary from user surveys, to technical assessments and lab experiments.

Acknowledgements. The research on which this paper reports has partly been funded by the Research Council of Norway project SECURIS (152839/220). It is influenced by discussions and interaction within the EU working group iTrust (IST-2001-34910). The authors thank Asbjørn Følstad and Ida Solheim for valuable input.

References

1. Australian/New Zealand Standard for Risk Management 4360:1999.
2. J. Ø. Aagedal, F. den Braber, T. Dimitrakos, B. A. Gran, D. Raptis, and K. Stølen. Model-based risk assessment to improve enterprise security. In *EDOC2002*, pages 51–62. IEEE Computer Society, 2002.
3. P. K. Bjørkeng, C. Haraldsen, and S. Stenseng. Strykkarakter til internett-bank. Aftenposten, August 2000.
4. A. Bouti and D. A. Kadi. A state-of-the-art review of FMEA/FMECA. *International journal of reliability, quality and safety engineering*, 1:515–543, 1994.
5. T. Dimitrakos, B. Ritchie, D. Raptis, J. Ø. Aagedal, F. den Braber, K. Stølen, and S.-H. Houmb. Integrating model-based security risk managament into ebusiness systems development: The coras approach. In *I3E2002*, pages 159–175. Kluwer, 2002.
6. F. N. Egger. Towards a model of trust for e-commerce system design. In *CHI 2000: Workshop Designing Interactive Systems for 1-to-1 E-commerce*, April 2000. `http://www.zurich.ibm.com/~mrs/chi2000/contributions/egger.html`.
7. F. N. Egger. *From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce*. PhD thesis, Eindhoven University of Technology, 2003.
8. B. Fogg, C. Soohoo, D. Danielson, L. Marable, J. Stanford, and E. R. Tauber. How do people evaluate a web sites credibility? Technical report, Stanford Persuasive Technology Lab, October 2002. `http://www.consumerwebwatch.org/news/report3_credibilityresearch/stanfordPTL_abstract.htm`.
9. B. J. Fogg. *Persuasive Technology. Using Computers to Change What We Think and Do*. Morgan Kaufman Publishers, December 2002.
10. B. J. Fogg and H. Tseng. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 80–87. ACM Press, 1999.
11. P. Herrmann. How to integrate trust management into a risk analysis process. Second Internal iTrust Workshop On Trust Management In Dynamic Open Systems, September 2003.
12. IEC 1025. *Fault Tree Analysis (FTA)*, 1990.
13. ISO/IEC TR 13335-1. *Information Technology - Guidelines for the management of IT Security - Part 1: Concepts and models for IT security*, 2001.

14. A. J. I. Jones. *The open agent society*, chapter 3; A logical framework. John Wiley & Sons, Chichester, UK, 2004. To be published.
15. S. Jones, M. Wilikens, P. Morris, and M. Masera. Trust requirements in e-business. *Communications of the ACM*, 43(12):81–87, 2000.
16. A. Jøsang and S. Knapskog. A metric for trusted systems. In *21st National Security Conference*, 1998.
`http://csrc.nist.gov/nissc/1998/proceedings/paperA2.pdf`.
17. M. Koufaris and W. Hampton-Sosa. Customer trust online: Examening the role of the experience with the web site. Technical Report #CIS-2002-05, Department of Statistics & Computer informations systems. Zicklin school of business, Baruch college, May 2002. CIS Working paper series.
18. M. S. Lund, I. Hogganvik, F. Seehusen, and K. Stølen. UML profile for security assessment. Technical Report STF40 A03066, SINTEF Telecom and informatics, December 2003.
19. R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995.
20. F. Redmill, M.Chudleigh, and J. Catmur. *Hazop and software hazop*. Wiley, 1999.