Modeling Public Response to Data Breaches

Completed Research Full Paper

Eric Bachura
University of Texas at San Antonio
eric.bachura@utsa.edu

Rohit Valecha
University of Texas at San Antonio
rohit.valecha@utsa.edu

Rui Chen
Iowa State University
ruichen@iastate.edu

H. Raghav Rao
University of Texas at San Antonio
hr.rao@utsa.edu

Abstract

In this document we describe a theoretical approach to modeling public emotional response cycles to crisis events. We also provide a preliminary theoretical approach to modeling crisis communication propagation that is counter intuitive to existing belief and literature regarding emotionally charged language and discussion artifact dispersion. The data set used to test these theories is contextualized by the OPM data breach of 2015 and consists of twitter data corresponding to the ensuing discussion following public notification that the breach had occurred. The resulting analysis reveals that an adapted Kübler-Ross model fits the aggregated public emotional response cycle and that emotionally charged language is negatively associated with messages the disperse more than average.

Keywords

Data breach, grief, emotional response, sentiment analysis, text analysis, OPM, crisis dialogue.

Introduction

Data breaches have resulted in over 900 million individual records being compromised since 2005 ("Privacy Rights Clearinghouse," 2016). This value is expected to more than double over the next five years (Wheatley, Maillart, & Sornette, 2016). The impact of these breaches has been of ongoing interest to the academic research community. While much of this research has focused on the economic costs of data breaches (Acquisti, Friedman, & Telang, 2006; Layton & Watters, 2014; Roberds & Schreft, 2009) others have examined trends in data breaches (Amorosi, 2011; Garrison, & Ncube, 2011; Gupta & Sharman, 2012; Holtfreter & Harrington, 2015) and practitioner oriented prevention and mitigation techniques (Brown, 2016; Bush, 2016; Fowler, 2016; German, 2016). What is missing in the existing body of research is an examination of the public and victim group reactions to data breaches. Developing prevention and mitigation techniques in the absence of a clear understanding of public and victim group responses leaves open the possibility of inadequate or poorly timed implementations of any mitigation approach that is developed. Furthermore, because reports indicate that fewer than half of those affected by data breach events take advantage of mitigation offers, such as credit monitoring offered by the breached organizations (PonemonInstitute, 2014), an understanding of why victims fail to take advantage of these protection mechanisms could lead to solutions designed to change this behavior. This in turn would lead to reductions in the economic costs associated with post-breach fraud. Understanding the reaction process from both the victim-group and public viewpoints allows organizations to understand and anticipate reactions to breach events. Such an understanding could inform public relation efforts designed to both meet the reacting groups' expectations while also protecting the organization's reputation. Though there have been many data breach events that have been widely discussed over a number of public forums, few have reached the level of awareness in the public dialogue as the data breach of the Office of Personnel Management (OPM) in 2015. The recent nature, scale, and scope of the OPM data breach presents an opportunity for researchers to explore the various dimensions of reaction to data breaches.

The OPM Data Breach

The 2015 OPM data breach currently stands as the single most impactful data breach in terms of personally identifiable information (PII) that was compromised with over 5.6 million fingerprint records stolen, 4.2 million personnel files, and 21.5 million individual background investigation files stolen (Chaffetz, Meadows, & Hurd, 2016). It is important to note that the 21.5 million individual background investigation files consist of data than can be found on what is known as an SF-86. This form is filled out by those seeking to obtain a security clearance. A copy of this form can be found here: https://www.gsa.gov/portal/forms/download/116390. Information found in these files goes well beyond what is normally considered traditional PII, such as social security numbers (SSN), home address, phone numbers, and email addresses. These forms contain a wide variety of information regarding the friends and family members of the individual requesting the clearance. As a result, the number of people impacted by this data breach extends into the broader public audience. Even if the PII that lends itself to fraud abuse is largely limited to the 21.5 million individuals whose records were stolen, the social network data represented in those records could potentially place many more people at risk of being targeted for criminal activity. This is demonstrated in FBI Director James Comey's remark regarding the OPM data breach when he said "My SF86 lists every place I've ever lived since I was 18, every foreign travel I've ever taken, all of my family, their addresses. So it's not just my identity that's affected. I've got siblings. I've got five kids. All of that is in there." (Chaffetz, Meadows, & Hurd, 2016).

Public notification of the OPM data breach occurred on June 4th, 2015. The two months following the public announcement saw a series of news events revealing increasing details about the size and scope of the data that was compromised as well as the events that lead to the data breach. The end of this two-month period following the initial announcement was marked by the resignation of OPM Director Katherine Archuleta following the acknowledgement that 21.5 million records of background investigation data were compromised. It took more than a year of investigation to determine the full nature of the circumstances leading up to the data breach, which began with warnings from the OPM Inspector General that the information stored and managed by OPM was vulnerable to hacking attempts as early as 2005 (Chaffetz, Meadows, & Hurd, 2016). This lack of responsiveness is pervasive throughout the entirety of the OPM data breach. It was not until six months after discovering data had been stolen that a public notification was made and it took two full months after that to release full information on the number of people affected and the type of data taken. The following is an outline of the events leading up to the OPM data breach notification (Chaffetz, Meadows, & Hurd, 2016):

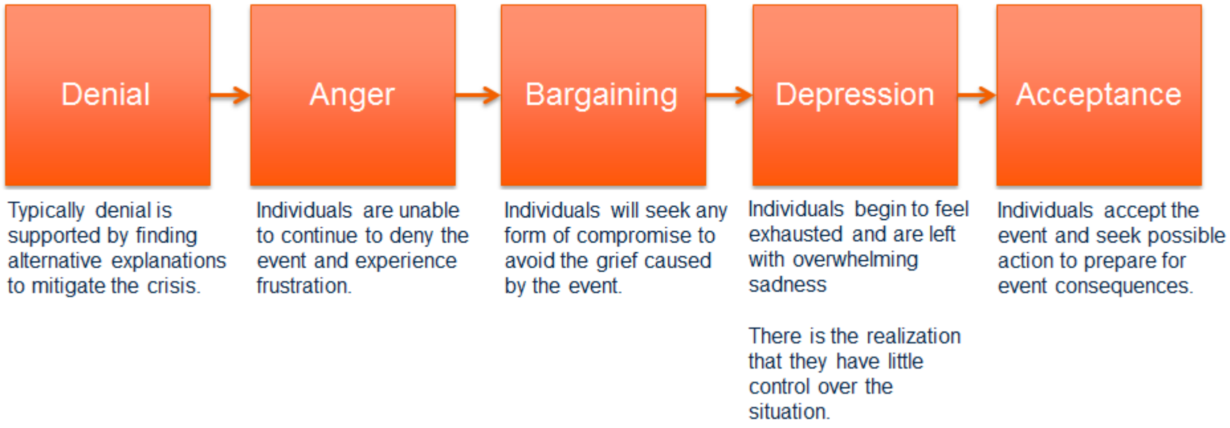
- 2005 to 2014 Inspector General rates OPM information security low and vulnerable to hackers
- 2012 first evidence of impending breach from malware installed on OPM server
- 2013 first active hacking efforts targeting OPM
- 2014 Inspector General upgrades OPM security issues from "material weakness" to "significant deficiency"
- 2014 OPM made aware of initial data loss by US-CERT
- 2014 OPM increases cybersecurity defensive activity but fails to disclose scope of data loss thus far to Senate committee
- 2014 hackers increase data exfiltration efforts, successfully targeting background investigation data and PII of all federal employees
- 2015 public notification occurs

Theories

This research effort provides preliminary evidence supporting two proposed theories seeking to: 1) Understand the public emotional response cycle with regard to crisis events such as data breaches; and 2) Model a possibly counterintuitive dialogue propagation process that frames information as a commodity.

Theory of Emotional Response to Crisis Events

The first of these is the theory of emotional response to crisis events. This theory is an adaptation of the Five Stages of Grief theory (Kübler-Ross, 2009). The Five Stages of Grief theory is often referred to as the Kübler-Ross stages of grief and can be described with the following figure:



Typically denial is supported by finding alternative explanations to mitigate the crisis.

Individuals are unable to continue to deny the event and experience frustration.

Individuals will seek any form of compromise to avoid the grief caused by the event.

Individuals begin to feel exhausted and are left with overwhelming sadness

There is the realization that they have little control over the situation.

Individuals accept the event and seek possible action to prepare for event consequences.

Figure 1 – Kübler-Ross Five Stages of Grief

We theorize that this model, originally proposed to explain behaviors observed in individuals facing individual crisis events (Kübler-Ross, 2009), can be adapted to explain the aggregate public response to a public crisis, which in this case is a data breach event. We approach the adaptation of this model by considering how each stage represented in the original model would be represented in a model designed to specifically model the emotional cycle of public response to a major crisis such as a data breach. First, it is expected that due to the overwhelming evidence and wide reporting of breaches such as the OPM data breach, the public will be unlikely to demonstrate denial of the event. Instead it is expected that there would be a brief period of uncertainty about the ramifications and context of the crisis at the early stages of reaction which would be indicated by heightened anxiety. Next it is expected that anger would be a prominent emotional response stage in the publics' reaction. The original theory describes the onset of anger as the result of frustration and an inability to continue denial (Kübler-Ross, 2009). Since the adapted model has replaced denial with anxiety and because anxiety is largely a function of uncertainty, once uncertainty begins to diminish, which is likely to happen quickly given the nature of modern communications and the 24-hour news cycles, anger will likely set in quickly. The bargaining stage from the original model relies on the individual's ability to develop a cognitive belief that they have bargaining power. However, in crisis events such as the OPM data breach, individuals have virtually no bargaining power nor is it likely they have the power to engage an organization in any negotiations regarding the effects of the data breach. Furthermore, crisis events such as data breaches are, by their very nature, well beyond any point of prevention. As a result, there is no expected conceptual alternative or replacement for the bargaining stage, thus it is dropped from the adapted model. The depression stage is one characterized by immense sadness. In the original Kübler-Ross model, this stage is the result of individuals having exhausted alternative emotional reactions. This stage is expected to be expressed in the adapted model through its primary emotional measure, sadness. The final stage of acceptance is an ideal end to the emotional cycle. It is important to recognize that in the original model this stage is not necessarily indicative of an end or even a solution to the underlying problem, but that it indicates an emotional shift from sadness and depression towards acceptance and resolutions that are both possible and realistic (Kübler-Ross, 2009). In the case of the adapted model, this would be expressed through a reduction in negative emotions and for those directly affected a move to seek mitigation and protection efforts. The adapted model can be seen in the figure below:

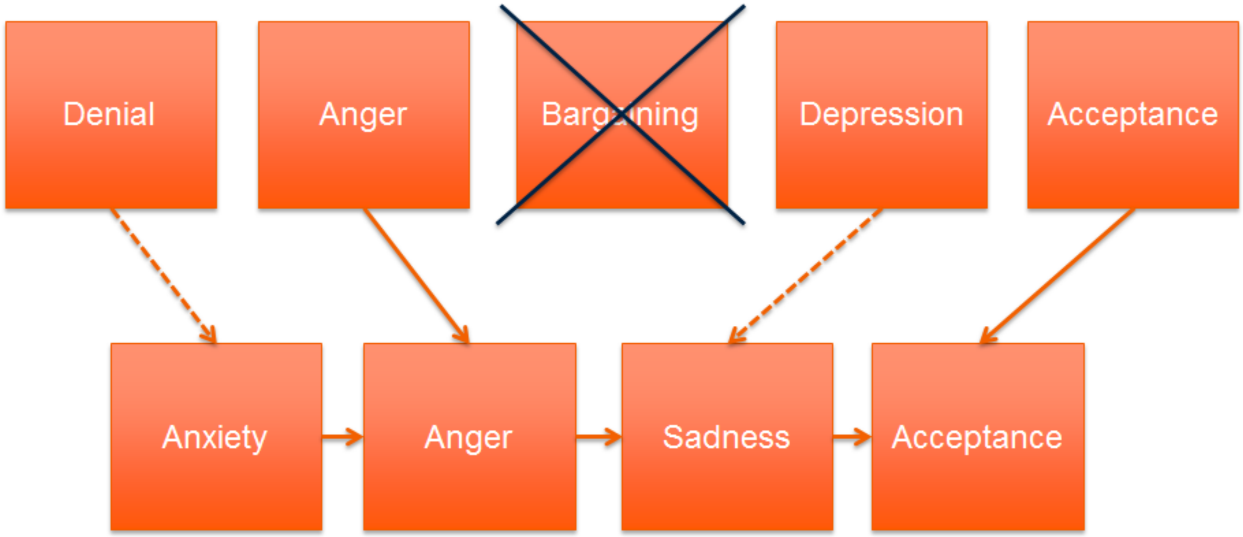


Figure 2 – Kübler-Ross Adapted Public Emotional Response Model

Evidence of this adapted model serving as a good explanation of the emotional response process to crisis events should be evidenced through a sentiment analysis of the ongoing dialogue associated with the breach. As time progresses, the aggregate sentiment values that compose dialogue activity should shift from the initial stages to the later stages.

Theory of Crisis Dialogue

Existing research indicates that messages in a group discussion medium such as twitter will have a higher likelihood of being copied and spread, furthering their influence, if they have a higher sentiment index, particularly of negative sentiments (Hansen, Arvidsson, Nielsen, Colleoni, & Etter, 2011). This aligns well with the conventional wisdom that there is no such thing as bad news (Burden, 2002) and that irrational and even incorrect information spreads faster than truth (Zubiaga, Liakata, Procter, Hoi, & Tolmie, 2016). The theory presented here stands in direct contrast to the existing research and could be considered counterintuitive at first glance. Discussions contextualized by crisis events contain two primary commodities: 1) Emotional expression; and 2) Information. Of these two commodities, emotional expression is supplied by individuals and carries a unique intrinsic value for each individual. Spreading another participant's emotional expression is of limited value because each individual has their own emotional expression and because it has lower intrinsic value since it is not their own personal expression. Meanwhile, unlike the abundant nature of emotional expression, information is limited. Thus, it is a higher value commodity due to both its scarcity and its inherent uncertainty reducing nature. The expectation is that a discussion contextualized by a crisis event will have participants seeking to express their own individual emotions corresponding to the data breach while simultaneously seeking information. Emotional expression serves as an outlet during each stage of the emotional cycle and facilitates progress along the emotional cycle as described in the theory of emotional response to data breaches. Information gathering reduces negative emotional responses and provides bounding parameters to the nature and effects of the crisis. As a result, it is expected that highly duplicated comments will contain lower emotional measures while comments that are not duplicated often will contain higher emotional measures. A figure of this theory can be seen below:

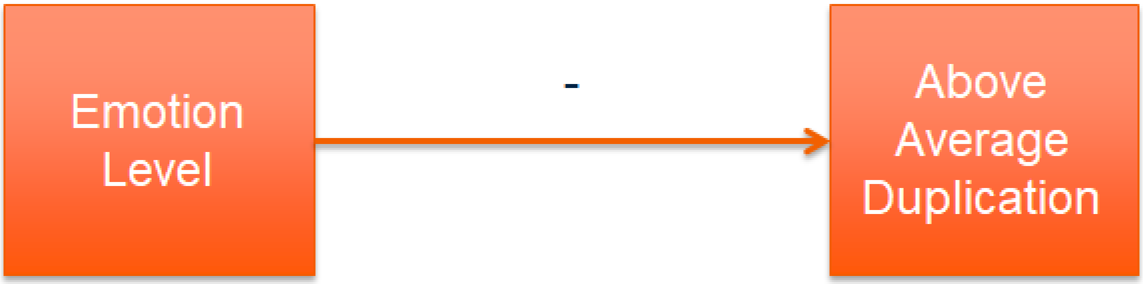


Figure 3 – Crisis Dialogue Theory

Evidence of the accuracy of this theory should be seen in the ability to discriminate highly duplicated comments from those not highly duplicated on their emotional scores. This would also be evidenced in significantly different mean scores across the two groups (highly duplicated versus not highly duplicated).

The Dataset

Within a day of the public announcement that OPM had been breached the hashtag #OPMHack started being used on Twitter. The nature of Twitter as a social dialogue medium, combined with the restricted tweet length of 140 characters, presents an ideal opportunity to conduct sentiment analysis of individual comments regarding the OPM data breach and aggregate them to determine the group sentiment levels. The reason the short nature of tweets is helpful in doing this is because it allows for discrete sentiment measures per comment. Using twitter data also has the advantage of using the retweet count of a message as a measure of its duplication count. The ideal timeframe to constrain the dataset is from the point of public notification up until the end of the month containing the resignation of OPM Directory Katherine Archuleta. What makes this timeframe ideal is the clustering of multiple news events within those first two months that represent a steady release of information regarding the nature and impact of the data breach. No such other cluster of events in close chronological proximity exists in the overall timeframe of the OPM data breach. This provides an opportunity to measure fluctuations in comment characteristics both in time and in response to regular events. These are the following events identified as being key events in the two-month time-frame of June 4th 2015 to July 31st 2015 (Chaffetz, Meadows, & Hurd, 2016; Sternstein & Moore, 2015):

- June 8th 2015 US-CERT determines information stolen contained many sources of PII, news outlets relay information to public citing anonymous sources
- June 12th 2015 OPM announces second data breach where security clearance data may have been compromised
- June 16th 2015 OPM Director acknowledges security clearance data compromised
- June 24th 2015 OPM Chief Information Officer (CIO) attempts to minimize significance of early warnings when security documents were stolen in 2014 OPM breach
- July 9th 2015 OPM reveals all background investigation data for 21.5 million individuals was compromised, broadening the scope of information stolen
- July 10th 2015 OPM Director Katherine Archuleta resigns

The resulting dataset was purchased from a third-party vendor of twitter data. The data purchased consisted of 18764 records (tweets) posted on twitter from June 4th 2015 to July 31st 2015, all containing the hashtag #OPMHack. This represents 1370 hours of data. The initial 19 parameters of the dataset include user ID, tweet ID, tweet content, retweet count, follower count, user profile data, user geo data, and tweet date-time stamp. 93 additional parameters were added to the dataset utilizing the Linguistic Inquiry and Word Count (LIWC) sentiment analysis tool. This tool is a widely used and accepted tool for textual analysis and is often used to extract emotional dimensions from text data in academic research (Pennebaker et al., 2015; Tausczik & Pennebaker, 2010). The 2016 version of this tool was used and at the time of this writing remains the latest version available. The primary measures of interest provided by

LIWC are the emotional dimensions of anxiety, anger, and sadness. These represent the emotional measures intended to test the theory of emotional responses to data breaches.

Additional measures corresponding to coping techniques and victimization were added to the dataset based off of the tweet content of each record. The two primary coping measures of rational thinking and emotional venting were adapted from Jin, Pang, and Cameron (2012). These represent the measures intended to test the theory of crisis dialogue by splitting messages into those perceived as informational and those perceived as emotional. Coding was performed on the dataset with the help of two graduate students. Following a coding instructional session and two consecutive rounds of interrater percentages exceeding a simple percentage of 70% and a Kohen's Kappa of 0.539 and 0.508 for rational thinking and emotional venting respectively was obtained. The graduate students were then each given 300 unique tweets to code which combined to 600 unique tweets and mapped to 5600 records in the overall dataset thanks to a number of messages being retweeted multiple times. Victimization was coded through a combination of keyword matching for a number of victimization phrase permutations (e.g. – "received my notification from OPM", "they lost my data", and "they let this happen to me") and manual review of over 1000 unique tweets for more nuanced indications of direct victimization by a graduate student. This resulted in identifying 140 victims out of 8379 unique users.

Analysis

To test for evidence of each of the theories discussed, a different analytic approach needs to be taken for each. The theory of emotional response to data breaches can be tested by analyzing the sentiment changes over time and in correspondence with consecutive breach events. The theory of crisis dialogue can be tested by performing discriminant analysis along the characteristics associated with the theory. Before proceeding to an analysis of each of these theories, some general analysis of the data is warranted. First is a confirmation of the relevance of the six identified events in the event cluster that occurred over the two-month period of the dataset. This can be seen in the aggregated tweet count graph presented in figure 4 below:

1. June 8th 2015 US-CERT determines personally identifying information (PII) was stolen, news outlets relay information to public citing anonymous sources
2. June 12th 2015 News outlets report SSN's of ***all*** federal employees compromised
3. June 16th 2015 OPM Director acknowledges background data compromised
4. June 24th 2015 OPM Chief Information Officer (CIO) attempts to minimize significance of early warnings when security documents were stolen in 2014 OPM breach
5. July 9th 2015 OPM reveals all background investigation data for 21.5 million individuals was compromised, broadening the scope of information stolen
6. July 10th 2015 OPM Director Katherine Archuleta resigns

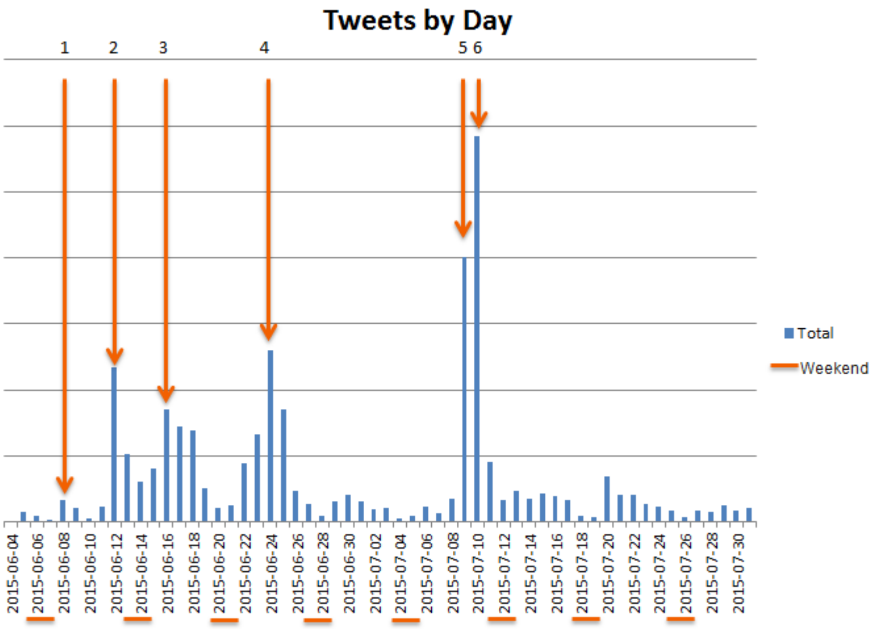


Figure 4 – Tweets by Day with Major Event List

It should be pointed out that the underlined dates correspond to weekend days (Saturday and Sunday) to emphasize the underlying cyclical fluctuations in tweet aggregations that result from behavioral changes corresponding to the traditional work week. It is important to note that despite these underlying fluctuations, the influence of the key news events related to the data breach have a profound effect on the discussion surrounding the data breach. While the correlation of these spikes in activity to the data breach events seems quite obvious, what is not as obvious is the dramatically different rates of drop off that occur when comparing the drop off following events five and six (which happen back to back) and all of the preceding events. The first day drop off activity rate for events one through 4 is a median of 35% and a mean of 35%, while the first day drop off activity rate for events five and size is 84%. Even before analyzing the sentiment fluctuations over time, this serves as an indicator that either breach fatigue or acceptance is possibly setting in following event six. This is because the quick drop off in engagement indicates either an acceptance of the breach event or an apathetic tendency towards it, as would be expected with the onset of breach fatigue. An analysis of the underlying sentiment could provide further clarification.

Testing the Theory of Emotional Response to Crisis Events

The sentiment measures of anxiety, anger, and sadness, aggregated over time reveal that the theory of emotional response to data breaches appears to model the data quite well. This can be seen in the figure below.

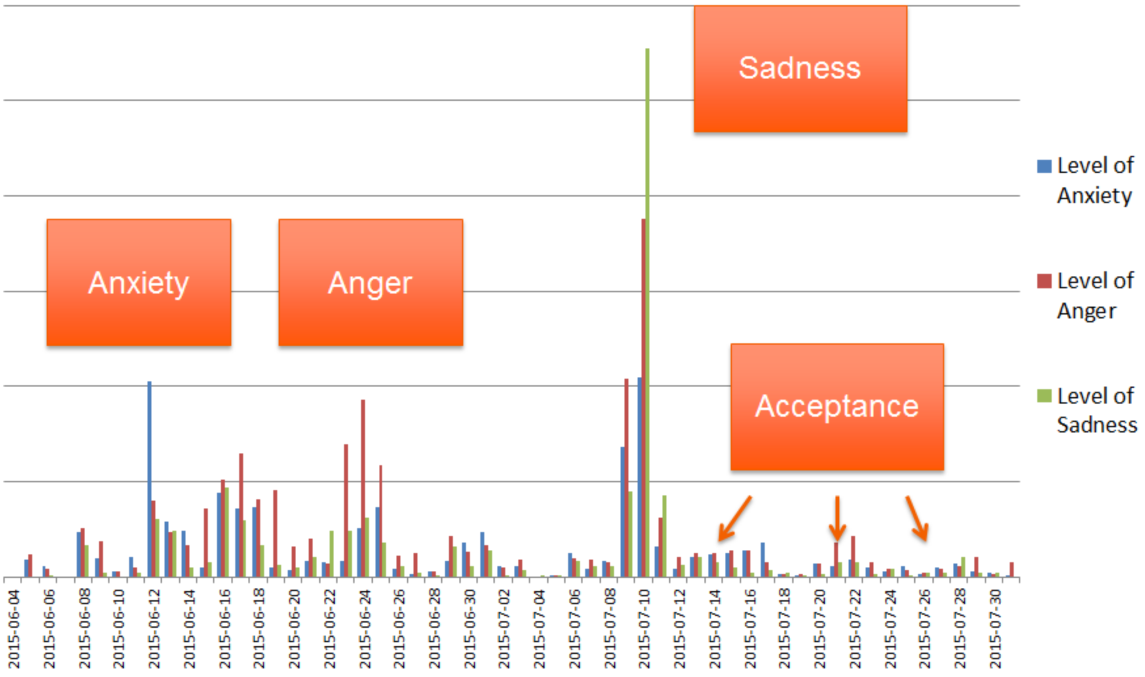


Figure 5 – Emotional Response Graph

It is clear that there are heightened levels of anxiety early on, followed by an extended period of heightened anger, and ending with a spike in sadness before a sustained drop in activity relative to prior engagement. These correspond nicely to the stages developed in the adapted Kübler-Ross model.

In addition to the aggregate public response, we are interested to determine the difference between the identified victim-group and rest of the dataset. To do this we procedurally scanned the dataset for instances of phrases indicating direct victimization such as references to receiving mail notification of data loss from OPM, individual specification of data loss due to the breach, and other statements of having been directly affected. We also tasked a graduate student to examine three separate samples of 500 tweets for more nuanced references to victimization. The former automated procedure identified 133 victims and the manual process was able to identify 7 more. In both cases we ensured that none of the tweets were retweets in an effort to further account for direct victimization. All tweets by these 140 users were then coded as a 1 under the victim parameter and the remaining tweets were coded as a 0. An ANOVA was then performed to identify differences between these groups, the results of which demonstrated no significant difference between the groups for anxiety and anger, but a significance of

0.042 for differences in sadness. Unlike the insignificant differences in anger and anxiety, sadness was higher in the victim group.

Testing the Theory of Crisis Dialogue

To test the theory of crisis dialogue, it is necessary to first dichotomize the dataset into two groups, one belonging to highly duplicated tweets and one belonging to the remainder. To enable this dichotomization, a mean retweet score of 30.66 is obtained. This is rounded up to 31 since a fractional retweet makes no sense. Using this value, all tweets with a retweet count of 31 or greater are coded as highly duplicated and all of the remainder are coded as not highly duplicated. The result is that 22.39% of the dataset is coded as highly duplicated (4201 tweets). A discriminant analysis performed using SPSS reaffirms that all 18764 cases are valid and far exceeds the 20 to 1 ratio that is preferred for conducting discriminant analysis. That is because in this case there are only 3 independent variables (anxiety, anger, sadness), which results in a ratio of approximately 6255 to 1 or roughly thirty times greater than the minimum preferred ratio. The smallest group also exceeds the minimum of 20 observations, with 4201 cases (which is the group corresponding to highly duplicated tweets). Discriminant analysis of this data results in significantly different means among the independent variables of anxiety, anger, and sadness as seen in the table below.

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
anx	0.996	76.167	1	18762	0
anger	1	7.205	1	18762	0.007
sad	0.988	225.937	1	18762	0

Table 1

The Box's M test is significant but due to the test's sensitivity to large sample sizes can sometimes be ignored. The Wilks's Lambda of 0.984 with a significance exceeding 0.001 indicates the functions ability to explain group membership better than chance. Unfortunately the canonical correlation value of 0.126 indicates the chosen independent variables explain only a small amount of the variation in group membership. The structure matrix indicates that sadness and anger were the more powerful indicators of group membership differences in comparison to anger (0.867, 0.503, and 0.155 respectively) and this is reinforced by the mean differences of those variables across groups (table 2).

Group Statistics					
RT31		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
0	anx	0.5047	1.73649	14760	14760
	anger	0.692	1.9837	14760	14760
	sad	0.5269	1.76863	14760	14760
1	anx	0.2541	1.02488	4004	4004
	anger	0.6011	1.55634	4004	4004
	sad	0.0987	0.6672	4004	4004
Total	anx	0.4512	1.61448	18764	18764
	anger	0.6726	1.90091	18764	18764
	sad	0.4355	1.60819	18764	18764

Table 2

Ignoring the Box's M significance, it would be clear that these independent variables do in fact differ across groups and would appear to support the proposed theory of crisis dialogue by demonstrating lower mean values across all emotional measures (table 8). Unfortunately, the lower explanatory power of these variables leaves the validity of this theory in question. Additional research into this possible phenomenon appears to be warranted though current results are inconclusive.

Future Research

Crisis Fatigue

We believe that further research into the effects of crisis fatigue (an emotional fatigue archetype of which Breach Fatigue is a subtype) on victim groups is the next logical step in this line of study. Crisis Fatigue builds on emotional exhaustion research (Barnes & Van Dyne, 2009) and contends that the inability to escape the news of data breach events leads to an inability to recoup emotional reserves and results in a fatiguing of emotional well-being. The primary distinction between the acceptance stage of our proposed Theory of Emotional Response to Crisis Events and a state of fatigue is that the former is a characteristic of the aggregate public response and the latter is a characteristic of the identified direct victim group (in this case, the group formed by those individuals directly affected by the data breach). We expect that evidence of Crisis Fatigue would be demonstrated by a lack of victims in the victim group taking advantage of post-breach protection offers such as credit report monitoring. It is our belief that this would be a circumvention of the natural emotional response process which should normally end in acceptance. Because crisis events such as data breaches are not expected to decline and since this has implications for economic activity, it would signify a need to intervene in the emotional response cycle to prevent either emotional exhaustion or to entice protective action earlier on in the emotional response cycle. Capturing data on measures that can be analyzed to verify this theory requires a degree of certainty about whether or not subjects affected by a particular breach (in this case the OPM hack) were offered post-breach protection and declined it.

Conclusion

The conclusion of this research effort is still incomplete due to a number of portions remaining in a state of ‘in-progress’. At the time of this writing there is a moderate level of confidence in continuing to pursue the theory of emotional responses to crisis events by gathering data on other crisis events and through mediums other than Twitter. The theory of crisis fatigue is currently unable to be directly tested and will remain that way until some of the collaborating parties involved are able to obtain the necessary data. In the meantime instruments are still being developed to adequately capture mitigation and post-breach protection efforts so that evidence for crisis fatigue can be verified or denied. Finally, the discriminant analysis performed for the initial testing of the theory of crisis dialogue resulted in a low canonical correlation value. This indicates that there are better models that can be developed to explain the differences between highly duplicated messages and all others. Continued theory development and data analysis will need to be conducted to determine what those indicators might be.

Acknowledgements

This research has been funded in part by NSF under grants 1651060 and 1554480. Usual disclaimer applies.

REFERENCES

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. *ICIS 2006 Proceedings*, 94.

Amorosi, D. (2011). Data Breach Spring. *Infosecurity*, *8*(3), 6–9. https://doi.org/10.1016/S1754-4548(11)70032-8

Barnes, C. M., & Van Dyne, L. (2009). `I'm tired': Differential effects of physical and emotional fatigue on workload management strategies. *Human Relations*, *62*(1), 59–92. https://doi.org/10.1177/0018726708099518

Brown, H. S. (2016). After the data breach: Managing the crisis and mitigating the impact. *Journal of Business Continuity & Emergency Planning*, *9*(4), 317.

Burden, B. C. (2002). When Bad Press Is Good News The Surprising Benefits of Negative Campaign Coverage. *The Harvard International Journal of Press/Politics*, *7*(3), 76–89.

Bush, D. (2016). How data breaches lead to fraud. *Network Security*, *2016*(7), 11–13. https://doi.org/10.1016/S1353-4858(16)30069-1

Chaffetz, J., Meadows, M., & Hurd, W. (2016). *Committee Releases Year-Long Investigative Report into OPM Data Breaches*. Retrieved from https://oversight.house.gov/release/committee-releases-year-long-investigative-report-opm-data-breaches/

Dobele, A., Lindgreen, A., Beverland, M., Vanhamme, J., & van Wijk, R. (2007). Why pass on viral messages? Because they connect emotionally. *Business Horizons*, *50*(4), 291–304. https://doi.org/10.1016/j.bushor.2007.01.004

Fowler, K. (2016). Data Breach Preparation and Response. *Network Security*, *2016*(10), 4. https://doi.org/10.1016/S1353-4858(16)30094-0

Garrison, C., & Ncube, M. (2011). A longitudinal analysis of data breaches. *Information Management & Computer Security*, *19*(4), 216–230. https://doi.org/10.1108/09685221111173049

German, P. (2016). A new month, a new data breach. *Network Security*, *2016*(3), 18–20. https://doi.org/10.1016/S1353-4858(16)30029-0

Gupta, M., & Sharman, R. (2012). Determinants of Data Breaches: A Categorization-Based Empirical Investigation. *Journal of Applied Security Research*, *7*(3), 375–395. https://doi.org/10.1080/19361610.2012.686098

Hansen, L. K., Arvidsson, A., Nielsen, F. A., Colleoni, E., & Etter, M. (2011). Good Friends, Bad News - Affect and Virality in Twitter. In J. J. Park, L. T. Yang, & C. Lee (Eds.), *Future Information Technology* (pp. 34–43). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-22309-9_5

Holtfreter, R. E., & Harrington, A. (2015). Data breach trends in the United States. *Journal of Financial Crime*, *22*(2), 242–260. https://doi.org/10.1108/JFC-09-2013-0055

Jin, Y., Pang, A., & Cameron, G. T. (2012). Toward a Publics-Driven, Emotion-Based Conceptualization in Crisis Communication: Unearthing Dominant Emotions in Multi-Staged Testing of the Integrated Crisis Mapping (ICM) Model. *Journal of Public Relations Research*, *24*(3), 266–298. https://doi.org/10.1080/1062726X.2012.676747

Kübler-Ross, E. (2009). *On Death and Dying: What the Dying Have to Teach Doctors, Nurses, Clergy and Their Own Families*. Taylor & Francis.

Layton, R., & Watters, P. A. (2014). A methodology for estimating the tangible cost of data breaches. *Journal of Information Security and Applications*, *19*(6), 321–330. https://doi.org/10.1016/j.jisa.2014.10.012

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. UT Faculty/Researcher Works.

PonemonInstitute. (2014). The Aftermath of a Mega Data Breach: Consumer Sentiment. Ponemon Institute.

Privacy Rights Clearinghouse. (2016, November 30). Retrieved December 12, 2016, from https://www.privacyrights.org/data-breaches

Roberds, W., & Schreft, S. L. (2009). Data breaches and identity theft. *Journal of Monetary Economics*, *56*(7), 918–929. https://doi.org/10.1016/j.jmoneco.2009.09.003

Sternstein, A., & Moore, J. (2015, June 26). Timeline: What We Know About the OPM Breach (UPDATED). Retrieved December 7, 2016, from http://www.nextgov.com/security/2015/06/timeline-what-we-know-about-opm-breach/115603/

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.

Wheatley, S., Maillart, T., & Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B*, *89*(1), 7. https://doi.org/10.1140/epjb/e2015-60754-4

Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS One*, *11*(3), e0150989.