Information Leakage as Diagnostic Communication

Research-in-Progress

Introduction

"The rumored leak photo of iPhone 7 actually looks pretty solid and more ruggard than the elegant iPhone 6, I must say" –Twitter User _annurrrrr (July 25, 2015)

The quote above encapsulates much of the challenge facing the technology companies which are trying to maintain control of communication with their customers in a digital marketplace. First, this customer's perception is based on product information that has been leaked to the public and has not been confirmed by the company. Second, the quote contains potentially valuable consumer feedback on the design of an upcoming smartphone. Information leaks—defined here as the improper secondary use or dissemination of private information (e.g. technology product information) by an authorized party—pose a serious threat to the organization's control of its own private information. This is especially true in the Big Data age where the spread of leaked information across communication networks is potentially viral and information can be rapidly copied and distributed via high-bandwidth one-to-many forums. Organizations operating in this environment need to develop communication strategies which balance the customer's need for information and the company's need for privacy and control.

This proposed research argues that while information leaks represent violations of an organization's privacy, advances in business intelligence/analytics technology can be used to analyze online communities' responses to the leaked information and regain control of communication with their customers. This ability helps organizations alleviate the often disastrous consequences of information leaks and possibly use the leak event as an opportunity to respond to customer perceptions, thus gaining competitive advantages. For example, an unfinished version of the superhero movie *X-Men Origins: Wolverine* was leaked to file sharing sites one month before the theater release of the film. On the first day, the movie was downloaded an estimated 100,000 times. Twentieth Century Fox Film Corporation detected the leak quickly and immediately responded with a communication strategy that responded to criticisms by emphasizing the incomplete nature of the film and encouraged fans to see the finished work in the theater. Some analysts have argued that the buzz generated by the leak and a quick response helped the film go on to gross $373 million worldwide despite the leak.[1]

Despite the increase in the prevalence of information leaks, this phenomenon has been largely overlooked in information systems (IS) research. A majority of IS-based privacy research is focused on privacy breaches and strategies for preventing or mitigating the significant negative consequences for the companies and consumers (Cavusoglu et al. 2004; Gatzlaff and McCullough 2010). Though the fallout from a leak can be just as damaging to the victim, leaks have not received much attention in the academic literature. Therefore, this research attempts to bridge this gap by understanding two fundamental questions: **(1) *How does leaked information travel through online communities? (2) How do different forms of diagnostic information (e.g., rumors, preannouncements, and leaked information) influence information sharing behaviors?***

[1] "FBI called in over Wolverine leak," http://news.bbc.co.uk/2/hi/entertainment/7978379.stm, (April 3, 2009).

This research begins by defining information leak and distinguishing it from similar concepts such as breaches and browse violations, two primary forms of privacy violations (Culnan and Williams 2009). We argue that information leaks represent a distinct form of organizational privacy violation. Then, we build a conceptual model based on Rumor Theory, Preannouncement Strategy, and Media Richness Theory, which together establish a foundation for understanding how consumers respond to leaked information. To test the model, we plan to conduct an exploratory study on the volume, rate and pattern of the spread of leaked technology product (Apple Watch and Apple iPhone 6s) information in Twitter.

This research contributes to the extant IS research in several ways. First, we propose a new type of privacy violation: information leak, which is important yet under-studied in IS privacy research. Second, we develop a new conceptual model to investigate how the spread of leaked product information and customer responses to the information differ from other types of product information (rumors and official preannouncements). Third, we will develop a data analytic approach to identify the most common patterns for spreading information leaks and thus contribute methodologically to IS research.

Theoretical Foundation

Information Leak and Privacy

Information leakage is rooted in the concept of privacy. While privacy has a long history in the social sciences, no consensus exists on what exactly privacy is (Pavlou 2011; Smith et al. 2011). The definition adopted by researchers is often determined by the context of the study (Dinev et al. 2013). The rise of digital technologies has led researchers to extend behavioral privacy to include information privacy (Bélanger and Crossler 2011; Clarke 1999; Smith et al. 2011). Belanger (2011, p. 1018) has defined information privacy as "the interest an individual has in controlling, or at least significantly influencing, the handling of data about themselves." Losses of control of an individual's private information can have economic and psychological consequences for the victim (Dinev et al. 2013). We argue that losses of control have similar impacts on organizations.

Table 1. Classification of Privacy Violations

		Secondary Use	
		Accidental	Volitional
Access	Authorized	I: Leak	II: Leak
	Unauthorized	III: Browse	IV: Breach

Privacy violations occur when an entity loses control of their own private information. According to Smith (1996), there are two salient factors which relate to issues of loss of control over private information: *secondary use* and *access of private information*. First, secondary use refers to the use of information for reasons other than the intent expressed when the data was originally collected (Culnan 1993; Culnan and Armstrong 1999). Second, privacy violators may be either authorized or unauthorized parties to the information they have exposed (Smith et al. 1996). Based on these two dimensions, we have identified four different types of privacy violations as show in Table 1. Quadrants III and IV represent the types of privacy violations that have been studied in

IS research, which has focused on unauthorized access to private information. Quadrants I and II represent a gap in IS research.

Much of the research on information privacy violations focuses on the impact of, prevention of, and technology for preventing *unauthorized access* to private information (Cavusoglu et al. 2004; Choobineh et al. 2007; Culnan and Williams 2009; Gatzlaff and McCullough 2010; Zafar and Clark 2009). Culnan and Williams (2009) refer to accidental access to unauthorized information as "browsing" (Quadrant III) and volitional access to unauthorized information as "information breaches" (Quadrant IV). A hacker using credit card information stolen from a company's servers to commit identity fraud would be classified as an information breach. Researchers have linked information breaches to financial losses for the organization (Cavusoglu et al. 2004) and its shareholders (Gatzlaff and McCullough 2010), and to reputational losses with the public (Lee et al. 2011).

Little prior research exists which investigates the secondary uses of private information in which the violators had authorized access to the data (Quadrants I and II). We refer to these quadrants as leaks, and define them as *the improper secondary use or dissemination of private information by an authorized party, accidentally or volitionally*. Quadrant I represents an accidental leak such as an occasion when a web developer publishes information to the company's website that had not been finalized or had not yet been approved for release. An example of a Quadrant II, volitional leak, would be an employee of a trusted member of the supply-chain intentionally releasing product information without their consent. This is a common problem within the motion picture industry where screener copies of unreleased movies are uploaded to websites that distribute pirated media. Leaks differ from browsing and breach violations in that technical security measures would be an ineffective safeguard because the violator has authorized access to information. In these instances, the owner of the information has trusted the authorized users to be competent and trustworthy stewards of the data. The leaker violates this trust and exposes the owner's private information to an unauthorized third-party. This research will focus information leaks, intentional or accidental, spread digitally through social media.

Diagnostic Communication

Consumers rely on diagnostic information to inform their purchase decisions (Biehal and Sheinin 2007). Diagnostic information is information that is "valuable for making decisions, and unbundled from physical objects, services, and execution" (Arora and Fosfuri 2005, p. 1092). There are two types of diagnostic information: product rumors and product preannouncement. Allport and Postman's (1947) argue that in the absence of legitimate communication on topics of interest to an individual or group, rumors often fill the information void. Rumors are commonly defined as "a specific (or topical) proposition for belief, passed along from person to person, usually by word of mouth, without secure standards of evidence being present" (Allport and Postman 1947, p. 9), and have been shown to depend on four conditions: general uncertainty, outcome-relevant involvement, personal anxiety and credulity (Rosnow 1991). This perspective suggests that rumors spread as a means of coping with uncertainty and reducing anxiety inherent in events for which few official details are available (Oh et al. 2013). A second form of diagnostic information is a product preannouncement. A preannouncement is defined as a "formal, deliberate communication before a firm actually undertakes a particular marketing action" (Eliashberg and Robertson 1988, p. 282) such as the release of a new technology product. Researchers have shown that preannouncements can act as an inducement to wait for new technology products (Farrell and

Saloner 1986) by reducing the perceived risk of purchasing a new product (Lee and Colarelli O'Connor 2003). Thus, product preannouncements and rumors serve a similar purpose: to reduce the anxiety and uncertainty of purchasing technology products.

We argue that information leaks may be a third form of diagnostic information, which occupies a distinct middle space between a fabricated rumor and an official preannouncement. We base our position on the commonalities shared between leaks, and rumors and preannouncements. First, all three serve to fill an information void. Second, both leaks and rumors represent unconfirmed information and would have similar levels of credulity. Rumors are often anonymous, and leakers are authorized internal agents who would be unlikely to advertise their responsibility for the leak for fear of reprisal. However, leaks and preannouncements would likely diverge from rumors on the amount of ambiguity present in the transmission. Rumors tend to be highly ambiguous while leaks and preannouncements would contain legitimate information and possibly very specific details. The commonalities between rumors, leaks and preannouncements suggest that all three can serve as diagnostic information.

Based on the similarities and differences of leaked information, rumors, and preannouncement, we argue that customers will value different diagnostic information differently, and that this effect will be enhanced by the richness of the media (Daft and Lengel 1986). Furthermore, we posit that the familiarity of the product will influence the value of diagnostic information such that customers will have higher product familiarity with established products (Apple iPhone) than with new products (Apple Watch) and need less information to counter any product uncertainty (Bone 1995). Finally, the various forms of diagnostic information will produce distinct information cascades as consumers encounter the information and then communicate it to their online communities (Leskovec et al. 2007; Leskovec et al. 2006). Specifically, we hypothesize that

H1. Leaked information will spread through in social media communities at a different (a)rate and (b)volume than would rumors or preannouncements.

H2. The richness of the leaked information will increase the customer's perceived value of the information and thus enhance the (a)rate and (b)volume of communication.

H3. The familiarity of the product will decrease the customer's product uncertainty and thus depress the (a)rate and (b)volume of communication.

H4. Leaked information will exhibit information cascade patterns which differ from other sources of diagnostic information.

Research Methodology

For this research we will conduct an exploratory study of Twitter data to assess the differential impact of rumors, leaks, and preannouncements on the rate and volume of communication within online social networks. We will use Radian6 to collect all relevant tweets which precede by six months the release of one new technology product (i.e., Apple Watch) and one established technology product (i.e., iPhone 6s). We selected these products because they are developed by a company which is well known for its high level of control over product communication. Also, the products represent two extremes for technology devices: new and established. The Apple Watch is an entirely new product and consumers interested in this device will have no prior experience with the device and will thus have high uncertainty regarding the attributes of the final product. The Apple iPhone 6s, though a new version, is the second edition of the iPhone 6 line and the 9th

iPhone. Customers interested in the newest iPhone will likely be less uncertain about the key attributes that will influence the adoption decision. Finally, Apple products generate large amounts of speculation and media attention: rumors of the Apple Watch preceded the announcement of the device by at least 26 months[2], and rumors of the Apple iPhone 6s began the day the Apple iPhone 6 was released.

We will then identify several releases of diagnostic information (rumors vs. leak vs. preannouncement) for each product which will provide a sample for comparisons of information types based on relevant tweets per hour (Asur and Huberman 2010) and tweet volume (Tumasjan et al. 2010). Finally, we will code tweets and any linked content such that tweets containing or linking to multimedia information will be classified as being high in media richness, while content lacking these features will be classified as being low in media richness. To test Hypothesis 1, we will use a multivariate analysis of variance (MANOVA) technique to assess the ability of diagnostic information (rumor vs. leak vs. preannouncement) to predict the tweet-rate and tweet volume. Then we will perform a simple slopes analysis to test the impact of different levels of media richness (H2) and product familiarity (H3) on these relationships. Finally, we will analyze the diffusion pattern of the three types of diagnostic to identify the most frequently occurring cascade shapes (Leskovec et al. 2007) which will determine whether the diffusion patterns differ between rumors, leaks and preannouncements (H4). Through these methods, we hope to gain a better understanding of the ways in which information leaks differ from other forms of diagnostic product information.

Discussion

Expected contributions

We expect to contribute to IS research in several ways. First, this project will render a systematic conceptualization of information leakage and distinguish leaks from similar concepts. We will continue enriching our understanding of information leaks in the course of this research. Second, our empirical test will show that leaked product information spreads faster and farther than official preannouncements. This suggests that though leaks are violations of an organizations privacy, they are valuable sources of diagnostic product information for consumers. Also, we will show that richer forms of information are more effective in their ability to reduce anxiety and uncertainty associated with the purchase of new technology products. Because leaks are potentially harmful for victims (i.e. organizations) and potentially helpful for recipients (i.e. consumers of innovative technologies), organizations need to develop strategies for leveraging the good and mitigating the bad. These findings should influence the ways in which organizations respond to leaked information.

This paper has implications for researchers and practitioners. We propose that between rumor and official preannouncement of product related information, there exists a third type of communication with customers: information leak by people who have authorized access to organizational information. Organizations can consider leveraging information leaks as an opportunity to collect customer feedback to their products and generate buzz about their forthcoming products. Upon confirmation, the findings from this research can inform

[2] "What Would an Apple Watch Do?" http://www.informationweek.com/mobile/mobile-devices/what-would-an-apple-watch-do/d/d-id/1108595?, (February 10, 2013).

organizations of how information leaks diffuse and the effects of information leaks, which are different from those of rumors and official preannouncements.

References

Allport, G. W., and Postman, L. 1947. "The Psychology of Rumor,").

Arora, A., and Fosfuri, A. 2005. "Pricing Diagnostic Information," *Management Science* (51:7), pp. 1092-1100.

Asur, S., and Huberman, B. 2010. "Predicting the Future with Social Media," *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*: IEEE, pp. 492-499.

Bélanger, F., and Crossler, R. E. 2011. "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS quarterly* (35:4), pp. 1017-1042.

Biehal, G. J., and Sheinin, D. A. 2007. "The Influence of Corporate Messages on the Product Portfolio," *Journal of Marketing* (71:2), pp. 12-25.

Bone, P. F. 1995. "Word-of-Mouth Effects on Short-Term and Long-Term Product Judgments," *Journal of business research* (32:3), pp. 213-223.

Cavusoglu, H., Mishra, B., and Raghunathan, S. 2004. "The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers," *International Journal of Electronic Commerce* (9:1), pp. 70-104.

Choobineh, J., Dhillon, G., Grimaila, M. R., and Rees, J. 2007. "Management of Information Security: Challenges and Research Directions," *Communications of the Association for Information Systems* (20:1), p. 57.

Clarke, R. 1999. "Internet Privacy Concerns Confirm the Case for Intervention," *Communications of the ACM* (42:2), pp. 60-67.

Culnan, M. J. 1993. "" How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use," *MIS quarterly*), pp. 341-363.

Culnan, M. J., and Armstrong, P. K. 1999. "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation," *Organization science* (10:1), pp. 104-115.

Culnan, M. J., and Williams, C. C. 2009. "How Ethics Can Enhance Organizational Privacy: Lessons from the Choicepoint and Tjx Data Breaches," *Mis Quarterly*), pp. 673-687.

Daft, R. L., and Lengel, R. H. 1986. "Organizational Information Requirements, Media Richness and Structural Design," *Management science* (32:5), pp. 554-571.

Dinev, T., Xu, H., Smith, J. H., and Hart, P. 2013. "Information Privacy and Correlates: An Empirical Attempt to Bridge and Distinguish Privacy-Related Concepts," *European Journal of Information Systems* (22:3), pp. 295-316.

Eliashberg, J., and Robertson, T. S. 1988. "New Product Preannouncing Behavior: A Market Signaling Study," *Journal of Marketing Research*), pp. 282-292.

Farrell, J., and Saloner, G. 1986. "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *The American economic review*), pp. 940-955.

Gatzlaff, K. M., and McCullough, K. A. 2010. "The Effect of Data Breaches on Shareholder Wealth," *Risk Management and Insurance Review* (13:1), pp. 61-83.

Lee, Y., and Colarelli O'Connor, G. 2003. "The Impact of Communication Strategy on Launching New Products: The Moderating Role of Product Innovativeness," *Journal of Product Innovation Management* (20:1), pp. 4-21.

Lee, Y. J., Kauffman, R. J., and Sougstad, R. 2011. "Profit-Maximizing Firm Investments in Customer Information Security," *Decision support systems* (51:4), pp. 904-920.

Leskovec, J., Adamic, L. A., and Huberman, B. A. 2007. "The Dynamics of Viral Marketing," *ACM Transactions on the Web (TWEB)* (1:1), p. 5.

Leskovec, J., Singh, A., and Kleinberg, J. 2006. "Patterns of Influence in a Recommendation Network," in *Advances in Knowledge Discovery and Data Mining*. Springer, pp. 380-389.

Oh, O., Agrawal, M., and Rao, H. R. 2013. "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises," *Mis Quarterly* (37:2), pp. 407-426.

Pavlou, P. A. 2011. "State of the Information Privacy Literature: Where Are We Now and Where Should We Go?," *MIS quarterly* (35:4), pp. 977-988.

Rosnow, R. L. 1991. "Inside Rumor: A Personal Journey," *American Psychologist* (46:5), p. 484.

Smith, H. J., Dinev, T., and Xu, H. 2011. "Information Privacy Research: An Interdisciplinary Review," *MIS quarterly* (35:4), pp. 989-1016.

Smith, H. J., Milberg, S. J., and Burke, S. J. 1996. "Information Privacy: Measuring Individuals' Concerns About Organizational Practices," *MIS quarterly*), pp. 167-196.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. 2010. "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment," *ICWSM* (10), pp. 178-185.

Warren, S. D., and Brandeis, L. D. 1890. "The Right to Privacy," *Harvard law review*), pp. 193-220.

Zafar, H., and Clark, J. G. 2009. "Current State of Information Security Research in Is," *Communications of the Association for Information Systems* (24:1), p. 34.