

When Organization Reply? Contentful and Emotional Factors Affect Management Response in Online Hacker Community

Yuanhong Ma, Beihang University, Beijing, China, yuanhongma@buaa.edu.cn

Liangqiang Li, Sichuan Agricultural University, Chengdu, China, lilq@sicau.edu.cn

Zhong Yao, Beihang University, Beijing, China, iszhao@buaa.edu.cn

Jing Zhang, Harbin Institute of Technology, Harbin, China, zhangjing20@hit.edu.cn

Yunzhong Cao, Sichuan Agricultural University, Chengdu, China, caoyz@sicau.edu.cn

Abstract

The growing cyber attacks and information security breaches make it necessary to explore the engagement of online hacker community. Applying the Fear Appeals Model and Protection Motivation Theory, this study examines the effect of exposure pressure and content quality on the management response for the voluntary vulnerability disclosure report in the online hacker community. The results show that the exposure pressure and content quality have a significantly positive effect on management response, while the exposure pressure has a greater influence than the content quality. Moreover, we build an emotion recognition approach using a word2vec-based LSTM algorithm. Based on the recognition outcome, we test the direct and moderating affect of emotional cues which are embedded in vulnerability disclosure report on management response. The results show that the effect of emotional cues on management response decision is limited. Finally, we also discuss the unexpected insignificance of emotional cues with rationalism and skepticism.

Proceedings of 2020 IFIP 8.11/11.13 Dewald Roode Information Security Research Workshop

Keywords: Online hacker community, Exposure pressure, Management response, Emotional cues, Text mining

1. Introduction

The increased incidence of cyber attacks and information security breaches have long become a major concern over the past few years. According to the Symantec Internet Security Threat Report(Symantec, 2019), the world's largest civil threat intelligence network, they intercept 142 million cyber attacks a day on average. As many as 86% of organizations in the world have experienced at least one cyber attack. The cyber attack has become one of the main reasons for the organization's loss. However, the increased system complexity and technical reliance on third-party parts (e.g., cloud services, open APIs, external programming libraries) make it difficult for in-house IT experts to perform sufficiently extensive and timely vulnerability discovery(Al-Banna et al., 2018). A number of organizations have opted for crowdsourced approaches to vulnerability discovery. There emerge two mainstreams crowdsourced approaches based on the existence of pre-promised rewards, namely, vulnerability rewards programs(VRPs) and voluntary vulnerability disclosure programs(VDPs) (Zhao et al., 2014).

The Vulnerability Rewards Programs(VRPs), also called bug bounty programs, which crowdsource their software security in public, as well as preset incentives to drive engagement and new bug discoveries(Maillart et al., 2017), such as Google, Facebook, and Github. The voluntary vulnerability disclosure programs(VDPs) refer that the benign hackers, also called white hat hackers, hunt for vulnerabilities and notify important stakeholder organizations spontaneously, as well as disclose their findings to public platforms, such as online hacker community(Zhao et al., 2014), e.g. Wooyun, HackerOne, BugCrowd, and Cobalt. The engagement motivation of the hackers on VDPs may be due to acquiring virtual gifts, reputation, social needs, self-fulfillment, knowledge sharing, and altruism in the

vulnerability discovery process. The organizational responses(i.e, management responses), who's vulnerabilities are disclosed in the online hacker community, and their attitudes are the key basis of these motivations. Therefore, the management response to vulnerability disclosure is important driver for the hackers' continuous engagement and community development.

Crowdsourced vulnerability discovery strategies have a few unique advantages, such as diversity of participant skills, high scalability, fast speed, and low cost (Al-Banna et al., 2018). Prior research has documented the importance of diversity in vulnerability discovery about the online hacker community, in which diversity encourages higher productivity of the vulnerability discovery process and more engagement behaviors(Zhao et al., 2014). Nevertheless, the diversity of participants might increase the managers' uncertainty to the vulnerability credibility due to the unsolicited vulnerability disclosure, which may expend the organization's resources to identify the authenticity of the vulnerability. Meanwhile, the online hacker community is not just a platform for vulnerability disclosure, it also exists crowd aggregation effect as a result of its crowd participants and social attribution. That is, the management response decision needs to consider the content quality of vulnerability disclosure and the exposure threat to other participants, such as their competitors and black hat hackers¹. For example, one hacker discloses the possible vulnerabilities in the online hacker community and discuss it with other hackers, but not send it to the firms directly. And the disclosed organization may perceive exposure pressure that the potential threat of their cyber may be attacked by the outgiving vulnerabilities they ignored. This may be the associated effect, or dark side, of VDPs for organizations. Besides, since the vulnerabilities

¹ The white hat is the hacker who can identify the security holes in the computer system or network system but do not exploit them maliciously. The black hat is the hacker who attack technological loopholes for illegal benefits. The grey hat is the hacker who may sometimes violate laws or typical ethical standards, but does not have the malicious intent typical of a black hat hacker.

were disclosed in the online hacker community, it's also unknown whether the emotional cues which are embedded in vulnerability disclosure report and dominated by social text will have impacts on management response decision. Above all, we try to address the following question:

- *Whether the content quality and exposure pressure of voluntary vulnerability disclosure influence management response decisions(i.e., response, timeliness, and rating) in the online hacker community?*
- *Whether the emotional cues which are embedded in vulnerability disclosure affect management response decision?*

According to the Fear Appeals Model (FAM), fear appeals refer to “persuasive messages designed to scare people by describing terrible things that will happen to them if they do not do what the message recommends”(Witte& Kim, 1992). The organizations might fear that the potential threat would happen if they overlooked the vulnerability of what the hacker disclosed in the online hacker community. Therefore, it's appropriate to be the theoretical support for exposure pressure. Closely related to FAM, Protection Motivation Theory(PMT) is a theory that was originally created to help clarify fear appeals. PMT proposes that people protect themselves based on four motivations: the perceived severity of a threatening event, the perceived probability of the vulnerability, the efficacy of the recommended preventive behavior, and the perceived self-efficacy(Rogers, 1975). The description of the vulnerability disclosure report depicts the severity and probability of vulnerability, so it may trigger organizations' motivation to protect their cybersecurity. Hence, we use PMT to support content quality.

Applying FAM and PMT, this study explores the management response to voluntary vulnerability disclosure in the online hacker community. First, we conducted an observational

study, spanning a four-year period from July 2010 to December 2013 and including 14,735 observations, to investigate the influence of exposure pressure(i.e., measured by the number of follow to a vulnerability report) and content quality(i.e., measured by the description length of vulnerability) on management response(i.e., response to disclosure, timeliness to disclosure, and rating to disclosure). The results indicate that the exposure pressure and content quality have prominent effects on management response, yet exposure pressure has a stronger impact than content quality. Second, to further elucidate the role of emotion, we mining the emotional cues which are embedded in the vulnerability report text by a machine learning method, i.e., Word2Vec based LSTM, and test their impacts on management response. The results show that the negative cues are much more than positive cues, but the emotional cues play a limited role in management response.

This research contributes to the existing literature as follows. First, we conducted a new perspective to unravel a negative effects in an online hacker community. Second, we reveal the role of emotional cues in the online hacker community invoking emotion as social information theory. Third, we applied a machine learning approach to extract emotional cues that are embedded in vulnerability text, which allows both practitioners and researchers to identify emotional cues on a large scale.

2. Empirical Model

On the basis of this cross-sectional data set, we used multiple regression models to examine the influence of exposure pressure and content quality on management response to voluntary vulnerability disclosure including response, timeliness, and rating. We let subscript i denote each disclosed vulnerability. Our independent variable, $response_i$ is evaluated by management response to disclosed vulnerability i . 0 if there is a reply, 1 if there is no reply. $Ln(Timeliness)_i$ is measured by taking the logarithm of duration from the time a hacker discloses a

vulnerability i to the time the organization responds, in particular, the value will be zero if the organization doesn't reply. $\ln(Rating)_i$ is the logarithm of score from organization to the disclosed vulnerability i , the value will be zero if the organization doesn't reply. As for dependent variables, $\ln(Follow)_i$ is used to measuring exposure pressure, which represents the logarithm of the number of follow to the disclosed vulnerability i . $\ln(DescriptionLength)_i$ is a classical indicator to evaluate content quality, which is calculated by the logarithmic value of the total word count in disclosed vulnerability i . We include three control variables. $UserLevel_i$ is represented by the level of user who discloses vulnerability i . $DamageLevel_i$ refers to the damage level that is given by the user who discloses vulnerability i . $Holiday_i$ refers to whether it is on national holiday or weekend in china when the user discloses vulnerability i . 0 if it's in a holiday, 1 if it's in a work day.

Equation (1):

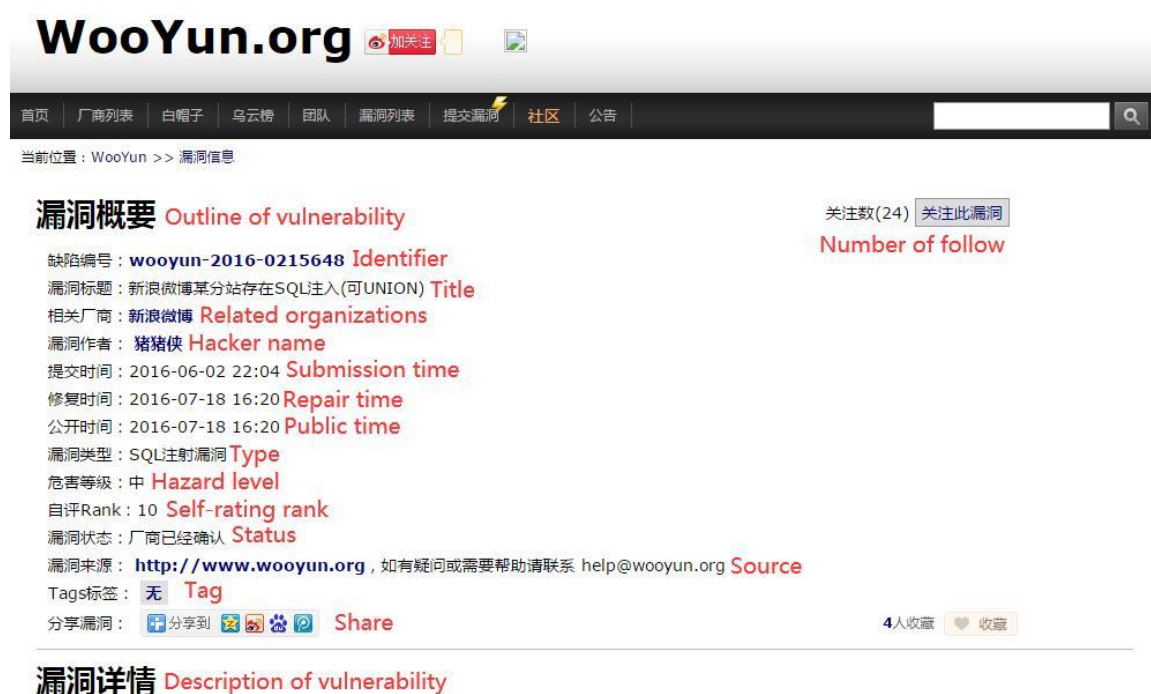
$$Reply^{\{Response, \ln(Rating), \ln(Timeliness)\}}_i = \beta_1 \ln(DescriptionLength)_i + \beta_2 \ln(Follow)_i + \lambda_1 UserLevel_i + \lambda_2 DamageLevel_i + \lambda_3 Holiday_i + \varepsilon_i$$

3. Empirical Analysis and Results

3.1 Data

To examine empirical model, we conduct a data set from Wooyun spanning from July 2010 to December 2013, including 767 hackers, 4211 organizations, and 14,735 vulnerability reports. Wooyun(wooyun.org) is a leading online hacker community in China which is launched in May 2010. Until 2015, it attracted 7,744 hackers who contributed 64,134 vulnerability reports related to 17,328 organizations(Zhao et al., 2015). Although it has been upgrading since July 2016, another similar online hacker community, namely HackerOne(Its online community is in Hacker Activity segment), in U.S. is burgeoning, which is launched in 2012, consist of approximately 200,000 researchers, and resolved 72,000 vulnerabilities until 2018.

Unlike HackerOne in some operational details, when a hacker finds a vulnerability, she/he can submit a disclosure report in the Wooyun community. After inspecting the report, Wooyun will inform the organization's administrators about the vulnerable object. Then, the organization will check and decide whether to respond and evaluate the report. And the progress will be disclosed to the public along with other hackers' engagement (i.e., comment, follow, or like) (Zhao et al., 2015). The common types of disclosed vulnerabilities include SQL injection, cross-site scripting (XSS), and logic errors/design flaws. The appearance of Wooyun community is shown in Figure 1.



The screenshot shows the Wooyun.org website interface. At the top is the header with the site name and navigation links. Below the header is a search bar and a list of categories. The main content area displays a vulnerability report with the following details:

- 漏洞概要** Outline of vulnerability
- 缺陷编号: **wooyun-2016-0215648** Identifier
- 漏洞标题: 新浪微博某分站存在SQL注入(可UNION) Title
- 相关厂商: 新浪微博 Related organizations
- 漏洞作者: 猪猪侠 Hacker name
- 提交时间: 2016-06-02 22:04 Submission time
- 修复时间: 2016-07-18 16:20 Repair time
- 公开时间: 2016-07-18 16:20 Public time
- 漏洞类型: SQL注入漏洞 Type
- 危害等级: 中 Hazard level
- 自评Rank: 10 Self-rating rank
- 漏洞状态: 厂商已经确认 Status
- 漏洞来源: <http://www.wooyun.org>, 如有疑问或需要帮助请联系 help@wooyun.org Source
- Tags标签: 无 Tag
- 分享漏洞: 分享到 Share
- 关注数(24) 关注此漏洞 Number of follow
- 4人收藏 收藏

Below the report details is a section for **漏洞详情** Description of vulnerability.



Figure 1. The Screenshot of Wooyun community

As shown in Table 1, the descriptive statistics of the main variables. *Follow*, *Length*, *Rating*, *Timeliness* are log-scaled. We also added one to these variables to avoid logarithms of zeroes.

Table 1. Descriptive Statistics of Main Variables

Variable	N	Mean	SD	Min	Max
User Level	14,735	1.789	0.576	0	3
Damage Level	14,735	2.24	0.763	1	3
Holiday	14,735	0.816	0.388	0	1
Ln(Follow)	14,735	1.203	1.081	0	5.505
Ln(Length)	14,735	5.327	1.445	0	11.093
Response	14,735	0.713	0.452	0	1
Ln(Rating)	14,735	1.523	1.074	0	2.996
Ln(Timeliness)	14,735	1.951	1.864	0	5.985

3.2 Empirical Results

Table 2 present the correlation matrix of our main model variables based on our 14,735 observations.

Table 2. Correlation Matrix

Variable	1	2	3	4	5	6	7	8
1.User Level	1							
2.Damage Level	0.053	1						
3.Holiday	-0.013	0.004	1					
4.Ln(Follow)	0.128	0.196	-0.088	1				
5.Ln(Length)	0.120	0.113	-0.021	0.229	1			
6.Response	0.104	0.086	0.003	0.078	0.006	1		
7.Ln(Rating)	0.119	0.255	0.007	0.150	0.053	0.900	1	
8.Ln(Timeliness)	-0.013	0.032	-0.111	0.003	-0.030	0.597	0.530	1

The regression results are reported in Table 3. The Model 1, Model 3, and Model 5 report the effect of control variables, including *Damage Level*, *Holiday*, *User Level*, on independent variables (i.e., *Response*, *Ln(Rating)*, *Ln(Timeliness)*) respectively. The *ln(follow)*, which represents exposure pressure, have a significant influence on *Response* ($\beta = 0.024$, $p < 0.001$) and *Ln(Rating)* ($\beta = 0.094$, $p < 0.001$), but unremarkable influence on *Ln(Timeliness)* ($\beta = 0.01$, $p = 0.516$). The *ln(length)*, which represents content quality, have a significant influence on *Response* ($\beta = -0.008$, $p < 0.05$) and *Ln(Timeliness)* ($\beta = -0.043$, $p < 0.001$), but unremarkable influence on *Ln(Rating)* ($\beta = -0.004$, $p = 0.487$). Especially, the significant coefficient of *Ln(follow)* is bigger than *Ln(length)*, which shows a stronger impact of exposure pressure than content quality on management response.

Table 3. Empirical Estimation Results

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Damage Level	0.048*** (0.005)	0.043*** (0.005)	0.081*** (0.020)	0.093*** (0.020)	0.351*** (0.011)	0.327*** (0.011)
Holiday	0.005 (0.010)	0.010 (0.010)	-0.535*** (0.039)	-0.540*** (0.039)	0.022 (0.022)	0.044 (0.022)
User Level	0.078*** (0.006)	0.076*** (0.006)	-0.054* (0.027)	-0.039 (0.027)	0.197*** (0.015)	0.178*** (0.015)
Ln(follow)		0.024*** (0.004)		0.010 (0.015)		0.094*** (0.008)
Ln(length)		-0.008** (0.003)		-0.043*** (0.011)		-0.004 (0.006)
Constant	0.462*** (0.018)	0.489*** (0.021)	2.301*** (0.073)	2.497*** (0.087)	0.366*** (0.041)	0.346*** (0.048)
R ² / Adjust R ²	0.017/0.017	0.021/0.020	0.014/0.013	0.015/0.015	0.076/0.076	0.084/0.084
F	86.49***	61.95***	67.86***	44.34***	404.36***	271.40***
N	14,735	14,735	14,735	14,735	14,735	14,735
Dependent variables	Response	Response	Ln(Timeliness)	Ln(Timeliness)	Ln(Rating)	Ln(Rating)

4. The Emotional Cues Embedded in Vulnerability Disclosure Report

Affective-as-information theory state that affect serves informational function and it is used as a kind of heuristic information for making evaluative judgments(Foo et al., 2009). So we further mining emotional cues that are embedded in the disclosure report with a machine learning method to explore management affective response. That is, the positive or negative emotions may affect management response reaction to vulnerability disclosure report

compared to the emotionless expression.

4.1 Emotion Recognition with Word2Vec-based LSTM

We used a supervised text mining method to recognize the emotion labels as positive, neutral, and negative with 1065 manually labeled data which is tagged by human intelligence. Before we selected the Word2Vec-based Long Short-Term Memory(LSTM), which is a kind of Recurrent Neural Network(RNN) that fuses time series features, we compared some common text mining methods. Specially, we applied four text feature extraction methods(namely, Bag-of-Word(BOW), Term Frequency-Inverse Document Frequency(TFIDF), Latent Dirichlet Allocation(LDA), and Word-to-Vector(Word2Vec)) with five classical classification algorithms, i.e., Naive Bayes, Logistic Regression, KNeighbors Classifier Random Forest Classifier, and Decision Tree Classifier.

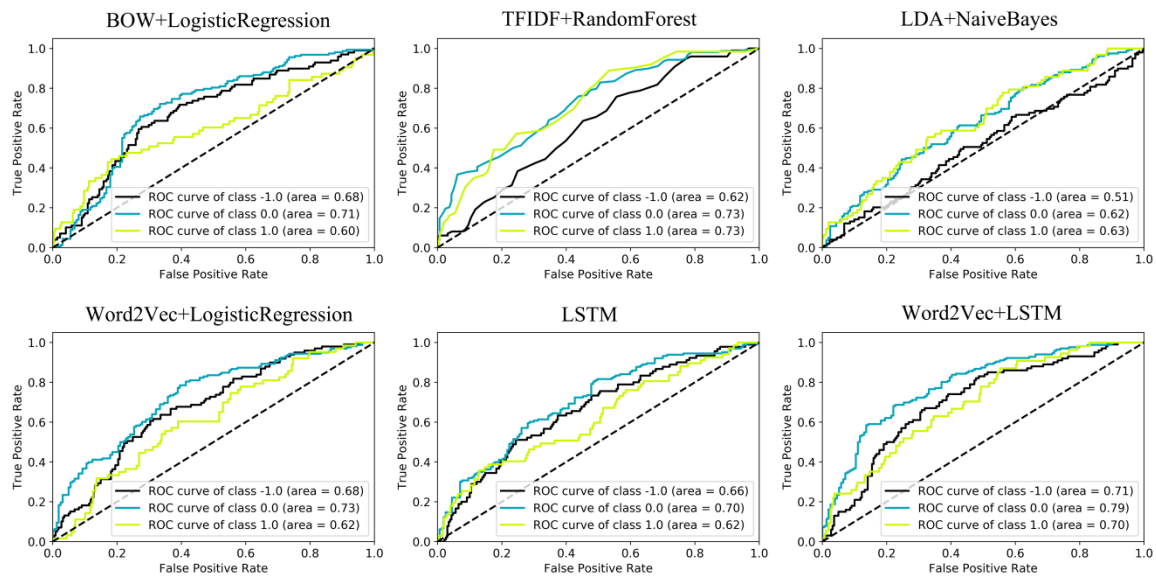


Figure 2. The ROC Curve of Different Algorithm

Finally, we report the results of four methods that perform best under each feature extraction

method with twenty possible combinations, i.e., BOW-based Logistic Regression, TFIDF-based Random Forest Classifier, LDA-based Naive Bayes, and Word2Vec-based Logistic Regression. As shown in Figure 2, it displays the Receiver Operating Characteristic(ROC) curve in which the closer the curve is to the upper left corner, the better the classification method is. Figure 2 also presents the results of pure LSTM and Word2Vec-based LSTM while we opt for the latter as our final emotional recognition method. In particular, we only removed meaningless interference symbols(such as HTML tag and punctuation) and did not use frequently-used stop words list, because, in sentiment analysis, some characters which were considered meaningless might affect the emotional tendency of the sentence(such as negative adverbs and degree words).

As shown in Figure 3, we observe a negative reporting bias in the vulnerability disclosure report, i.e., in addition to neutral statements, the hackers express more negative emotion than positive emotion in the disclosure report. The motivation of the over-expression of negative emotions negative is maybe to emphasize the severity of the vulnerability and affect management response.

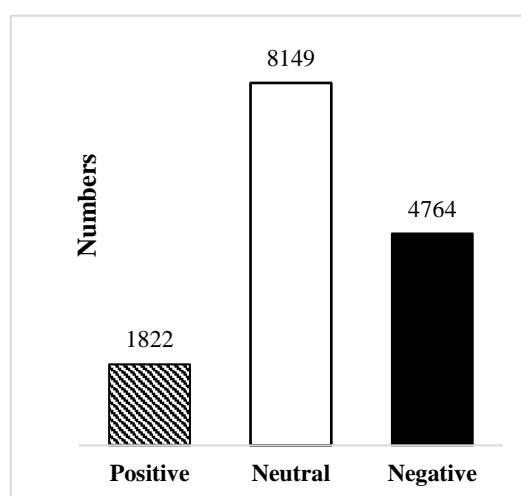


Figure 3. The Distribution of Emotional Cues

4.2 The Role of Emotional Cues on Management Response

Using Equation (1) as our baseline model, we add emotional cues into the model as a direct and moderating role on management responses, as shown in Equation (2). Borrowing from the results of sentiment analysis above, we test the direct effect of emotional cues(i.e., positive, neutral, and negative) and the interaction effect of emotional cues with exposure pressure(i.e., $\ln(Follow)$) and content quality(i.e., $\ln(DescriptionLength)$) on management response(i.e., $Response$, $\ln(Rating)$, $\ln(Timeliness)$).

Equation (2):

$$Reply^{(Response, \ln(Rating), \ln(Timeliness))}_i = \beta_1 \ln(DescriptionLength)_i + \beta_2 \ln(Follow)_i + \beta_3 Emotion_i + \beta_4 Emotion_i * \ln(DescriptionLength)_i + \beta_5 Emotion_i * \ln(Follow)_i + \beta_6 UserLevel_i + \beta_7 DamageLevel_i + \beta_8 Holiday_i + \epsilon_i$$

Table 3 shows the results of emotional cues on management response. Only two relationships are significant. Firstly, emotional cues have a significantly positive effect on management response($\beta = 0.012$, $p < 0.01$). That is, the more positive the emotion in vulnerability disclosure report, the more likely it is to get management response. Then, the moderating effect of emotional cues and exposure pressure(i.e., $Emotion * \ln(Follow)$) on management rating is significantly positive($\beta = 0.022$, $p < 0.01$). That is, the emotional cues positively regulate the effect of exposure pressure on management rating, in other words, the more positive the emotion in vulnerability disclosure report is, the greater the effect of exposure pressure on management rating will be.

We try to explain the unexpected insignificance of emotional cues with rationalism and skepticism. The rationalism means that, compared to individuals, the managers are more likely to engage in a rational mindset and make response decision with overlooking the emotional cues. The skepticism refers that the emotion-embedded vulnerability disclosure reports may trigger the manager's skeptical psychology when they run into the emotional reports with persuasive intention and doubt hackers' motivation for the vulnerability disclosure.

Table 3. Results of Emotional Cues on Management Response

Variables	Baseline	Model 7	Baseline	Model 8	Baseline	Model 9	Model 10
Damage Level	0.043*** (0.005)	0.043*** (0.005)	0.093*** (0.020)	0.092*** (0.020)	0.327*** (0.011)	0.326*** (0.011)	0.271*** (0.011)
Holiday	0.010 (0.010)	0.010 (0.010)	-0.540*** (0.039)	-0.541*** (0.045)	0.044 (0.022)	0.044 (0.022)	-0.242*** (0.022)
User Level	0.076*** (0.006)	0.075*** (0.006)	-0.039 (0.027)	-0.039 (0.027)	0.178*** (0.015)	0.177*** (0.015)	0.219*** (0.015)
Ln(follow)	0.024*** (0.004)	0.025*** (0.004)	0.010 (0.015)	0.010 (0.015)	0.094*** (0.008)	0.094*** (0.008)	
Ln(length)	-0.008** (0.003)	-0.008** (0.003)	-0.043*** (0.011)	-0.044*** (0.011)	-0.004 (0.006)	-0.005 (0.006)	
Emotion		0.012** (0.006)		0.017 (0.024)		0.019 (0.014)	-0.042** (0.014)
Emotion*Ln(follow)		0.004		0.002		0.022**	

		(0.004)		(0.018)		(0.010)	
		0.004		-0.026		0.011	
Emotion*Ln(length)		(0.005)		(0.022)		(0.012)	
Constant	0.489***	0.494***	2.497***	2.501***	0.346***	0.353***	0.394***
	(0.021)	(0.021)	(0.087)	(0.088)	(0.048)	(0.049)	(0.041)
R ² / Adjust R ²	0.021/0.020	0.021/0.021	0.015/0.015	0.015/0.015	0.084/0.084	0.085/0.085	0.061/0.060
F	61.95***	39.68***	44.34***	27.93***	271.40***	170.95***	237.03***
N	14,735	14,735	14,735	14,735	14,735	14,735	14,735
Dependent variables	Response	Response	Ln(Timeliness)	Ln(Timeliness)	Ln(Rating)	Ln(Rating)	Ln(Follow)

5. Conclusion

In this study, we examine the effect of exposure pressure and content quality on the management response about the voluntary vulnerability disclosure report in the online hacker community. The results show that the exposure pressure and content quality have a significantly positive effect on management response decision, while the exposure pressure has a greater influence than the content quality. Furthermore, we build an emotion recognition approach using a word2vec-based LSTM algorithm. And based on the outcomes, we test the direct and moderating affect of emotional cues in vulnerability disclosure reports on management response. The results show that the effect of emotional cues on management response decision is limited.

Our conclusions are instructive for practitioners and researchers on the online hacker community. Future works can conduct more state-of-the-art text mining methods to extract the emotional cues. Other online hacker communities can also be studied to enhance reliability and generalization.

References

- Al-Banna, M., Benatallah, B., Schlagwein, D., Bertino, E., & Barukh, M. C. (2018). Friendly Hackers to the Rescue: How Organizations Perceive Crowdsourced Vulnerability Discovery. In PACIS (p. 230).
- Foo, M. D., Uy, M. A., & Baron, R. A. (2009). How do feelings influence effort? An empirical study of entrepreneurs' affect and venture effort. *Journal of Applied Psychology*, 94(4), 1086.
- Maillart, T., Zhao, M., Grossklags, J., & Chuang, J. (2017). Given enough eyeballs, all bugs are shallow? Revisiting Eric Raymond with bug bounty programs. *Journal of Cybersecurity*, 3(2), 81-90.
- Rogers, R. W. . (1975). A protection motivation theory of fear appeals and attitude change¹. *The Journal of Psychology Interdisciplinary and Applied*, 91(1), 93-114.
- Witte, & Kim. (1992). Putting the fear back into fear appeals: the extended parallel process model. *Communication Monographs*, 59(4), 329-349.
- WEI, T. Y., Wang, Q. H., & Hui, K. L. (2019). See no evil, hear no evil? Dissecting the impact of online hacker forums. *MIS Quarterly*, 43(1), 73.
- Zhao, M., Grossklags, J., & Chen, K. (2014, November). An exploratory study of white hat behaviors in a web vulnerability disclosure program. In *Proceedings of the 2014 ACM workshop on security information workers* (pp. 51-58).