# The Effects of Security Management on Security Events

Frank Nagle
Marshall School of Business, University of Southern California, Los Angeles, CA 90089, naglef@marshall.usc.edu

Sam Ransbotham
Carroll School of Management, Boston College, Chestnut Hill, MA 02467, sam.ransbotham@bc.edu

George Westerman
Massachusetts Institute of Technology, Cambridge, MA 02139, georgew@mit.edu

Although the desired outcome of security management is better security, empirical evidence for this link is scarce. The scarcity arises from lack of data at the firm level for either security posture or incidents across a broad sample of companies. To address this, we use a novel dataset of daily firm-level security information for 480 of the Fortune 500 enterprises that consists of over 33 million security events. The dataset, obtained from a security monitoring company, contains daily measures of security management and negative security events for a 294 day period, yielding 133,248 firm/day observations. Empirical analysis finds that the number of open ports in a firm is associated with higher incidences of botnet activity, potential exploitation, and unsolicited communications, with some analyses also showing a link to malware activity. The findings are robust to several alternative specifications using hidden Markov and hierarchical linear models. This paper thus finds empirical evidence for a fundamental assumption of security practice — a link between security management and improved security.

**Acknowledgment:** We thank the anonymous provider of the detailed security data. Sam Ransbotham is grateful to the NSF for support provided through NSF CAREER award 1350061.

## 1. Introduction

How does security management affect security outcomes? While the question is important, it has been surprisingly difficult to analyze empirically across a broad sample of enterprises. The threat environment is constantly evolving, with large spikes in activity that vary by day, industry, company, and technical specifics of vulnerabilities. Technical solutions are important, but not sufficient (Dhillon and Backhouse 2000, Zhang et al. 2014, Ransbotham and Mitra 2009). For example, users are seen as the weakest link in security (Laszka et al. 2013) and need to be incentivized properly (August and Tunca 2006, August et al. 2014, Acquisti et al. 2016). Furthermore, while successful attacks generate high levels of visibility, prevented attacks do not. The difficulty of connecting actions to outcomes leads companies to focus their security investment decisions on process-oriented

frameworks rather than on outcome-based measures (Moore et al. 2015). Yet this can lead to a proliferation of potential security investments that lack direct evidence of their effectiveness. To paraphrase John Wanamaker's views on advertising, "Half of my security investments are wasted. I just don't know which half."

The difficulty of empirically linking security practices to security outcomes extends to academic research. Researchers use analytical modeling (e.g. Kannan and Telang 2005, Gupta and Zhdanov 2012) or study the actions of security professionals within firms (Mahmood et al. 2010) or use data based on publicly disclosed attacks (Sarabi et al. 2016, Liu et al. 2015). However, data on company-level security practices combined with granular data on attacks against the firm beyond a single firm are rare (e.g. Edwards et al. 2016, Ransbotham and Mitra 2009, Mitra and Ransbotham 2015, Ransbotham 2010). Prior studies do show the relationship between specific security techniques and outcomes in test environments or small samples of firms (e.g. Kholidy and Baiardi 2012, Vieira et al. 2010, Wang et al. 2010). Other studies examine managers' stated levels of security management, or their intention to practice strong security management, but not their actual practices (e.g. Gordon et al. 2005, Albrechtsen and Hovden 2009). Others focus on external attacks or company vulnerabilities, but not both. While each area has made important contributions to the literature on security management, they are not able to empirically examine the relationship, across a large number of companies, between security management practices *in situ* and the actual incidents these firms experience.

We address the need to empirically link actual practices to actual incidents using a novel dataset. The data, from a security monitoring service, includes 133,248 daily observations from 480 of the Fortune 500 firms over a 294 day period. We examine five specific measures of security for each firm. To proxy for proper security management, we use the number of open ports with known security vulnerabilities on a day. Open ports are a well-known threat vector, and closing unnecessary ports does not require advanced security management expertise; rather, it requires managerial attention to processes of deterrence, prevention, detection, and remedies (Straub and Welke 1998). In particular, it is cheaper to close ports than to invest in security software. Therefore, closing ports is a simple and inexpensive security step to take. As such, the number of open ports help proxy for basic security management practice (Siponen and Oinas-Kukkonen 2007). For dependent variables, we use four types of negative security events: botnet activity, malware activity, potential exploitation, and unsolicited communication. In aggregate, these four events are important indicators of the possible negative security events that a firm can experience.

Due to the daily granularity of the observations in our dataset, we can use strict models that include firm fixed effects such that we compare a firm to itself while controlling for other potential covariates at time and incident levels. This allows for a more causal interpretation of the results

since other aspects of the firm (e.g. size, number of IT security employees, location, etc.) are all held constant. Using hierarchical linear modeling to examine across 133,248 firm-day observations, we find that the number of ports open in a firm on a given day is an accurate predictor of a negative security outcome. Considering the negative events separately, we find that this effect is driven primarily by three of the four event types: botnet activity, potential exploitation, and unsolicited communications. These findings are robust to additional specifications of the hierarchical linear model as well as hidden Markov modeling. Furthermore, we show the value of a data source which can capture the daily security posture of hundreds of companies.

## 2. Empirical Methods

### 2.1. Data

One main difficulty in empirically linking *in situ* corporate security practices to security incidents is data availability. It is difficult to obtain granular information on the two main constructs across a large number of firms. While detailed information may be available from within a single firm, it is difficult to know if results are idiosyncratic to the firm. Alternatively, while some data (such as surveys) may be available across a number of firms, it can be difficult to reconcile perceptions of activity with actual behavior by attackers or firms. As a result, empirical findings thus far are limited (e.g. Edwards et al. 2016, Ransbotham and Mitra 2009, Mitra and Ransbotham 2015, Ransbotham 2010, Ransbotham et al. 2012).

Using a novel dataset from a security monitoring service, we address both of the data challenges for this research question. This third party dataset contains longitudinal measures relating to both security practices and security outcomes, measured on a daily basis for a large number of firms. The data is collected via global sensor networks that monitor external network traffic and communications to and from firms based on their assigned IP address space. The accuracy of this data is independently verified to confirm it accurately reflects the internal state of the firm's networks. Furthermore, the data has been shown to have a high degree of correlation with public announcement of breaches at the firms. The dataset consists of daily observations of 480 of the Fortune 500 firms for the period from 1 July 2014 until 20 April 2015. (However, some measures are occasionally unavailable and thus we do not have a perfectly balanced panel of every firm on every day for every event type.) This results in a set of 133,248 firm/day observations. Each observation contains counts of basic security management activities and four types of negative security incidents, among other measures.

We use the number of *open ports* with known vulnerabilities as a proxy for security management effectiveness. Identifying open ports via port scanning is often a first step an attacker will make for reconnaissance (Ransbotham and Mitra 2009). Therefore, closing unnecessary ports, especially

those that have known vulnerabilities, is an important security management practice (Baldwin et al. 2012) that reduces the chance of opportunistic paths of compromise (Ransbotham and Mitra 2009). It is also one of the first steps recommended for security management because of its relative ease of implementation. Hence, although the measure is certainly not a complete measure of security practices, the number of open ports is a "tip-of-the-iceberg" for identifying a firm's security management practices.

We measure attacker activity through four types of negative security incidents:

a.) Botnet activity: Messaging between the focal firm and botnet command and control servers.

b.) Potential exploitation: Internet communication designed to indicate that a system at the focal firm has potentially been exploited.

c.) Unsolicited communication: Suspicious or irregular Internet traffic originating from the focal firm (e.g., communication with darknets).

d.) Malware activity: Malware activity originating from the focal firm indicating that malware has compromised a system.

The dataset contains 33,305,880 events. Each day averages 113,285 events with a minimum of 20,526 events and maximum of 325,744 events. Table 1 describes the breakdown by risk types and date.

**Table 1 Security Events by Date**

| Type | Total | Min | Max | Mean | Median | St. Dev. |
|---|---|---|---|---|---|---|
| Botnet activity | 24,950,007 | 16,883 | 291,743 | 84,864 | 75,888 | 45,815.58 |
| Potential exploitation | 8,276,413 | 8 | 130,480 | 28,151 | 21,978 | 23,870.47 |
| Unsolicited communication | 74,047 | 0 | 1,098 | 252 | 223 | 152.02 |
| Malware activity | 5,413 | 0 | 34 | 18 | 19 | 6.51 |

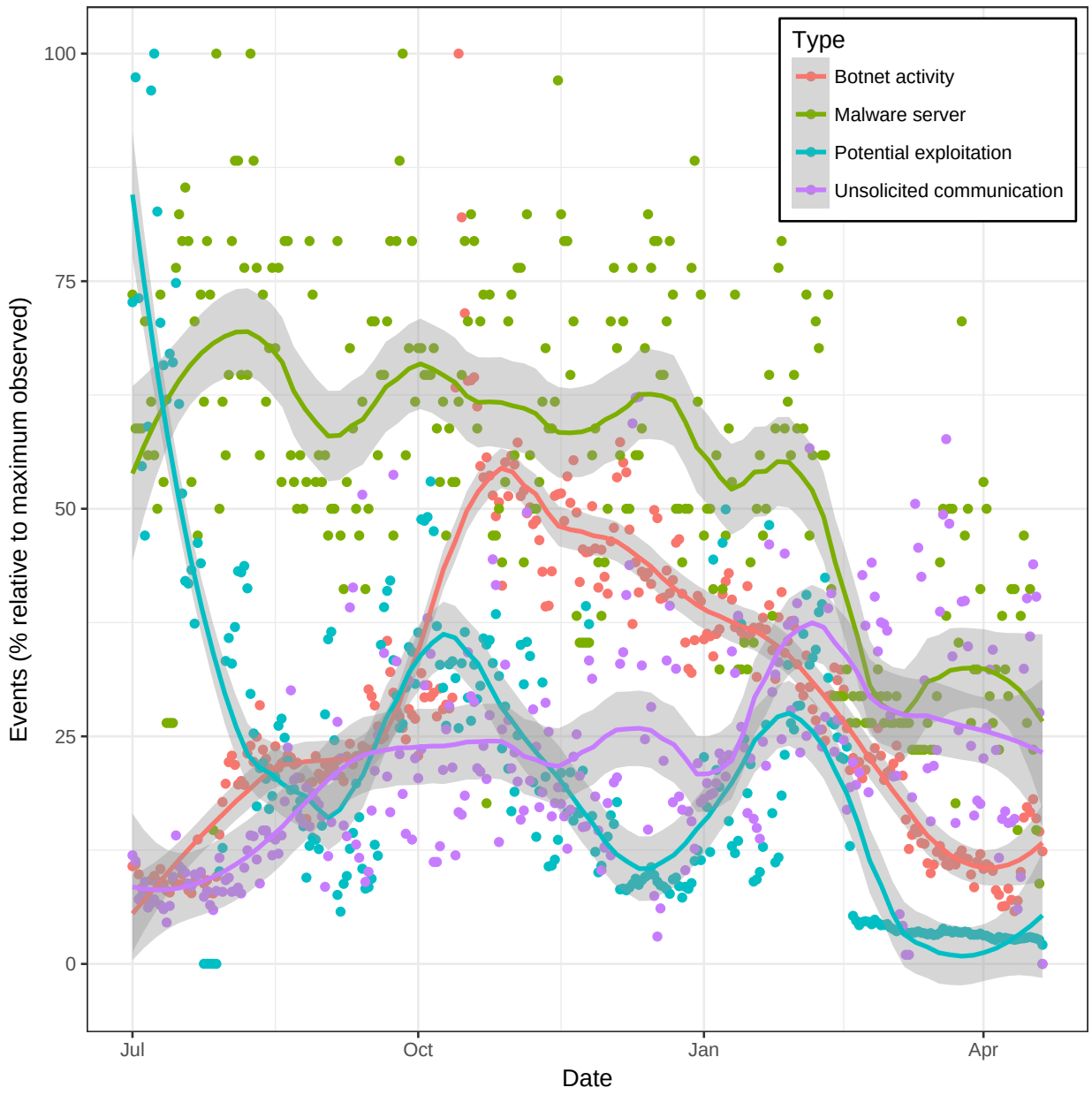
The number of events varies widely across firms and days (see Table 2). Events also show systematic variation across days (see Figure 1). This correlated daily variation (Table 3) creates identification challenges which we address in the analysis.

**Table 2 Descriptive Statistics for Security Management and Events by Firm and Date**

| Statistic | Min | Max | Mean | Median | St. Dev. |
|---|---|---|---|---|---|
| Botnet activity | 0 | 282,326 | 187.24 | 0 | 4,281.29 |
| Potential exploitation | 0 | 120,615 | 62.11 | 0 | 1,532.69 |
| Unsolicited communication | 0 | 891 | 0.56 | 0 | 8.28 |
| Malware activity | 0 | 32 | 0.04 | 0 | 0.80 |
| Open ports | 0 | 1,974,000 | 2654.67 | 14 | 49,132.59 |

133,248 observations of 480 firms

**Figure 1 Security Events by Date**




*Note.* Lines represent a smoothed trend line (LOESS) for each type of event.

## 2.2. Model Specification

**2.2.1. Hierarchical Linear Modeling** Each unit of observation in our analysis is the number of security events of each type, for each firm, for each day. Due to the nature of the phenomena being measured – many zero states, punctuated occasionally by extreme non-zero values when events occur – we focus the regression analysis on the binary presence of each type of event in the focal firm on the focal day. We use a logit specification because of the outcome variables are binary. (However, later analysis in Table 8 uses continuous measures and finds consistent results.) However,

**Table 3 Variable Correlations**

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Botnet activity | 1.000 | | | | | | | |
| 2. Potential exploitation | 0.635 | 1.000 | | | | | | |
| 3. Unsolicited communication | 0.311 | 0.238 | 1.000 | | | | | |
| 4. Malware activity | 0.055 | 0.060 | 0.161 | 1.000 | | | | |
| 5. Open ports (log) | 0.169 | 0.144 | 0.241 | 0.195 | 1.000 | | | |
| 6. Employees | 0.118 | 0.103 | 0.107 | 0.218 | 0.331 | 1.000 | | |
| 7. Assets (current) | 0.086 | 0.089 | 0.321 | 0.165 | 0.434 | 0.361 | 1.000 | |
| 8. Revenue per employee | −0.009 | −0.008 | −0.008 | −0.018 | −0.138 | −0.169 | 0.124 | 1.000 |
| 9. Market value | 0.103 | 0.099 | 0.231 | 0.110 | 0.386 | 0.434 | 0.771 | 0.070 |

Pearson product moment correlation using 133,248 observations of 480 firms

because of the correlated nature of events, as well as the systematic variation of events across days and firms, we include fixed effects for each individual day, each firm, and each type of event. Errors for each of these effects are unlikely to be independent. Therefore, we use a hierarchical logit model with crossed date, firm, and type to incorporate this lack of independence (Figure 2).

Each firm $f$ is in the set of all firms $F$; each day $t$ is in the set of all dates $T$; and each type of security event $s$ is the set of all security events $S$. The presence of a security event $y_{f,t,s}$ is 1 if an event of type $s$ was observed in firm $f$ on day $t$ and 0 otherwise. Our focal variable, $open\_ports_{f,t}$ is the number of open ports observed in a scan of firm $f$ on day $t$. $X_f$, $X_t$, and $X_s$ are indicator variables for each firm, day, and event type (respectively). The variable $y_{f,t,s}$ is models as

$$y_{f,t,s} = \beta_{0,f} + \beta_{1,t} + \beta_{2,s} + \beta_3 \times \log(open\_ports_{f,t} + 1) + \epsilon_{f,t,s}, \forall f \in F, \forall s \in S, \forall t \in T \quad (1)$$

with firm, date, and event type specific effects estimated as

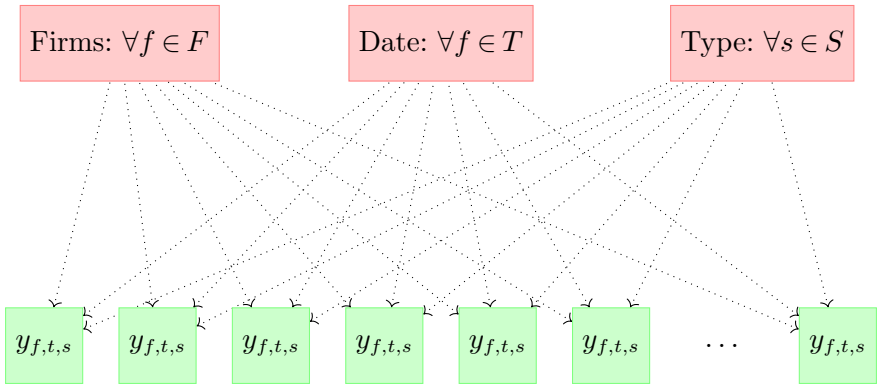
$$\beta_{0,f} = \beta_0 + \beta_f \times X_f + \epsilon_{0,f}, \forall f \in F \quad (2)$$

$$\beta_{1,t} = \beta_1 + \beta_t \times X_t + \epsilon_{0,t}, \forall t \in T \quad (3)$$

$$\beta_{2,s} = \beta_2 + \beta_s \times X_s + \epsilon_{0,s}, \forall s \in S. \quad (4)$$

This crossed structure allows observation level error distributed as a standard logistic distribution, $\epsilon_{f,t,s}$, as well as correlated errors at the level of the firm ($\epsilon_{0,f}$), day ($\epsilon_{0,t}$), and event type ($\epsilon_{0,s}$). The model is simple conceptually, but the crossed effects control for any unobserved temporal trends (affecting all firms), any firm specific effects (that are unchanging), and any event specific effects.

**Figure 2 Hierarchical crossed model of security events**



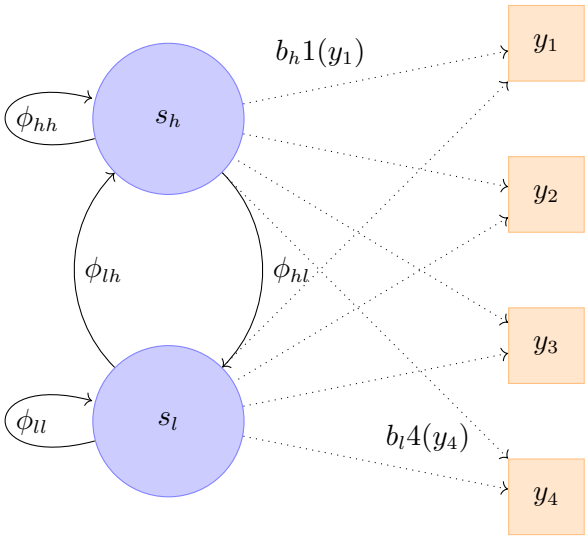

*Note.* Each observation $y_{f,t,s}$ is influenced by clustered firm level ($f$), date level ($t$), and event type level ($s$) effects.

**2.2.2. Hidden State Models** The longitudinal nature of our data allows us to conduct additional analyses of the latent link between security management and security outcomes. While the granular daily data is useful, security practices and outcomes may be slow to react; instead, each may contribute to a general trend toward more or less security. From this perspective, we can consider our data as signals of unobservable latent attributes. Security is an inherently unobservable combination of many factors. Neither security management practices nor attacker practices are directly observable. They are complex combinations of conditions, some of which we can measure and others which we cannot. However, longitudinal analysis can infer the presence of latent states ranging from higher levels of security to lower levels. Hidden Markov models identify these latent security states and, importantly, the effects of the security management covariate on changes in the firm latent security status. This is particularly important in our context where we only have a subset of the possible security events and possible managerial activities.

Formally, the model (illustrated in Figure 3) consists of the following components. $F$ is the set of all firms in the dataset. A specific firm $f$ is observed daily starting at time $t = 0$ (1 July 2014); all firm observations begin on the same date. Activity related to the firm is observed until 20 April 2015 resulting in a uniform number of periods for each firm ($T = 294$ days).

A finite set $S$ of $n$ discrete states, $\{S_1 \dots S_n\}$, abstracts the level of security in a firm. The initial probabilities $(\pi_s, s \in S)$ of these discrete states cover all options ($\sum_{s \in S} \pi_s = 1$). By convention, the base state, $S_1$, reflects a high level of security within the firm. While latent security in an individual firm is idiosyncratic, the set $S$ covers all possible security levels. For example, firm $f_a$ likely has a different latent level of security than firm $f_b$ on the same day but both have the same possible set of state options. There may be many latent level of security for a firm; for simplicity of exposition, we present results that consider two simple states— high and low. In the high state, the firm shows

**Figure 3 Hidden Markov Model of Low and High Security States**




*Note.* An HMM with two states $\{S_h, S_l\}$ which can emit four discrete symbols $\{y_1, y_2, y_3, y_4\}$. $\phi_{ij}$ is the probability to transition from state $S_i$ to state $S_j$. $b_j(y_k)$ is the probability to emit symbol $y_k$ in state $s_j$.

few signs of malicious activity and in the low state, the firm shows evidence of a negative security event that day. Later analysis considers a larger number of states.

These discrete states of firm security are not directly observed. Instead, a measurement model $(B_s, s \in S)$ relates the hidden state to observable measures based on $m$ emitted symbols, $\{y_1 \dots y_m\}$. For our model, we focus on four emitted symbols. First, $y_1$ measures the presence (binary, yes or no) of *botnet_activity* in a firm on a day. Second, $y_2$ measures the presence (binary, yes or no) of *potential_exploitation* in a firm on a day. Third, $y_3$ measures the presence (binary, yes or no) of *unsolicited_communication* from a firm on a day. Fourth, $y_4$ measures the presence (binary, yes or no) of *malware_activity* in a firm on a day.

Finally, a transition model $(\phi)$ reflects the probability $(\phi_{ij})$ of transition from state $S_i$ in time $t$ to state $S_j$ in time $t+1$. For each source $(i)$ and destination state $(j)$, the transitions probabilities sum to 1; $\sum_{i \in S} \phi_{ij} = 1, \forall j \in S$. The hidden Markov model allows the transition probabilities to depend on the number of *open ports*, reflecting security management activity, where larger values indicate less effective security management. These transition probabilities are the focus of the empirical analysis and estimate the influence of security management on the hidden security state of the firm.

## 3. Results

### 3.1. Regression Results

The likelihood of an occurrence of each different type of event may differ considerably. To evaluate this, Table 4 examines the hierarchical linear models with distinct subsamples for each individual

security event type. The models include crossed effects for *day* and *firm* to control for temporal effects that act on all firms as well as any unchanging firm-specific effects. Coefficients for *Open ports* represent the change in log likelihood of the focal security event attributable to an increase in open ports in a firm on a given day. Positive significant coefficients indicate that weaker security management activities, as indicated by more open ports, are associated with more events for botnet activity (Model T1, $\beta = 0.082$, $p < 0.001$), potential exploitations (Model T3, $\beta = 0.325$, $p < 0.001$), and unsolicited communication (Model T4, $\beta = 0.468$, $p < 0.001$) but not for malware activity (Model T2, $\beta = 0.048$, $p = 0.19$). The insignificance of the coefficient for malware activity may be due to the sparsity of malware events in the dataset.

**Table 4 Effect of Security Management on Individual Event Types**

| | **T1** Botnet activity | **T2** Malware activity | **T3** Potential exploitation | **T4** Unsolicited communication |
|---|---|---|---|---|
| Fixed effects estimates | | | | |
| Constant | $-2.497^{***}$ | $-15.112^{***}$ | $-3.811^{***}$ | $-12.182^{***}$ |
| | (0.207) | (0.919) | (0.154) | (0.001) |
| Open ports (log) | $0.082^{***}$ | 0.048 | $0.325^{***}$ | $0.468^{***}$ |
| | (0.018) | (0.037) | (0.019) | (0.001) |
| Random effect variance ($\sigma^2$) | | | | |
| Day level | 0.607 | 0.544 | 2.626 | 0.389 |
| Firm level | 17.506 | 107.562 | 5.491 | 35.516 |
| Fit statistics | | | | |
| Log likelihood | $-36,145.760$ | $-1,270.592$ | $-36,118.750$ | $-6,436.343$ |
| Akaike information | 72,299.530 | 2,549.185 | 2,549.185 | 12,880.690 |

Mixed effect generalized binomial (logit) models of the effect of security management activity on four security event types (botnet activity, malware activity, potential exploitations, and unsolicited communication). All models vary intercepts by day and firm (crossed). 133,248 observations in 480 firms from 1 July 2014 until 20 April 2015; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; standard errors in parentheses

This analysis, however, treats each event type independently and does not take advantage of the correlated nature of the events themselves. The presence of an event in a firm on a given day is not independent of the other events. Ineffective security management can leave a firm open for many types of threats. Therefore, we conducted additional analyses to account for the non-independence of errors at the level of event types.

Table 5 summarizes results from the hierarchical linear logit model with crossed date, firm, and type. Model M0 includes fixed effects for event type nested within date level (Equation 3) effects. Model M1 adds the focal variable, *open_ports*, and finds a positive relationship ($\beta = 0.636$,

$p < 0.001$) between the number of open ports and security events. Model M2 adds the firm level (Equation 2) effects to the date level (Equation 3) effects. Finally, Model M3 contains firm level (Equation 2), date level (Equation 3), and event type (Equation 4) effects. The relationship between open ports and security events is positive ($\beta = 0.252$, $p < 0.001$).

**Table 5 Effect of Security Management on Security Events**

| | M0 | M1 | M2 | M3 |
|---|---|---|---|---|
| Fixed effects estimates | | | | |
| Constant | $-2.671^{***}$ | $-5.092^{***}$ | $-5.151^{***}$ | $-5.843^{***}$ |
| | (0.198) | (0.832) | (0.936) | (0.991) |
| Open ports (log) | | $0.636^{***}$ | | $0.252^{***}$ |
| | | (0.003) | | (0.011) |
| Daily $t$ fixed effects | yes | yes | yes | yes |
| Event type $s$ fixed effects | yes | yes | yes | yes |
| Firm $f$ fixed effects | | | yes | yes |
| Random effect variance ($\sigma^2$) | | | | |
| Day level | 0.1943 | 0.839 | 0.452 | 0.626 |
| Event type level | 2.747 | 4.266 | 10.029 | 10.036 |
| Firm level | | | 7.983 | 6.372 |
| Fit statistics | | | | |
| Log likelihood | $-169,574.000$ | $-139,044.300$ | $-97,965.270$ | $-97,713.910$ |
| Ratio relative to control ($\chi^2$) | | $61059.49^{***}$ | | $502.73^{***}$ |
| Akaike information | 339,154.000 | 278,096.600 | 195,938.500 | 195,437.800 |

Mixed effect generalized binomial (logit) models of the effect of security management activity on four security event types (botnet activity, malware activity, potential exploitations, and unsolicited communication). Model M0 is control model varies intercepts by day; Model M1 adds *Open ports*; Model M2 varies intercepts by day and firm (crossed); Model M3 varies intercepts by day, firm, and event type (crossed). 532,992 observations in 480 firms from 1 July 2014 until 20 April 2015; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; standard errors in parentheses

For brevity, the fixed effect estimates for the 294 dates and 480 firms are not included in the table. However, Figure 4 (for firms and dates) and Figure 5 (for types) illustrate the variation in each of the fixed effects.

The ROC curve in Figure 6 indicates how much predictive value is added by the models. The control model, Model M0, has an area under curve (AUC) of 0.81 using only date and event type effects. The AUC increases to 0.89 with the addition of security management (open ports). The final model including crossed date, firm, and event type effects has an AUC of 0.95.

### 3.2. Hidden State Modeling

Hidden Markov Models (HMM) build upon the repeated (daily) observations of individual firms to identify two or more latent states of security. This examines the extent to which the measure

**Figure 4 Variation in fixed effects**

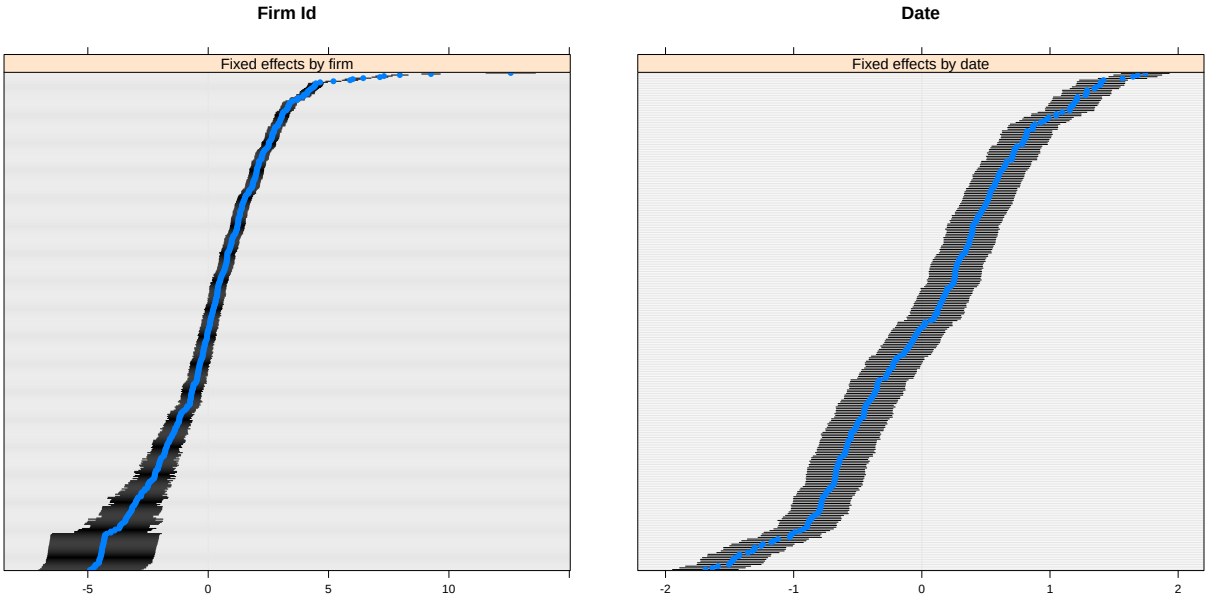
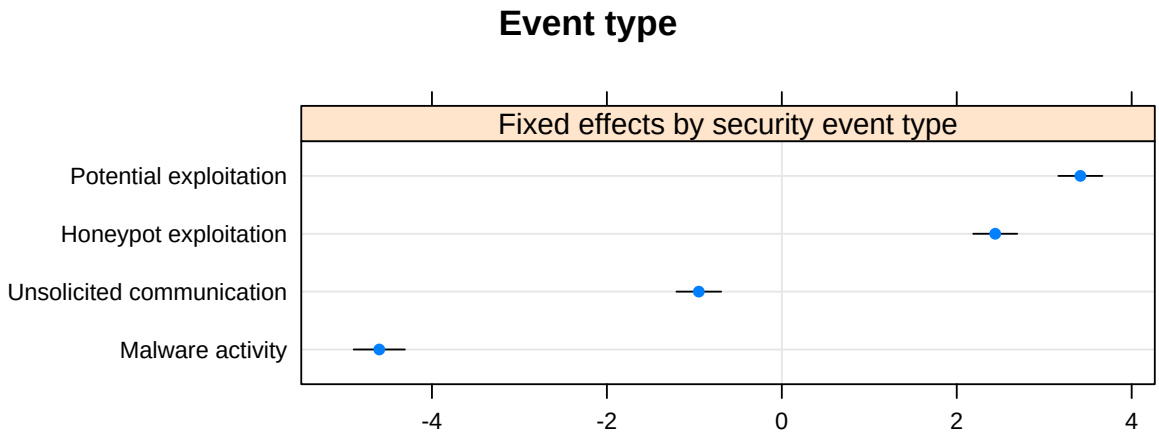



**Figure 5 Variation in fixed effects by security event type**

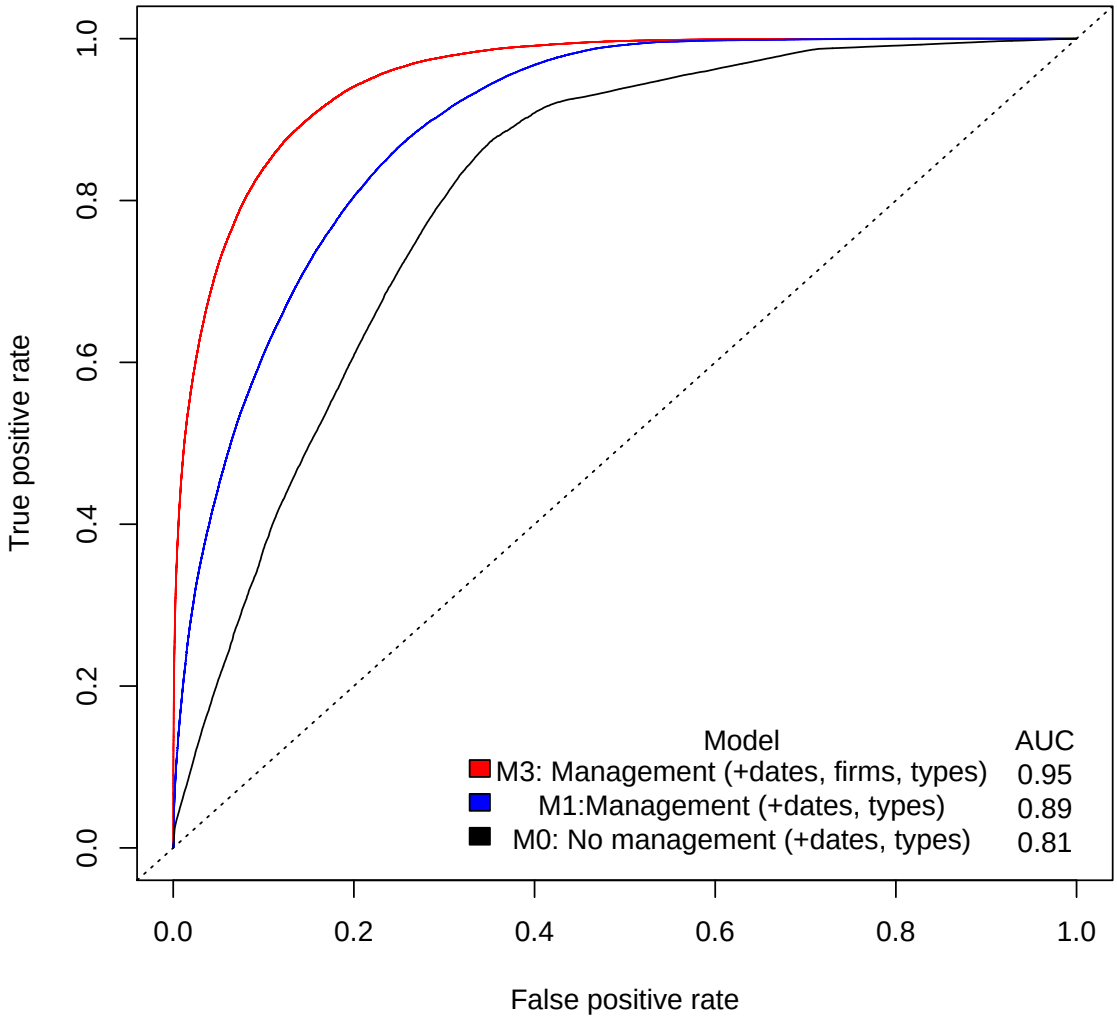



of basic security management, *open ports*, increases the likelihood of transitioning between higher security and lower security states. This analysis complements our prior analyses using alternative methods.

Table 6 estimates a transition matrix between the states and how security management influences the transitions. The first finding is that security states are relatively sticky; the probabilities of transitioning between states is low, as shown in the transition matrix. On average, the likelihood of transitioning from the high security state to the low one in a given day is only 1 percent.

Second, poor security management increases the likelihood of transitioning from the high to the low security state. The addition of the security management proxy improves the model fit;

**Figure 6 Receiver operating characteristic curves**




*Note.* Model M2 is excluded; it is similar to M3.

for example, AIC reduces from 224,942 to 224,577 ($\chi^2 = 368.96$, $p < 0.001$). The transition matrix indicates that a larger number of open ports increases the likelihood of transitioning from the high security state to the low security state ($\phi_{hl} = 0.340$). Additionally, a larger number of open ports decreases the likelihood of transitioning back from a low security state to a high security state ($\phi_{hl} = -0.139$). The results are consistent with the earlier regression analyses.

The analysis in Table 6 considers only two states. However, a less dichotomous scenario is likely. To investigate, we consider additional numbers of latent states. Table 7 summarizes those analyses. Adding additional states does increase model fit and reduce AIC and BIC. With each increasing number of states, the addition of security management measures increases the model

fit significantly. We focus on the two state model (Table 6) for simplicity of interpretation, but transition matrix results are consistent with larger numbers of latent states.

**Table 6 Hidden Markov Model Transitions Model Analysis**

| Variable | Model HMM0 | | Model HMM1 | |
|---|---|---|---|---|
| Initial States | | | | |
| | State $S_h$ | State $S_l$ | State $S_h$ | State $S_l$ |
| Probability | 0.634 | 0.366 | 0.631 | 0.369 |
| Transition Matrix | | | | |
| | State $S_h$ | State $S_l$ | State $S_h$ | State $S_l$ |
| State 1 ($S_h$) | 0.989 | 0.011 | 0.995 | 0.005 |
| State 2 ($S_l$) | 0.025 | 0.975 | 0.046 | 0.954 |
| Effects of Covariates on Transition Probabilities | | | | |
| | $S_h \to S_l$ | $S_l \to S_h$ | $S_h \to S_l$ | $S_l \to S_h$ |
| Constant | −3.573 | −1.916 | −5.358 | −3.027 |
| Open ports (log) | | | 0.340 | −0.139 |
| Loglikelihood | | -112,460.3 | | -112,275.8 |
| AIC | | 224,942.5 | | 224,577.5 |
| BIC | | 225,050.3 | | 224,704.9 |
| Loglikelihood Ratio $\chi^2$ | | | | 368.96*** |

Two state hidden Markov model where $S_h$ is the more secure state; $S_l$ is the less secure state. Significance levels: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; AIC is Akaike information criterion; BIC is Bayesian information criterion.

**Table 7 Effect of Varying the Number of States in Hidden Markov Models of Security Events**

| Hidden | Control Model | | | Security Management | | | |
|---|---|---|---|---|---|---|---|
| States | LogL | AIC | BIC | LogL | AIC | BIC | LLR |
| 2 | −112,460 | 224,943 | 225,050 | −112,276 | 224,578 | 224,705 | 368.96*** |
| 3 | −101,844 | 203,728 | 203,924 | −101,752 | 203,556 | 203,811 | 183.77*** |
| 4 | −88,587 | 177,236 | 177,540 | −88,313 | 176,712 | 177,133 | 548.38*** |
| 5 | −85,560 | 171,208 | 171,639 | −85,668 | 171,465 | 172,092 | −217.26 |

Hidden Markov models (with and without transition based on the number of open ports) modeling an increasing number of latent states. LogL is LogLikelihood; AIC is Akaike information criterion; BIC is Bayesian information criterion; LLR is the LogLikelihood Ratio.

Table 8 considers several alternatives to the focal results (Model M3 in Table 5). Given the sparseness of some attacks, the results could be driven by firms or dates that have little activity. Model R1 excludes 10% of the least attacked firms and continues to find that security management affects activity ($\beta = -5.223$, $p < 0.001$). Model R2 excludes 10% of the dates with the least activity and continues to find that security management affects activity ($\beta = -5.753$, $p < 0.001$). Additional covariates are available for many of the firms in our dataset. Model R3 includes additional firm covariates of the number of employees ($\beta = 0.409$, $p < 0.001$), firm assets ($\beta = 0.493$, $p < 0.01$), sales

per employee ($\beta = -0.469$, $p < 0.01$), and market value ($\beta = 0.362$, $p < 0.05$). Model R4 interacts market value and security management and finds that security management has a stronger effect in larger firms ($\beta = 0.057$, $p < 0.001$). Model R5 log transforms the number of security events as a dependent variable (instead of a dichotomous outcome) and finds that security management has a relationship with the log of the number of security events ($\beta = 0.033$, $p < 0.001$). All models vary intercepts by day, firm, and event type (crossed).

## 4. Discussion

While the practice of security depends on the assumption that security management improves security outcomes, the literature on security struggles to provide evidence of the assumption. While companies reveal security incidents in response to regulatory requirements, they are understandably reticent to share more than necessary. As a result, it can be difficult to empirically assess the links between security management and security events outside of labs or a single company. Further confounding the situation is the Hawthorne effect (Rothlisberger and Dickson 1939) — when researchers place attention on a change in security management, security practitioners pay more attention to what they are doing. A better approach would be to examine the effect of security management practices when nobody is looking, and to do so across a large number of firms.

Our study does exactly that. Our data, gathered on a daily basis by a security monitoring firm, consistently measures the basic level of security management in each firm, as signified by the number of open ports detected on each day. It also measures the extent to which the firm experiences four types of negative security events on that day. The power of our dataset rests in the consistent measurement of this data across 480 large firms.

While the data is powerful, it also has complexity. Namely, the prevalence of each event type is correlated across companies on a given day due to changing patterns of attack. They are correlated within a firm across days, since firms rarely resolve security events instantaneously. Furthermore, the prevalence of a given event type can be correlated with that of the other event types, either within or across firms. Therefore, we used two distinct methods to improve estimates of the link between security management and security events.

Hierarchical linear modeling of open ports against specific event types, using crossed days and firms to address lack of independence of errors, finds evidence of a link for three of the four events: botnet activity, potential exploitation, and unsolicited communications. (The event which did not show a link, malware activity, is also the one least present in our dataset and also likely to be associated with open ports on a company's server.) However, this analysis did not account for the correlation between events. An additional HLM analysis using crossed day, firm, and event type finds an effect of security management on all four event types. As an additional test, Hidden

**Table 8 Effect of Security Management on Adverse Events (Robustness)**

| | **R1** Excluding least attacked firms | **R2** Excluding least attacked dates | **R3** Firm covariates | **R4** Firm covariates | **R5** Continuous outcome (log) |
|---|---|---|---|---|---|
| Fixed effects estimates | | | | | |
| Constant | $-5.223^{***}$ | $-5.753^{***}$ | $-5.436^{***}$ | $-5.588^{***}$ | 0.134 |
| | (0.901) | (1.074) | (1.150) | (0.961) | (0.123) |
| Open ports (log) | $0.251^{***}$ | $0.234^{***}$ | $0.305^{***}$ | $0.261^{***}$ | $0.033^{***}$ |
| | (0.011) | (0.012) | (0.012) | (0.013) | (0.001) |
| Employees (std.) | | | $0.409^{***}$ | $0.431^{***}$ | |
| | | | (0.117) | (0.119) | |
| Assets (std.) | | | $0.493^{**}$ | $0.495^{**}$ | |
| | | | (0.167) | (0.167) | |
| Sales / employee (std.) | | | $-0.469^{**}$ | $-0.441^{**}$ | |
| | | | (0.147) | (0.149) | |
| Market value (std.) | | | $0.362^{*}$ | 0.105 | |
| | | | (0.170) | (0.172) | |
| Market value × Open ports (std.) | | | | $0.057^{***}$ | |
| | | | | (0.005) | |
| Random effect variance ($\sigma^2$) | | | | | |
| Day level | 0.623 | 0.526 | 0.663 | 0.622 | 0.004 |
| Firm level | 4.173 | 6.554 | 4.017 | 4.108 | 0.260 |
| Event type level | 9.994 | 7.214 | 2.686 | 10.047 | 0.058 |
| Fit statistics | | | | | |
| Observations | 474,444 | 476,468 | 411,264 | 411,264 | 564,220 |
| Log likelihood | −97,261.0 | −89,651.9 | −80,578.5 | −80,508.4 | −459,995.8 |
| Akaike information | 194,531.9 | 179,313.7 | 161,175.0 | 161,036.8 | 920,003.7 |

Mixed effect generalized binomial (logit) models (R1, R2, R3, and R4) or mixed hierarchical linear model (R5) of the effect of security management activity on four security event types (botnet activity, malware activity, potential exploitations, and unsolicited communication). Model R1 excludes 10% of the least attacked firms; Model R2 excludes 10% of the dates with the least activity; Model R3 includes additional firm covariates; Model R4 interacts market value and security management; Model R5 uses a log transformation of the number of security event types. All models vary intercepts by day, firm, and event type (crossed). Observations in 480 firms from 1 July 2014 until 20 April 2015; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; standard errors in parentheses. "(std)" indicates standardized variables for presentation.

Markov models exploit the longitudinal daily nature of our data. A two-state model shows that poor security management significantly increases the likelihood of a firm's transition from high security to a low security state. Additional tests exploring higher numbers of latent states consistently demonstrate this pattern.

Our study is not without limitations. Data on security management and security outcomes is difficult to obtain across large numbers of firms. Our measure of security management, namely open ports, does not cover the full range of security management. It is a proxy for companies doing basic levels of security management. Our study is also limited by the data gathering mechanisms of the security monitoring service provider. Events such as DDOS or practices such as default admin passwords cannot be detected by this provider and thus are beyond the scope of this research. Finally, while the study examines practices and events at each firm each day, it does not examine relationships to broad levels of attack across industries or to any events that could cause industry-wide attention to security management. In our models, these effects would be addressed by the crossed day, firm and event controls, but not examined directly.

Further research can examine the effect of particular high profile events on the security management of companies. It could also examine the extent to which event levels experienced in a firm relate to announced incidents, or even to financial outcomes. Finally, we could supplement this secondary data with primary data on security management, perhaps through surveys or other methods. We are currently pursuing these research avenues.

## 5. Conclusion

Increasing costs of security management are leading to increased skepticism among senior executives about the need to provide higher and higher levels of security funding. While the assumption that security management improves security outcomes has face validity, it has been surprisingly difficult to show empirically across a broad range of firms. Our study provides evidence that the core assumption behind the practice of security is valid. Better security management does improve security outcomes.

This finding is comforting, if not surprising. The main contribution of our work is that we are the first researchers to show this link empirically in steady state across a large number of firms. An additional contribution is showing the value of a data source which can capture the daily security posture of hundreds of companies. We believe the methods and findings in this paper are an important addition to the literature on security management. We hope our study will serve as a foundation to further advance the theory and practice of security management.

## References

Acquisti, A., I. Adjerid, R.H. Balebako, L. Brandimarte, L.F. Cranor, S. Komanduri, P.G. Leon, N. Sadeh, F. Schaub, M. Sleeper, S. Wang, Y. amd Wilson. 2016. Nudges for privacy and security: Understanding and assisting users choices online. Tech. rep. Https://ssrn.com/abstract=2859227.

Albrechtsen, E., J. Hovden. 2009. The information security digital divide between information security managers and users. *Computers & Security* **28**(6) 476–490.

August, T., R. August, H. Shin. 2014. Designing user incentives for cybersecurity. *Communications of the ACM* **57**(11) 43–46.

August, T., T.I. Tunca. 2006. Network software security and user incentives. *Management Science* **52**(11) 1703–1720.

Baldwin, A., I. Gheyas, C. Ioannidis, D. Pym, J. Williams. 2012. Contagion in cyber security attacks. *Journal of the Operational Research Society* .

Dhillon, G., J. Backhouse. 2000. Technical opinion: Information system security management in the new millennium. *Communications of the ACM* **43**(7) 125–128.

Edwards, B., J. Jacobs, S. Forrest. 2016. Risky business: Assessing security with external measurements. Tech. rep.

Gordon, L.A., M.P. Loeb, W. Lucyshyn, R. Richardson. 2005. 2005 CSI/FBI computer crime and security survey. *Computer Security Journal* **21**(3) 1.

Gupta, A., D. Zhdanov. 2012. Growth and sustainability of managed security services networks: an economic perspective. *MIS Quarterly* **36**(4) 1109–1130.

Kannan, K., R. Telang. 2005. Market for software vulnerabilities? Think again. *Management Science* **51**(5) 726–740.

Kholidy, H.A., F. Baiardi. 2012. Cids: A framework for intrusion detection in cloud systems. *Information Technology: New Generations (ITNG), 2012 Ninth International Conference on*. IEEE, 379–385.

Laszka, A., B Johnson, P. Schöttle, J. Grossklags, R. Böhme. 2013. Managing the weakest link. *European Symposium on Research in Computer Security*. Springer, 273–290.

Liu, Yang, Armin Sarabi, Jing Zhang, Parinaz Naghizadeh, Manish Karir, Michael Bailey, Mingyan Liu. 2015. Cloudy with a chance of breach: Forecasting cyber security incidents. *USENIX Security*. 1009–1024.

Mahmood, M.A., M. Siponen, D. Straub, H.R. Rao, T.S. Raghu. 2010. Moving toward black hat research in information systems security: an editorial introduction to the special issue. *MIS quarterly* **34**(3) 431–433.

Mitra, S., S. Ransbotham. 2015. Information disclosure and the diffusion of information security attacks. *Information Systems Research* **26**(3) 565–584.

Moore, T., S. Dynes, F.R Chang. 2015. Identifying how firms manage cybersecurity investment. *Available: Southern Methodist University. Available at: http://blog. smu. edu/research/files/2015/10/SMU-IBM. pdf (Accessed 2015-12-14)* **32**.

Ransbotham, S., S. Mitra. 2009. Choice and Chance: A conceptual model of paths to information security compromise. *Information Systems Research* **20**(1) 121–139.

Ransbotham, S., S. Mitra, J. Ramsey. 2012. Are markets for vulnerabilities effective? *MIS Quarterly* **36**(1) 43–64.

Ransbotham, Sam. 2010. An empirical analysis of exploitation attempts based on vulnerabilities in open source software. *WEIS*.

Rothlisberger, F.J., W.J. Dickson. 1939. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago*. Harvard University Press.

Sarabi, Armin, Parinaz Naghizadeh, Yang Liu, Mingyan Liu. 2016. Risky business: Fine-grained data breach prediction using business profiles. *Journal of Cybersecurity* **2**(1) 15–28.

Siponen, M.T., H. Oinas-Kukkonen. 2007. A review of information security issues and respective research contributions. *ACM Sigmis Database* **38**(1) 60–80.

Straub, D., R. Welke. 1998. Coping with systems risk: Security planning models for management decision making. *MIS Quarterly* **22**(4) 441–469.

Vieira, K., A. Schulter, C. Westphall, C. Westphall. 2010. Intrusion detection techniques in grid and cloud computing environment. *IT Professional, IEEE Computer Society* **12**(4) 38–43.

Wang, C., L. Fang, Y. Dai. 2010. A simulation environment for scada security analysis and assessment. *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, vol. 1. IEEE, 342–347.

Zhang, Jing, Zakir Durumeric, Michael Bailey, Mingyan Liu, Manish Karir. 2014. On the mismanagement and maliciousness of networks. *NDSS*.