Microsoft

Enhancing Cybersecurity with Big Data: Challenges & Opportunities

Independently Conducted by
Ponemon Institute LLC
Sponsored by
Microsoft Corporation
November 2014

CONTENTS

2 **Introduction**

3 **The Emergence of Big Data**

How organizations use
big data for cybersecurity 5

6 **Challenges to Using Big Data for Cybersecurity**

Privacy 7
Complexity 8
Cost 8

9 **Recommendations**

10 **Appendices**

Appendix A: Methods 10
Appendix B: Research limitations 11

INTRODUCTION

Today, the volume and variety of data being created by the world's many devices has reached unprecedented levels and will continue to escalate. This big data has created profound business and social opportunities in every field, enabling the discovery of previously hidden patterns and the development of new insights to inform and guide decisions.

At the same time, protecting the information of individuals and organizations from online threats has become an urgent priority, and big data tools and techniques to enhance cybersecurity are a natural development. For example, an organization might aggregate and analyze log data from all of its computing devices to identify malicious activities. In the absence of big data techniques, the task of storing, processing, and analyzing vast amounts of data is, for all but a few organizations, simply not feasible.

Microsoft commissioned a study from the Ponemon Institute to understand whether and how organizations are using big data to improve cybersecurity, and to identify the challenges they face, including security and privacy considerations. The study surveyed more than 100 executive-level respondents in the United States and Europe representing their organization's IT security, privacy, and compliance functions such as chief information security and privacy officers.

The results suggest that while most companies have a strong interest in using big data to improve cybersecurity, concerns about the cost and complexity of big data solutions as well as privacy concerns, give them pause in deploying it. Specifically, many respondents noted a conflict between privacy and security, coupled with ambiguity of how privacy will be protected in these applications.

KEY FINDINGS:

- **The use of big data to protect companies from cyberthreats is growing** among the organizations represented in this research. According to the findings, 54 percent of respondents say their companies either use big data analytics as part of their cybersecurity defenses or plan to do so.
- **Barriers to using big data for cybersecurity include cost** (insufficient budget, 42 percent), complexity (half of respondents), and privacy concerns (56 percent).
- **Personal information must be protected** in the use of big data analytics. Although 58 percent of respondents believe that personal information used in big data analytics for cybersecurity can be protected to minimize harm or risk to individuals (29 percent say no, 13 percent unsure), 64 percent believe that clearer rules about the use of personal information are needed.

54%

RESPONDENTS THAT SAY THEIR COMPANIES USE BIG DATA ANALYTICS FOR CYBERSECURITY DEFENSE OR PLAN TO DO SO

64%

RESPONDENTS THAT BELIEVE CLEARER RULES ABOUT PERSONAL INFORMATION USE ARE NEEDED

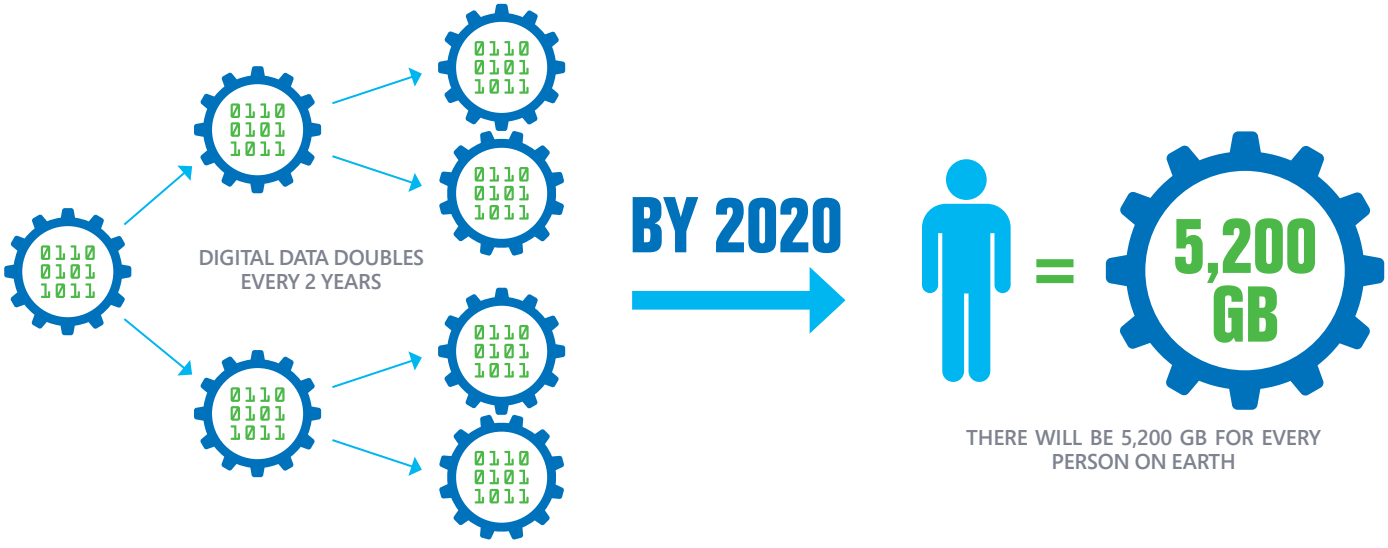
THE EMERGENCE OF BIG DATA

Big data most commonly refers to increasingly large and complex data sets. It refers not only to the volume of data, but also its variety and the velocity at which it is created, linked, and altered. This explosion of data is the result of the dramatically expanding universe of sensors, information technology services, and connected devices, all producing data.

"Big data tools offer astonishing and powerful opportunities to unlock previously inaccessible insights from new and existing data sets. Big data can fuel developments and discoveries in health care and education, in agriculture and energy use, and in how businesses organize their supply chains and monitor their equipment. Big data holds the potential to streamline the provision of public services, increase the efficient use of taxpayer dollars at every level of government, and substantially strengthen national security." [1]

The massive increases in the amount of data, coupled with increased complexity contributed by cloud computing, Bring Your Own Device (BYOD) policies, and the Internet of Things, have tested traditional information security processes. No longer is it feasible, or even possible, for human analysts to sift through this vast data and make sense of the complex interrelationships among data and systems.

All digital data created, replicated, or consumed is growing by a factor of 30, doubling every two years. By 2020, there will be over 40 trillion gigabytes of digital data—or 5,200 gigabytes for every person on earth.[2]



[1] "Big Data: Seizing Opportunities, Preserving Value," White House Big Data Report, May 2014. aka.ms/WhiteHouse-BigData

[2] "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," IDC, April 2014. aka.ms/IDC-IoT

Big data analytics focus on data discovery using data science, advanced statistical functions and algorithms, and visualization tools. These techniques can accelerate analysis, leading to insights from both traditional and nontraditional data sources. The value to an organization is the ability to discover previously unseen patterns and to develop actionable insights about their businesses and environments, including cybersecurity.

Applying such analytics to the challenge of cybersecurity has the potential to advance beyond the detection of anomalies to suggesting where analysts should look for suspicious activity, or even predicting security incidents so they can be mitigated before they occur. If big data analytics can predict hardware failures[3], similar techniques can be used to safeguard the security of critical business and government systems and enable:

- Capturing, processing, and refining data from a wide variety of services, devices, and networks.
- Visualizing large streams of data to quickly identify anomalous or suspicious events.
- Applying machine learning techniques to identify new security events based on historical patterns.
- Correlating unsupervised data sets into specific insights, detections and the identification of risks.

"Growing demands for mobility, BYOD, and the pervasive use of cloud resources diminish our ability to identify anomalous or gratuitous Internet traffic."

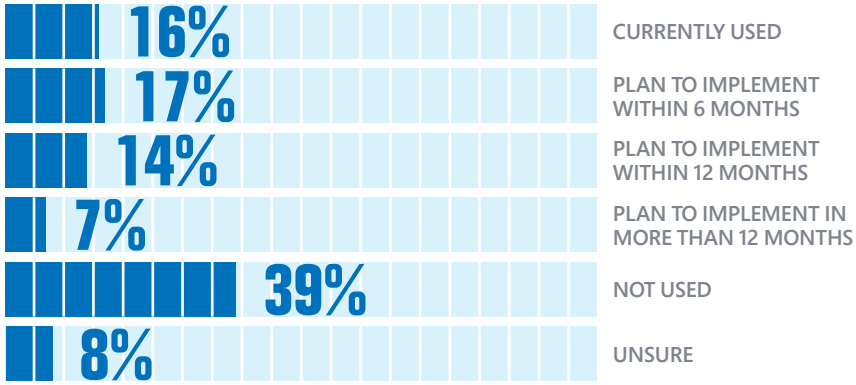
- IT director, United States

[3] "Analyzing Massive Machine Maintenance Data in a Computing Cloud," IEEE Computer Society, October 2012. aka.ms/IEEE-BigData

HOW ORGANIZATIONS USE BIG DATA FOR CYBERSECURITY

The use of big data to protect companies from cyberthreats is growing among the organizations represented in this research. According to the findings shown to the right, 54 percent of respondents say their companies either use big data analytics as part of their cybersecurity defenses or plan to in the future.

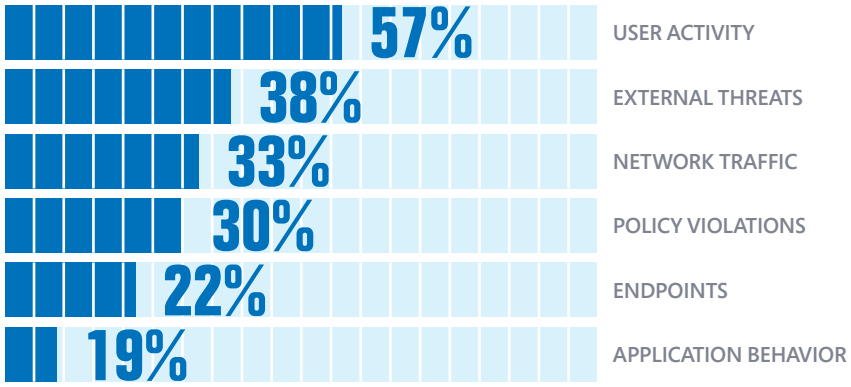
Current and planned use of big data analytics for organizational cybersecurity



The strong interest shown by executives in the use of big data analytics to secure information systems appears to be driven by a desire to better understand behavior on their systems. In particular, as shown to the right, they are interested in more information about user activity, external threats, and better understanding patterns in network traffic.

Where more data is needed to secure information systems

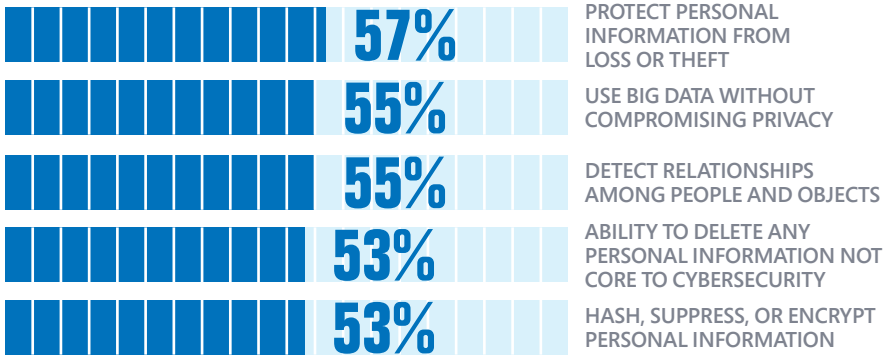
Two choices permitted



As to the right, almost 60 percent of the respondents believed that it is either essential or very important for any system using big data analytics for cyberdefense to protect personal information from loss or theft. More than half of all respondents have other requirements as well, including the ability to detect relationships among people and objects and to use big data without compromising privacy. Interestingly, the ability to detect relationships between people and objects such as a listing of all the Internet sites a user has visited to determine if they had visited a malicious site can challenge the privacy policies of many organizations.

Important capabilities in performing big data analytics for cyberdefense

"Essential" and "Very important" responses combined

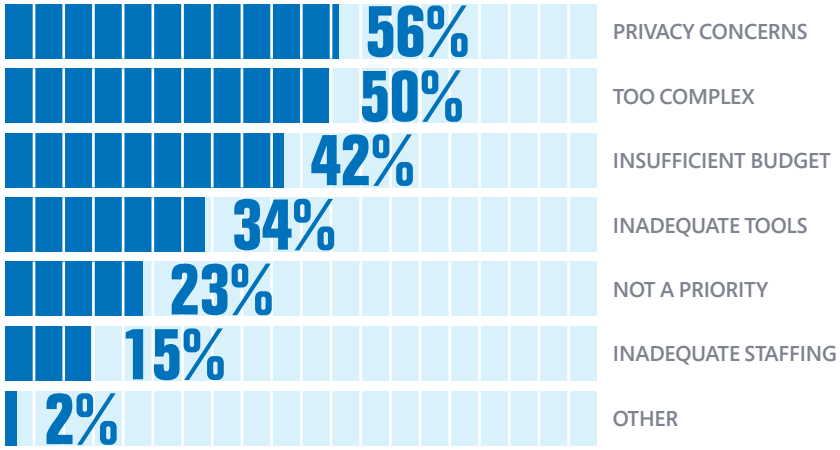


CHALLENGES TO USING BIG DATA FOR CYBERSECURITY

Sixty percent of respondents say the use of big data analytics is essential or very important to their cybersecurity defense, yet only 16 percent of those surveyed have actually implemented big data solutions. There are clear obstacles to the use of big data for cybersecurity. Among the 47 percent of respondents who have no plans to adopt big data for cybersecurity or are unsure about its use, the primary reasons for not doing so are privacy concerns, worries about solutions that are too complex, and insufficient budget.

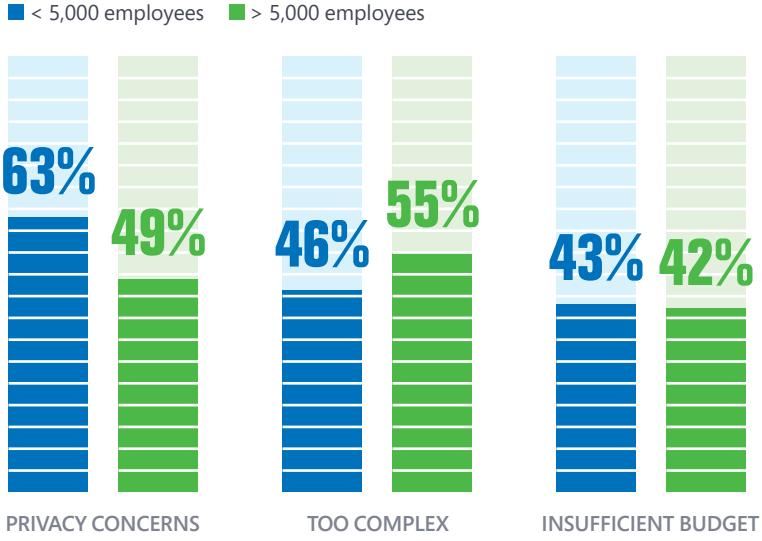
Why organizations are not using big data analytics for cyberdefense

More than one choice permitted



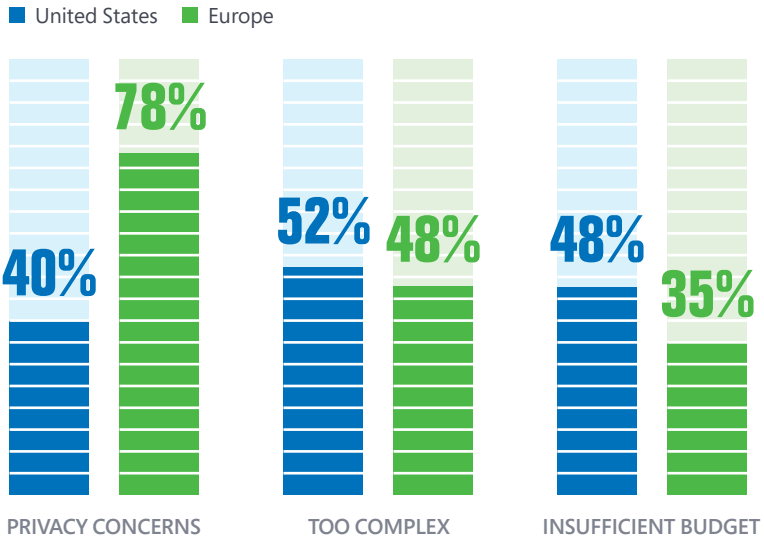
As shown to the right, concerns about the use of big data solutions for cyberdefense vary with the size of the organization. Complexity overtakes privacy as the fundamental impediment to adoption in organizations with 5,000 or more employees, and privacy concerns are key for organizations with fewer than 5,000 employees.

Comparisons of implementation concerns: large versus small organizations



Responses in the United States and Europe show a significant difference related to privacy and an appreciable gap related to budget. Forty percent of US respondents say privacy is a concern preventing them from using big data analytics, compared to 78 percent of European respondents. This contrast is not surprising when one considers the differing cultural views of privacy in the United States and Europe. When asked whether privacy and security are carefully balanced within their organizations, slightly over half of US respondents answered negatively or that they are unsure compared to one-third in Europe.

Comparing reasons for not deploying big data analytics for cybersecurity—United States versus Europe



PRIVACY

Organizations using big data to gain insights into potential threats must overcome several privacy challenges. Among the 48% of respondents who either do not have a plan to implement big data analytics or are not sure, concerns about privacy and the impact on personal information was identified as one of the top two barriers. More specifically, respondents' statements suggest that an ambiguous regulatory picture is preventing adoption of new big data technologies. This is unsurprising given that big data solutions, especially for security, often require the collection and processing of large data sets which may contain personal information. One respondent stated, "We decided to hold off on all big data analytical solutions until there are clear rules about privacy in the big data environment." Several respondents state that their enterprise's privacy policies that govern the collection and use of data are problematic when it comes to implementing big data in cybersecurity. A manager from the financial sector in the E.U. stated, "I don't see how we would be able to take advantage of big data solutions for cybersecurity with my company's current policies." If big data solutions are able to deliver meaningful security gains one may expect an evolution of privacy policies to accommodate this use case.

In addition, the challenge of obtaining consent from a user prior to using PII in a cybersecurity context is cited as a major barrier. One respondent well-described the challenge stating, "A big data solution for cybersecurity requires some degree of stealth and secrecy. How can this be accomplished with a requirement to collect consent in advance of use?"

> *"We decided to hold off on all big data analytical solutions until there are clear rules about privacy in the big data environment."*
>
> ■ Financial executive, Germany

> *"Intelligence about our organization's IT environment requires us to spot unusual and suspicious behaviors. This type of big data is likely to reveal, at a minimum, some individual identifiers."*
>
> ■ Services executive, United States

COMPLEXITY

Complexity is inherent in the adoption, implementation, and maintenance of big data technologies. About half of the respondents cite complexity as the main barrier to deploying big data analytics to enhance their enterprise's cyberdefense. Respondents do not know how a big data solution would affect their legacy technology environment, and who has the right expertise to manage the new technologies.

Successfully implementing a big data solution does require sophisticated technologies that will store, organize, and further analyze vast and varied data sets. Interoperability among existing data environments and new technologies is contingent upon choosing the right technologies and having the right expertise to implement them. Moreover, enterprises both large and small lack specially trained analysts to design these big data systems and use the results of the analysis.

"We are still living in the dark ages. Our company relies on configuration changes and activity logs to determine unusual behavior."

- Retail executive, United States

"We are in dire need of enterprise solutions that make us smarter about net[work] flows. Frankly, we are not there yet."

- Technology executive, United Kingdom

COST

The increasing stealth and sophistication of cyberattacks can put a strain on even the most generous security budgets, which are already spread thin addressing risk: insecure mobile devices and apps, including personally owned devices and apps entering the workplace (BYOD); non-compliance with regulations; data breaches; social engineering tactics; insider negligence; and use of insecure cloud services.

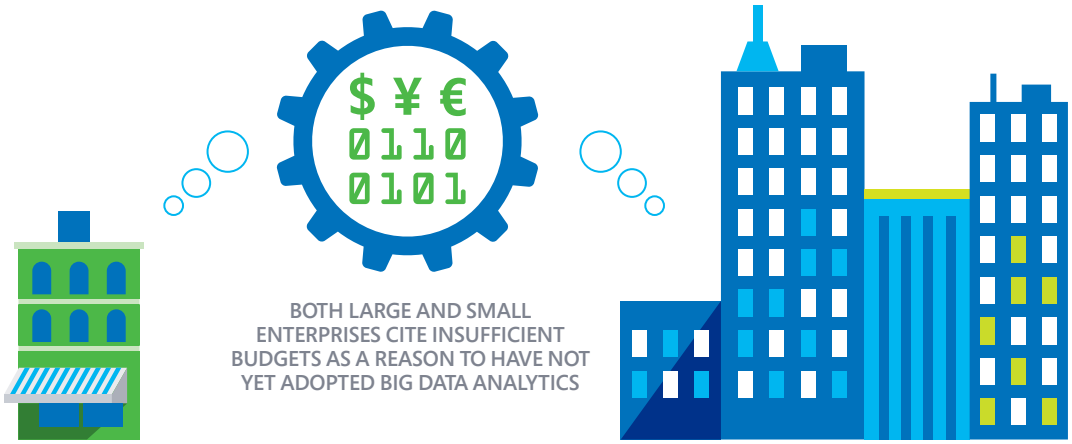
In this survey, almost identical percentages of respondents from both large and small enterprises cite insufficient budgets as a reason they have not yet adopted big data analytics. The various costs of deploying a big data solution can be many, and often include storage, computers, data tools, and data visualization frameworks. However, the emergence of big data solutions offered as cloud services, combined with a growing number of firms offering these services, may be seen as an indication that costs will fall and adoption rates will increase. Beyond the capital costs of adopting big data tools are the opportunity costs. Will purchasing new technologies affect an enterprise's ability to invest in and maintain other technologies crucial to their security, such as firewalls and detection software?

"We know how to better secure our company from security breaches. We have invested in the assessments, audits and consultants to know where are vulnerabilities. But senior leadership won't give us the funds we need."

- Director, IT security, Germany

"Fortunately, our leadership has had a wake-up call and they are easing up on the purse strings. For the first time we are close to having an adequate budget for both security and operations."

- VP legal, Netherlands



BOTH LARGE AND SMALL ENTERPRISES CITE INSUFFICIENT BUDGETS AS A REASON TO HAVE NOT YET ADOPTED BIG DATA ANALYTICS

RECOMMENDATIONS

The explosive use of big data across many different industries has arrived just as they are facing serious cybersecurity threats. Fortunately, big data tools and techniques have the potential to help organizations manage cybersecurity risks. The Ponemon Institute, whose mission is to conduct "empirical studies on critical issues affecting the management and security of sensitive information about people and organizations," offers the following recommendations:

1. **Create an overall strategy** for using big data to enhance cybersecurity that specifies the desired outcomes, stakeholders, and data required for analysis.

2. **Inventory the available internal and external data sources** to understand the volume of data, frequency of updates, and any usage restrictions that may apply.

3. **Define a policy** for how to manage any personal information that may be contained in data sets analyzed by a big data solution.

4. **Evaluate cloud computing technologies** to provide cost-effective capabilities for storing and processing large amounts of data.

5. **Investigate the application of big data tools** along with existing machine-learning algorithms to reduce complexity.

APPENDICES

APPENDIX A: METHODS

A sampling frame of 1,164 senior privacy, compliance, and security executives in the United States and five European countries was selected as interview participants. Out of those, 103 interviews were completed for an 8.8 percent response rate.

TABLE 1.
SAMPLE RESPONSE

NUMBER	PERCENTAGE
TOTAL CONTACTS	1,164	100.0%
CONFIRMED	128	11.0%
COMPLETED	103	8.8%
US SAMPLE	59	5.1%
EUROPE SAMPLE	44	3.8%

TABLE 2.
POSITION WITHIN THE ORGANIZATION

UNITED STATES	EUROPE
SENIOR EXECUTIVE	14%	14%
VICE PRESIDENT	36%	36%
DIRECTOR	34%	36%
MANAGER	12%	11%
SUPERVISOR	3%	2%
OTHER	2%	0%
TOTAL	**100%**	**100%**

TABLE 3.
EUROPEAN COUNTRIES WHERE INTERVIEWS WERE CONDUCTED

EUROPE
UNITED KINGDOM	34%
GERMANY	27%
FRANCE	18%
NETHERLANDS	16%
BELGIUM	5%
TOTAL	**100%**

As shown to the right, 72 percent of US respondents are from organizations with a headcount greater than 1,000. Seventy percent of European respondents are from organizations with a headcount greater than 1,000.

TABLE 4.
ORGANIZATION SIZE

UNITED STATES	EUROPE
FEWER THAN 500	14%	16%
501 TO 1,000	14%	14%
1,001 TO 5,000	22%	27%
5,001 TO 10,000	22%	20%
10,001 TO 25,000	14%	9%
25,001 TO 75,000	8%	9%
MORE THAN 75,000	7%	5%

APPENDIX B: RESEARCH LIMITATIONS

There are inherent limitations to survey research that must be carefully considered before drawing inferences from findings.

- **Non-response bias**. The current findings are based on a sample of survey returns. Surveys were sent to a representative sample of individuals, resulting in a large number of usable returned responses. Despite non-response tests, it is always possible that individuals who did not participate are substantially different in terms of underlying beliefs from those who completed the instrument.
- **Sampling-frame bias**. The accuracy is based on contact information and the degree to which the list is representative of individuals who are senior privacy, compliance and security executives. We also acknowledge that the results may be biased by external events such as media coverage. Finally, because we used a web-based collection method, it is possible that non-web responses by mailed survey or telephone call would result in a different pattern of findings.
- **Self-reported results**. The quality of survey research is based on the integrity of confidential responses received from subjects. While certain checks and balances can be incorporated into the survey process, there is always the possibility that a subject did not provide accurate responses.

If you have questions or comments about this executive summary or would like to receive the full report, please contact us:

Ponemon Institute LLC

Attn: Research Department

2308 US 31 North

Traverse City, Michigan 49686 USA

1.800.877.3118

research@ponemon.org

Ponemon Institute

Advancing Responsible Information Management

Ponemon Institute is dedicated to independent research and education that advances responsible information and privacy management practices within business and government. Our mission is to conduct high quality, empirical studies on critical issues affecting the management and security of sensitive information about people and organizations.

As a member of the Council of American Survey Research Organizations (CASRO), we uphold strict data confidentiality, privacy and ethical research standards. We do not collect any personally identifiable information from individuals (or company identifiable information in our business research). Furthermore, we have strict quality standards to ensure that subjects are not asked extraneous, irrelevant or improper questions.