Rev Account Stud (2018) 23:1177–1206
https://doi.org/10.1007/s11142-018-9452-4

Do firms underreport information on cyber-attacks? Evidence from capital markets

Eli Amir[1] · **Shai Levi**[1] · **Tsafrir Livne**[2]

Published online: 19 June 2018

Abstract Firms should disclose information on material cyber-attacks. However, because managers have incentives to withhold negative information, and investors cannot discover most cyber-attacks independently, firms may underreport them. Using data on cyber-attacks that firms voluntarily disclosed, and those that were withheld and later discovered by sources outside the firm, we estimate the extent to which firms withhold information on cyber-attacks. We find withheld cyber-attacks are associated with a decline of approximately 3.6% in equity values in the month the attack is discovered, and disclosed attacks with a substantially lower decline of 0.7%. The evidence is consistent with managers not disclosing negative information below a certain threshold and withholding information on the more severe attacks. Using the market reactions to withheld and disclosed attacks, we estimate that managers disclose information on cyber-attacks when investors already suspect a high likelihood (40%) of an attack.

Keywords Cyber attacks · Data breaches · Disclosure

Jel classification M41 · G14

✉ Shai Levi
shailevi@tau.ac.il

Eli Amir
eliamir@post.tau.ac.il

Tsafrir Livne
tsafrir@unc.edu

[1] Coller School of Management, Tel Aviv University, Tel Aviv 6997801, Israel

[2] Kenan-Flagler Business School, University of North Carolina, Chapel
Hill 27516-3490 NC, United States

1 Introduction

Firms must disclose cyber-attacks that materially damage their businesses (Securities and Exchange Commission 2011, 2018). However, because investors cannot discover most cyber-attacks independently, and because managers often have incentives to withhold negative unobservable information from investors, firms may underreport cyber-attacks. In this study, we estimate the extent to which publicly traded firms withhold information on cyber-attacks. Specifically, we identify cyber-attacks that firms disclosed and attacks that were withheld and later independently discovered. We then use the differential market reaction to these attacks to estimate the extent of underreporting.

Reviewing data on cyber-attacks between 2010 and 2015 suggests many disclosures on the attacks are made after investors discover them. Data breaches are revealed to the market, for example, by customers whose information is stolen or by the hackers themselves.[1] In addition, the number of cyber-attacks public companies disclosed, about 300 during that period, seems low in comparison to the thousands reported by independent sources.[2]

The extent of information withholding is unobservable, and we are aware only of data breaches that are eventually revealed either by the attacked firms or sources outside the firm. We estimate the extent of withholding from the market reaction to revealed attacks, where market reaction approximates the damage caused by cyber-attacks. We find that, in cases where firms immediately disclosed the cyber-attack, their equity values declined by 0.33%, on average, in the three days after disclosure and by 0.72% in the month after disclosure. In comparison, the decline in market values was much larger in cases where firms did not disclose the attack and parties outside the firm later discovered it: 1.47% in the three days after the discovery of the attack, and 3.56% in the month afterward. These findings suggest firms withhold more severe cyber-attacks from investors. From the differential market reaction to disclosed and withheld attacks, we estimate that managers disclose cyber-attacks when investors already believe that, with a 40% chance, an attack has occurred; when uncertainty about the existence of a cyber-attack is higher, managers withhold the information.

Using alternative estimates of damage caused by cyber-attacks, we also find that information about more severe attacks is withheld. Specifically, we use damage estimates released by the attacked firms and an objective index that measures the severity of cyber-attacks, based on type of data breached, the number of records stolen, and the source of the breach. Both damage estimates show that firms withhold information about more severe attacks, whereas milder attacks are more likely to be disclosed by the firm.

[1] For example, Target, the US retailer, experienced a data breach involving millions of its customers' credit and debit cards and, after customers and credit card companies revealed the breach, the firm confirmed it. In some cases, the hackers themselves may reveal the breach. For example, hackers breached the LinkedIn network and stole a database containing 6.5 million users' encrypted passwords in June 2013. The hackers later published the attack, hoping to receive help from fellow hackers in cracking the encrypted passwords. After the hackers published the passwords, LinkedIn acknowledged this breach.

[2] According to Verizon (2015), more than 20,000 data breaches occurred in the US private sector during that period.

In support of the relation between chances of discovery and withholding, we find withholding firms have less analyst coverage, weaker corporate governance, and lower litigation risk than disclosing firms. Investors follow more closely firms with greater analyst coverage, and the chance of discovery in these firms is higher. In addition, firms with stronger governance are less likely to conceal negative news from their investors. Specifically, firms with less entrenched management (Bebchuk et al. 2009) and fewer material weaknesses reported following Section 404 of the Sarbanes-Oxley Act are more likely to disclose information on cyber-attacks. Using membership in the high-tech industry as a proxy for litigation risk, we find disclosing firms are more likely than withholding firms to be in high-tech industries. High litigation risk increases the expected loss from withholding information, increasing the attractiveness of disclosure (e.g., Skinner 1994, 1997).

We contribute to the literature by using disclosure theory to explain the market effects of cyber-attacks. Studies that examine the stock price reaction to cyber-attacks find mixed results. For instance, Cavusoglu et al. (2004) find data breaches have a statistically significant negative effect on stock prices. By contrast, Campbell et al. (2003) and Kannan et al. (2007) find the market effect of breaches is generally insignificant.[3] Gordon et al. (2011) report a decrease in the effect of breaches on stock prices over time. They conjecture that, with increased media reporting of data breaches without apparent devastating effects on targeted corporations, investors lowered their assessment of the costs of data breaches. Kvochko and Pant (2015) review recent cases in which large data breaches had a small impact on stock prices.[4] Consistent with the latter studies, we find the negative reaction to most cyber-attacks in our sample (2010–2015) is quite small. However, unlike prior researchers, we distinguish between cyber-attacks that were voluntarily disclosed and those that were withheld from investors and later independently discovered, and we find that, in the latter cases, the market reaction is negative and significant. These results suggest cyber-attacks that are unknown to investors are more likely to be severe and that the market reaction reported elsewhere understates the damage cyber-attacks cause.

We also contribute to the literature that examines the different timing of good- and bad-news disclosures. For example, Kasznik and Lev (1995) examine whether firms warn investors of upcoming negative earnings surprises. Amir and Ziv (1997) find firms delay the adoption of new accounting standards with negative financial effects. Chambers and Penman (1984) find late earnings announcements contain, on average, worse news than early announcements. Kothari et al. (2009) find the magnitude of negative stock price reaction to bad news is greater than the magnitude of positive stock

[3] Mixed results exist also for specific types of data breaches. For example, Hovav and D'arcy (2003) and Kannan et al. (2007) find denial-of-service attacks have an insignificant effect, whereas Ettredge and Richardon (2003) find this kind of attack has a significant negative impact on the market value of firms. For further review of this literature, see Spanos and Angelis (2016).

[4] The market reaction to data breaches is, on average, not different from zero also according to Hilary et al. (2016) and Gordon et al. (2010) find firms gain market value when they voluntarily disclose information on items pertaining to information security.

price reaction to good news and infer from their evidence that managers accumulate and withhold bad news up to a certain threshold but leak and immediately reveal good news.[5] In the case of cyber-attacks, however, withheld information will likely never be revealed to investors. In addition, for cyber-attacks that are eventually revealed, the data indicate when the firm learned of the attack and therefore whether information withholding occurred. This setting and data enable us to distinguish between cases of disclosing and withholding and show that, consistent with theory, managers withhold more negative information and voluntarily disclose less severe attacks. This setting also allows us to examine the different market reactions to withholding and disclosure and to estimate when withholding information is worthwhile for managers. Using market reactions to withheld and disclosed attacks, we show that managers disclose cyber-attacks only when the likelihood that investors believe an attack is imminent is high.

2 Hypothesis development

When investors know managers possess and withhold negative information, they will reduce share price to reflect the worst possible news, which, in turn, will drive managers to make full disclosure (Grossman and Hart 1980; Grossman 1981). However, when investors are uncertain about whether managers possess negative information, a partial disclosure equilibrium will emerge where some firms find it beneficial to withhold bad news (Dye 1985; Jung and Kwon 1988).

Cyber-attacks are often unobservable to the public when they occur, and thus managers can withhold information about attacks from investors without being discovered. Without disclosure, investors will reduce stock prices only by the expected value of the bad news withheld, which in the case of cyber-attacks equals the probability that an attack occurred (and the manager is withholding the information) times the average damage. Because the number of attacks discovered by investors is small, investors usually do not have a reason to suspect that, in the absence of disclosure, the likelihood a breach occurred is high.[6] Therefore the expected loss from not disclosing is low, and withholding is an attractive option for managers.[7]

To develop our main hypothesis—firms will withhold information on the more severe cyber-attacks and voluntarily disclose the milder ones—we use a setting similar to that used by Dye (1985).[8] Assume a cyber-attack on the firm with a probability p and a loss x. Only the manager learns of the attack and the damage, x, whereas investors know the ex-ante distribution of the damage, $\tilde{x}$.

[5] Baginski et al. (2018) show that managers' career concerns lead them to delay disclosure of bad news.

[6] As we will show, public firms reported only dozens of data breaches over our six-year sample, and thus the probability of significant attacks seems low.

[7] Litigation costs that can deter withholding are also expected be low. Litigation follows almost every data breach (Southwell et al. 2017)—breaches that are voluntarily disclosed by firms as well as those withheld by firms and later discovered by third parties. It seems that firms withholding information avoid (in case the withheld breach is not discovered) the almost automatic litigation that follows, and therefore their expected litigation costs are not necessarily higher than those of firms voluntarily disclosing the breach (White 2014).

[8] Dye (1985) assumes firm owners wish to maximize current share price and provide managers with incentives to withhold negative information. The assumption that, in general, managers wish to maximize share prices is reasonable because their career and reputation are often linked to share prices.

Managers will withhold information on the damage, x, when the loss from disclosing (equal to x) is higher than the expected loss from withholding. Because investors know the ex-ante distribution of the damage, $\tilde{x}$, and the probability of cyber-attacks in the case of no disclosure, they can estimate the probability that the decision not to disclose is due to withholding, *prob(withholding)*, and the expected loss in the case of no disclosure, which is $prob(withholding)E(\tilde{x}|withholding)$. Managers are aware that, in the absence of disclosure, investors will adjust stock prices down by this expected loss. Then, when managers observe the actual damage from a cyber-attack, x, they will decide to withhold the information in case $x < prob(withholding)E(\tilde{x}|withholding)$.

Dye ([1985]) shows that, in such a setting, a disclosure threshold, $\underset{\sim}{x}$, exists above which managers will disclose information, which equals $prob(withholding)E(\tilde{x}|\tilde{x} < \underset{\sim}{x}) = \underset{\sim}{x}$. The disclosure threshold, $\underset{\sim}{x}$, equals the probability, *prob(withholding)*, that the manager has news and withholds it times the expected value of the news withheld. Managers withhold bad news if the damage from the attack is worse (i.e. lower) than $\underset{\sim}{x}$. It follows that managers will withhold information on the more severe cyber-attacks—those that will cause a loss in stock prices below the disclosure threshold, $\underset{\sim}{x}$.[9] Because the expected value of the bad news withheld, $E(\tilde{x}|\tilde{x} < \underset{\sim}{x})$, is negative, when investors believe the probability of managers holding negative information, *prob(withholding)*, is higher, the disclosure threshold, $\underset{\sim}{x}$, will be lower and managers will disclose more negative news.[10] Using market reactions, we empirically estimate the probability of withholding, *prob(withholding)*, at which managers choose to withhold information about a cyber-attack.

The probability of withholding equals $\frac{\underset{\sim}{x}}{E(\tilde{x}|\tilde{x}<\underset{\sim}{x})}$. To estimate this probability, we need empirical proxies for the disclosure threshold, $\underset{\sim}{x}$, and the expected value of bad news withheld, $E(\tilde{x}|\tilde{x} < \underset{\sim}{x})$. We use the average stock returns in the withholding cases that are discovered by investors as a proxy for the expected value of bad news withheld, $E(\tilde{x}|\tilde{x} < \underset{\sim}{x})$.[11] We assume the average damage of the discovered attacks represents the damage in withheld cases. To estimate the disclosure threshold, $\underset{\sim}{x}$, we use the average return reaction in the cases in which managers disclosed the cyber-attack. However, managers disclose losses whenever they are low enough (above the threshold), and we observe the average loss. To estimate the threshold loss, we assume, as Dye (1985) does, the loss is uniformly distributed on the interval $[x,0]$, where the threshold x is a negative number.[12] The expected loss disclosed is hence $\frac{\underset{\sim}{x}}{2}$. It follows that the probability that managers are withholding bad news on cyber-attacks is

[9] In practice, the probability of independent discovery by investors may affect the disclosure policy. However, Dye's (1985) model does not consider the probability of independent discovery of bad news by investors (cyber-attack, in our case). He assumes that investors cannot discover bad news that the manager withheld. Because the probability of discovery of cyber-attacks by investors is practically very small, Dye's (1985) model adequately describes disclosure of cyber-attacks, as we empirically demonstrate.

[10] Jung and Kwon (1988) show how, in this setting, an increase in the probability with which investors believe managers have negative information will lower the disclosure threshold and will trigger the release of information managers would otherwise withhold.

[11] As discussed below, the average market reaction to attacks that are discovered may be a biased estimate of the damage. Specifically, the decrease in price upon discovery may be larger than the damage due to the negative reputation effects and litigation risk associated with withholding. In this case, our withholding-probability estimate will be downward biased.

[12] Dye (1985) uses the same assumption in the illustrative example of his theorem (p. 129). As discussed below, even if the loss is not uniformly distributed, we can still estimate the minimal probability of withholding, because the disclosure threshold will not be higher than the actual return reaction in the cases in which firms disclosed the cyber-attack.

$$prob(withholding) = \frac{2{*}Return\ reaction\ to\ immidiate\ disclosure}{Return\ reaction\ to\ discovery\ of\ \ withholding} \qquad (1)$$

As we show below, the average market reaction is, for example, −0.72% in the month following an immediate disclosure of the breach by firms and − 3.56% when the breach was not disclosed but investors later discovered it. These estimates imply managers disclose cyber-attacks when investors already believe that, with a probability of 40%, an attack has occurred.

3 Data

We combine two data sources that report details on cyber-attacks. Our first data source is the AuditAnalytics cyber-attacks database, which documents 186 incidents between 2010 and 2015. For 162 of these incidents, we obtain stock returns from the Center for Research in Security Prices (CRSP). The second data source is the VCDB VERIS community database, which contains thousands of documented incidents, of which only a small fraction relates to public companies. A description of the VCDB VERIS database is available, for example, in the Verizon (2015) Data Breach Investigations Report. According to the report, the database includes information on data breaches collected by Verizon during its "paid external forensic investigation" services and by 70 other cyber-security companies and organizations. We match company names in VCDB with those of US publicly traded company names in CRSP using fuzzy-matching, and then manually verify the results.[13] We identify 158 additional data breaches of public firms between 2010 and 2015 that are not included in the AuditAnalytics database. We then validate the information in each entry of the combined dataset (e.g. the accuracy of disclosure dates) through the reference links that are provided in the dataset, as well as through searches in news sources and company filings.

The identity of most firms on the database is unknown. The anonymity of breaches balances the need of VCDB VERIS users for relevant data with the privacy of firms. Most private firms do not share information on their operations with third parties, and many of the anonymous records are likely of private firms. Public firms will also be reluctant to reveal negative information to competitors and investors, so a large number of anonymous breaches in the database can therefore contain records of public firms.

Finally, for the purpose of using the same sample throughout the analysis, we exclude firms that are missing data on material weakness following Section 404 of the Sarbanes-Oxley Act of 2002 and data necessary for calculating Bebchuk et al.'s (2009) index—see details below.[14] Combining our data sources, we obtain data for 276 incidents involving 156 publicly traded companies between 2010 and 2015, of which 58 firms had more than one cyber-attack.

[13] Fuzzy matching is a textual search-algorithm that provides a score for the likelihood that a pair of text strings is similar. For instance, 'Microsoft Corporation', and 'Microsoft corp.' will receive a very high matching score by the algorithm.

[14] This sample-selection criterion does not change the results, and results with all 320 incidents are similar to those presented below.

Table 1 Sample selection

Year	Number of Cyber-Attacks	Number of Firms	Attack Type Availability	Confidentiality	Integrity
2010	16	14	3	12	1
2011	30	24	5	24	0
2012	45	39	13	29	2
2013	50	40	15	34	1
2014	86	68	11	71	2
2015	49	42	11	36	2

The table presents the number of cyber-attacks and firms by year and type of attack. “Availability” includes breaches that stop the business from making its services available to customers (also known as denial of service). “Confidentiality” incudes breaches that allow unauthorized users access to confidential information such as bank account credentials, credit card data, medical records, insurance history, usernames, or passwords. “Integrity” includes breaches that compromise the reliability of a database or a website. The incidents occurred between 2010 and 2015

For descriptive purposes, we classify cyber-attacks into three categories following Gordon et al. (2011). The first category—availability—includes breaches that stop the business from making its services available to customers (also known as denial of service). We include in this category cases in which the breach can jeopardize availability, for example, cases in which hackers gain access and can disrupt main systems or steal intellectual property. The second category—confidentiality—are breaches that allow unauthorized users access to confidential information, such as bank account credentials, credit card data, medical records, insurance history, usernames, or passwords. The third category—integrity—includes breaches that compromise the reliability of a database or a website. We classify all breaches in the sample into one of the three categories, except for four breaches for which the information is not available in the data and the attack type is unknown. Table 1 summarizes information on our sample by year and attack type.

4 Results

4.1 Univariate analysis

We classify sample firms into three groups according to their disclosure policy. We classify an incident as “disclosing” if the firm disclosed the cyber-attack before an outside party discovered it or concurrently with a discovery of the incident by outsiders (143 cases, 51.8%). We classify an incident as “withholding” if the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack (47 cases, 17.0%).[15]

[15] We classify cases as “withholding” only if the firm clearly learned of the attack before a party outside the firm discovered it. In many cases, firms eventually disclose the date on which they learned of the attack; AuditAnalytics, the data vendor, provides this date, and we collect this date for VCDB VERIS data cases.

Disclosing the information shortly after the breach does not amount to withholding. Firms may have legitimate reasons to delay disclosure of a breach. For example, fixing security vulnerabilities is the reason that Sony PlayStation Network used to explain its short delay in disclosing a breach of 77 million of its user accounts on April 2011.[16] Indeed, firms may need a few days to reasonably assess the attack and repair it. However, as the withholding time increases, fixing security vulnerabilities becomes a less plausible explanation. For example, Target experienced a data breach involving 40 million of its customers' credit and debit cards in November 2013. Only three weeks later, after third parties discovered the breach, Target disclosed it to shareholders, saying "Target is partnering with a leading third-party forensics firm to conduct a thorough investigation of the incident," but not directly stating that fixing vulnerabilities is the reason for the withholding.[17] The reasons companies provide for withholding information are often not verifiable and usually driven by litigation concerns (White 2014). We therefore use an objective criterion to determine withholding—if the firm was aware of the breach and did not disclose it until its discovery by third parties, we define the breach as withheld. When we alternatively define withholding as a case in which the firm did not disclose the breach for at least 14, 21, or 30 days after it learned of the attack and then investors discovered the attack only from third parties, we find similar results under all alternative definitions.[18] The Securities and Exchange Commission requires firms to disclose material negative events to investors, and nondisclosure of a material cyber-attack is not a legitimate choice for firms (e.g., White 2014).[19] Firms are not required to disclose attacks with immaterial effects. In 86 cases, we find that, after discovery of the breach, the firm clarified it was immaterial.[20] Hence we classify these 86 cases (31.2%) as "immaterial."

According to our hypothesis, withholding information is more likely when the damage of the cyber-attack is larger. We use three measures of the cost of the cyber-attack. *Damage* is an estimate, made by the attacked firm, of the damage the cyber-attack caused, divided by the market value of equity at the beginning of year. We obtained 38 such damage estimates provided by the attacked firms in their financial statements after the attack. The second variable, *severity*, is an index taking values from 0 to 10, depending on the severity of the cyber-attack (0 = low damage and 10 = very high damage). Gemalto (an international digital security company) created the index to measure the severity of cyber-attacks. It rates the severity of data breaches based on the type of data breached, the number of records stolen, the source of the breach, and

[16] Stringer, H. (2011, May 5). A Letter from Howard Stringer. *Sony Corporation*. Retrieved from http://blog.us.playstation.com.

[17] Target (2013, December 19). Target Confirms Unauthorized Access to Payment Card Data in U.S. Stores. Retrieved from https://corporate.target.com.

[18] We find stronger results when we define withholding as a case in which the firm did not disclose the breach for longer periods after it learned of it. For example, for firms that did not disclose the breach for at least 14 days, the returns in the month after the discovery is −4.83%, compared with −3.56% reported in Table 4.

[19] Only enforcement agencies that investigate an attack can require a firm not to disclose the breach to allow them time to complete the investigation. We did not find any such requests in the withholding cases included in our sample.

[20] In 30 out of the 86 immaterial cases, the firm ignored reports on cyber-attacks. For legal purposes, a nonresponse is considered a statement that the event was immaterial. These cases were indeed minor and occurred in large companies. Omitting these 30 cases from the sample does not change the results in any meaningful way.

whether the hackers used the stolen data. We calculated the index for the entire sample. *Severity* and *damage* are highly correlated (Pearson correlation of 0.49).

The third measure is the market reaction to the cyber-attack, *Ret(−1,3)*, which is the cumulative risk-adjusted returns from one trading day before until three trading days after the date a cyber-attack became known to investors. The market reaction should reflect the damage the firm incurs from the attack. However, the market reaction may also reflect the negative reputation the firm incurs from withholding information. In the multivariate analysis below, we control for this endogeneity and show our results are robust.

We adjust stock returns for risk using the value-weighted market return reported by CRSP (VWRETD). Specifically, we compute the difference between the buy-and-hold returns of the stock and the buy-and-hold-returns of the value-weighted market portfolio.[21] We use this relatively simple risk adjustment because we can apply it to all the cases in our sample, thus maximizing sample size.[22]

Consistent with our hypothesis, Panel A of Table 2 shows the average severity of attacks in the withholding cases, 4.92, is larger than in the disclosing cases, 4.32, at the 0.09 level. The decrease in stock price at the discovery of withholding cases, −1.47%, is also larger than in disclosing cases, −0.33%, at the 0.04 level.[23] We find that damage in the withholding cases, which is 1.81% of the market value of equity, is larger than in the disclosing cases, 0.62% of market value of equity. However, we only have 13 withholding cases and 19 disclosing cases. Although the damage in the withholding cases is about three times larger than in the disclosing cases, on average, the difference is significant only at the 0.11 level. Overall, the results using the alternative measures of attack severity support the hypothesis that firms withhold information on larger attacks and voluntarily disclose smaller attacks. Panel B of Table 2 compares disclosing to immaterial cases. The results show the three alternative measures of attack severity (*severity*, *Ret(−1,3)*, and *damage*) are larger in disclosing than immaterial cases. This result is not surprising because immaterial attacks, by their nature, are small.

Next, we compare the characteristics of withholding and disclosing firms. Firms are more likely to disclose cyber-attacks when the likelihood of outside parties discovering the breach is larger. As a measure of outside monitoring by investors, we use the number of analysts following the firm. Firms with greater analyst coverage are followed more closely by investors and thus more likely to disclose negative information, such as data breaches. We measure analyst coverage (*analysts*) as the number of analysts on I/B/E/S during the year. As Panel A shows, firms that disclosed cyber-attacks are followed by 14.11 analysts, on average, whereas withholding firms are followed by 9.86 analysts, on average, and the difference is statistically significant at the 0.02 level. Panel B of Table 2 shows the average number of analysts following firms with immaterial cyber-attacks was 18.57, which is larger than the number of analysts following disclosing firms, at the 0.01 level. These results suggest withholding firms are followed by fewer analysts, which is consistent with less monitoring by investors and hence the lower probability of independent discovery.

[21] This approach is equivalent to using a beta equal to 1, as firm-specific beta estimates are noisy (Fama and French, 1996).

[22] Using alternative risk adjustments for smaller samples, we find similar results. See Table 9.

[23] Negative reputation from withholding can also affect stock returns, and we control for this endogenous effect in Table 7 below.

Table 2 Univariate analysis

Variable	(a)			Exp. Relation	(b)			(a) – (b)	
	Disclosure				Withholding			***P*-values**	
	#Obs.	Mean	Median		#Obs.	Mean	Median	*t*-test	*W-test*
Panel A: Disclosing vs. Withholding									
Severity	143	4.32	4.00	<	47	4.92	5.60	(0.09)*	(0.08)*
Ret(−1,3)	143	−0.326	−0.086	>	47	−1.470	−1.304	(0.04)**	(0.02)**
Damage	19	0.624	0.129	<	13	1.812	0.274	(0.11)	(0.20)
Analysts	143	14.11	16.25	>	47	9.86	7.83	(0.02)**	(0.02)**
SOX404	143	0.08	0.00	<	47	0.66	0.00	(<.01)***	(<.01)***
Entrenchment	143	1.36	1.00	<	47	1.77	2.00	(0.02)**	(0.03)**
Hi-Tech	143	0.22	0.00	>	47	0.09	0.00	(0.02)**	(0.02)**
ROA	143	0.05	0.05	?	47	0.05	0.04	(0.42)	(0.40)
MV	143	48,244	12,852	>	47	27,001	5905	(0.08)*	(0.06)*
Panel B: Disclosing vs. Immaterial									
Severity	143	4.32	4.00	>	86	3.08	2.10	(<.01)***	(<.01)***
Ret(−1,3)	143	−0.326	−0.086	<	86	0.274	−0.260	(0.15)	(0.23)
Damage	19	0.624	0.129	>	6	0.027	0.003	(0.14)	(0.01)***
Analysts	143	14.11	16.25	?	86	18.57	20.21	(0.01)***	(0.02)**
SOX404	143	0.08	0.00	?	86	0.09	0.00	(0.45)	(0.42)
Entrenchment	143	1.36	1.00	?	86	1.15	1.00	(0.09)*	(0.04)**
Hi-Tech	143	0.22	0.00	?	86	0.26	0.00	(0.29)	(0.29)
ROA	143	0.05	0.05	?	86	0.06	0.04	(0.07)*	(0.29)

Table 2 (continued)

Variable	**(a)**			**Exp. Relation**	**(b)**			**(a) – (b)**	
	Disclosure				**Withholding**			***P*-values**	
	#Obs.	**Mean**	**Median**		**#Obs.**	**Mean**	**Median**	***t*-test**	***W-test***
MV	143	48,244	12,852	?	86	84,735	29,541	(0.01)***	(<.01)***

[a] The table presents characteristics of firms in three subsamples: disclosing, withholding, and immaterial. “Disclosing” are cases in which the firm disclosed the cyber-attack before an outsider discovered it. “Withholding” are cases in which the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack. “Immaterial” are cases in which an outsider discovered the attack, but the firm communicated that the attack caused no material damage

[b] Variables: *Severity* is an index that takes the values 0-10 based on the severity of the cyber-attack. *Ret(−1,3)* is cumulative risk-adjusted returns from one trading day prior to the discovery date until three days after the discovery date in percent. *Damage* is the dollar damage disclosed by the attacked firm divided by the market value of equity. *Analysts* is the number of analysts covering the firm. *SOX404* is the number of material weakness in the preceding five years, where a larger value indicates weaker corporate-governance quality. *Entrenchment* is an index of managers’ entrenchment, where a larger entrenchment value indicates weaker corporate-governance quality. *Hi-Tech* is an indicator variable that is equal to 1 for firms in the high-tech industry. *ROA* is net income divided by total assets in the year before the attack. *MV* is the market value of equity in millions at the beginning of the year

[c] The sample includes 276 cyber-attacks between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the *p*-values reported in the parentheses

Firms with stronger corporate governance are less likely to withhold negative news from their investors, because stronger governance is associated with stronger fiduciary responsibility. We use the number of material weaknesses the firms reported under Section 404 of the Sarbanes-Oxley Act of 2002 in the five years preceding the breach (*SOX404*) as a measure of governance strength. Section 404 requires all publicly traded companies to establish internal controls and procedures for financial reporting, and its purpose is to reduce the possibility of corporate fraud. Reports of material weaknesses occur when deficiencies in controls create a reasonable possibility of misstatements in the firm's financial statements (Ge and McVay 2005). Although a material-weakness report does not mean a material misstatement has occurred, it means internal controls may not be strong enough to detect or prevent a material misstatement on a timely basis. However, the existence of material weaknesses in controls increases the likelihood that firms withhold information on losses associated with cyber-attacks. Data on material weaknesses are available on the AuditAnalytics database.

As Panel A of Table 2 shows, withholding firms had more material weaknesses than disclosing firms in the years prior to the withholding. The average of *SOX404* is 0.66 and 0.08 for withholding and disclosing firms, respectively, and the difference is statistically significant at the 0.01 level. We do not find any difference between the average *SOX404* of disclosing firms and that of firms that experienced immaterial cyber-attacks (Panel B of Table 2).

We also use Bebchuk et al.'s (2009) entrenchment index as a governance metric. Larger index values suggest weaker corporate governance. As Panel A of Table 2 shows, disclosing firms have lower entrenchment-index values (average 1.36) than withholding firms (average 1.77), and the difference is significant at the 0.02 level. According to Panel B, the entrenchment index of disclosing firms is higher than that of firms with immaterial damage (average 1.36 vs. 1.15, respectively, significant at the 0.09 level). According to Bebchuk et al.'s (2009) entrenchment index, firms that disclose information on cyber-attacks have stronger corporate governance than firms that withhold information on cyber-attacks.

We expect that firms with higher litigation risk will disclose information on cyber-attacks (Skinner 1994). Similar to Kasznik and Lev (1995), we use membership in high-tech industries as a proxy for high litigation risk. We use a high-tech indicator that equals 1 for firms in drugs (SIC codes 2833–2836), R&D services (8731–8734), programming (7371–7379), computers (3570–3577), and electronics (3600–3674) and 0 otherwise. As Panel A of Table 2 shows, 22% of disclosing firms are in these high-tech industries, whereas only 9% of withholding firms are in the high-tech sectors (difference is statistically significant at the 0.02 level). As Panel B shows, the percentage of firms with immaterial cyber-attacks that are in the high-tech sectors (26%) is similar to the percentage of high-tech firms in the disclosing subsample. In the context of our research, high-tech firms may also have greater technical capability to quickly discover and remedy cyber-attacks. Thus technical capability, not litigation risk, could be the reason they are more likely to disclose.

Table 2 also presents the profitability (*ROA*) in the year before the attack, measured as net income, divided by total assets. Firms may time the disclosure of negative information based on their overall profitability. For instance, firms withhold the negative news in good years, and clean the slate and disclose the negative information in periods with weaker profitability (Levitt 1998). As the table shows, disclosing and withholding firms report similar *ROA*s, and the differences between the groups are not statistically significant. Therefore differences in profitability do not explain the decision of firms to disclose or withhold information on cyber-attacks.

Finally, we examine whether the three subsamples differ from each other in terms of firm size, measured as the market value of equity at the beginning of the year (*MV*). In line with the findings on the number of analysts, we find disclosing firms have larger market values than withholding firms; larger firms are often followed by more analysts. In addition, firms with immaterial cyber-attacks have larger market values than disclosing firms, as discussed above.

4.2 Multivariate analysis

Table 3 provides the results for testing whether firms with more severe cyber-attacks are more likely to withhold than disclose the attack. We use a multivariate logistic regression of the following form:

$$Disclosing_{it} = a + b_1 Severity\ of\ Attack_{it} + controls + \varepsilon_{it} \qquad (2)$$

The dependent variable—disclosing—equals 1 for disclosing cases and 0 for withholding cases. We estimate the model with the three alternative proxies for attack severity defined above (*severity*, *Ret(−1,3)*, and *damage*). Control variables are the firm characteristics described in Table 2 and year fixed effects.

As Panel A of Table 3 shows, and consistent with our hypothesis, the severity of the attacks that are withheld is larger than the severity of those that are immediately disclosed. This result holds for the three severity measures.[24] Specifically, the coefficient on *severity* (model 1) is −0.055 (*p*-value = 0.09), and the coefficient on *damage* (model 3) is −0.805 (p-value = 0.07). Disclosing is also associated with lower stock price decreases upon discovery of the attack—the coefficient on *Ret(−1,3)* is 0.088 (p-value = 0.03). The larger price decreases associated with withholding reflect both the greater severity of withheld attacks and the negative reputation caused by withholding—we analyze this issue further below.

An examination of the control variables reveals that poor corporate-governance metrics—higher *SOX404* and *entrenchment*—are associated with less disclosing and more withholding of information on cyber-attacks, with significance levels of 0.02 and 0.12, respectively, in Model 1. This result is consistent with the claim that stronger corporate governance leads to more timely disclosure of negative information. Finally, membership in high-tech industries, which serves as a proxy for higher litigation risk, is positively associated with disclosure (p-value = 0.03 in Model 1).

In Panel B of Table 3, we compare disclosing cases with cases with immaterial cyber-attacks. As expected, we find the severity index is higher for disclosing than for immaterial cases (at the 0.01 level, Model 1). However, *damage* and *Ret(−1,3)* of disclosing cases are not significantly different from those of immaterial cases (Models 2

[24] Data on *severity* and *Ret(−1,3)* are available for the entire sample, whereas the *damage* variable is available only for a small subsample of firms. Note that we get similar results when we perform the analysis with the same subsample for *damage*, *severity*, and *Ret(−1,3)*.

Table 3 Multivariate analysis

	A: Disclosing vs. Withholding				B: Disclosing vs. Immaterial			
	Sign	Model 1	Model 2	Model 3	Sign	Model 1	Model 2	Model 3
Severity	–	−0.055			+	0.243		
		(0.09)*				(<.01)***		
Ret(−1,3)	+		0.088		–		−0.036	
			(0.03)**				(0.15)	
Damage	–			−0.805	+			9.579
				(0.07)*				(0.36)
Analysts	+	0.005	0.015	−0.046	?	0.012	0.003	0.137
		(0.32)	(0.22)	(0.25)		(0.23)	(0.42)	(0.14)
SOX404	–	−0.221	−0.358	4.556	?	−0.185	−0.188	1.299
		(0.02)**	(0.03)**	(0.48)		(0.25)	(0.24)	(0.49)
Entrenchment	–	−0.109	−0.153	−0.725	?	−0.003	−0.024	−0.254
		(0.12)	(0.16)	(0.08)*		(0.49)	(0.44)	(0.35)
Hi-Tech	+	0.736	1.208	5.357	?	0.179	0.156	2.935
		(0.02)**	(0.04)**	(0.06)*		(0.34)	(0.35)	(0.48)
ROA	?	−0.561	−1.397	−33.81	?	1.326	1.444	24.534
		(0.37)	(0.32)	(0.04)**		(0.31)	(0.29)	(0.31)
LogMV	+	0.024	0.013	0.355	?	−0.409	−0.366	−1.402
		(0.37)	(0.46)	(0.25)		(<.01)***	(<.01)***	(0.24)
# Observations		190	190	32		229	229	25
# Disclosure		143	143	19		143	143	19
R^2 (LRI)		9.81%	10.47%	47.64%		14.26%	9.78%	56.03%

[a] The table presents results for testing the hypothesis that firms withhold more severe cyber-attacks. We estimate a logistic regression:

$Disclosing_{it} = a + b_1 Severity\ of\ Attack_{it} + controls + \varepsilon_{it}$ (2)

In Panel A, the dependent variable equals 1 for disclosing and 0 for withholding. In Panel B, the dependent variable equals 1 for disclosing and 0 for immaterial. “Disclosing” are cases in which the firm disclosed the cyber-attack before an outsider discovered it. “Withholding” are cases in which the firm did not disclose the cyber-attack for at least two days after it had learned of its occurrence and a party outside the firm consequently discovered the attack. “Immaterial” are cases in which an outsider discovered the attack, but the firm communicated that the attack caused no material damage

[b] We estimate the model with three alternative proxies for the severity of the attack. *Severity* is an index between 0 and 10 depending on the severity of the cyber-attack. *Ret(−1,3)* is cumulative risk-adjusted returns from one trading day prior to the discovery date until three days after the discovery date. *Damage* is the dollar damage disclosed by the attacked firm as a percentage of the market value of equity

[c] We include control variables, which are defined in Table 2. We estimate each model with year fixed effects

[d] The sample includes 276 cyber-attacks between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. P-values are reported in the parentheses

and 3). This last finding suggests the overall damage caused by disclosed attacks is not materially higher from the severity of attacks classified as immaterial attacks and thus supports the claim that the firms disclose, on average, smaller cyber-attacks. We also find that larger firms are more likely to experience an immaterial cyber-attack (at the 0.01 level).

Overall, the results in Table support our hypothesis that the severity of withheld cyber-attacks is larger than the severity of disclosed cyber-attacks. Additionally, stronger governance and higher litigation risk are associated with more disclosure and less withholding of information on cyber-attacks.

4.3 Market reaction to cyber-attacks

In efficient markets, the return reaction to a cyber-attack should reflect the damage to the firm from the attack but, in cases of withholding, also the negative reputation associated with withholding information; investors may conclude, from firms withholding information on a cyber-attack, that management is not completely forthcoming about other potential problems. We first perform a univariate analysis of the return reaction (Table 4), and then in a multivariate analysis, we disentangle the direct effect of the attack and the reputation effect.

Table 4 Market reaction to cyber-attack disclosures

Panel A: Market reaction to withholding and disclosure			
Cyber-Attack Disclosures	**#Obs**	***Ret (−1,3)***	***Ret (−1,30)***
Disclosing	143	−0.33%	−0.72%
		(0.35)	(0.37)
Withholding	47	−1.47%	−3.56%
		(<.01)***	(0.01)***
Immaterial	86	0.27%	0.95%
		(0.57)	(0.29)
Panel B: Implied withholding probability			
	Ret (−1,3)		***Ret (−1,30)***
prob(withholding)	45%		40%
p-value	(<.01)***		(<.01)***

[a] Panel A presents the stock market reaction to withholding and disclosing information on cyber-attacks. “Withholding” are cases in which the firm did not disclose the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack. “Disclosing” are cases in which the firm disclosed the cyber-attack before an outsider discovered it. Finally, we present market reaction for cases without material damage (“immaterial”); in these cases, an outsider discovered the attack, but the firm communicated that the attack caused no material damage

[b] We present cumulative risk-adjusted returns from one trading day prior to the discovery date until three and 30 days after the discovery date, labeled *Ret(−1,3)* and *Ret(−1,30)*, respectively. Returns are risk-adjusted using value-weighted market returns

[c] Panel B presents the implied probability of withholding, *prob(withholding)*, the cyber-attack from investors: $prob(withholding) = \frac{2*Return\ reaction\ to\ Disclosing}{Return\ reaction\ to\ Withholding}$

[d] The sample includes 276 cyber-attacks between 2010 and 2015

[e] *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the *p*-values reported in the parentheses

Table 4 presents the cumulative risk-adjusted returns surrounding the date investors learned about the cyber-attack. In the main tests, we adjust the returns for risk using the value-weighted market return reported by CRSP (variable *VWRETD*). We present return reaction for two return windows: (i) a short window from one trading day prior to the disclosure until three trading days after the disclosure, denoted as *Ret(−1,3)*, and (ii) a long window from one trading day prior to the disclosure until 30 trading days after the disclosure, denoted *Ret(−1,30)*.

Focusing on the short window, we find the average market reaction to disclosing is −0.33% but not statistically different from zero. This result suggests data breaches that firms disclosed did not have a significant marginal effect on the stock value. These findings coincide with negative and insignificant stock returns that other studies find around data breaches (e.g., Campbell et al. 2003; Kannan et al. 2007). By contrast, we find the average market reaction to the 47 cases in which firms withheld information on cyber-attacks is −1.47% (significant at the 0.01 level); that is, stock value decreased 1.47% from one day prior to disclosure until three days after investors independently discovered the breach. In addition, in cases of immaterial cyber-attacks (86 cases), we find the average market reaction is 0.27% but not statistically different from zero.

Within 30 days of the discovery date, stock prices continued to decline for withholding firms. Specifically, returns 30 trading days after discovery were, on average, −3.56% (significant at the 0.01 level). This result suggests investors take a few days to understand the firm withheld material negative information and to fully respond to the information.

These results support the hypothesis that news on withheld cyber-attacks is more negative than news on disclosed cyber-attacks. Consistent with prior studies, cyber-attacks, in general, have a low negative effect on the market value of equity; however, we find that, in cases where firms withheld information and investors eventually revealed the breach, the market reaction was negative and significant. The findings are consistent with our hypothesis that firms withhold negative information below a certain threshold, disclose information on less severe cyber-attacks, and keep from investors more severe cyber-attacks that may significantly affect stock prices.

Figure 1 presents the cumulative risk-adjusted returns from one trading day prior to the discovery date until 60 days after the discovery date for withholding, disclosing, and immaterial cyber-attack cases. The results show a clear pattern: the stock price decrease in the withholding cases is larger than in cases in which firms immediately disclose the breach to investors. The negative reaction to withholding information is not temporary; it persists long after the discovery of the cyber-attack. In comparison, cumulative returns of the disclosing portfolio stay insignificantly different from zero over the 60 trading days. Returns in longer windows may be driven not only by the cyber-attack but also by other events, and the power of the test will therefore be lower, especially in small samples like ours.

If firms announce earnings during the 30 days after the discovery of the cyber-attack, the earnings announcements and not the cyber-attack could affect market reaction. We therefore exclude from the return calculation the three days around the announcement, from a day before to a day after the quarterly earnings announcement.[25] We find similar

[25] Less than 8% of the attack-discovery dates exactly coincide with the earnings announcements, and when excluding these observations, we get similar results.

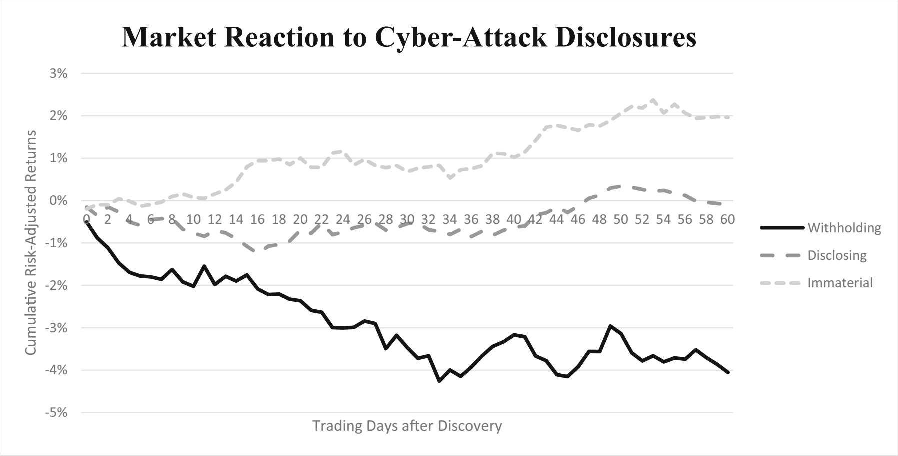

Fig. 1 The figure presents the stock market reaction to withholding and disclosing information on cyber-attacks. "Withholding" are cases in which the firm did not disclose the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack. "Disclosing" are cases in which the firm disclosed the cyber-attack no later than its discovery by investors. We also present market reaction for cases with immaterial damage ("immaterial"); in these cases, an outsider discovered the attack, but the firm declared that the attack caused no material damage. The cumulative risk-adjusted returns from one trading day prior to the discovery date until 60 days after the discovery date is calculated in each case, and the figure presents the mean returns for stocks in each of the three portfolios. The sample includes 276 cyber-attacks between 2010 and 2015

results to those presented in Table 4. We also get similar results when we exclude in the same manner seasoned equity offerings and dividend distributions, for which we get data from CRSP.

4.4 Implied probability of withholding

We estimate the implied probability of cyber-attack withholding using Eq. (1) and the market reaction to disclosing and withholding information on cyber-attacks. Panel B of Table 4 shows the results. Based on the return reaction in the three days after the discovery date, *Ret(−1,3)*, we estimate the probability of withholding to be about 45%, which is twice the return reaction of −0.33% in disclosing cases, divided by the return reaction of −1.47% in withholding cases. If these return reactions indeed capture the damage caused by the cyber-attack, managers will disclose the cyber-attack only when investors believe the probability that managers hold negative information is higher than 45%.

When we use a long return window after the discovery date, we get similar estimates of the probability of withholding. As Panel B shows, cost estimates based on the 30-day return window suggest the probability of withholding is 40%. That is, only when the chance that investors already know of the cyber-attack is at least 40% do firms choose to disclose the information.

As discussed in section 2 above, our measure of the implied probability of withholding assumes the damage is uniformly distributed and therefore the disclosure threshold in disclosure cases is estimated to be twice the average returns. The distribution of loss may be different, but in any case, the disclosure threshold will not be higher than the actual returns in the disclosing cases. Therefore, at the minimum, the implied probability of withholding is 20%, according to the 30-day window. Our estimate of the loss in withholding cases may be also biased. We assume the average returns in the withholding cases that are discovered represent the damage. However, empirically, the decrease in price upon discovery may be larger than the damage, because of negative reputation effects and litigation risk that can be associated with withholding, in which case, our withholding-probability estimates are downward biased. Another assumption we make is that the return reaction to the cyber-attack starts on the discovery date. To validate this assumption, we check and find that the cumulative risk-adjusted returns between day −10 and day −2 before the discovery date is 1.00%. If investors had started suspecting managers were withholding negative information, prices would have declined before the discovery date.

To estimate the statistical significance of the withholding-probability estimates and specifically that the withholding probability triggering disclosure is higher than zero, we assume it is a proportion that is distributed between 0 and 1 for a sample 143 observations. We find the withholding-probability estimates are greater than zero at least at the 0.01 level. Results are similar when we use bootstrapping (e.g., Chernick 2007) and use 100 random samples with replacement from the original sample of return reactions to estimate the standard deviation.

4.5 Effect of withholding on returns

Next, we examine whether the results in Table 4 reflect the market reaction to the withholding decision after controlling for the damage caused by the cyber-attack. Market reaction is driven by direct damage the cyber-attack causes, but may also reflect the negative reputation associated with withholding. We use a multivariate regression to control for the direct damage caused by the attack and test the additional reputation effects of withholding.

We use the direct-damage estimates disclosed by the attacked firms. In our sample, we find that 38 firms reported the dollar value of the damage caused by the cyber-attack in a press release or in subsequent financial statements. To control for the effect of the damage on the return reaction to the cyber-attacks, we use the following OLS regression:

$$Ret(-1,3)_{it} = \beta_0 + \beta_1 Disclosing_{it} + \beta_2 Withholding_{it} + \beta_3 Damage_{it} + \varepsilon_{it} \quad (3)$$

Ret(−1,3) is the cumulative risk-adjusted returns from one day prior to discovery until three days after discovery, *disclosing* is an indicator variable that equals 1 for disclosing cases, *withholding* is an indicator variable that equals 1 for withholding cases, and *damage* is the damage estimate provided by the firm, divided by the market value at the beginning of year. A negative slope coefficient on *withholding* will suggest negative reputation is associated with withholding information from investors.

Table 5 presents the results from estimating Eq. (3). We find that, after controlling for damage, the return reaction to withholding is negative and significant upon the discovery of the attack. In model 2, the coefficient on *withholding* is −2.325 (significant at the 0.08 level), suggesting withholding is associated with a decrease of 2.325% beyond the direct damage caused by the cyber-attack. The coefficient on *damage* is

Table 5 Market reaction after controlling for damage reported by firms

Independent Variables	Sign	Ret(−1,3)		Ret(−1,30)	
		(1)	(2)	(3)	(4)
Intercept		0.819	0.830	−0.339	−0.308
		(0.25)	(0.25)	(0.42)	(0.43)
Disclosing	–	−1.022	−0.796	−1.325	−0.616
		(0.23)	(0.28)	(0.26)	(0.38)
Withholding	–	−3.001	−2.325	−4.909	−2.791
		(0.04)**	(0.08)*	(0.05)**	(0.16)
Damage	–		−0.378		−1.186
			(0.01)***		(<.01)***
# Observations		38	38	38	38
R^2		9.21%	15.30%	8.28%	26.59%

[a] The table presents estimation results of Eq. (3):

$Ret(-1,3)_{it} = \beta_0 + \beta_1 Disclosing_{it} + \beta_2 Withholding_{it} + \beta_3 Damage_{it} + \varepsilon_{it}$ (3)

[b] Dependent variable: *Ret(−1,3)* is the cumulative risk-adjusted returns from one day before to three trading days after the date the market learned of the attack. We also estimate the regression using as a dependent variable the cumulative risk-adjusted returns from one trading day prior to the discovery date until 30 trading days after the discovery date, denoted *Ret(−1,30)*

[c] Independent variables: *Disclosing* is an indicator variable that equals 1 when the firm disclosed the cyber-attack before an outsider discovered it. *Withholding* is an indicator variable that equals 1 when the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence, and a party outside the firm consequently discovered the attack. *Damage* is the dollar damage disclosed by the attacked firm, divided by the market value of equity

[d] The sample includes 38 firms that eventually disclosed a damage estimate. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the *p*-values reported in parentheses, and errors are clustered by firm

−0.378 (significant at the 0.01 level), suggesting that, within three trading days, investors do not fully react to the direct damage caused by the attack. When extending the return window to 30 days after the discovery (model 4), the coefficient on *damage* is closer to −1 (−1.186, with *p*-value <.01). A coefficient of around −1 on *damage* suggests investors consider the damage estimate as accurate. The results suggest equity values of withholding firms decreased beyond the damage estimates provided by their managers after the attack. In addition, a decrease in value may relate to the decision to withhold information. If investors eventually learn of the cyber-attack from other sources, they are likely to update their beliefs on the integrity and quality of management. Whether managers withheld information or just failed to monitor their information systems and identify the attack, investors will take the lack of timely disclosure as a negative signal. Furthermore, firms that withhold bad news may face litigation once it is discovered (Skinner 1994, 1997; Kasznik and Lev 1995), which will also negatively affect equity value.

Because the regression results suggest that decrease in equity values upon discovery of withholding is partly driven by the negative reputation effects associated with withholding, the withholding-probability estimates based on returns will be downward biased. For example, when we use the dollar-damage figures that are reported by firms (see Table 2), instead of the market reaction to calculate the probability of withholding, we get that, following Eq. (1), the probability of withholding is 69% (2*0.624/1.812).

Next, we examine the sensitivity of our results to the inclusion of the firm-characteristic variables introduced in Table 2. The purpose of this analysis is to alleviate concerns that firm characteristics make certain firms more vulnerable to cyber-attacks and hence more likely to be included in the sample. We use the following regression model:

$$\begin{aligned} Ret(-1,3)_{it} &= \delta_0 + \delta_1 Disclosing_{it} + \delta_2 Withholding_{it} + \delta_3 Severity_{it} + \delta_4 Analysts_{it} \\ &+ \delta_5 HiTech_{it} + \delta_6 SOX404_{it} + \delta_7 Entrenchment_{it} + \delta_8 ROA_{it} + \delta_9 LogMV_{it} + \varepsilon_{it} \end{aligned} \tag{4}$$

The results in Table 6 show a negative coefficient on withholding (−1.695, p-value less than 0.01), suggesting returns decreased by 1.695%. The coefficient on disclosing is also negative (−0.503) but not statistically significant from zero, indicating the returns in the disclosing cases are not different than in the immaterial-damage cases. In addition, the coefficient on withholding is lower than the coefficient on disclosing at the 0.01 level. The probability of withholding based on these regression estimates is 45% for the *Ret(−1,3)* window, which is comparable to the univariate-based results presented in Table 4.

The coefficient on *severity* is also negative, as expected, and statistically significant at the 0.09 level. *Severity* proxies for the damage, and the low significance of the coefficient relative to the coefficient on the actual-damage variable used in Table 5 may be attributed to the noisy nature of this proxy.

We also compute the probability of withholding for different attack severities using the results in Table 6. According to the regression in Table 6, disclosed cyber-attacks decrease share price by 0.503%, on average, whereas withheld attacks decrease share prices by 1.693%. Moreover, an increase of one unit in *severity* decreases prices by an additional 0.118% (for the severity scale that goes from 1 to 10). These results suggest, for example, that the return reaction to the most severe attacks, *severity* = 10, that are disclosed is −1.683% (= − 0.503-10*0.118) and, for those withheld, −3.376% (= − 0.503-1.693-10*0.118).

Based on these estimates, the probability of withholding, $prob(withholding) = \frac{2{*}Return\ reaction\ to\ Disclosing}{Return\ reaction\ to\ Withholding}$, is 99.7% for the most severe attacks (*severity* = 10) and 53.7% for the least severe (*severity* = 1).

4.6 Sensitivity analyses

First, we control for the effect of self-selection. The decision to disclose or withhold may be driven by unobservable factors unrelated to the severity of the attack, thus inserting bias into the results reported above, because the withholding/disclosing decision is not an exogenous variable. To deal with this problem, we use an instrumental variable approach.

We use the state of incorporation as an instrumental variable. Some US states require firms to disclose attacks, regardless of their severity. For example, California firms must notify customers or individuals whose private information was breached, and if the number of individuals affected is greater than 500, the company must also notify the Attorney General of California (Cal. Civ. Code §

Table 6 Market reaction after controlling for firm characteristics

Independent Variables	Sign	Ret(−1,3)		Ret(−1,30)	
		(1)	(2)	(3)	(4)
Disclosing	–	−0.598	−0.503	−1.852	−1.082
		(0.15)	(0.20)	(0.07)*	(0.19)
Withholding	–	−1.832	−1.695	−4.862	−3.436
		(<.01)***	(<.01)***	(<.01)***	(<.01)***
Severity	–		−0.118		−0.280
			(0.09)*		(0.08)*
Analysts	?		−0.064		−0.085
			(0.01)***		(0.10)*
Hi-Tech	?		0.385		1.200
			(0.27)		(0.23)
SOX404	?		−0.049		−0.346
			(0.30)		(0.25)
Entrenchment	?		−0.297		−0.455
			(0.13)		(0.21)
ROA	?		4.181		2.949
			(0.19)		(0.40)
LogMV	?		0.094		0.677
			(0.29)		(0.04)**
# Observations		276	276	276	276
R^2		3.19%	6.32%	4.24%	6.76%

[a] The table presents the market reaction to withholding of information on cyber-attacks, after controlling for firm characteristics and with year fixed effects:

$Ret(-1,3)_{it} = \delta_0 + \delta_1 Disclosing_{it} + \delta_2 Withholding_{it} + Controls + \varepsilon_{it}$ (4)

[b] Dependent variable: *Ret(−1,3)* is the cumulative risk-adjusted returns from one day before to three trading days after the date the market learned of the attack. We also estimate the regression using as a dependent variable the cumulative risk-adjusted returns from one trading day prior to the discovery date until 30 trading days after the discovery date, denoted *Ret(−1,30)*

[c] Independent variables: *Disclosing* is an indicator variable that equals 1 when the firm disclosed the cyber-attack before an outsider discovered it. *Withholding* is an indicator variable that equals 1 when the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence, and a party outside the firm consequently discovered the attack. For the definitions of the remaining variables, see Table 2

[d] The sample includes 276 observations between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the p-values reported in parentheses, and errors are clustered by firm

1798.82). Therefore incorporation in California is expected to affect the decision to disclose but not the damage; that is, incorporation in California will not bring about more severe attacks on firms. This fact allows us to use the state of incorporation as a valid instrument in our analysis.

We identify 25 states, plus the territory of Puerto Rico, that require firms to notify the state attorney general of certain breaches as high-disclosure states. These are: CA, CT, FA, HI, IN, IA, LA, ME, MD, MA, MO, MT, NE, NH, NJ, NY, NC, ND, OK, OR, RI, SC, VT, VA, WA, and PR.[26] HDState is an indicator variable equal to 1 for these states and 0 for other states of incorporation. We use the following 2SLS estimation:

$$Withholding_{it} = \alpha + \beta HDState_{it} + \varepsilon_{it} \tag{5a}$$

$$Ret(-1,3)_{it} = \alpha + \beta_1 \overline{Withholding_{it}} + \beta_2 Severity_{it} + Controls + \theta_{it} \tag{5b}$$

where the variables are similar to those in Eq. (4) above and regressions are estimated with year fixed effects. In the first stage, we estimate (5a) and use the expected value, $\overline{Withholding_{it}}$, in the second stage in estimating (5b).

Of the 51 cyber-attacks against firms incorporated in high-disclosure states, only 11.8% were withheld by the firms versus 19.0% of the attacks against firms incorporated in other states. Estimating Eq. (5a) with year fixed effects, we find the coefficient on HDState is −0.075 and lower than zero at the 0.10 level. This result suggests withholding is less frequent in high-disclosure states. Using the expected value from estimation of Eq. (5a), we estimate (5b) and present the estimation results in Table 7.

Table 7 includes two regressions—OLS and 2SLS. As the table shows, the coefficients on *severity* are negative and significant in both regressions, suggesting the severity of the attack reduces stock prices. In addition, the coefficients on all the control variables are of similar magnitude and significance levels. The main difference between the two regressions relates to the coefficient on *withholding*. Specifically, the coefficient on *withholding* is negative, −1.265 (*p*-value = 0.02) in the OLS regression, but once 2SLS is used, the coefficient on $\overline{Withholding_{it}}$ is positive, 20.22 (p-value = 0.03). The results in the OLS regression suggest withholding information produces negative reputation effects. However, once the self-selection is removed, we find withholding by itself does not have a negative effect on stock returns.

We also perform the analysis using the longer return window (−1,30) and find that the coefficient on the withholding instrument, $\overline{Withholding_{it}}$, is negative, −11.70 but not statistically significant (*p*-value of 0.34). Although the sign of the coefficient on withholding instrument is different for the *Ret(−1,3)* and *Ret(−1,30)* return windows, the conclusion is similar—after controlling for endogeneity, withholding does not have a negative effect on returns.[27]

Another source of self-selection relates to the assumption that ex-post discovered attacks represent the population of withheld attacks. If not all cyber-attacks are discovered or if those undiscovered attacks differ materially from those included in our

[26] The fact individuals can access a firm's website over the Internet from other states is not sufficient to give these states jurisdiction over the firm (Rosenblatt 1999). We therefore use state of incorporation as an instrument for the disclosure level to which the firm is obligated.

[27] The large coefficients on the withholding instrument do not necessarily suggest withholding has a larger effect in the 2SLS estimation. The distributions of the withholding variable (used in the OLS regression) and that of the withholding instrument differ. The withholding variable in the OLS regression is an indicator variable with a standard deviation of 0.433, whereas the withholding instrument, $\overline{Withholding_{it}}$, is the expected value of withholding (a continuous variable) from the first stage of a 2SLS model, with a standard deviation of 0.046. One standard deviation change in the 2SLS withholding instrument does not necessarily lead to greater effects than a one standard deviation change in the OLS withholding variable. Moreover, when adding instrumental variables to the first stage of the 2SLS model (the two governance metrics, *SOX404* and *entrenchment*), we find similar results. Hence our findings are unlikely to be driven by model specification.

Table 7 Controlling for endogeneity

Independent Variables	Sign	Ret(−1,3)		Ret(−1,30)	
		OLS	2SLS	OLS	2SLS
Withholding	–	−1.265		−2.480	
		(0.02)**		(0.04)**	
$\overline{Withholding}$	?		20.22		−11.70
			(0.03)**		(0.34)
Severity	–	−0.141	−0.191	−0.297	−0.313
		(0.06)*	(0.02)**	(0.09)*	(0.08)*
Analysts	?	−0.051	−0.046	−0.116	−0.114
		(0.06)*	(0.08)*	(0.07)*	(0.07)*
Hi-Tech	?	0.206	0.593	0.837	1.118
		(0.39)	(0.22)	(0.34)	(0.29)
SOX404	?	−0.030	−0.058	−0.228	−0.421
		(0.37)	(0.29)	(0.34)	(0.21)
Entrenchment	?	−0.301	−0.338	−1.233	−1.333
		(0.13)	(0.11)	(0.03)**	(0.02)**
ROA	?	4.276	4.686	−0.780	−1.897
		(0.19)	(0.18)	(0.48)	(0.45)
LogMV	?	−0.039	−0.003	0.595	0.587
		(0.42)	(0.49)	(0.10)*	(0.10)*
# Observations		190	190	190	190
R^2		8.38%	8.37%	9.30%	8.23%

[a] To control for endogeneity, the fact that severity of the attack can affect firms' decision to disclose or withhold, we use 2SLS estimation with year fixed effects:

$Withholding_{it} = \alpha + \beta HDState_{it} + \varepsilon_{it}$ (5a)

$Ret(-1,3)_{it} = \alpha + \beta_1 \overline{Withholding_{it}} + \beta_2 Severity_{it} + Controls + \theta_{it}$ (5b)

HDState is an indicator variable that equals 1 for high-disclosure states. *Ret(−1,3)* is the cumulative risk-adjusted returns from one day before to three trading days after the date the market learned of the attack. *Ret(−1,30)* is the cumulative risk-adjusted returns from one trading day prior to the discovery date until 30 trading days after the discovery date

[b] *Withholding* is an indicator variable that equals 1 when the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack. $\overline{Withholding}$ is the expected value from Eq. (5a). See Table 2 for the definitions of the remaining variables

[c] The sample includes 190 observations of attacks classified as withholding and disclosing between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the *p*-values reported in parentheses, and errors are clustered by firm

sample, our results may be biased. To address this issue, we use a procedure similar to that of Heckman (1979).

To estimate the extent of self-selection in the sample of discovered cyber-attacks, we use data from the Verizon Data Breach Investigations Report. The report includes information on data breaches collected by more than 70 cyber-security companies and organizations. The identity of most firms on the report is unknown. The cyber-security companies collected the data during "paid external forensic investigation" services, and although they contributed data on the breaches to the report, they did not reveal the identity of most of the firms that were attacked. Figures 2 of Verizon's 2013, 2014, and 2015 Data Breach Investigations Reports provide the total number of data breaches in each industry in each year, respectively, 2013, 2014, to 2015.

To estimate the extent of self-selection in our data, we divide the number of known breaches in the industry during 2013–2015 (which are also part of our database) by the total number of breaches indicated by the reports for the industry during those three years. The result is an estimate of the probability that a cyber-attack will be revealed, conditional on industry membership. For example, about 38% of the attacks in the information industry are known (industry no. 51 according to the two-digit North American Industry Classification System), which is about eight times higher than the average rate of discovery according to the Verizon data. As discussed above, high-tech firms face greater litigation risk and may also have greater technical capability to discover and remedy breaches and therefore are expected to more frequently disclose data breaches. Based on these probabilities of discovery, which in essence are the result of the first stage of the Heckman procedure, we calculate the inverse Mills ratio (IMR), and estimate the following regression with year fixed effects:

$$Ret(-1,3)_{it} = \alpha + \beta_1 Withholding_{it} + \beta_2 Severity_{it} + Controls + IMR + \theta_{it} \quad (6)$$

As the results in Table 8 show, the estimation results of Eq. (6) are similar to those presented in the main analysis. Specifically, we find the coefficients on *withholding* and *severity* are negative and significant.

In the main tests, we adjust stock returns for risk using the value-weighted market return. We apply this relatively simple risk adjustment to all 276 data breaches in our sample to maximize the sample size. Table 9 presents the results with adjustments for size, book-to-market, and momentum quintile portfolios as in Daniel et al. (1997), applied to 215 breaches, and CRPS size-decile portfolios, applied to 247 breaches. These two alternative risk adjustments yield similar results.

Our results are similar when we control for the type of attack. Using Gordon et al.'s (2011), classification we add three indicator variables to Eq. (4)—one for each attack type (availability, integrity, and confidentiality). The main results (not tabulated) are similar to those reported in Table 6. Specifically, the coefficient on *withholding* is −1.710 (*p*-value <0.01), and the coefficient on *disclosing* is −0.483 (p-value = 0.21). Also, none of the attack-type indicators is significantly different from zero at the 0.10 level.[28] These results suggest the type of attack does not drive the effect of withholding on the market reaction.

We perform additional robustness tests to rule out alternative explanations to our results. First, to rule out the possibility that marketwide effects are driving our results, we perform the analysis using raw returns instead of market-adjusted returns. For

[28] On a univariate level, availability, confidentiality, and integrity attacks are associated with returns, *Ret(−1,3)*, of −0.77%, −0.30%, and − 0.04%, respectively. Gordon et al. (2011) similarly find that availability attacks are associated with larger damages than confidentiality attacks, and integrity attacks are associated with the lowest damages. Once we control for the damage, the attack type does not provide any additional explanatory power.

Table 8 Controlling for selection bias

Independent Variables	Sign	Ret(−1,3)		Ret(−1,30)	
		(1)	(2)	(3)	(4)
Withholding	–	−1.274	−1.268	−3.288	−2.600
		(0.01)***	(0.02)**	(0.01)***	(0.03)**
Severity	–		−0.140		−0.265
			(0.06)*		(0.08)*
Analysts	?		−0.051		−0.127
			(0.07)*		(0.06)*
HiTech	?		0.225		1.529
			(0.39)		(0.26)
SOX404	?		−0.031		−0.276
			(0.37)		(0.31)
Entrenchment	?		−0.304		−1.347
			(0.13)		(0.02)**
ROA	?		4.300		0.108
			(0.18)		(0.50)
LogMV	?		−0.040		0.577
			(0.42)		(0.11)
IMR		−0.452	0.255	5.382	9.253
		(0.44)	(0.47)	(0.24)	(0.13)
# Observations		190	190	190	190
R^2		4.93%	8.38%	5.42%	9.97%

[a] To control for selection bias (discovered attacks may be a biased sample of the withheld attacks), we use Heckman's correction. We estimate the following regression with year fixed effects:

$$Ret(-1,3)_{it} = \alpha + \beta_1 Withholding_{it} + \beta_2 Severity_{it} + Controls + IMR + \theta_{it} \quad (6)$$

Ret(−1,3) is the cumulative risk-adjusted returns from one day before to three trading days after the date the market learned of the attack. *Ret(−1,30)* is the cumulative risk-adjusted returns from one trading day prior to the discovery date until 30 trading days after the discovery date

[b] *Withholding* is an indicator variable that equals 1 when the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack. *IMR* is the inverse Mills ratio based on the probability that cyber-attacks will be discovered in the industry in which the firm operates. See Table 2 for variable definitions

[c] The sample includes 190 observations of attacks classified as withholding and disclosing between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the p-values reported in parentheses, and errors are clustered by firm

example, we may underestimate the cost of the cyber-attack due to information spillover (e.g., a discovery of an attack on one firm may have an effect on other firms), and by subtracting market returns, we underestimate the cost of the cyber-attack. Using raw returns as the dependent variable in Eq. (4) yields similar results (not tabulated for brevity). The coefficient on *withholding* is −1.708 and significant (*p*-value of 0.01); the coefficient on *disclosing* is −0.638 and statistically insignificant (p-value of 0.20). The results indicate marketwide factors do not drive the different effect of disclosing and withholding on market reaction.

Table 9 Market reaction with alternative risk adjustments

Cyber-Attack Disclosures	#Obs	*Ret (−1,3)*	*Ret (−1,30)*
Panel A: Size-adjusted returns			
Disclosing	122	−0.56%	−1.04%
		(0.06)*	(0.11)
Withholding	43	−1.32%	−2.98%
		(<.01)***	(0.01)***
Immaterial	82	0.29%	1.07%
		(0.29)	(0.12)
Panel B: Returns adjusted for size, book-to-market, and momentum			
Disclosing	106	−0.36%	−0.46%
		(0.12)	(0.28)
Withholding	36	−1.17%	−2.35%
		(<.01)***	(0.05)**
Immaterial	73	0.36%	0.84%
		(0.21)	(0.23)

[a] Panel A presents the stock market reaction with size-adjusted returns, and Panel B, with stock returns adjusted for risk using size, book-to-market, and momentum portfolios. We present cumulative risk-adjusted returns from one trading day prior to the discovery date until three and 30 days after the discovery date, labeled *Ret(−1,3)* and *Ret(−1,30)*, respectively

[b] "Withholding" are cases in which the firm had not disclosed the cyber-attack for at least two days after it learned of its occurrence and a party outside the firm consequently discovered the attack. "Disclosing" are cases in which the firm disclosed the cyber-attack before an outsider discovered it. "Immaterial" are cases in which an outsider discovered the attack, but the firm communicated that the attack caused no material damage

[c] The sample includes 276 cyber-attacks between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively, for the p-values reported in the parentheses

Second, several firms experienced multiple cyber-attacks, and investors may be reacting differently to a first cyber-attack than to a second or third. To control for multiple attacks, we add to Eq. (4) indicator variables for the first attack on firms, an indicator variable for the second attack on firms, and so on. Our results and inferences do not change. The coefficient on *withholding* is −1.684 and significant (p-value of <.01), and the coefficient on *disclosing* is −0.535 and statistically insignificant (p-value of 0.21).

Third, Gordon et al. (2011) report a decline in the market reaction to cyber-attacks over time. To examine whether change in market reaction over our sample period drives the results, we estimate Eq. (4) with a time-ordinal variable and without the year fixed effects. The coefficient on this variable is not statistically different from zero (−0.002, p-value = 0.22). Also, the results do not change: the coefficient on *withholding* is −1.587 (p-value = 0.01), and the coefficient on *disclosing* is −0.482 (p-value = 0.21). The results of this sensitivity analysis suggest change in market reaction over time does not drive the different effect of disclosing and withholding on market reaction that we report.

In addition, we find that a delay in disclosure by itself is not associated with larger negative market reaction. The average return reaction, *Ret(−1,30)*, for cases disclosed and delayed by at least two weeks (20 cases) is 0.46%, whereas for other disclosed breaches (123 cases) the return reaction is −0.90%; the difference between the two groups is not significant (p-value is 0.55). These findings suggest that delayed disclosures are not associated with larger negative market reaction, as long as the firm itself eventually disclosed the attack.

However, delays are associated with more negative market reaction when third parties discover the attacks. The average return reaction for attacks discovered by third parties more than two weeks after the firm learned of the attacks is −4.83% (30 cases), whereas the reaction is only −1.30% for attacks discovered less than two weeks after the firm (17 cases) and statistically higher at the 0.10 significance level (p-value of 0.09). Longer delays from when the firm learns of the attack and when third parties discover it suggest that the firm intentionally withheld the information. While firms may need some time to examine and react to the attack, as withholding time grows longer, the motivation for withholding information is more likely hiding information from investors.

To validate our results, we test whether the decision to withhold information on cyber-attacks is associated with managers' incentives to do so. Specifically, we examine whether managers who have a larger equity stake in the firm are more likely to withhold severe cyber-attacks from investors. Using data from ExecuComp, we calculate for each firm the value of the options and stock of the top five executives minus their salaries (stock compensation) in the year prior to their decision to withhold or disclose the cyber-attack.[29] Of the 190 withholding and disclosing cases described in Table 3 (Panel A, Model 1), we find compensation data in 133 cases. We expect to find more withholding cases when managers have higher equity stakes. When we split the sample on stock compensation each year, we find that 33% of firms with above-median stock compensation (high stock compensation firms) withheld information on cyber-attacks—23 withholding versus 46 disclosing cases—whereas only 23% of firms with below-median stock compensation (low stock compensation firms) withheld information on cyber-attacks—15 withholding versus 49 disclosing cases. Consistent with our prediction, we find that managers withhold more severe attacks. The severity of the attacks withheld by high stock compensation firms (on average, 5.50) is higher than that of disclosed attacks (on average, 4.57) at the 0.10 level. In contrast, the severity of attacks withheld by managers of low stock compensation firms was, on average, 4.31, versus average severity of 4.39 of disclosed attacks. To further test the effect of stock compensation on withholding in a multivariate setting, we estimate the following logistic regression:

$$Disclosing_{it} = \alpha + b_1 Comp_{i,t-1} + b_2 Severe_{it} + b_3 Comp_{i,t-1} \times Severe_{it} + controls + \varepsilon_{it} \qquad (7)$$

The dependent variable (*disclosing*) equals 1 for disclosing cases and 0 for withholding cases. *Comp* is an indicator variable that equals 1 for firms with above median stock Compensation for the year. *Severe* is an indicator variable that equals 1 for firms with above median severity of cyber-attack. We include control variables and year fixed-effects as in Table 2. Negative coefficient on the interaction variable $Comp_{i,\ t-1} \times Severity_{it}$, $b_3 < 0$, will support our prediction, which suggests managers with compensation incentives will withhold severe cyber-attacks. As Table 10 shows, the coefficient on the interaction variable is negative and significant at the 0.05 level, supporting the prediction.

[29] We calculate the value of managers' stocks and options based on Coles et al. (2006).

Table 10 Effect of management compensation

	Sign	Model 1	Model 2
Comp	?	0.247	−0.278
		(0.34)	(0.35)
Severe	?	0.041	0.027
		(0.47)	(0.48)
*Comp*Severe*	–	−1.229	−1.540
		(0.07)*	(0.05)**
Analysts	+		0.012
			(0.32)
SOX404	–		−1.061
			(0.04)**
Entrenchment	–		−0.146
			(0.26)
Hi-Tech	+		1.049
			(0.12)
ROA	?		0.007
			(0.50)
LogMV	+		0.368
			(0.03)**
# Observations		133	133
# Disclosure		95	95
R^2 (LRI)		5.18%	20.43%

[a] The table tests whether withholding of severe attacks is more likely when managers have higher equity stake in the firm. We estimate the following logistic regression:

$$Disclosing_{it} = a + b_1 Comp_{i,\ t-1} + b_2 Severity_{it} b_3 Comp_{i,\ t-1} \times Severity_{it} + controls + \varepsilon_{it} \quad (7)$$

where the dependent variable equals 1 for disclosing and 0 for withholding. "Disclosing" are cases in which the firm discloses the cyber-attack before an outsider discovers it. "Withholding" are cases in which the firm did not disclose the cyber-attack for at least two days after it had learned of its occurrence and a party outside the firm consequently discovered the attack. *Comp* is an indicator variable that equals 1 for firms with above median equity compensation, where equity compensation is the market value of stock and options held by the top five executives of the firm relative to their salary. *Severe* is an indicator variable that equals 1 for firms with above median severity of cyber-attack. We include control variables, which are defined in Table 2. We estimate each model with year fixed effects

[b] The sample includes 133 cyber-attacks between 2010 and 2015. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. P-values are reported in the parentheses

Finally, to demonstrate the market reaction to news announcements of withholding and disclosing firms in general does not differ, we estimate Eq. (4) for earnings-announcement days. For each firm attacked in our sample, we include the market reaction to the four quarterly earnings during the year, from one trading day before to one trading day after the announcement. We find the coefficients on the withholding and disclosing dummy variables are not different from zero, suggesting the effect of the cyber-attack and not a general earnings effect drives the different market reaction we record for disclosing and withholding firms.

5 Conclusion

Cyber-attacks are one of the main risks firms must manage. Studies raised doubts on whether cyber-attacks are indeed so harmful. In particular, studies used the market reaction to cyber-attacks to show the loss from cyber-attacks is small and decreasing.

The source of information on cyber-attacks, in most studies, is the firm itself. However, managers may have strong incentives to withhold information on cyber-attacks, especially when the occurrence of the cyber-attack and the damage caused are uncertain. Unlike prior studies, we classify cyber-attacks into two main groups: cyber-attacks the attacked firms disclosed and cyber-attacks that were withheld and later independently discovered by sources outside the firm. We show the market reaction to disclosed attacks is indeed small, but the market reaction to withheld attacks is negative and significant.

Using market reactions to cyber-attacks that were disclosed and cyber-attacks that were withheld and later discovered, we estimate the extent to which firms withhold information on cyber-attacks. We find managers disclose less severe attacks and withhold information from investors on attacks that cause greater damage. The evidence is consistent with the theory that managers will not disclose negative information below a certain threshold when investors are uncertain about whether the firm possesses negative information.

The proportion of the market reaction to withheld and disclosed cyber-attacks also implies managers disclose cyber-attacks only when investors already suspect that, with a 40% chance, an attack has occurred. When the likelihood of independent discovery by external parties is lower, managers withhold the information. Overall, our analyses suggest voluntary disclosure of cyber-attacks is rare. If regulators wish to ensure information on cyber-attacks reaches investors, they should consider imposing stricter mandatory disclosure rules regarding cyber-attacks and clearer materiality thresholds.

Acknowledgments We thank Peter Easton (Editor), Eti Einhorn, Tsahi Versano, two anonymous referees, and seminar participants at the 2017 American Accounting Association annual meeting in San Diego, 2017 European Accounting Association annual meeting in Valencia, Bar Ilan University, Ben Gurion University, ESSEC, Exeter University, Hebrew University of Jerusalem, INSEAD, University of Padua, and Tel Aviv University for useful comments. We also thank the Blavatnik Interdisciplinary Cyber Research Center, the Jeremy Coller Foundation, and Henry Crown Institute of Business Research for financial support.

References

Amir, E., & Ziv, A. (1997). Recognize, disclose or delay; Timing the adoption of SFAS No. 106. *Journal of Accounting Research, 35*(Spring), 61–81.

Baginski, S. P., Campbell, J. L., Hinson, L. A., & Koo, D. S. (2018). Do career concerns affect the delay of bad news disclosure? *The Accounting Review, 93*(2), 61–95.

Bebchuk, L., Cohen, A., & Ferrell, A. (2009). What matters in corporate governance? *Review of Financial Studies, 22*(2), 783–827.

Campbell, K., Gordon, L., Loeb, M., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security, 11*, 431–448.

Cavusoglu, H., Mishra, B., & 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*, 69–104.

Chambers, A., & Penman, S. (1984). Timeliness of reporting and the stock price reaction to earnings announcements. *Journal of Accounting Research, 22*(1), 21–47.

Chernick, M. (2007). *Bootstrap methods: A guide for practitioners and researchers* (2nd ed.). New York: Wiley.

Coles, J. L., Daniel, N. D., & Naveen, L. (2006). Managerial incentives and risk-taking. *Journal of Financial Economics, 79*, 431–468.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance, 52*(3), 1035–1058.

Dye, R. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research, 23*(1), 123–145.

Ettredge, M., & Richardson, V. (2003). Information transfer among internet firms: The case of acker attacks. *Journal of Information Systems, 17*, 71–82.

Fama, E., & French, K. (1996). The CAPM is wanted, dead or alive. *Journal of Finance, 51*(5), 1947–1958.

Ge, W., & McVay, S. (2005). The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons, 19*(3), 137–158.

Gordon, L. A., Loeb, M. P., & Sohail, T. (2010). Market value of voluntary disclosures concerning information security. *MIS Quarterly, 34*, 567–594.

Gordon, L., Loeb, M., & Zhou, L. (2011). The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security, 19*, 33–56.

Grossman, S. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics, 24(3)*, 461–483.

Grossman, S., & Hart, O. (1980). Disclosure laws and takeover bids. *Journal of Finance, 35(2)*, 323–334.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153–161.

Hilary, G., Segal, B., & Zhang, M. (2016). Cyber-risk disclosure: Who cares? *Georgetown McDonough School of Business Research Paper No. 2852519*, p. 59.

Hovav, A., & D'Arcy, J. (2003). The impact of denial-of-service attack announcements on the market value of firms. *Risk Management and Insurance Review, 6*, 97–121.

Jung, W., & Kwon, Y. (1988). Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research, 26*(1), 146–153.

Kannan, A., Rees, J., & Shridhar, S. (2007). Market reactions to information security breach announcements: An empirical analysis. *International Journal of Electronic Commerce, 12*, 69–91.

Kasznik, R., & Lev, B. (1995). To warn or not to warn: Management disclosures in the face of an earnings surprise. *Accounting Review, 70*(1), 113–134.

Kothari, S. P., Shu, S., & Wysocki, P. (2009). Do managers withhold bad news? *Journal of Accounting Research, 47*(1), 241–276.

Kvochko, E., & Pant, R. (2015). Why data breaches don't hurt stock prices. *Harvard Business Review, March, 31*, 2015.

Levitt, A. (1998). The numbers game. *The CPA Journal, 68*(12), 14–19.

Rosenblatt, B. (1999). Principles of jurisdiction. Harvard University, Berkman Klein Center for Internet & Society. Retrieved from https://cyber.harvard.edu.

Securities and Exchange Commission (2011). Division of corporation finance, CF disclosure guidance, Topic no. 2 – Cybersecurity, October 13, 2011. Securities and Exchange Commission. Retrieved from http://www.sec.gov.

Securities and Exchange Commission (2018). Commission statement and guidance on public company cybersecurity disclosures, February 26, 2018. Securities and Exchange Commission. Retrieved from http://www.sec.gov.

Skinner, D. (1994). Why firms voluntarily disclose bad news? *Journal of Accounting Research, 32*(1), 38–60.

Skinner, D. (1997). Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics, 23*, 249–282.

Southwell, A., Vandevelde, E., Bergsieker, R., & Bisnar-Maute, J. (2017). Gibson Dunn Reviews U.S. Cybersecurity and Data Privacy, February 3, 2017. *The* CLS Blue Sky Blog, Columbia Law School. Retrieved from http://clsbluesky.law.columbia.edu.

Spanos, G., & Angelis, L. (2016). The impact of information security events on the stock market: A systematic literature review. *Computers & Security, 58*, 216–229.

Verizon Enterprise Solutions (2015). Verizon 2015 Data Breach Investigations Report. Verizon Enterprise Solutions. Retrieved from http://www.verizonenterprise.com.

White, M. J. (2014). Opening Statement at SEC Roundtable on Cybersecurity, March 26, 2014. Securities and Exchange Commission. Retrieved from http://www.sec.gov.