Perceived information security risk as a function of probability and severity

T.Sommestad, H.Karlzén, P.Nilsson and J.Hallberg

Swedish Defence Research Agency, Linköping, Sweden
e-mail: teodor.sommestad@foi.se

Abstract

Information security risks are frequently assessed in terms of the *probability* that a threat will be realized and the *severity* of the consequences of a realized threat. In methods and manuals, the product of this probability and severity is often thought of as the *risk* to consider and manage. However, studies of human behavior and intentions in the field of information security suggest that in general, this is not the way security is perceived. In fact, few studies have found an interaction (i.e., a multiplicative relationship) between probability and severity. This paper describes a study where the ratings of risk and the two variables probability and severity were collected on 105 security threats from ten individuals together with information about the respondents' expertise and cognitive style. These ten individuals do not assess risk as the product of probability and severity, regardless of expertise and cognitive style. Depending on how risk is measured, an additive model explains 54.0% or 38.4% of the variance in risk. If a multiplicative term is added, the mean increased variance is only 1.5% or 2.4%, and for most of the individuals the contribution of the multiplicative term is statistically insignificant.

Keywords

Information security risk assessment, Risk perception, Perceived severity, Perceived probability

1. Introduction

It is widely accepted and uncontroversial to view information security in terms of perceived risks. Information security risks are, in many of the most widely accepted definitions, assessed in terms of the *probability* that a threat will be realized and the *severity* of the consequences of a realized threat. For instance, the following literature describe security risk as a combination of probability (in other contexts termed likelihood or frequency) and consequence (or impact or magnitude) (NIST, 2012)(Club de la Sécurité de l'Information Français, 2011)(Karabacak and Sogukpinar, 2005) and (Lund et al., 2011).

In the literature the relationship between severity, probability and risk is also clear – risk is defined as the product of the severity and the probability. Thus, rational and balanced security decisions require that risk is assessed as the product of probability and severity. The rationale for this is clearest in the extreme cases – with no negative effect (severity zero) the probability should be irrelevant, and with no possibility of

happening (probability zero) the severity should be irrelevant. But it is also clear in-between these extremes – if a bad thing is twice as likely or twice as severe as another bad thing, the expected costs will be twice as large.

A multiplicative relationship is well established in decision making theory regarding information security. However, results from both information security and from other domains suggest that people do not multiply the two in practice. For example, in the original formulation of the Protection Motivation Theory it was proposed that an interaction of perceived vulnerability and perceived severity influenced behavioral intentions (Maddux and Rogers, 1983; Rogers, 1983). However, this interaction has been abandoned for a simpler additive model on empirical grounds – empirical data does not offer firm support of a multiplicative relationship (Das et al., 2003)(Pechmann et al., 2003)(Cismaru and Lavack, 2007). One possible explanation for these results is that humans are incapable or unwilling to adhere to reason and mathematical stringency and prefer to simply combine (add) a percentage with a cost into a risk value. Another possible explanation is that studies fail to observe the multiplicative relation for one reason or another. There are several reasons to expect that this is the case.

First, some studies have measured the intentions to engage in protective behaviour rather than assessing actual risk. Clearly, the effectiveness and costs of the protective behavior is also a factor to consider in such protective decisions, and this may have distorted the results – at least when both factors are present. Second, the scales to measure probability and severity used in many of the studies of information security behavior (e.g., (Posey et al., 2011)) are not suited for multiplication. A multiplicative operation requires that two ratio scales are used, which is seldom the case in the research. For instance, a Likert scale with questions asking if the respondent *Completely Agree* or *Completely Disagree* does not produce a ratio, and multiplications with such variables are questionable if not outright invalid. Third, it is possible that some persons multiply probability and severity to calculate risk and others do not or that some interpret scales differently and, e.g., do not start their severity ratings at 0 but do start at 0 for probability and risk. This would distort the results of between-subject designs. Fourth, when a fairly homogenous group of respondents are asked to assess one or few incidents in a between-subject design, a large portion of the variance may be because of measurement errors, i.e., it comes from unreliable responses rather actual differences in perceptions. To discover an interaction term when most of the observed variance between subjects' perceptions is due to error requires considerable sample size. Fifth, the incidents may be too homogenous resulting in only part of the scales being used.

Only one study was found that addresses the relationship between probability, severity and risk while isolating risks from remedies, using scales allowing multiplication, and using a within-subject-design. This study, by Weinstein (2000), comprised a convenience sample of 12 individuals who assessed 201 health risks, covering the entire probability-severity matrix, on two occasions. The respondents first assessed risk (R) by prioritizing events and valuing hypothetical insurances. After 1 to 2 weeks they assessed probability (P) and severity (S) for the same events. A clear multiplicative effect was found in the sample. A function with only a

multiplicative term (i.e., R=P*S) explained approximately 90% of the variance explained by a function that also included the additive terms (i.e., R=P+S+P*S), i.e., the additive function did not add much. However, the interaction between probability and severity appears to vary with the magnitude of these two. For example, considering events with high probability and high severity, the severity matters most, but for events with low probability and high severity, the multiplicative relationship is highly significant. The results also suggest that there are considerable individual differences between how people assess health risks. For instance the respondents are, on average, insensitive to health risks with moderate to high (P>40%) probability, but the sensitivity varies between respondents.

This paper performs a study similar to Weinstein's (2000), but in the information security domain. A within-subject design is used and meaningful scales are used to test the risk equation in the minds of ten individuals. More specifically, the following hypothesis is tested.

> H1: Perceived information security risk is determined as the product of its perceived probability of occurrence and perceived severity.

In addition, it is reasonable to suspect that people who are more used to the subject matter will be more inclined to multiply. Thus, the present study adds a between-subject design by investigating if the tendency to multiply severity and probability to obtain risk is higher among respondents who 1) are proficient in risk assessments, 2) possess information security expertise, or 3) have a rational decision making style rather than an intuitive one. The following hypotheses are tested.

> H2: The tendency to assess risk as a product of probability and severity is related to risk assessments experience.
>
> H3: The tendency to assess risk as a product of probability and severity is related to information security expertise.
>
> H4: The tendency to assess risk as a product of probability and severity is related to cognitive decision making style.

Section 2 of the paper describes the method. Section 3 describes the results and section 4 discusses these results.

2. Method

The study design is heavily influenced by the one used by Weinstein (2000). The sections below describe the participants, the survey instrument and the data collection procedure.

2.1. Participants

The survey was distributed to a strategic sample of 10 researchers active in the areas of information security, IT security, IT management or human factors. All respondents are from the Swedish Defence Research Agency (as are the authors of

this paper), possess university degrees, are in the age range 29-54 and work as researchers. In order to test H2 and H3, pertaining to security expertise and experience in risk assessments, five of the respondents were drawn from the information security research group and five of the respondents were drawn from the research group called "Human, technology organization", which specializes in requirements engineering and human-machine-interaction. Thus, whereas the participants are a convenience sample drawn from the authors' own organization, the sample is designed to test the hypotheses in question. Furthermore, the questionnaire had all participants answer questions on both probability and severity rather than separating the two factors. This reflects the common situation where experts conduct the entire risk analysis process from threat elicitation to countermeasure recommendations. Thus, stakeholders might be involved with asset elicitation but are unlikely concerned with specific threats, or qualified to determine potential consequences.

2.2. Material and scales

Two paper based questionnaires were used to conduct the study. The first questionnaire comprised two parts: one part asking questions about the respondent and one part asking the respondent to assess the probability and severity of 105 incidents. The second questionnaire repeated some of the probability and severity questions in the first questionnaire to allow reliability tests, but focused on measuring the perceived risk associated with the 105 incidents.

2.2.1. Incidents and scenarios

The 105 potential incidents (or scenarios) were designed to be meaningful for the target population. For example, they used information objects and threats that are relevant for the organization. Some examples include:

- "A computer virus extracts all documents related to cooperation with foreign states in the office network and shares this with a foreign intelligence service."
- "Spyware is introduced into the organization's office network by an international defence corporation".
- "Employees intentionally violate policies related to the storage of secret documents."
- "A scientist's USB-stick with five years of collected (unclassified) material is stolen at an international conference."

The incidents were constructed to cover the whole range of possible assessments. In other words, they were designed to be assessed as all combinations of low probability, high probability, low severity, and high severity. Fortunately, identifying incidents of high probability and severity turned out to be difficult.

2.2.2. Perceived probability and severity

In the first questionnaire, the respondents were asked to provide the severity and probability of each incident. The perceived *severity* of incidents was indicated by marking a line stretching from 0 (Minimal, no harm at all) to 10 (Greatest harm). In the questionnaire, it was emphasized that the worst of all 105 incidents should be rated a 10 and that other ratings should be proportional to this (e.g., that 5 is half as harmful a 10). The perceived *probability* of an incident occurring during the next ten years was provided by marking a line with endpoints 0% (Minimal, completely unlikely) to 100% (Maximal, guaranteed to happen).

Anchors were present along this line, however, respondents were free to mark any point on the line. The corresponding value (e.g., severity 1.6 or probability 16%) was measured using a ruler. To enable tests of reliability, i.e., that answers were stable over time, the second questionnaire asked the respondents to provide probability and severity assessments for twelve randomly selected incidents a second time.

2.2.3. Perceived risk

In the second questionnaire, the respondents were asked to provide the overall perceived risks associated with the incidents in two ways to increase confidence in the results. Both of these methods are supposed to reflect the perceived risk associated with an incident, without considering how easy or difficult it would be to lower the risk.

First, the respondents were presented with the hypothetical scenario that they would have the power to eliminate some of the risks corresponding to the 105 incidents. They were asked to mark the *priority* of eliminating the risks by putting a mark on a line stretching from 0 (Not at all prioritized) to 10 (Absolutely highest priority). Second, the respondents were asked to indicate the *expected costs* of the incident in monetary terms. More concretely, respondents were asked to write how much they would be prepared to pay to insure the organization against the risk if they were in charge of the budget. As in the study by Weinstein (2000), an anchor and an upper limit were used to simplify the assessment. The respondents were told that no risk was worth more than 10 million SEK (approximately 1 million EUR) and that protection against incidents involving lost or stolen USB-sticks ever happening (this is an acceptable deviation of the standard definition of insurance, also shared by Weinstein (2000)) was worth about 30% of the maximum amount.

2.2.4. Decision making style

Cognitive style was measured using eight items. These items are direct translations of the items presented by (McShane, 2006), which in turn is inspired by (Scott and Bruce, 1995) and the Cognitive Style Index (Allinson and Hayes, 1996). Four items measure the tendency to be rational, i.e., to ignore gut instinct when it contradicts objective information and to make decision based on facts and logical analysis. Four items measure the tendency to be intuitive, i.e., to make decision based on inner feelings or instinct rather than to rely on rational choices conflicting with intuition.

2.2.5. Expertise and experience

Expertise and experiences were obtained from self-ratings by the respondents, which were validated against dichotomous classifications made by the investigators based on organisational department. Self-ratings were provided on the format "Completely agree" to "Completely disagree" for the following statements: "I work with security assessments or risk assessments", "I work with information security", and "My colleagues think that I am an IT-security expert or information security expert".

2.3. Data collection procedure

Respondents were provided the second questionnaire one to two weeks after they had answered the first questionnaire. One week was expected to remove the opportunity of simply recollect their previous responses and multiply them to obtain responses for the second questionnaire. In addition, after the first questionnaire they were asked to remove all copies or notes related to their responses. Furthermore, to avoid influencing the respondents' risk assessment procedure (or combination procedure) they were not told what the test actually was about. They were only told that the aim was to investigate how risk perceptions vary between people and why they vary.

2.4. Validity and reliability measurement

In the study, the items on cognitive style had a Cronbach's alpha of 0.810 and the items on security expertise had a Cronbach's alpha of 0.962, i.e., they were highly internally consistent. As expected, the five participants who belonged to the information security research group considered themselves to have high security expertise while the other five participants evaluated themselves much lower (means 4.533 compared to 1.733 on the scale 1-5).

The repeated questions of the second survey showed 8 participants to be highly reliable with Pearson correlations larger than 0.767 ($p<0.001$), whereas the reliability of two participants was statistically non-significant. Thus, the tests and retests suggests that all but two respondents reasoned about incidents in a similar way when answering questions on probability and consequence and questions about risk. Furthermore, the two measures for risk used in the second questionnaire were highly internally consistent with an overall standardized Cronbach's alpha of 0.776, with the figure for each respondent being above 0.7, i.e., showing sufficient consistency for all respondents.

3. Results

The risk equation used by the respondents is inferred within-subjects and presented in section 3.1, which presents the test of H1. As will be seen, the results of this test made it difficult to test H2-H4. Section 3.2 describes this further.

3.1. The risk equation

As by Weinstein (2000), the hypothesis is tested by modeling the relationship between answers in the first questionnaire (on probability and severity) as predictor

variables for answers in the second questionnaire (on priority and insurance premium) in a linear regression model. Table 1 provides the figures of the regression models for risk as priority (upper half of the table) and risk as insurance premium (lower half of the table). R^2(S, P) is the coefficient of determination for the linear (non-interaction) model, indicating the fit of that model. ΔR^2(SxP) describes how much the fit improves when considering an interaction model (multiplicative term). Four rows (p) indicate the significance (*) or non-significance (ns) of R^2, the severity (S), the probability (P) and ΔR^2, respectively.

As the table shows, few of the respondents show a tendency to multiply probability and severity to obtain the remediation priority or the insurance fee, and thus there is little support for H1. Considering the priority, the interaction-term is significant for three of the respondents; considering the insurance premium, the interaction-term is significant for two of the respondents. Furthermore, the contribution of the interaction term is small in the regression models for all respondents. At most, the interaction term adds 0.096 (statistically non-significant) explained variance to a regression model which explains 0.193 of the variance (participant #5) and 0.082 (statistically significant) of explained variance to a model which explains 0.453 of the variance (participant #8). Overall, the mean additional variance obtained by introducing the interaction term is 0.015 for priority and 0.024 for insurance premium. This should be related to an additive model, which explains 0.540 and 0.384 of the variance.

Participant		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean**
Risk as priority	R^2(S, P)	0.381	0.683	0.493	0.545	0.352	0.724	0.544	0.540	0.434	0.544	0.540
	pR^2	*	*	*	*	*	*	*	*	*	*	
	pS	*	*	*	*	*	*	*	*	*	*	
	pP	ns	ns	*	*	ns	ns	*	ns	ns	*	
	ΔR^2(SxP)	0.003	0.018	0.008	0.015	0.000	0.014	0.009	0.072	0.000	0.008	0.015
	$p\Delta R^2$	ns	*	ns	ns	ns	*	ns	*	ns	ns	
Risk as insurance premium	R^2(S, P)	0.108	0.464	0.505	0.357	0.193	0.657	0.406	0.453	0.281	0.415	0.384
	pR^2	*	*	*	*	*	*	*	*	*	*	
	pS	ns	*	*	*	*	*	*	*	*	*	
	pP	ns	ns	ns	ns	ns	ns	*	*	*	*	
	ΔR^2(SxP)	0.000	0.013	0.007	0.014	0.096	0.003	0.004	0.082	0.024	0.001	0.024
	$p\Delta R^2$	ns	ns	ns	ns	ns	*	ns	*	ns	ns	

Table 1: Regression analyses with linear and interaction models

It should be added that the insignificance of the multiplicative term is not because the additive terms are present. The mean variance in risk (priority) explained by a model with only the multiplicative term is 0.049, and it is only statistically significant for the three respondents (as it was with the additive terms in the model). Furthermore, it

is worth noting that these results hold within all quadrants of the probability-severity-spectrum, i.e. for high/low, low/high or low/low probability and severity.

3.2. Variables related to the tendency to multiply

There were no statistically significant correlations between expertise and either risk as priority or risk as insurance. Nor were there any statistically significant correlations between cognitive style and either risk as priority or risk as insurance. However, as described above, there was no general tendency to multiply probability and consequence in the studied population. As a consequence, identifying variables that relate to this tendency (i.e., H2-H4) is doomed to fail.

4. Discussion

Most of the respondents seem to have an idea of probabilities and severities associated with information security incidents. For eight out of ten respondents, the responses provided at different weeks had very strong correlations (>0.75). This idea is also, to some extent, shared among the respondents. Between-subjects correlations are above 0.50 for both probabilities and severities. Thus, their responses seem to stem from some partially shared perception of the information security threats. This suggests that the survey is able to measure the perceptions it set out to measure. Nevertheless, there are many possible reasons for the fact that our result – in contrast to Weinstein (2000) – does not support a multiplicative relationship between severity and probability in people's minds when calculating risk. The results indicate that information security risk assessments are determined by the severity.

Similarly to Weinstein we used a limited sample non-random sample. Our participants were more homogenous in terms of profession and slightly more homogenous in terms of age and gender than the sample of Weinstein. Any of these factors may explain the focus on incident severity and the insignificance of the multiplicative terms in this test. However, it is unclear to the authors why they should. On the contrary, it is hard to see how and why a population of researchers, of which many had considerable risk assessment experience, should be unable or unwilling see risk as a product of probability and severity.

The scales and measurement procedure used in this test is different from the ones used by Weinstein (2000) in several ways. First, Weinstein's first survey concerned (compound) risk where he let half of the participants prioritise the incidents and the other half estimate the insurance premiums. We instead measured (compound) risk in the second survey, with probability and severity in the first. This may have caused our participants to be more prone to thinking of risk as a product of probability and severity, so this is not an issue considering our results. Second, we let all the participants rate risk both by priority and insurance premium. This made it possible to verify that the two measurements correlated strongly and it is hard to see why this will remove the tendency to multiply probability and severity. Third, we measured priority with the slightly different phrasing "stop the incidents from happening or render them harmless if they do". While this phrasing is different form Weinstein's ("If you purchase insurance against a particular problem, you are guaranteed that it

will never happen to you"), it is unclear to us why this would remove the tendency to multiply. Fourth, it is possible that it was harder for our participants to reason in terms of monetary loss for an organization rather than hundreds of dollars for a personal insurance premium. However, as risk as priority and risk as insurance premium correlated, it is hard to see this as a possible reasons for the insignificant multiplicative term. Also, there were no substantial differences between those of our respondents used monetary risk and relative risk, so difficulties understanding scales is unlikely to be an issue. Fifth, we further imposed restrictions on risk as priority and severity, with both max values defined by the "worst" among our incidents for risk and severity respectively. But this would only lead to our measurements being off by a (scale-converting) constant, which is no problem in regression models.

Perhaps the most important difference between our study and Weinstein's (2000) study – and indeed between information security and health – are the topics of the incidents. In our case, the incidents relate to the participants' organisation rather than the participants themselves and our incidents are less well-known than say pneumonia or rash from poison ivy. Weinstein partly based his incidents on a standard compendium of diseases, while we constructed our own. This may have led to incidents that were more difficult to interpret with greater variance between subjects. However, our results suggest that the respondents' assessments agreed and the performed test-retests suggest that most respondents understood the questions well enough to answer them similarly. Thus, the scenarios were clearly comprehendible. Also, the answers for each respondent showed no more absolute correlation between probability and severity than those Weinstein reported (-0.56). This correlation should be expected to be negative, as few incidents have high values for both probability and severity.

Another significant difference to Weinstein's (2000) survey is probabilities were (implicitly) restricted in that incidents should happen in the respondents' remaining lifetime. For an organisation, there is no natural upper time limit so to avoid infinite possibilities. We used a ten year limitation, and we do not anticipate any issues with our results due to this.

In conclusion, it is doubtful that information security experts are any better at risk assessments than novices, at least concerning the combination of severity and probability to form risk. For this reason, it is straightforward to recommend appropriate risk matrices which force the assessor to adhere to the established definition of risk as the mathematical product of probability and severity.

5. Acknowledgements

The authors would like to thank Neil Weinstein for his thoughts on why the multiplicative relationship is hard to detect.

This research was funded by the Swedish Civil Contingencies Agency.

6. References

Allinson, C. W., & Hayes, J. (1996). The Cognitive Style Index: A Measure of Intuition-Analysis For Organizational Research. *Journal of Management Studies*, *33*(1), 119–135.

Cismaru, M., & Lavack, a. M. (2007). Interaction effects and combinatorial rules governing Protection Motivation Theory variables: a new model. *Marketing Theory*, *7*(3), 249–270.

Club de la Sécurité de l'Information Français. (2011). MEHARI 2010 Processing guide for risk analysis and management. Paris: Club de la Sécurité de l'Information Français.

Das, E. H. H. J., de Wit, J. B. F., & Stroebe, W. (2003). Fear appeals motivate acceptance of action recommendations: evidence for a positive bias in the processing of persuasive messages. *Personality & Social Psychology Bulletin*, *29*(5), 650–64.

Karabacak, B., & Sogukpinar, I. (2005). ISRAM: information security risk analysis method. *Computers & Security*, *24*(2), 147–159.

Lund, M. S., Solhaug, B., & Stolen, K. (2011). *Model-driven risk analysis: the CORAS approach*. *Media*. Springer Verlag.

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*, *19*(5), 469–479.

McShane, S. L. (2006). Activity 8.8: Decision Making Style Inventory. In *Canadian Organizational Behaviour* (Sixth edit.). McGraw-Hill Education. Retrieved from http://highered.mheducation.com/sites/0070876940/student_view0/chapter8/activity_8_8.html

NIST. (2012). NIST Special Publication 800-30 Revision 1 Guide for Conducting Risk Assessments. Gaithersburg, USA: NIST.

Pechmann, C., Zhao, G., Goldberg, M. E., & Reibling, E. T. (2003). What to Convey in Antismoking Advertisements for Adolescents:The Use of Protection Motivation Theory to Identify Effective Message Themes. *Journal of Marketing*, *67*(2), 1–18.

Posey, C., Roberts, T., Lowry, P. B., Courtney, J., & Bennett, R. J. (2011). Motivating the insider to protect organizational information assets: Evidence from protection motivation theory and rival explanations. In *Proceedings of the Dewald Roode Workshop in Information Systems Security 2011* (pp. 1–51). Blacksburg, Virginia, September 22–23, pp.: IFIP WG 8.11 / 11.13.

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. Cacioppo & R. Petty (Eds.), *Social Psychophysiology*. New York, New York, USA: Guilford Press.

Scott, S. G., & Bruce, R. A. (1995). Decision-Making Style: The Development and Assessment of a New Measure. *Educational and Psychological Measurement*, *55*(5), 818–831.

Weinstein, N. D. (2000). Perceived probability, perceived severity, and health-protective behavior. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association*, *19*(1), 65–74.