Cyentia INSTITUTE

Information Risk Insights Study

A Clearer Vision for Assessing the Risk of Cyber Incidents

IRIS 2022

WITH SPONSORSHIP FROM
THE CYBERSECURITY & INFASTRUCTURE SECURITY AGENCY

Introduction

"Risk comes from not knowing what you're doing."

~Warren Buffett

The Information Risk Insights Study (IRIS) is a research series dedicated to clearing away the fog of FUD (fear, uncertainty, and doubt) that has clouded cyber risk decisions for far too long. Each report leverages real-world data and rigorous analysis focused on key aspects and challenges of managing cyber risk.

We're extremely excited to be able to share the IRIS 2022 with you. There's a stigma that sequels are rarely better than the original, but we've worked hard to improve this edition in every way. We examine more data than ever, our analytical techniques have matured, and we've greatly enriched the data to yield insights that would have been impossible two years ago.

Come join us as we focus the IRIS on 77K cyber events resulting in $57 billion in total reported financial losses and 72 billion data records compromised over the last decade.

Acknowledgements

The Cyentia Institute wishes to acknowledge and thank the Cybersecurity Division and the Office of the Chief Economist at the Cybersecurity and Infrastructure Security Agency (CISA), and the Lawrence Livermore National Laboratory for sponsoring this study. It is our sincere hope that this research will lead to better cyber risk assessments and cyber risk reduction decisions for years to come.

Table of Contents

The Cyentia Institute is a research and data science firm working to advance cybersecurity knowledge and practice. We pursue this goal through our data-driven products and joint research publications like this study.

IRIS 2022 Key Findings

Cybersecurity incidents are growing in frequency. The average number of events publicly reported each month has increased 44% over the last decade.

The Healthcare and Finance sectors claim the most incidents. They have 76X more events on public record than the least-breached industries of Mining and Agriculture.

In terms of the likelihood of experiencing at least one cyber event in a single year, however, the Hospitality and Information Services sectors top the list.

Large organizations with over $100B in annual revenue are 32X more likely to have multiple security incidents in a single year than smaller firms.

The relative impact of incidents on smaller firms is much greater, however. SMBs were the victim in 89% of all cyber loss events that exceeded 10% of annual revenues.

Typical financial costs reported for a cyber event stand at $266K but the top 5% of loss events balloon to $52M. The largest in our dataset is a whopping $12B.

Despite common belief to the contrary, financial losses attributed to cyber events have NOT increased over the past 20 years.

The Information Services sector has the largest typical incident cost of $476K. The biggest extreme loss (95th percentile) belongs to the Transportation sector at $177M.

System intrusions accounted for nearly half of all events and more than half of total losses recorded over the last decade. Perhaps surprisingly, ransomware ranks #4 in frequency.

Valid Accounts, Phishing, and Exploit Public-Facing Applications are the three most common MITRE ATT&CK initial access techniques observed across all incidents.

Like what you see? *Join the vision!*

We intend to continue the IRIS in the future to discover even more insights for managing information risk. If you'd like to join in that effort by contributing relevant data or sponsoring, please reach out to us at research@cyentia.com.

119
Cy
cyentia

What is "Information Risk?"

Since these studies are read by audiences from different backgrounds with varying working definitions of "risk," we think it necessary to make sure we're all on the same page. A quick web search will find many definitions of risk, and we're not going to attempt to pick just one or tender yet another. While definitional variations abound, most agree at some level that risk involves the frequency and impact of adverse events. Thus, information (or cyber) risk deals with the occurrence and cost of events that adversely affect information systems.

Unfortunately, reliable data about the frequency and impact of cyber events[1] has historically been difficult to obtain. This lack of data presents a serious challenge for decision makers, causing many to fall back on subjective judgments and qualitative ratings. We know that struggle well, and that's why we're so excited about the IRIS series. Our extensive analysis in this latest study yields objective data on the frequency and financial impact of cybersecurity incidents on organizations of all types and sizes. We hope it helps many teams escape the qualitative quagmire of information risk assessments.

DATA USED IN THIS STUDY

This study leverages a vast dataset spanning over 77,000 cyber events experienced by 35,000 organizations over the last decade. This dataset is drawn from Advisen's Cyber Loss Data, which contains over 138,000 cyber events collected from publicly verifiable sources.

This dataset is widely used, with three features that make it ideal for this research:

1. It is the most comprehensive list of historical cyber incidents we've found.
2. It tracks losses publicly disclosed in the wake of those incidents.
3. It includes supplemental firmographic information on organizations affected by cyber events and the broader economy.

Additional information about Advisen's Cyber Loss Data and our analysis of it can be found in the Methodology and Firmographics section.

Find out more at Advisen.

Advisen
A ZYWAVE COMPANY

[1]We use the terms cyber event, loss event, and incident interchangeably in this study to generally refer to adverse events that impact the confidentiality, integrity, or availability of a firm's information assets. This terminology encompasses common events such as data breaches, ransomware infections, DDoS attacks, insider misuse, and lost or stolen laptops.

Incident Frequency Analysis

"Could it happen to us?"
~Every risk manager ever

It's undoubtedly one of the most common questions confronting risk managers of any stripe, including those of the cyber variety. Since a handwavy "maybe" isn't up to snuff for this study, we approach the question from several different angles in this section. We'll start off with a simple count of historical incidents and work our way up to developing a frequency distribution to support probabilistic statements. Sound like your jam? Great; let's do this!

Historical cyber loss events

"Nothing is more dangerous than a man who knows the past."
~Gleeman, The Wheel of Time

Past events aren't a perfect predictor of future trends, but they're certainly not irrelevant occurrences either. Figure 1 tallies publicly known[2] cyber loss events each month over the last decade. Keep in mind that incident reporting often lags behind by months (or years) as events progress from discovery to disclosure, which explains the apparent falloff toward the end of the time period.

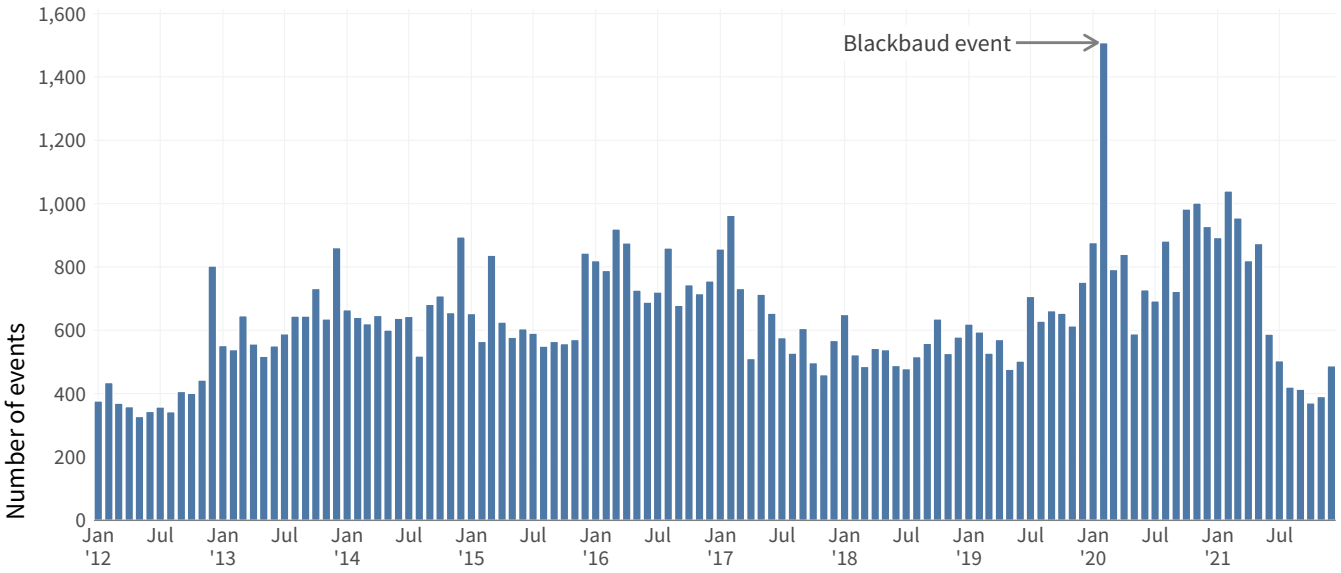


Figure 1: Number of publicly reported cyber loss events each month from 2012 to 2021

The most noticeable pattern in Figure 1 is the recurring spikes, which initially appear to hit in January of each year. Perhaps cybercriminals make aggressive New Year's resolutions but soon break them just like the rest of us. In actuality, the spikes land in December and are a result of the 12/31/YYYY date assigned to incidents that cannot be tied to a particular date except the year in which they occurred.

Another noticeable feature is the tallest spike in early 2020. That's associated with the Blackbaud ransomware breach, which caused spillover events for over 800 organizations.

"Are incidents occurring more often of late?" A common question, and the answer appears to be "Yes."

[2] The distinction of publicly known events is important because we're not claiming that Figure 1 (or this report) reflects all incidents that occurred during this timeframe. We can only analyze those that make their way into the public record, through outward signs or impacts, mandatory reporting, voluntary disclosure, etc. Advisen and Cyentia closely monitor such events and have high confidence that this dataset is representative of significant cyber events.

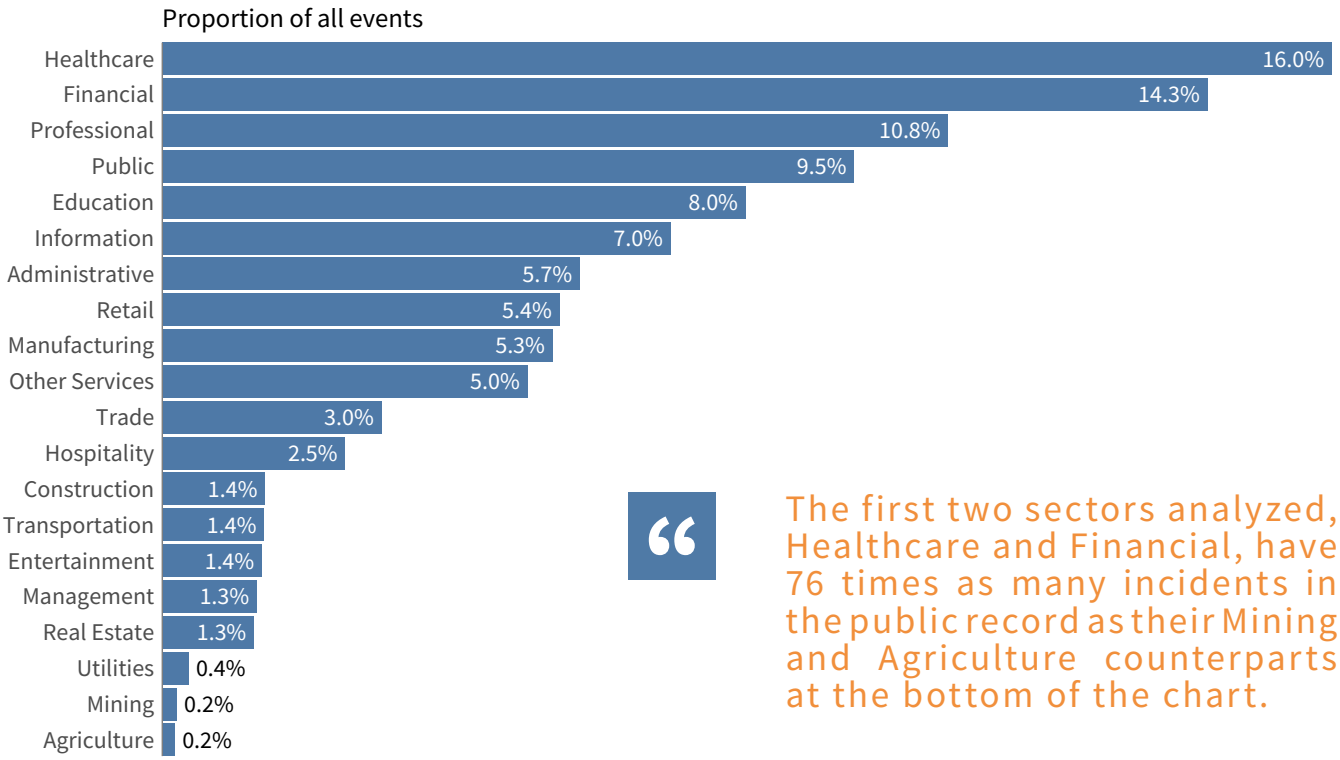
Beyond the spikes, there's also a less immediately discernible but more strategically important trend exhibited in Figure 1. There's a period of increasing event counts in the first few years, followed by a plateau and falloff and then a steady rise into 2021 (until the aforementioned reporting-related falloff).

"Are incidents occurring more often of late?" is a common question and one prompted by the data here. The answer appears to be "Yes." The geometric mean[3] (geomean) of the monthly incident count in 2012–2013 was 496 compared to 718 for 2020–2021. That's an increase of 44.7% over the last 10 years.

REPORTED INCIDENTS PER SECTOR

An overall tally and trending of security incidents provides a starting point for assessing frequency, but it doesn't help much with the "us" part of "Could it happen to us?" Organizations of all types are represented in the dataset, so the logical next step is to examine an industry-level breakdown of loss events in Figure 2.

The industries presented in this study are based on top-level sectors, as defined by the North American Industry Classification System (NAICS). Full sector labels, definitions, and subsector listings can be found at Census.gov



> The first two sectors analyzed, Healthcare and Financial, have 76 times as many incidents in the public record as their Mining and Agriculture counterparts at the bottom of the chart.

Figure 2: Proportion of publicly known incidents attributed to each sector

It probably won't surprise anyone to learn that some industries experience more security incidents than others. Nor is it unexpected that regulated sectors with strict mandatory disclosure requirements, such as Healthcare, Financial, and Public (government), would have more than their fair share. The first two sectors have 76X as many incidents in the public record as their Mining and Agriculture counterparts at the bottom of the chart.

Despite the disproportionality, take care not to draw any hasty conclusions about which industries are more/less risky than others based on what you see here. Each sector differs in multiple ways, including the number of active firms, regulatory obligations to report incidents, business models, technology portfolios, and the distribution of organization sizes. All that means many things other than cyber risk posture contribute to the number of publicly known incidents shown in Figure 2.

[3]The geometric mean is a more accurate measure of central tendency than the arithmetic mean (the average) for highly skewed data.

REPORTED INCIDENTS PER ORGANIZATION

Although incident frequency at the industry level is interesting, enterprise cyber risk models generally focus on individual firms. We begin the transition in Figure 3 by investigating the number of incidents attributed to each organization in our dataset over the 10-year timeframe.

Over three-quarters of the firms in our dataset[4] experienced only one publicly known event over the last decade. The other 22% recorded multiple incidents, with 4% suffering five or more. This is not the last chart you'll see in this study depicting the long tail of cyber risk.

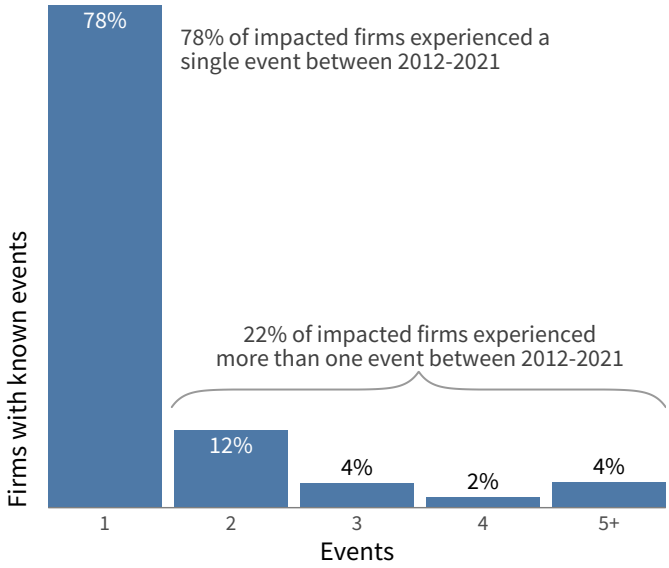


Figure 3: Number of publicly known incidents per organization

Modeling loss event frequency

> "I'm a model; do you know what I mean?"
> Right Said Fred (obviously punning the over-reliance on averages in risk modeling)

Studying incident frequency across 10 years provides a useful perspective, but most cyber risk managers seek to answer forward-looking questions like, "What's the likelihood we'll suffer a loss event in the next 12 months?" This section develops an answer to this question with a suitable statistical model.

To derive annualized event frequency on a per-firm basis, we could tally the number of incidents each year for each organization in the data. However, that would yield just 10 observed periods for each firm and result in very erratic measurements. Instead, we divided our dataset into 12-month rolling windows and counted the events for each organization. This gave us up to 107 observations per firm[5], a larger sample that we could employ more confidently to model the annualized loss event frequency.

We then treated these sliding windows as samples from an underlying probability distribution and used maximum likelihood estimation to find the parameters that best fit the data for a number of candidate distributions. Using the Kolmogorov-Smirnov test and the Cramér-von Mises statistical tests[6], we examined whether we could reject the null hypothesis that samples were drawn from the fitted distribution. We found that one particular distribution, the Poisson log-normal distribution, passed the requisite statistical tests and provided realistic estimates of multi-event years.

Frequency parameters: Poisson log-normal		
Type	Mean (μ)	Standard deviation (σ)
Upper Bound	-2.284585	0.8690759
Lower Bound	-6.394251	1.7831914

Table 1: Event frequency model parameters

What's the upshot of this statistical pedantry? We have a nice, closed-form representation of the probability of an organization facing a certain number of events in a one-year time span. Adventurous souls who would like to implement their own version of this model, will find the requisite parameters in Table 1. Plug and chug in the risk modeling tool of your choice.

[4]Since we don't have a reliable count of all active firms around the world, we can't say how many had zero incidents. Only firms with at least one publicly known cyber event are in our dataset.
[5]Subject to the firm being in operation over the entire 10-year period, a fact we account for in our data preparation.
[6]Distributions we tried: Poisson, negative binomial, geometric, their zero-inflated versions and the Poisson log-normal.

LOWER AND UPPER BOUND ESTIMATES

Estimating the probability of incidents requires a known sample of firms on which to base calculations. Unfortunately, we don't have a reliable count of relevant, active firms around the world. But we have a couple of proxies that can be used as a basis for reasonable lower and upper bound estimates.

LOWER BOUND: This includes all registered organizations in the United States according to Dun & Bradstreet (because we don't have numbers for the whole world). This assumes that incident frequency among the U.S. firms is similar to that everywhere else, which is certainly not the case. But it's a good starting point, even if you don't work for a U.S. firm. We call this the lower bound because it assumes that all registered firms engage in activities that subject them equally to the kinds of incidents found in this dataset. We don't believe that to be the case.

UPPER BOUND: This includes all organizations recorded in our dataset, which means these organizations have experienced a known incident at some point in the past. While that's clearly not the case for all organizations, this upper bound approach is based on the premise that not all firms are equally subject to the kinds of incidents contained in this dataset (i.e., perhaps they don't use IT or aren't subject to incident disclosure regulations). This assumes that all firms prone to incidents have already had one incident, thus likely resulting in overestimation.

The "just right" (Goldilocks) zone is, of course, somewhere in the middle. It's impossible for us to know exactly where your organization falls between the lower and upper bounds, so we've opted to share both to support your assessment. We keep things simple by presenting upper-bound charts and including lower-bound values in the tables. In general, the upper-bound offers a more risk-averse view (higher values). Choose one or fuse both to suit your organization's risk posture and tolerance.

For the rest of us, Figure 4 presents the upper-bound observed values from historical data (in gray) and modeled estimates (in blue) for annualized loss event frequency. Note that the observed and modeled values align pretty well, pointing to a good model that fits the data.

Having a version of Figure 4 for YOUR organization would obviously be ideal, but producing that would require a host of factors that aren't in our data to study. However, those desiring more tailored event frequency models for particular types of organizations may find them via IRIS Risk Retina. We couldn't possibly generate frequency models for every desired firmographic permutation in a public report!

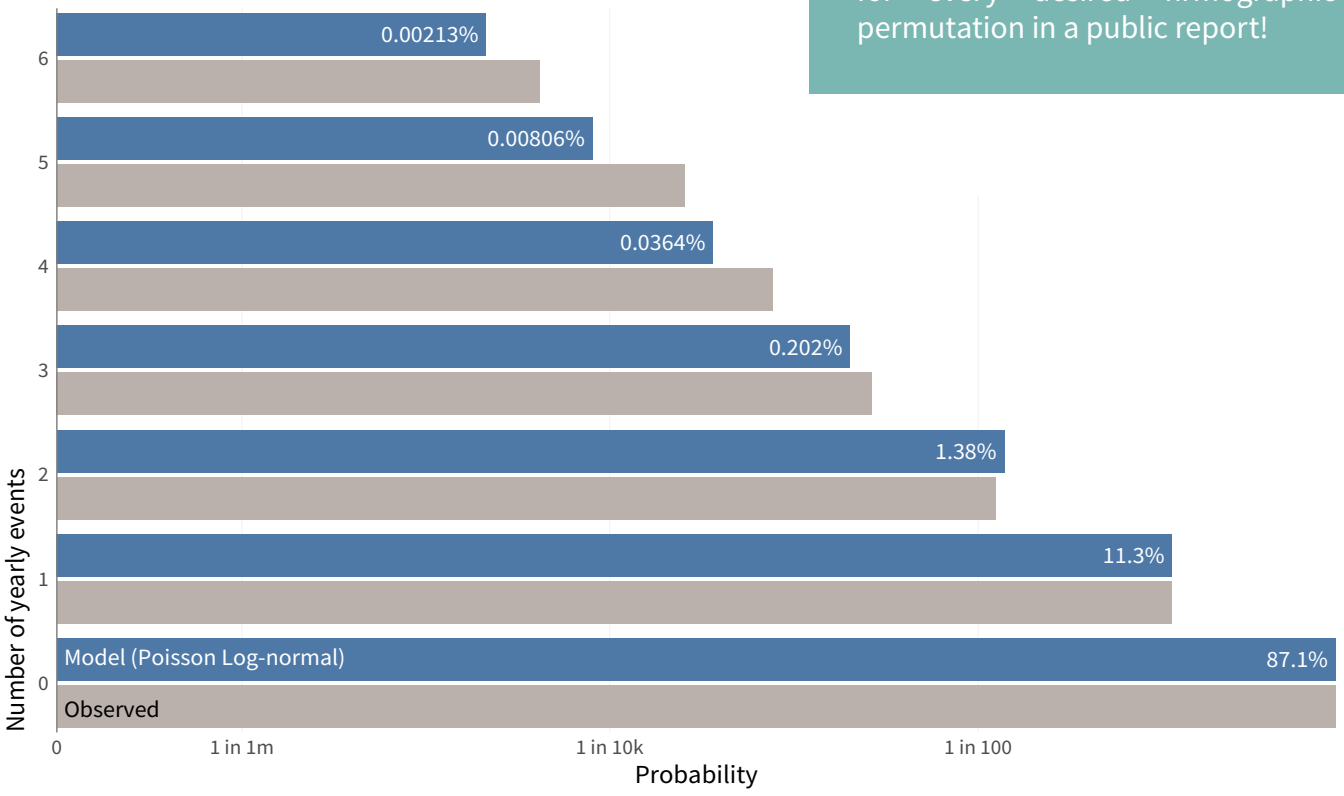


Figure 4: Upper bound observed and modeled annualized loss event frequency

EVENT FREQUENCY BY ORGANIZATION SIZE

Because the IRIS 20/20 demonstrated that event frequency differs by organization size, we repeated the process outlined above to develop separate models for groups ranging in log increments from $10M to over $100B in annual revenue[7]. Table 2 presents the output of those models for each revenue bracket and serves as a handy pocket reference guide for cyber risk managers who need a quick answer when asked about loss event frequency.

Probability of a firm experiencing a given number of events			
Revenue category	One or more	Two or more	Three or more
Upper Bound			
More than $100B	29.33%	9.32%	3.56%
$10B to $100B	21.93%	4.91%	1.28%
$1B to $10B	17.04%	3.09%	0.71%
$100M to $1B	12.95%	1.56%	0.23%
$10M to $100M	11.53%	1.12%	0.11%
Lower Bound			
More than $100B	29.30%	9.31%	3.49%
$10B to $100B	14.20%	2.73%	0.71%

Table 2: Quick reference for loss event frequency estimates

Per Table 2, the upper bound probability of at least one cyber event is close to 2.5X higher for large enterprises than for small to midsize firms. That disparity grows to nearly 9X for two or more events and 32X for three or more. Thus, we conclude that assessing the likelihood of multiple related or unrelated incidents is critical for managing cyber risk in enterprise-class organizations.

Risk management is all about understanding and managing the unlikely outcomes, and the above analysis should help better plan for such scenarios.

Overall though, the most likely outcome for organizations of any size is that they won't experience any incidents over the next year that bubble up to public knowledge (87%). Good to know, but certainly not supportive of an "it would never happen to us" conclusion. Risk management is all about understanding and managing the unlikely outcomes, and the above analysis should help better plan for such scenarios.

WHY DIDN'T YOU USE A BETAPERT DISTRIBUTION?

While BetaPERT distributions are commonly used in cyber risk quantification, they're not appropriate for estimating the expected number of events. If you're modeling the probability of an event, the BetaPERT will work fine because it's a continuous distribution. But as seen above, organizations can and do experience more than one incident in a year. Thus, a discrete distribution must be used to estimate frequency as we've done here.

[7]We decided not to develop a model for organizations under $10M in revenue for various reasons related to data quality and model reliability. Anytime we show stats for loss event frequency, they refer to organizations over $10M in revenue.

LOSS EVENT FREQUENCY BY INDUSTRY

We suspect that the reaction of many to the event frequency statistics in Table 2 is "But what about my organization?" Well, to put it bluntly, we don't know anything about your organization. But we can share some insight into where your organization's industry stands relative to others. Figure 5 compares the modeled probability estimate of firms in various industries suffering at least one security incident relative to the public sector[8].

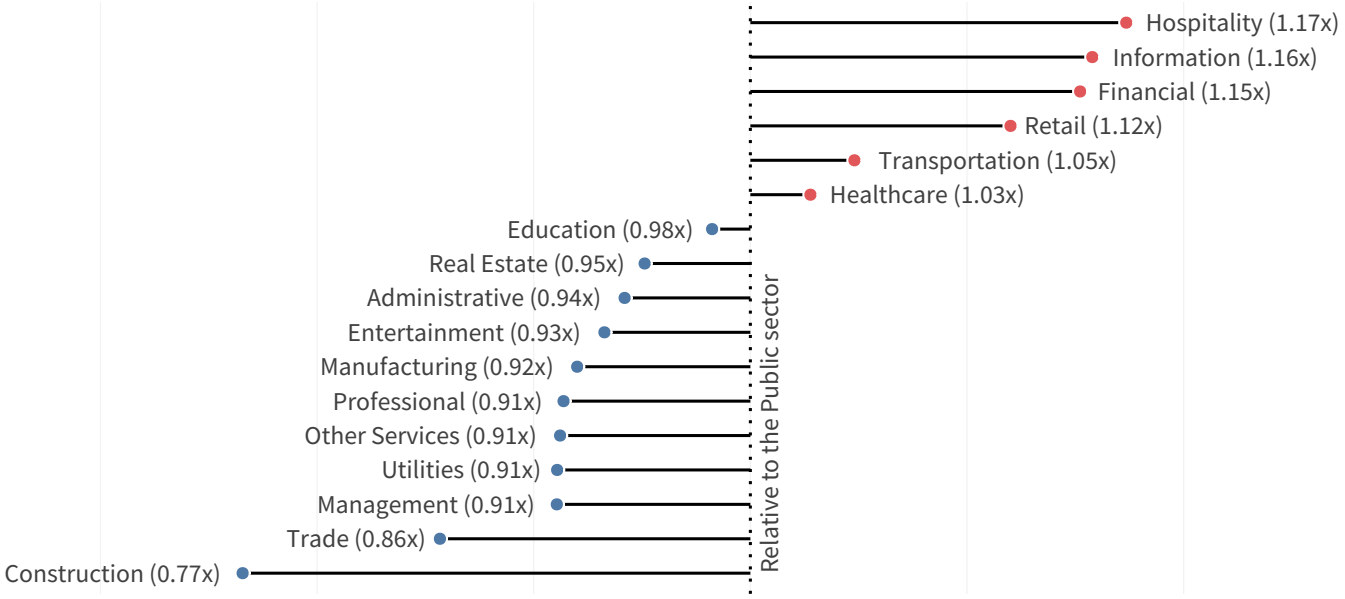


Figure 5: Relative probability of one or more loss events among sectors

Over half of sectors fall below the loss event frequency attributed to the public sector. At the same time, note that the variation isn't huge among those with the lowest and highest rates. This stems from the fact that the variation in event frequency among organizations within industries is much greater than we see when comparing across the industry means.

That said, flipping between Figures 2 and 5 does reveal some interesting differences. The Hospitality sector resides in the lower half for the proportion of all events in Figure 2 but jumps to first place when assessing the probability of experiencing a cyber loss event. Despite having more incidents in the historical record than any other, Healthcare doesn't make the top five in terms of annualized likelihood. We'll leave you to review and compare as desired while we move on to the analysis of financial losses in the next section.

[8]We chose the Public sector as the reference point here for two reasons: First, these relative comparisons aid CISA in their primary mission of prioritizing private sector engagements relative to Federal Civilian Executive Branch (FCEB) Agencies. Second, there's a lot of publicly-available information on public sector's incidents, which provides a good baseline for comparison.

Loss Magnitude Analysis

"He who wishes to fight must first count the cost."
Obligatory quote of Sun Tzu to make this a legit cybersecurity report

Having established estimates and distribution parameters for loss event frequency, we now turn to the task of counting the costs incurred when an organization suffers from cyber events. We'll start with observed losses from our historical dataset and then construct a model that best fits those values. We also include an analysis of data records exposed and how to leverage this oft-misused measure to estimate overall financial impact.

RISK PRO TIP:

While reading this section, keep in mind that not all losses for all incidents become public. Certain types of losses are easier to identify from public records, such as class action suits and SEC Filings. Other forms of loss can be difficult to quantify and/or get absorbed internally rather than resulting in outward expenditures. We suspect the losses from highly public, major incidents are more complete than those from minor events due to increased scrutiny and public records. Thus, we hold that our recorded losses suitably reflect known financial losses from publicly visible cyber incidents.

Historical loss events

Financial losses tend to be less reported than other data points for cyber events. There are many reasons for this, but the result is that the majority of incidents in our dataset do not include any information about losses. But there are enough of them (over 1,800 to be exact) that do include losses to form a well-supported quantitative understanding of the size and shape of those losses over the last decade.

In the IRIS 20/20, we took some extra steps to make the point that loss magnitude for cyber events doesn't follow the familiar bell-shaped "normal" distribution that applies neatly to many other things. A simple linear-scale chart of losses reveals a strong skewing toward lower values, with an exceedingly long tail of rare-but-extreme losses extending to the right. We're not including that chart this time.We'll go right into Figure 6, which presents a distribution of historical cyber event losses on a log scale. If you're thinking it looks like the aforementioned normal bell curve, you've just made an astute and important observation. Loss magnitude closely follows a log-normal distribution[9].

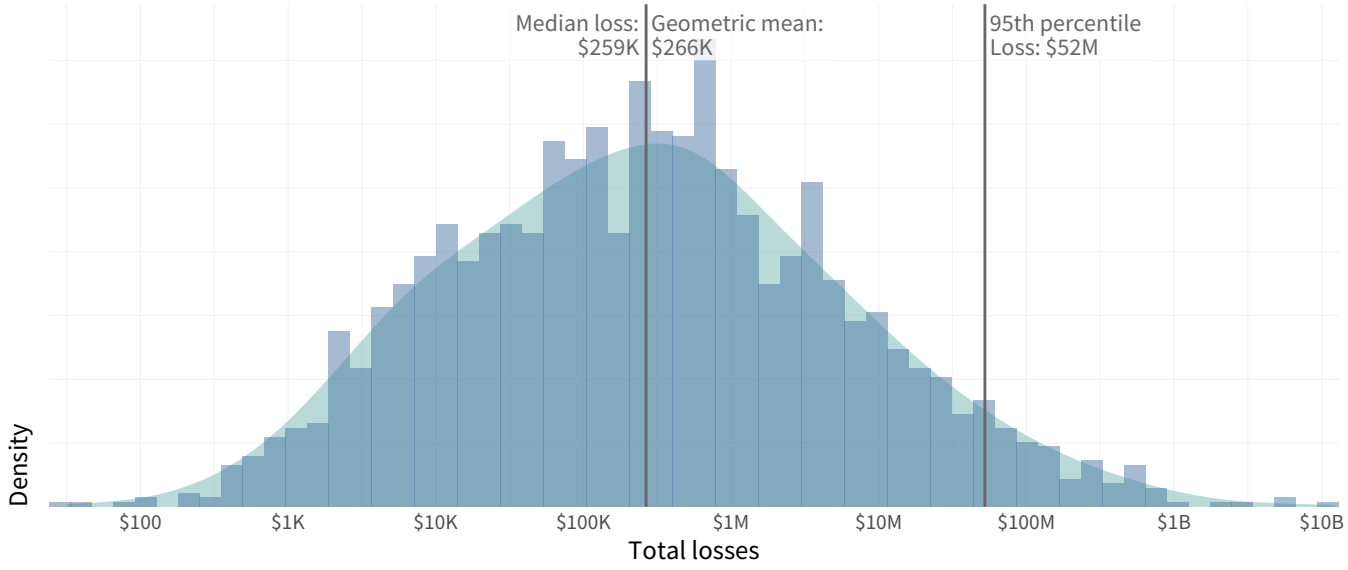


Figure 6: Distribution of reported losses for security incidents from 2012 to 2021

[9] If this mention of a log-normal distribution has you running off to find your college statistics textbook, don't bother. A log-normal distribution is just a normal (Gaussian) distribution, which ensures that if we take the log of every point in our dataset, we can apply all the same properties and techniques from a normal distribution to this collection of log-transformed points. Isn't math fun?

Overall, recorded losses range from less than a hundred bucks, to well over a billion USD. That span is probably a wee bit too wide to satisfy executives and board members asking that dreaded question, "How much is this going to cost us?" But take heart—statistics have our backs here.

Remember that we're looking at events over a 10-year period. Recent news stories may have you wondering about the effects of inflation[10] on these numbers and our ultimate results. We've taken this into account and adjusted all monetary amounts via the U.S. annual inflation result to put things into today's dollars.

As noted in Figure 6 and Table 3, the typical cost of a security incident is about $263K (as measured by the median and geometric mean). The average (arithmetic mean) is over $25M; however, it's not a good measure of typical losses due to the long-tailed distribution. If you're looking to convey what a really bad cyber event might cost, we suggest using the 95th percentile value of $52M. The worst-case scenario—according to public records—currently stands at $12B. Beyond that, feel free to venture deep into "What If?" territory, but we have no data sherpa to guide you. But models can help navigating realms beyond the edge of data; hence, stay tuned.

Loss summary							
Minimum	First quartile	Geometric mean	Median	Third quartile	95th percentile	Maximum	Total events
$32	$29K	$266K	$259K	$2M	$52M	$12B	1,893

Table 3: Loss magnitude summary statistics

> We find it surprising how much the typical loss magnitude varies among industries. For example, Information Services is nearly 8 times that of the Agricultural sector.

In Table 4, you'll find statistics on the magnitude of *typical* (measured by the geometric mean) and major (95th percentile) loss events for each primary sector. It's expected that the top 5th percentile of losses would vary substantially—from a low of $3M in Agriculture to a high of $177M in Transportation. But we do find it surprising how much the typical loss magnitude varies among industries (Information Services is nearly 8X that of Agriculture). It's another example of how narrowing key measures to peer organizations will improve the accuracy and utility of your cyber-risk assessments.

Table 4 (Right): Loss magnitude summary statistics by sector

Losses observed per sector		
Sector	Geometric mean	95th percentile
Administrative	$183K	$50M
Agriculture	$61K	$3M
Construction	$66K	$6M
Education	$139K	$5M
Entertainment	$468K	$92M
Financial	$437K	$88M
Healthcare	$211K	$13M
Hospitality	$217K	$52M
Information	$476K	$108M
Management	$472K	$136M
Manufacturing	$467K	$108M
Mining	$2M	$8M
Other Services	$103K	$13M
Professional	$384K	$91M
Public	$145K	$14M
Real Estate	$131K	$4M
Retail	$354K	$52M
Trade	$317K	$12M
Transportation	$369K	$177M
Utilities	$298K	$19M

[10]Strictly speaking, the Consumer Price Index for All Urban Consumers as published by the Federal Reserve

LOSS MAGNITUDE RELATIVE TO REVENUE

The distribution of reported losses presented in the above section makes no distinction between firmographic and other factors that influence loss magnitude. One organizational factor that received a good deal of attention following the publication of IRIS 20/20 is organization size. In a nutshell, we found that the relative impact of cyber events on SMBs is substantially worse than that of larger enterprises. Figure 7 updates that analysis and reaffirms the same takeaway.

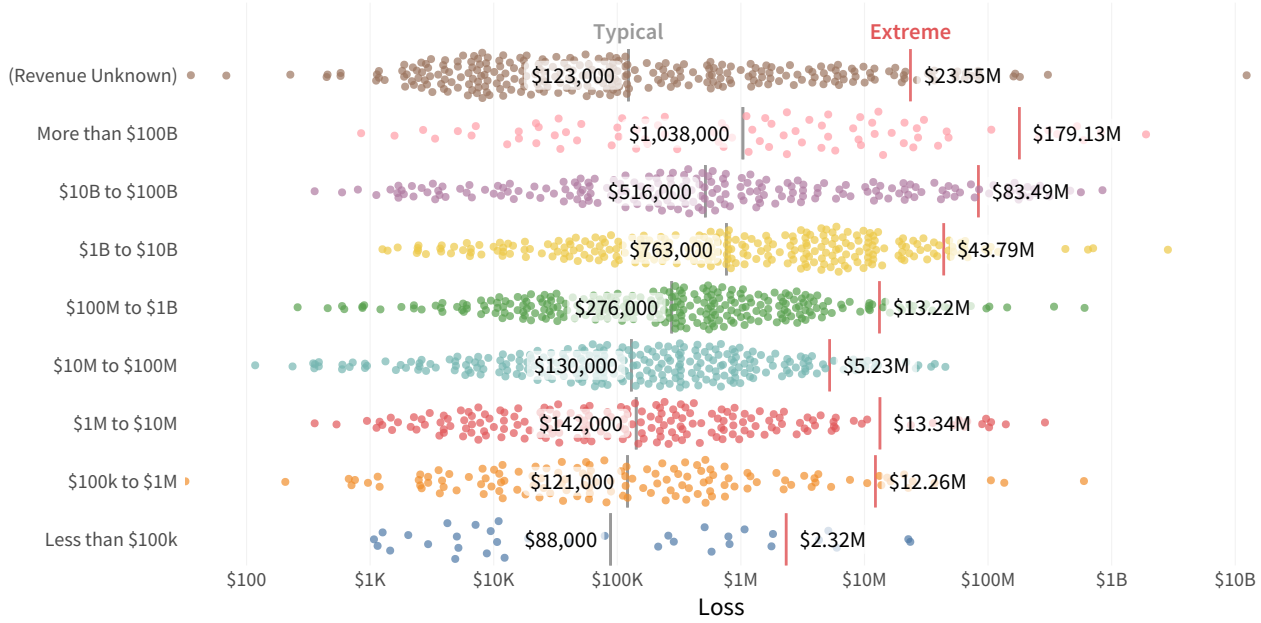


Figure 7: Distribution of reported cyber event losses by firm size (in revenue)

On the surface, the absolute costs of a typical or extreme loss event for large organizations exceed those of small companies by more than 10X. That's certainly worth incorporating into enterprise cyber-risk assessments. But some simple math yields another important finding lurking just under the surface. A $10B enterprise hit with the typical (geomean) loss amount for that size tier of $516K can expect a cost that represents 0.00516% of annual revenues. A small shop that brings in $100K per year could lose nearly its entire annual earnings in a typical loss event ($88K)!

Diving even deeper into the topic of relative impact, Figure 8 plots historical event losses as a percentage of annual revenue. There, we see that the reported losses for two-thirds of all publicly known security incidents fall below 1% of revenue (and most of those far below that mark). A little over a quarter of incidents fall in the span between 1% and 100%, while 6% actually exceed the organization's yearly income. What's more, some events exceed revenue by 100X!

Figure 8: Distribution of event losses as a percentage of annual revenue

The colors applied to Figure 8 bring us back to the discussion of the relative impact of cyber events on small vs. larger organizations. Gartner defines a small business as one having less than $50M in annual revenue. So, that's the distinction that appears here in red. It's clear that the majority of loss events involving midsize and large firms (in blue) fall below 1% of their income, while the higher ratios on the right side of the spectrum are almost entirely populated by small businesses. Here's a sobering stat: SMBs were the primary victim in 89% of all cyber loss events that exceeded 10% of revenue.

ARE LOSSES GROWING OVER TIME?

When presenting our analysis of loss magnitude from IRIS 20/20, it was quite common for someone to ask whether loss magnitude was increasing over time. "We haven't looked into that" was our standard answer, but we knew we couldn't play that card indefinitely. So, we decided to take up that question in this 2022 update.

To do things right, we're widening our standard 10-year timeframe all the way back to the start of the millennium when Y2K bugs were top of mind for most of us in the field at the time. Figure 9 uses the same inflation-adjusted approach from the above section to plot publicly recorded losses for cyber events occurring each year (blue-gray dots). We've added box plots to draw attention to the key stats and make it easier to track what's happening to the median and variation of loss magnitude over time.

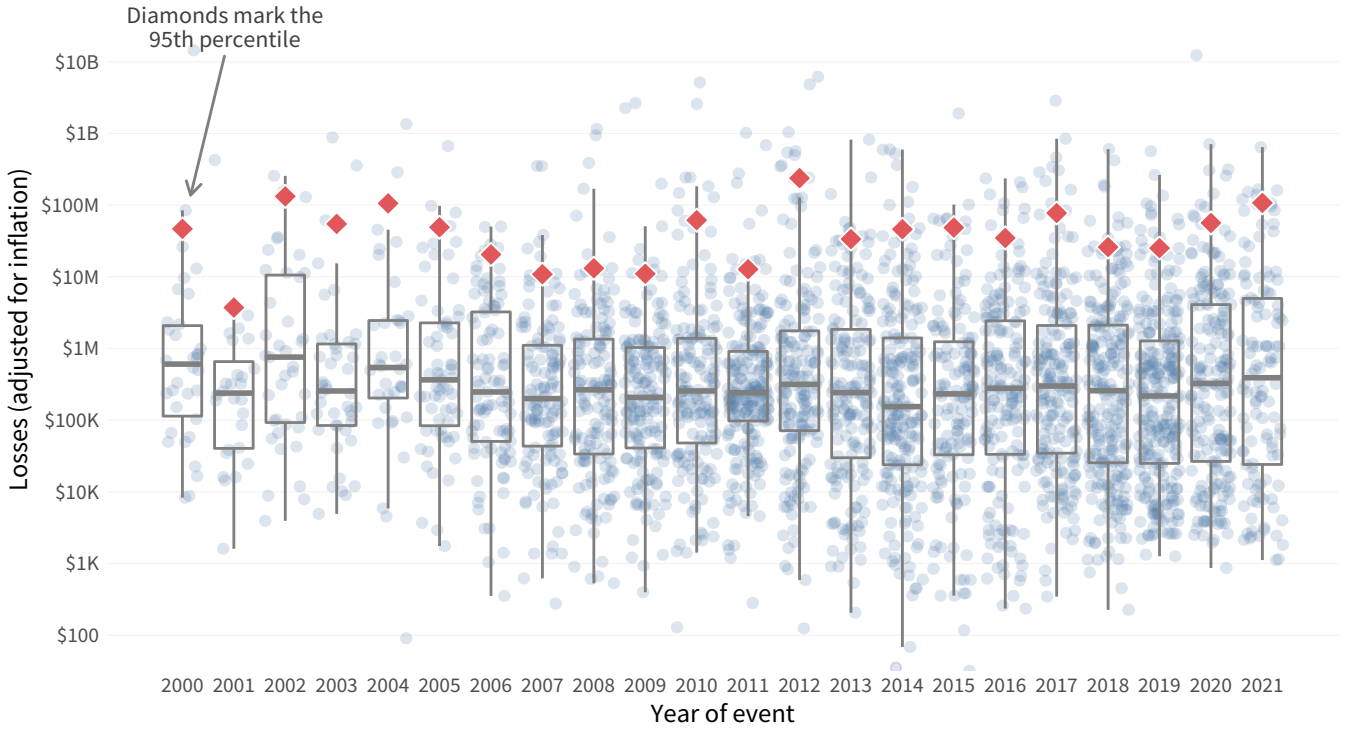


Figure 9: Inflation-adjusted reported losses from cyber events per year

If you're having trouble discerning whether the median loss values in Figure 9 are increasing or decreasing over time—don't fret. Your eyes are not deceiving you. There has been actually no statistically significant trend either up or down for median losses over the last two decades. We suspect that will be met with surprise or skepticism by many because there's a general sense that everything in cybersecurity is getting worse.

Yes, there are a thousand "buts" that could be added here, many of which are valid. It's noteworthy that the upper range of extreme losses has been trending up over the last several years. On the lower end, expanded mandatory disclosure laws probably result in larger numbers of minor incidents becoming public knowledge (note the increased density of dots in the lower 25th percentile starting in the mid-2000s). There appears to be more variation in loss magnitude (note the wider interquartile range in the latter years). Perhaps soft costs and brand damage are increasing but aren't reflected in public loss reporting (though studies are mixed regarding the lasting reputational impact of cyber events).

All this (and more) is certainly worth considering. But let's not outright dismiss what 20 years of data on a huge number of cyber loss events reveal here. And maybe, just maybe, it would be okay to acknowledge that every morning in cybersecurity might not be a more Terrible, Horrible, No Good, Very Bad Day than the one before.

Modeling loss magnitude

In this section, we identify a generalized distribution and associated parameters for cyber event losses based on historical data. While frequency involves a discrete number of events (with a probability of occurrence), losses can be fractions of dollars. This shift from a discrete to a continuous space broadens the options for applicable distributions.

We tested several continuous distributions to determine which achieved the best fit for observed losses among all incidents in our 10-year sample and landed on log-normal. That's convenient because it means that we can apply all the same techniques from a normal distribution by taking the log of every point in our loss data.

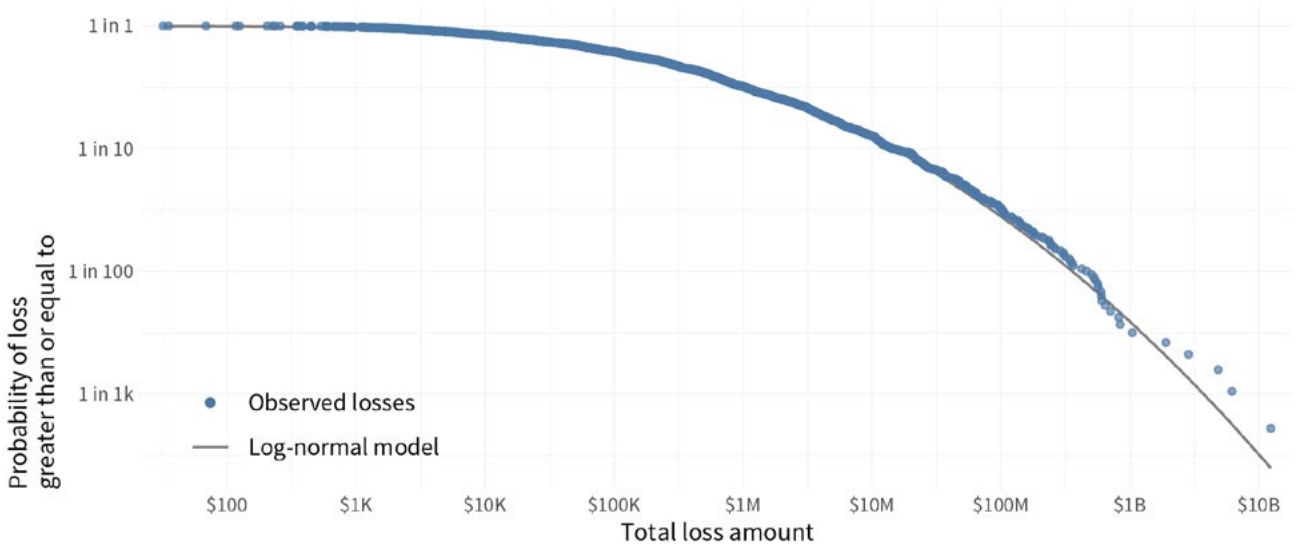


Figure 10: Log-normal distribution fit to historical cyber event losses

Figure 10 fits the log-normal distribution (black line) to historical losses (blue dots). The model fits pretty well until around the $1B mark, where it appears to under predict observed losses. But it's not surprising that things would get a little wild and unpredictable way out in the fringe of the observed data points. And because of that, we'll examine those extreme tail losses more closely in a later section and offer another approach to anticipating such events.

Table 5 provides the relevant parameters associated with the derived distribution to plug into your cyber loss modeling tool of choice. Similar parameters can be produced for a BetaPERT model, but the fit is not nearly as good as what's shown above for the log-normal distribution.

Loss parameters: Log-normal	
Mean (μ)	Standard deviation (σ)
12.55949	3.068723

Table 5: Loss magnitude model parameters

More straight talk on per-record loss estimates

> "It is a tale...full of sound and fury, signifying nothing."
> ~Macbeth

We devoted a fair amount of space in the IRIS 20/20 to set things straight to estimating losses using a flat cost-per-record approach. We felt it was necessary because the approach has a strong following among cyber risk practitioners, despite being a terrible predictor of loss magnitude. We won't harp on it as much this time, but we still feel compelled to revisit—and continue to improve—cost-per-record estimates in light of the latest data.

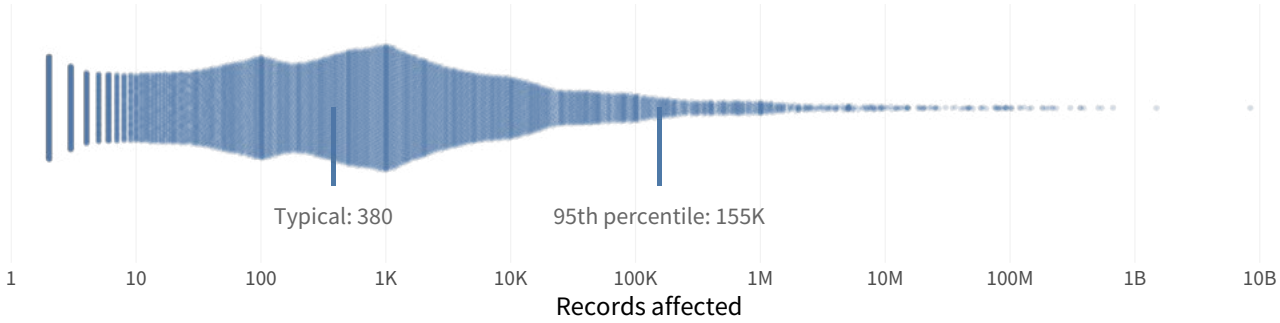


Figure 11: Number of data records affected by security incidents from 2012 to 2021

Let's begin by viewing the number of data records compromised by security incidents during our 10-year period of study. It's clear from Figure 11 that, similar to financial losses, record counts exhibit a long-tailed distribution. That's important to know because it means multiplying the number of records by an average cost per record won't yield valid estimates. The next chart illustrates why.

Figure 12 calculates the per-record costs for each incident in the dataset. If the relationship between record count and loss magnitude were linear, these events would converge around the horizontal dotted line representing the average cost per record. Clearly, that's not happening here. Instead, losses appear much higher for small data breaches and plummet to pennies per record for mega breaches. Thus, using flat cost-per-record results in routine under- and over-estimates that stray upwards of $100B off the actual reported cost.

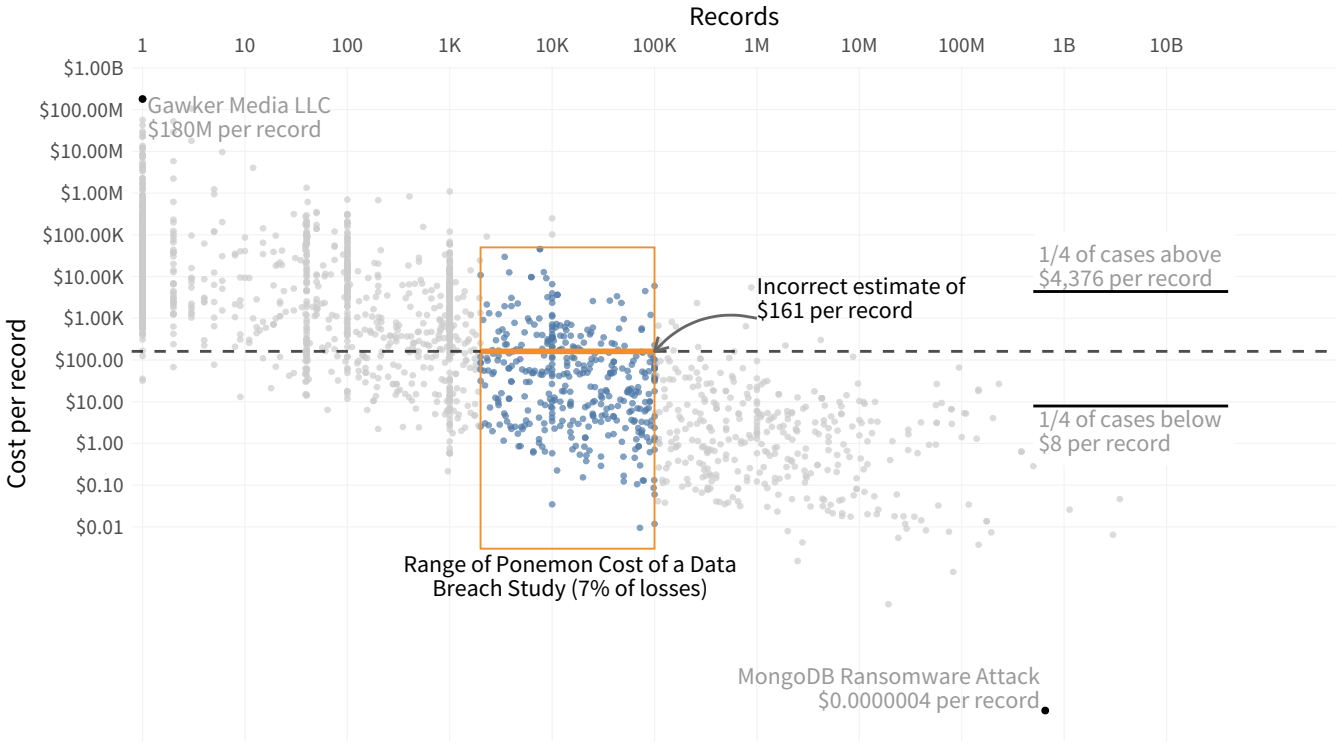


Figure 12: The fallacy of a flat cost per record for estimating cyber event losses

The orange box in Figure 12 marks the (very small) range of events to which the Ponemon Cost of a Data Breach Study "average cost per record" metric can be applied. It's always good to state such limitations, but that doesn't prevent people from ignoring them anyway. One example of overestimation based on misapplying per-record cost estimates claimed $5T in losses from cloud misconfigurations[11]. No study can fully prevent the misuse of findings, but this metric seems particularly prone to misapplication

The good news is that we can create a fairly simple (non-linear) model that does a much better job of predicting a range of losses from the number of records compromised in a breach. Use Table 6 with our (and, more importantly, the data's) blessings to replace the flat cost-per-record method that's long outlived its usefulness.

	Probability of at least this much loss					
Records	$10K	$100K	$1M	$10M	$100M	$1B
10K	91.5%	69.2%	35.6%	10.8%	1.8%	0.1%
100K	94.9%	77.8%	45.8%	16.5%	3.2%	0.3%
1M	97.1%	84.8%	56.2%	23.8%	5.7%	0.7%
10M	98.5%	90.1%	66.2%	32.6%	9.3%	1.4%
100M	99.2%	94.0%	75.2%	42.5%	14.5%	2.7%
1B	99.6%	96.5%	82.7%	52.9%	21.3%	4.8%
10B	99.8%	98.1%	88.6%	63.1%	29.7%	8.0%

Table 6: Probable losses based on the number of records affected in a cyber event

Table 6 works like this: Pick the number of records for which you're trying to estimate losses. The percentages in that row denote the probability of losses in the amounts shown in each column. So, for example, a breach of 100K records will almost certainly (95% chance) cost at least $10K, but it probably won't (3.2% chance) exceed $100M. Although it is not quite as simple as $161 per record, it's a whole lot more accurate for better risk assessments.

RISK PRO TIP:

We realize that converting data records affected by probable financial losses in the manner espoused above is more complicated than a flat cost-per-record approach. It may even garner pushback from executives who "just want a number." However, Table 6 is a far more honest representation of the large degree of uncertainty involved in the records-to-dollars conversion. Helping decision-makers understand that reality and incorporate it into their planning might be a short-term battle, but it will be a long-term win. And it's perfectly fine to start with simple statements like, "A breach of 100M records has a median loss of about $1M, but there's a small chance (3%) that it could cost 1000 times that amount."

[11]Derived by multiplying the then estimated $150 cost per record times 33B records involved. Using the current $161 published estimate, another $300 billion in losses would have been added to this total.

Quantifying Risk Exposure

> "Risk is a function of how poorly a strategy will perform if the 'wrong' scenario occurs."
> ~Michael Porter

Event frequency and loss magnitude are good things to know in and of themselves, but many risk managers have questions like, "What's the likelihood we'll lose $X over the next year?" One way of answering such questions is to create an exceedance probability curve (EP curve), more commonly known as a loss exceedance curve (LEC), among cyber risk professionals. The purpose of LECs is to demonstrate the probability of experiencing a minimum amount of loss in a given time period. Combined with an understanding of a firm's risk appetite, LECs are great for exploring whether additional mitigation efforts are warranted.

In Figure 13, we combine frequency and loss parameters into a simulation to produce an overall LEC for a single firm. Trace any point on the curve to the x and y intercepts to determine exposure. For example, there's less than a 2% chance that any given organization will suffer cyber event losses exceeding $10M in a year.

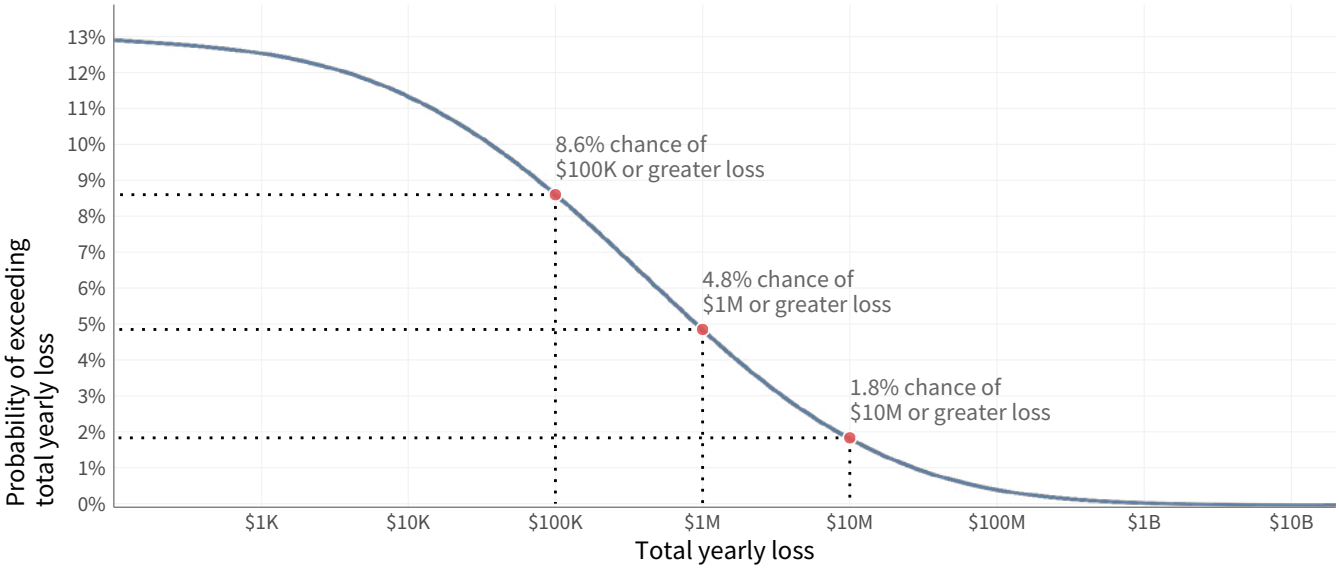


Figure 13: Example cyber loss exceedance curve for a "typical" organization (Upper Bound)

The modifier of "any given" or "typical" organization is important because this curve makes no distinction between the particulars of your organization. Your organization's risk exposure might be higher or lower due to any number of factors, including external profile, location of operation, business model, IT environment, and security posture.

We encourage you to consider the information presented in Figure 13 in light of such factors.

Extreme event and tail risk analysis

As indicated in the previous section, cyber risk has a long tail of rare but highly impactful events. Most executives and risk managers worry far more about that tail than the bulk of more predictable loss scenarios. We studied 100 of the most impactful cyber incidents in the IRIS Xtreme and sought to understand the actors, techniques, and contributing factors. We won't reproduce those analyses here, but we do want to develop a general understanding of extreme loss events and methods to analyze so-called tail risk.

Figure 14 compares the number of incidents exceeding one of the thresholds used in the IRIS Xtreme (roughly, the upper 10th percentile in publicly recorded financial losses and/or data records compromised). It's clear that certain types of organizations seem to have a higher propensity for extreme cyber loss events than others. And if your organization is one of those listed toward the top, analyzing the risk of long-tail events becomes even more important.

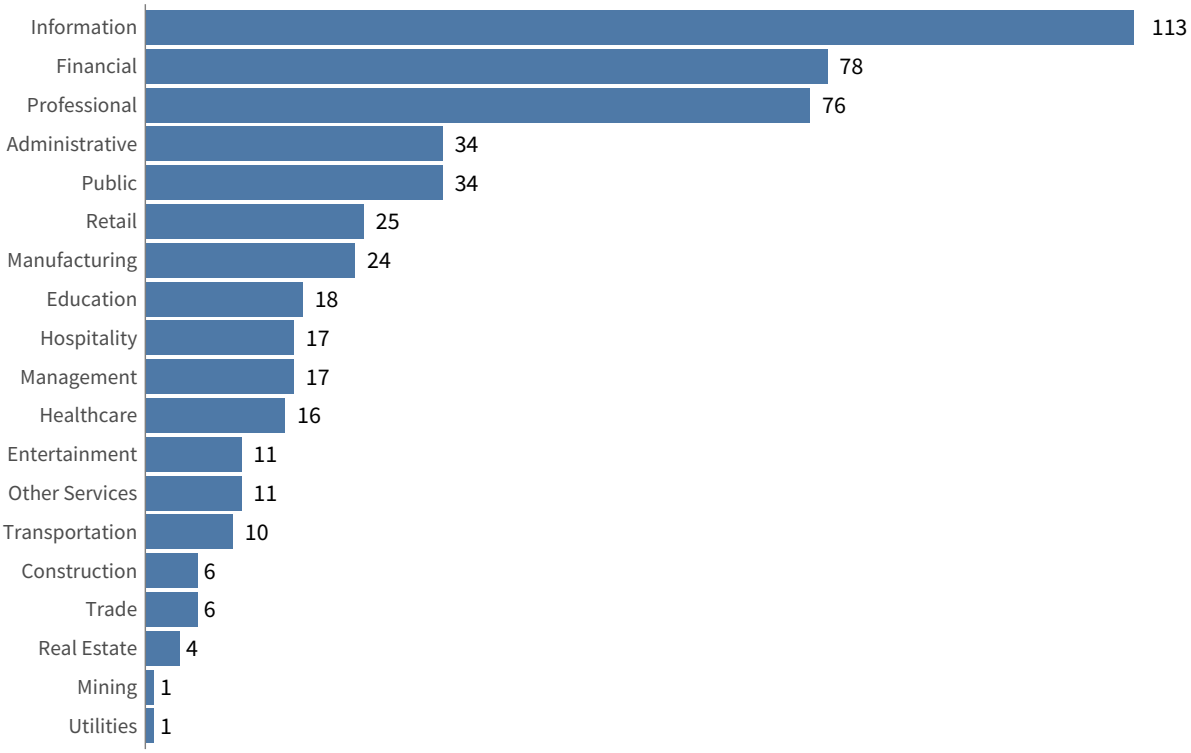


Figure 14: Number of extreme historical loss events per sector

As an example of how organizations can practically conduct tail risk analysis, we turn to a technique frequently employed by actuaries called tail value at risk (TVaR) or sometimes conditional tail expectation (CTE). TVaR is a simple concept: Given the top X% of losses a firm is likely to experience in a year, what is its average value? TVaR is useful because it helps illustrate how scary the heavy tail we see in losses can be.

	90%	95%	99%
Upper Bound			
More than $100B	$120.43M	$236.25M	$1.04B
$10B to $100B	$77.69M	$153.41M	$692.21M
$1B to $10B	$55.15M	$109.49M	$504.41M
$100M to $1B	$41.00M	$81.72M	$385.77M
$10M to $100M	$35.26M	$70.36M	$334.71M
Lower Bound			
More than $100B	$120.41M	$236.20M	$1.04B
$10B to $100B	$53.57M	$106.66M	$499.74M
$1B to $10B	$17.96M	$35.91M	$174.19M
$100M to $1B	$4.84M	$9.69M	$48.33M
$10M to $100M	$561.22K	$1.12M	$5.61M

Table 7: TVaR analysis

Table 7 highlights the upper- and lower-bound TVaR estimates for each revenue grouping. While the 95th percentile for all loss events might be a mere $52M (see Table 3), the upper-bound 95% TVaR exceeds $91M for all organization sizes. This means that the minimum loss magnitude (marked by the 95th percentile) of a once-in-20-year kind of event might not seem so bad, but averaging the full length of the long tail becomes very pricey indeed.

The relatively low probability of high annual losses may surprise some, especially those who regularly hear that the digital sky is falling around them. But for most of us at least, the sky remains where it's supposed to be, despite the ardent cries of cyber-Chicken Littles. The fact is that most organizations will not suffer a security incident, and those that do probably won't experience a worst-case scenario. But some will, and that possibility must be managed realistically. Information like this offer a far better starting point for doing so than the wavey hands driven by FUD.

The fact is that most organizations will not suffer a security incident, and those that do probably won't experience a worst-case scenario.

But some will, and that possibility must be managed realistically.

Incident Patterns and TTPs

"Pay attention to the patterns and we can presage what comes next."
~Gaal Dornick, Foundation

So far, we've treated all cyber loss events the same. That's fine at the macro level when our main goal is to quantify the overall risk to the organization. But when the goal turns to mitigating that risk, it helps to know something about the types of incidents that are most common and costly.

To help with this, we categorize incidents into a manageable number of patterns based on common threat actors, techniques, vectors, and technical impacts, as defined in the list below. These patterns are intended to represent the high-level scenarios we often see on risk registers for assessment and reporting purposes.

IRIS Incident Patterns:

All security incidents in our historical dataset are assigned one of the following patterns using a combination of natural language processing techniques and human expert assessment.

DOS ATTACK: Any attack intended to render online systems, applications, or networks unavailable, typically by consuming processing or bandwidth resources.

ACCIDENTAL DISCLOSURE: Data stores that are inadvertently left accessible to unauthorized parties, typically through misconfigurations on the part of the data custodian.

SCAM OR FRAUD: Any incident that primarily employs various forms of deception to defraud the victim of money, property, identity, information, and so on.

SYSTEM INTRUSION: All attempts to compromise systems, applications, or networks by subverting logical access controls, elevating privileges, deploying malware, and so on.

INSIDER MISUSE: Inappropriate use of privileged access, either by an organization's own employees and contractors or a trusted third party.

PHYSICAL THREATS: Threats that occur via a physical vector, such as device tampering, snooping, theft, loss, sabotage, and assault.

RANSOMWARE: A broad family of malware that seeks to encrypt data with the promise to unlock upon payment or seeks to completely eradicate data/systems without the pretense of collecting payment.

SYSTEM FAILURE: All unintentional service disruptions resulting from system, application, or network malfunctions or environmental hazards.

Figure 15 compares the percentage of the total recorded events and losses associated with each incident pattern. The proportion of compromised data records is also reflected by the size of the dot. The intent is to enable managers to focus on the riskiest types of loss events. Anything above the dotted line has a higher percentage of financial losses relative to its frequency of occurrence. Patterns below the line are relatively less costly.

Being the sole pattern in the dreaded upper-right quadrant, system intrusions are far and away the riskiest incident pattern. They account for about half of all events as well as half of total losses recorded over the last decade. While not part of this current analysis, our studies of extreme and massive multi-party events point to exploitation of valid accounts (T1078) and public-facing applications (T1190) as the primary initial access techniques for system intrusions.

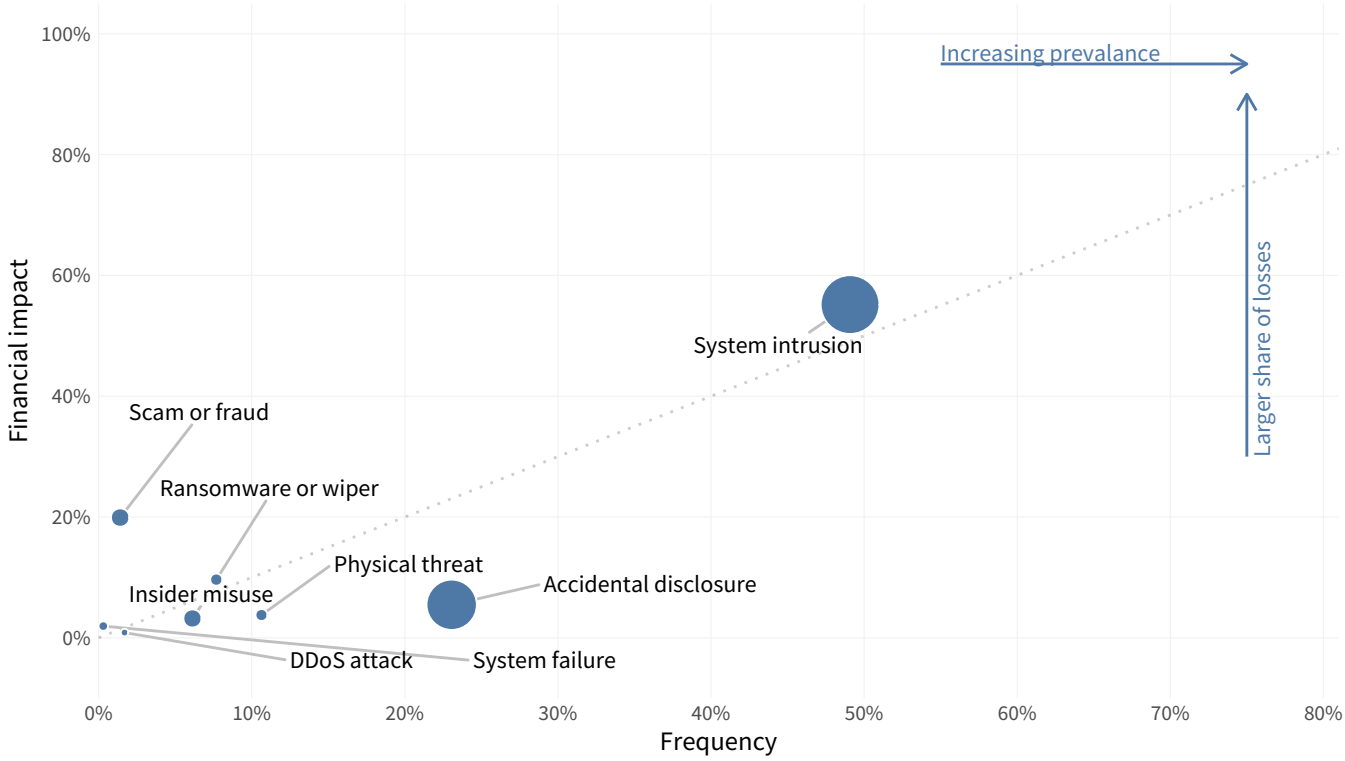


Figure 15: Relative frequency and losses associated of common incident patterns

Given all the buzz around ransomware over the last several years, you may expect to find it in a prominent position in Figure 15. You'll actually find it in the lower left amid a cluster of several other patterns at about 6% of all incidents. It does punch above its weight, though, accounting for 15% of financial losses. For those looking for tips on managing this risk, we refer you to CISA's Ransomware Guide as a starting point.

Accidental disclosure clocks in at #2 for incident frequency and #2 for data records compromised over the last 10 years. This one's a bit sad because, unlike system intrusions or ransomware, it's something we do to ourselves. Whether your organization collects customer information, develops cutting-edge IP, or handles other forms of data, you must take care of how it's stored and shared. Cloud storage is often cheap and convenient—a very attractive combo in times when cash is tight for many—but it's also prone to misconfiguration that leaves your data open to any stranger on the internet.

“ Figure 15 compares frequency and losses associated with each incident pattern. The intent is to enable managers to focus on the riskiest types of loss events.

Depending on the question being asked, it may be useful to have exact values for the proportion of events, losses, and records shown in Figure 15. Table 8 captures all of this, along with a ranking for each measure. We could dive into each of these incident patterns individually, but that's beyond the scope of what we could fit into this study. But we do hope this new addition to the IRIS helps organizations better categorize and understand their own risk scenarios.

	Frequency		Financial impact		Records affected	
	Percentage	Overall rank	Percentage	Overall rank	Percentage	Overall rank
Accidental disclosure	23.3%	2nd	5.2%	4th	38.64%	2nd
DoS attack	1.6%	6th	0.8%	8th	0.08%	8th
Insider misuse	6.3%	5th	2.9%	6th	2.46%	4th
Physical threat	11.0%	3rd	3.5%	5th	0.50%	6th
Ransomware	6.5%	4th	7.0%	3rd	0.51%	5th
Scam or fraud	1.4%	7th	18.4%	2nd	2.59%	3rd
System failure	0.3%	8th	1.8%	7th	0.16%	7th
System intrusion	49.6%	1st	60.2%	1st	55.05%	1st

Table 8: Ranking of common incident patterns with relative frequency, financial loss, and data loss statistics

A breakdown of incident patterns is certainly useful for assessing risk—otherwise, we wouldn't bother doing it—but the patterns do not specify how the events occurred. Take the #1 incident pattern of system intrusion as an example. How did the perpetrators gain access to the victim's systems? What did they do once inside? What controls could have thwarted those actions? All of these are valid questions that would greatly assist risk managers in prioritizing defenses.

When we look at the #1 incident pattern of system intrusion, several questions that would greatly assist risk managers in prioritizing defenses are, rightly, raised:

How did the perpetrators gain access to the victim's systems?

What did they do once inside?

What controls could have thwarted those actions?

Well, the good news is that we can now begin answering questions like those. The bad news is that we only have room to answer one of them. But one is better than none, right?

Since the publication of IRIS 20/20, we've developed a combination of analytical capabilities to identify tactics, techniques, and procedures (TTPs) used in a cyber event[12]. Specifically, we associate threat actions from the Vocabulary for Event Recording and Incident Sharing (VERIS) framework (the basis for Verizon's Data Breach Investigations Report (DBIR)) as well as techniques defined in MITRE ATT&CK. We focus on the ATT&CK Initial Access tactic here because it directly addresses the above question of how perpetrators gain access to the victim's systems.

[12] Incident patterns, VERIS actions, and MITRE ATT&CK techniques are not native fields in the Advisen Cyber Loss Data. We employ human analysis, machine learning, natural language processing, and external data sources to accomplish this. We're happy to nerd out on the details if you want to know more about our incident data enrichment capabilities and how we incorporate them into our research and services beyond the IRIS.

Figure 16 reveals the top initial access techniques identified across historical security incidents for each sector. Full definitions and real-world examples for each technique can be found on the ATT&CK website. MITRE is also kind enough to provide a list of recommended mitigations for these techniques to help you develop a more threat-informed defense.

	Rank								
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Administrative	Valid Accounts	Phishing	Trusted Relationship	Exploit Public-Facing Appli...	Drive-by Compromise	External Remote Services	Hardware Additions	Replication Through Removab...	Supply Chain Compromise
Agriculture	Phishing	Drive-by Compromise	Exploit Public-Facing Appli...	External Remote Services	Replication Through Removab...	Trusted Relationship			
Construction	Phishing	Valid Accounts	Drive-by Compromise	Trusted Relationship	Exploit Public-Facing Appli...	External Remote Services	Supply Chain Compromise		
Education	Valid Accounts	Phishing	Trusted Relationship	Exploit Public-Facing Appli...	Drive-by Compromise	External Remote Services	Replication Through Removab...	Supply Chain Compromise	
Entertainment	Valid Accounts	Exploit Public-Facing Appli...	Phishing	Trusted Relationship					
Financial	Valid Accounts	Trusted Relationship	Phishing	Exploit Public-Facing Appli...	Drive-by Compromise	External Remote Services	Hardware Additions	Replication Through Removab...	
Healthcare	Valid Accounts	Trusted Relationship	Phishing	Exploit Public-Facing Appli...	Drive-by Compromise	External Remote Services	Hardware Additions	Replication Through Removab...	Supply Chain Compromise
Hospitality	Valid Accounts	Phishing	Trusted Relationship	Drive-by Compromise	Exploit Public-Facing Appli...	Hardware Additions	External Remote Services		
Information	Valid Accounts	Exploit Public-Facing Appli...	Trusted Relationship	Phishing	Drive-by Compromise	External Remote Services	Replication Through Removab...		
Management	Valid Accounts	Phishing	Exploit Public-Facing Appli...	Drive-by Compromise	Trusted Relationship	External Remote Services	Replication Through Removab...		
Manufacturing	Phishing	Valid Accounts	Exploit Public-Facing Appli...	Drive-by Compromise	Trusted Relationship	External Remote Services	Replication Through Removab...	Hardware Additions	
Mining	Phishing	Exploit Public-Facing Appli...	Trusted Relationship	Drive-by Compromise	External Remote Services	Valid Accounts			
Other Services	Valid Accounts	Trusted Relationship	Exploit Public-Facing Appli...	Phishing	Drive-by Compromise	External Remote Services	Replication Through Removab...		
Professional	Valid Accounts	Phishing	Trusted Relationship	Exploit Public-Facing Appli...	Drive-by Compromise	External Remote Services	Replication Through Removab...		
Public	Phishing	Trusted Relationship	Valid Accounts	Exploit Public-Facing Appli...	Drive-by Compromise	External Remote Services	Replication Through Removab...		
Real Estate	Phishing	Drive-by Compromise	Trusted Relationship	Valid Accounts	Exploit Public-Facing Appli...	External Remote Services	Hardware Additions		
Retail	Exploit Public-Facing Appli...	Valid Accounts	Phishing	Trusted Relationship	Drive-by Compromise	External Remote Services	Hardware Additions	Replication Through Removab...	Supply Chain Compromise
Trade	Phishing	Exploit Public-Facing Appli...	Drive-by Compromise	Valid Accounts	Trusted Relationship	External Remote Services	Replication Through Removab...	Supply Chain Compromise	
Transportation	Valid Accounts	Phishing	Exploit Public-Facing Appli...	Trusted Relationship	Drive-by Compromise	External Remote Services			
Utilities	Trusted Relationship	Phishing	Valid Accounts	Drive-by Compromise	Exploit Public-Facing Appli...	External Remote Services			

Figure 16: Ranking of ATT&CK initial access techniques across sectors

We suspect most readers will focus on the top techniques listed for their sectors, which is exactly the point of including Figure 16[13]. For our part, we'll share a few sector-agnostic observations:

Not all techniques are observed for all sectors. Some of those stem from the limited visibility of events (we're pulling only from public data sources). But we think it also to some extent relates to the diversity and complexity of attacks against organizations in each sector.

Phishing and valid accounts rank among the top three techniques for most industries. If you want to keep threat actors out of your systems (and who doesn't?), prioritizing detections and defenses for those vectors should be very high on your list.

We're modifying the definition of 'trusted relationship' to include insiders who abuse their trust or privileges during an event. Official ATT&CK documentation uses that only for trusted third parties and does not include techniques specific to insider misuse. We needed a place to put in those actions, and we felt this was the best fit. That's one of the reasons it's more prevalent in Figure 16 than you'll likely see from other sources.

Overall, we see the results in Figure 16 as a good reminder that if the threat actor isn't a trusted party already (an insider or business partner), chances are good they'll try to target one (via phishing) or become one (via valid accounts). We hope that this analysis provides impetus and sparks ideas to impede their ability to do that and more!

[13] A special thanks to the crowd at FAIRcon22 for their eagle eyes in detecting a glitch in the preview version of this chart.

Methodology and Firmographics

All incidents and losses analyzed in this report are from Advisen's Cyber Loss Data. For those new to this data set, Advisen maintains a repository of well over 100,000 cyber events, with events ranging back as far as the mid-20th century. They compile this valuable information through publicly available sources, such as breach disclosures, company filings, litigation details, and Freedom of Information Act requests. The dataset also goes through a rigorous process of matching events to known company IDs (e.g., D&B and S&P). This enables the many firmographic views we share in this report.

For this report, we are working off the July 2022 release of the Advisen data feed, focusing on a 10-year window ranging from 2012 to 2021. In addition to Advisen's standard fields, we further enriched the dataset through a combination of natural language processing techniques and manual analysis. We also removed incidents that are exclusively privacy-related, as these are dominated by issues that most information security practitioners do not typically include in their response plans (items such as telephone privacy). Our 10-year period of analysis includes 77K cyber events, $57B in total reported financial losses, and 72 billion compromised records.

A tally of the industries represented by incidents in our analysis is given back in Figure 3, so we won't reproduce that here. Another one we won't bother showing is a regional breakdown. The sample contains organizations predominantly from the United States (74%), Europe (9%), and Latin America (9%). Since some of our analysis incorporates organizational revenues, we've included that along with the employee count below.

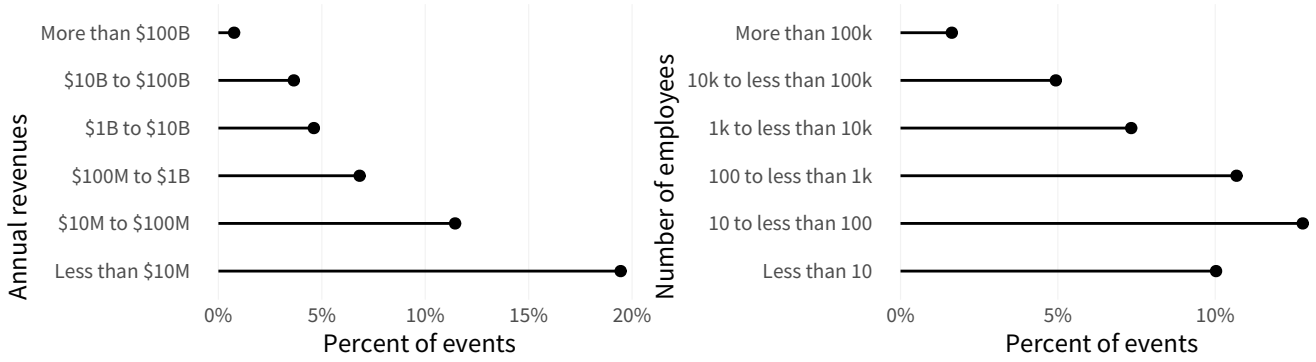


Figure 17: Annual revenue and employee count for firms affected by incidents in our sample