Risk Communication in Security Using Mental Models

Debin Liu, Farzaneh Asgharpour, L. Jean Camp
Indiana University, Bloomington, IN, USA

Abstract. In computer security, risk communication refers to a mechanism used to inform computer users against a given threat. Efficacy of risk communication depends not only on the nature of the risk, but also alignment between the conceptual model of the risk communicator and the user's perception or mental model of the risk. The gap between the mental model of the security experts and non-experts could lead to ineffective and poor risk communication. Our research shows that for a variety of the security risks self-identified security experts and non-experts have different mental models. We propose that the risk communication methods should be designed based on the non-expert's mental models with regard to each security risk.

Keywords: Mental model, Pile sorting, risk communication

1 Introduction

The mental models approach is a risk communication method based on the conceptual models of recipients of the communication. A mental model is an internal conception for how something works in the real world [1]. This notion is very case specific and is subject to change due to experience, schema segments, perception, and problem-solving strategies.

The mental model approach in risk communication has effectively been used in environmental [2] as well as medical [3] risk communication. While this has been done for privacy [4] it has not been introduced to information security [1]. This work is grounded in mental models as it has been developed in environmental risk communication. The goal of mental models in environmental research is to enhance risk communication about household toxics [1]. Like computer security, environmental risks can be much more problematic at home than at the work place. For instance, paint stripper and other chemical hazards are, like computers, more easily regulated in the work place than home. As the mental models have not been investigated in security, we begin with a quantitative approach to evaluate the five mental models proposed by Camp [5]. Camp enumerates five possible mental models from the computer security literature: physical security, medical infections, criminal behavior, warfare and economic failure.

Risk communication is typically a message formulated by the security experts to warn a community of non-experts against a set of threats. The difference between the mental model of the experts and non-experts with regard to the risk can decrease the efficacy of the risk communication. This difference is often a consequence of two different levels of knowledge about the subject matter. One may think that since the experts have access to the technical definition of the risks and know all the catalysts and consequences of each threats, their mental model is more reliable for designing risk communication instruments. The key point is that the purpose of a risk communication is not conveying the total “truth” to the users, but just prompting them to take an appropriate action to defend their system against a certain danger. Even though mitigation of a specific risk requires knowledge of the nature of the risk, efficacy of the risk communication requires the experts to understand their target group. For example in warning a five-year old child against electrical shock, it is much more beneficial to the child to explain a simplified warning using the child’s terminology and imagination, rather than explaining the danger with reference to electromagnetic field theory.

In this work, we define a distance measure between different mental models. Using our proposed measure we estimate the distance between the security experts’ and non-experts’ mental models. We also propose some non-expert’s mental models for each security risk. The details of our experiment design are explained in Section 2. Section 3 covers the data analysis. Section 4 concludes the paper.

2 Experiment Design

With reference to [5] we consider five possible mental models for describing computer security risks as mentioned in Section 1. We designed a pile sorting experiment [6] to identify computer users’ risk perception of various cyber security risks.

Suppose that the set $\Re = \{R1, R2, \dots , Rn\}$ presents all the security risks given in our experiment. We consider two levels of expertise in security: expert (E) and non-expert (NE). By E we mean someone who knows all the technical definitions of the security-related words. We defined NE as someone who dose not know the technical definition of the security terminologies and at most knows some practical aspects of the risks. The main purpose of the experiment is to estimate and compare the experts’ and non-experts’ perception, or mental model, for each member of $\Re$. To classify our participants as experts and non-experts we provided the definition of the expert and non-expert in the instruction section of the experiment and asked the participants to declare their level of expertise.

In this experiment, we gave a set of 66 words (Appendix A - Table 3) to the participants and asked them to cluster the words into groups of similar words. In [6] pile sorting experiment is performed by sorting cards, with some given words typed on them, into different piles. We performed the experiment online and asked the participants to mark similar words with the same colors. The participant determined the similar words according to his/her personal perception of each word, and used some given colors to cluster the words into different groups. Appendix B - Figure 1 shows a screenshot of the pile sorting experiment.

The wordlist (Appendix A-Table 3) contained the name of various major risks, some common security related words and some words directly related to each of the following mental models: physical security, medical infections, criminal behavior, economic failure, and warfare. The words related to each mental model are all driven from Webster’s Thesaurus. Given that there is always the possibility that a certain word might be unknown to the participant, we also specified a color for the words which might not be familiar to the participant. Finally, to leave enough room for other possible mental models, we gave one color for words which in a participant's view might not belong to any of the above categories.

We used the following correspondence between mental models and colors: physical security-green, medical infection-blue, criminal behavior-orange, warfare-red, economic failure-yellow. We also considered the color purple for words which did not match with any of the above mental models, according to the participants' perception, and, gray for the words not familiar for the participant. To be able to keep track of the participants' mental models and to maintain consistency in associating colors with different mental models, we provided instructions on how to associate colors with words. Due to various cultural color interpretations, we decided not to follow any specific cultural pattern in associating colors, as for instance color green is associated with peace for some people and with the environment for some others. The unintuitive and arbitrary color selection made the participants to refer to the instructions more frequently and therefore to be more careful in assigning colors to words.

We used macromedia Flash and PHP to present the pile-sorting experiment as an online experiment. We used SPSS and Matlab for the multidimensional scaling and data analysis.

3 Data Analysis

Presenting our online experiment to the faculty and students of various disciplines and levels of knowledge in computer security, we end up with 74 data entries in total. Out of 74 participants, 25 are self-declared experts and 49 were self-declared non-experts. For each group of participants we first find the matrix of intra-similarity between the words, regardless of their associated mental models, and then, based on the sorted piles, assign the correlated mental model to each pile.

The original data are first tabulated and interpreted as proposed in [6]. Every time a participant marks a pair of words with the same color, we count that as a vote for similarity between the two words. Therefore, as an example, if most of the participants mark the words "trade" and "stock" with the same color, then we can say these two words are highly similar in people's perception. In contrast, if only a few participants assign the words "war" and "fever" with the same color, we interpret this result as these two words are not very similar. This way, we have two 66×66 matrices, one for experts and one for non-experts. We name these two matrices as Expert's Choice Matrix and Non-expert's Choice Matrix, and show them by ECM and NCM In order to reveal underlying perceptual dimensions that participants use to distinguish among these words, we present the symmetric matrix via multidimensional scaling map [7] and locate the expert's and non-expert's choice matrices into a two dimensional space.

Before applying the multidimensional scaling method to map words in a two dimensional space, we define a function to measure the distance between each pair of words in our wordlist. We also use this function to measure the distance between security risks and mental models and finally to assign a mental model to each risk.

Considering matrix ECM , the distance $d_E(w_i, w_j)$ between two words w_i and w_j is defined as $d_E(w_i, w_j) = 1 - \frac{ECM_{i,j}}{n}$ in which n is the number of data entries in the expert data set and $ECM_{i,j}$ is the element of the matrix ECM located in the row i and the column j . $d_{NE}(w_i, w_j)$ is defined similarly.

Having all the above distances, we replace the elements of the two matrices with the corresponding distances and present the distance matrices instead of similarity matrices. Figures 2 and 3 in Appendix B show the map of the multidimensional scaling of the distance matrices ECM and NCM .

To find the mental models of experts and non-experts with regard to a risk R , we define a function to measure the distance between R and each of the mental models. Then, considering all the distances we will assign the mental model, or models, with minimum distance from R .

Table 4 (appendix A) shows a list of three words under each mental model. For each of these words, under a given mental model, around 75% of our participants have grouped the word with the other related words within the same group. We call these words obvious words and refer to each set of obvious words under a certain mental model as an obvious mental model. For a given risk R , we define the expert-distance between R and an obvious mental model $M = \{w_1, w_2, w_3\}$ as $D_E(M, R) = \frac{1}{3}\sum_{w \in M} d_E(w,R)$ where $d_E(w, R)$ is the expert-distance between w and R . Similarly the non-expert-distance $D_{NE}(M, R)$ could be defined for the non-expert community.

To each risk R we assign at least one expert (EM_R) and one non-expert (NM_R) mental model according to the following definition. EM_R (NM_R) is the mental model corresponding to the obvious mental model with minimum expert-distance (non-expert-distance) from R .

Appendix A - Table 1 (Appendix A - Table 2) presents D_{NE} (D_E) between security risk and the mental models. Based on these tables, we find the expert and non-expert mental models indicated in the Appendix A - Table 5.

As one can see in Table 5, some of the probabilities are very low. The reason is that in average 34.3% of non-experts, and 40.5% of the expert participants considered an arbitrary mental model, other than our five suggested mental models throughout the risks. In other words, more than one third of the participants found computer security risks were not consistent with the mental models the previous work had found in the computer security literature. This implies that naive users and computer security experts may be even further apart than suggested by this work. This fact also suggests the need for qualitative study, such as interviews, to find other possible mental models.

Our methodology shows that for 10 out of 29 risks, the expert and non-expert communities have different mental models (risks corresponding to the checked boxes in the last column of the table 5). We are uncertain if our definition of an expert has somehow affected the results. We are currently studying if changing the definition of the expert to the “security specialists who have been either teaching or studying in computer security for at least 5 years” has any affect on our final outcomes.

One can also see that the medical mental model is chosen by experts four times whereas just once by the non-experts. On the other hand physical security is selected 7 times by the non-experts but only 4 times by experts. This suggests that the medical mental model is not a very good candidate for risk communication towards non-expert community, whereas physical security potentially could be an appropriate mental model for this purpose. As an example, Appendix B - figure 4 shows the distribution of the functions D_E and D_{NE} over all the 29 security related words with regard to the “physical security” mental model. For almost all the risks the experts have more distance from the “physical security” compare to non-experts.

4 Conclusion

This paper reports an initial experiment to verify the mental models of the security experts and non-experts with regard to security risks. Previously these models had been implicit in security risk communication. The first task was to use the security literature to extract implicit mental models, as done in [5]. This work uses pile sorting to test both the similarity between experts and non-experts and the coverage of those implicit mental models. Our experiment illustrates that for 70% of the security risks non-expert community has either physical security or criminal mental model. We also show that computer security risks are more distant from medical threats for non-experts than for experts. Our statistics, shows that in average 40.5% of the expert community and 34.3% of the non-expert participants marked an arbitrary mental model as their mental model for all the security risks. This implies that none of the mental models implicit in the security literature fit the understanding or the impression of the related risk. Further research includes conducting qualitative interviews either with individuals or within some focus groups to expand the mental models. We would repeat the pile sorting experiment with these models. We also propose that the efficacy of the security risk communication could be increased by adjusting the risk communications with the mental models of non-expert community. Narrowing down the definition of the expert to “security specialists” one can repeat the experiment and measure the mental model of each group. We expect to receive even more difference between the security specialists’ and non-experts’ mental models.

Acknowledgements

We would like to thank professor Youn-Kyung Lim for her helpful comments on this paper and Christian Briggs for his suggestions on our experiment’s interface.

References

[1] Morgan, M.G., Fischhoff, B, Bostrom, A., Atman, C.J.: Risk Communication: A Mental Models Approach. Cambridge University Press. Cambridge, UK (2001)

[2] Ronnfeldt, Carsten F.: Three Generations of Enviroment and Security. Jornal of Peace Research, Vol. 34, No. 4 (1997) 473-482

[3] Jungermann, H., Schutz, H., Thuring, M.: Mental models in risk assessment: informing people about drugs. Risk Analysis. Blackwell (1981)

[4] Diesner, J., Kumaraguru, P., Carley, K. M.: Mental Models of Data Privacy and Security Extracted from Interviews with Indians. 55th Annual Conference of the International Communication Association. New York, NY (2005)

[5] Camp, L. Jean: Mental Models of Security. IEEE Technology and Society (2006)

[6] John Gatewood: Pile sorts. http://www.analytictech.com/borgatti/etk3.htm#N_9_

[7] Kruskal, J., Wish, M.: Multidimensional Scaling. Sage Publication (1978)

Appendix A: Tables

Table 1- *Non-expert distances* ($D_{NE}(M, R)$) between security risks and mental models

	Criminal	Physical	Medical	Market	Warfare
Adware	0.8980	0.7143	0.9524	0.8571	0.9048
Spyware	0.7483	0.7891	0.9524	0.8844	0.9184
Phishing	0.7075	0.8435	0.9592	0.8844	0.8912
Identity theft	0.3537	0.8912	0.9660	0.9456	0.8367
Spam	0.7279	0.8571	0.9184	0.8231	0.8639
Hijackers	0.4286	0.9048	0.9524	0.9660	0.6327
Cookies	0.8844	0.7823	0.9660	0.7279	0.8844
DoS attack	0.8163	0.8163	0.8912	0.8095	0.8571
Download	0.8027	0.8707	0.9524	0.8299	0.9320
Trojan	0.6803	0.8503	0.9388	0.9048	0.6939
Keystroke	0.7075	0.6667	0.9524	0.8299	0.8639
Junk mail	0.7483	0.8299	0.9388	0.8639	0.8435
Virus	0.7755	0.9116	0.5646	0.9592	0.8231
Worm	0.6735	0.8503	0.8231	0.8844	0.8095
Hacking	0.4830	0.8367	0.8980	0.8844	0.7279
Binder	0.9252	0.8231	0.9864	0.7347	0.9524
Exploit	0.6735	0.9592	0.9388	0.7415	0.8435
Zombie	0.8367	0.8980	0.8571	0.7891	0.7823
Authentication	0.9184	0.5510	0.9796	0.8980	0.9524
Click fraud	0.5578	0.8231	0.9252	0.8571	0.8980
Password	0.9456	0.5646	0.9932	0.8639	0.9388
User ID	0.9524	0.6259	1.0000	0.7891	0.9524
Firewall	0.8844	0.5102	0.9660	0.8707	0.8707
Back door	0.7143	0.7347	0.9660	0.8571	0.8571
Blacklist	0.8027	0.6735	0.9592	0.8571	0.9116
Spoofing	0.6463	0.8912	0.9592	0.8231	0.8844
Dropper	0.8231	0.9524	0.8231	0.7823	0.9184
Address book	0.9796	0.8503	0.9660	0.6939	0.9592
Honey pot	0.9388	0.8776	0.9524	0.7211	0.9592

Table 2- E*xpert distances* ($D_{NE}(M, R)$) between security risks and mental models

	Criminal	Physical	Medical	Market	Warfare
Adware	0.9333	0.9600	0.9600	0.6267	0.9600
Spyware	0.6533	0.9600	0.9333	0.8933	1.0000
Phishing	0.4800	0.9733	0.9333	0.9333	0.9600
Identity theft	0.3467	0.9467	0.9333	0.9733	0.9467
Spam	0.8933	0.9733	0.9733	0.7067	0.9200
Hijackers	0.4133	1.0000	0.9467	1.0000	0.7600
Cookies	0.9333	0.9467	0.9067	0.7067	0.9733
DoS attack	0.7200	0.9200	0.9067	0.9067	0.7867
download	0.7733	0.9467	0.9733	0.8267	0.9200
Trojan	0.7467	0.9733	0.8400	0.9067	0.7867
Keystroke	0.7333	0.8400	0.9200	0.8800	0.9733
Junk mail	0.9200	0.9600	1.0000	0.6133	0.9067
Virus	0.8800	0.9733	0.5733	0.8933	0.8667
Worm	0.8267	0.9333	0.8133	0.8933	0.9067
Hacking	0.6400	0.9733	0.9867	0.9733	0.7067
Binder	0.9467	0.8800	0.9067	0.6667	0.9467
Exploit	0.7200	0.9333	0.9467	0.7067	0.8800
Zombie	0.8933	0.9733	0.7600	0.8667	0.9067
Authentication	0.9733	0.6533	0.9067	0.8800	1.0000
Click fraud	0.4533	0.9733	0.8933	0.9733	0.9600
Password	0.9733	0.7333	0.9333	0.8800	0.9733
User ID	0.9867	0.7867	0.9600	0.8000	0.9733
Firewall	0.9733	0.5600	0.9867	0.8667	0.9467
Back door	0.7200	0.7333	0.9600	0.9200	0.9067
Blacklist	0.8533	0.8267	1.0000	0.8000	0.9733
Spoofing	0.5600	0.9200	0.9467	0.9867	0.9467
Dropper	0.8400	0.9467	0.7867	0.8667	0.9067
Address book	0.9333	0.8933	0.9333	0.6800	1.0000
Honey pot	0.9200	0.8933	0.9867	0.7600	0.9733

Table 3. List of all the words used in pile sorting experiment

Crime	Medical	Physical security	Warfare	Economical	Security
Fingerprint	Cancer	Fence	Bombing	Distribute	Adware
Counterfeit	Detoxification	Door-lock	Attack	Exchange	Spyware
Robbery	Nausea	Shield	Destroy	Export	Phishing
Theft	Sore	Inviolability	War	Trade	Identity theft
Mugging	Inflammation	Invulnerability	Suicide	Advertise	Spam
Housebreaking	Fever		Terror	Endorse	Hijackers
Kidnapping	Illness			Stock	Cookies
Vandalism	Contagious			Risk	DoS attack
Injection	Epidemic				Drive-by-download
					Trojan
					Keystroke logger
					Junk mail
					Virus
					Worm
					Hacking
					Binder
					Exploit
					Zombie
					Authentication
					Click fraud
					Password
					User ID
					Firewall
					Back door
					Blacklist

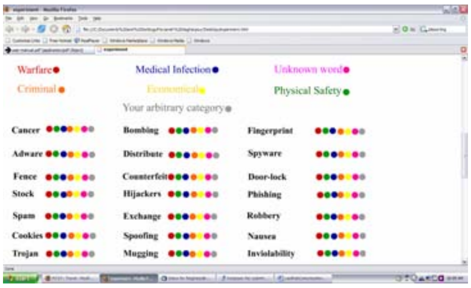
Table 4. *Obvious Mental Model*. Each word belongs to the related mental model with at least 75% probability.

Criminal	Medical	Physical security	Warfare	Economical
Theft	Epidemic	Fence	Bombing	Export
Housebreaking	Fever	Door-lock	Destroy	Trade
Kidnapping	Illness	Shield	War	Stock

Table 5. *Non-expert and expert mental models.* $P(NM_R)$ is the probability that the non-expert community select M as its mental model for the risk R. $P(EM_R)$ is defined similarly. These probabilities are chosen based on our original data entries for the non-expert and experts participants. For instance the probability of having "*physical security*" as the *non-experts' mental model* for the risk "*Adware*" is *0.28*. The reason for this is that 28% of non-experts have assigned this mental model to the risk "*Adware*".

	NM_R	$P(NM_R)$	EM_R	$P(EM_R)$	*Different MM*
Adware	Physical	28.57%	Market	28%	√
Spyware	Criminal	26.53%	Criminal	36%	
Phishing	Criminal	20.41%	Criminal	56%	
Identity Theft	Criminal	59.18%	Criminal	72%	
Spam	Criminal	22.45%	Market	16%	√
Hijackers	Criminal	48.98%	Criminal	72%	
Cookies	Market	6.12%	Market	16%	
DOS	Market	14.29%	Criminal	32%	√
Download	Criminal	10.20%	Criminal	24%	
Trojan	Criminal	22.45%	Criminal	28%	
Keystroke	Physical	28.57%	Criminal	24%	√
Junk Mail	Criminal	16.33%	Market	12%	√
Virus	Medical	38.78%	Medical	48%	
Worm	Criminal	26.53%	Medical	20%	√
Hacking	Criminal	36.73%	Criminal	40%	
Binder	Market	8.16%	Market	12%	
Exploit	Criminal	28.57%	Market	20%	√
Zombie	Warfare	18.37%	Medical	28%	√
Authentication	Physical	51.02%	Physical	36%	
Click Fraud	Criminal	42.86%	Criminal	60%	
Password	Physical	48.98%	Physical	28%	
User ID	Physical	38.78%	Physical	24%	
Firewall	Physical	46.94%	Physical	48%	
Back Door	Criminal	20.41%	Criminal	28%	
Blacklist	Physical	36.73%	Market	12%	√
Spoofing	Criminal	32.65%	Criminal	52%	
Dropper	Market	10.20%	Medical	20%	√
Address Book	Market	4.08%	Market	12%	
Honey Pot	Market	12.24%	Market	8%	

Appendix B: Figures



Warfare ● Medical Infection ● Unknown word ●

Criminal ● Economical ● Physical Safety ●

Your arbitrary category ●

Cancer	●●●●●●●	Bombing	●●●●●●●	Fingerprint	●●●●●●●
Adware	●●●●●●●	Distribute	●●●●●●●	Spyware	●●●●●●●
Fence	●●●●●●●	Counterfeit	●●●●●●●	Door-lock	●●●●●●●
Stock	●●●●●●●	Hijackers	●●●●●●●	Phishing	●●●●●●●
Spam	●●●●●●●	Exchange	●●●●●●●	Robbery	●●●●●●●
Cookies	●●●●●●●	Spoofing	●●●●●●●	Nausea	●●●●●●●
Trojan	●●●●●●●	Mugging	●●●●●●●	Inviolability	●●●●●●●

Fig. 1. Screenshot of the pile sorting experiment

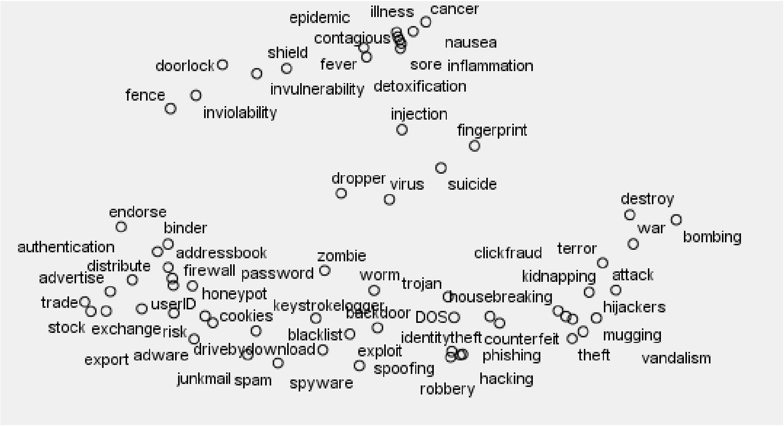


Fig. 2. Multidimensional Scaling Map for *ECM*

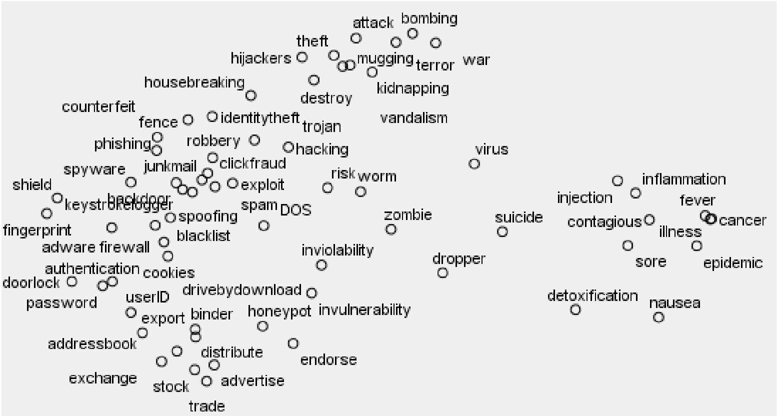


Fig. 3. Multidimensional Scaling Map for *NCM*

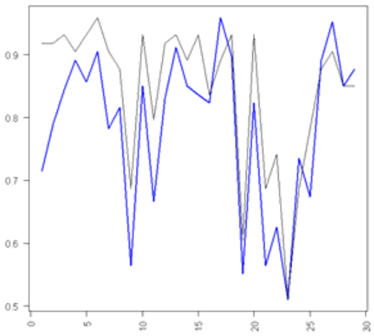


Fig. 4. Graph of Distance between each security risk (horizontal axis) and “Physical safety” mental model, blue=non-expert, black=experts. The vertical coordinate of the edges of the graphs represent the distance between the security risk and “physical safety” mental model