Verizon Data Breach Investigations Report[1]

by M. E. Kabay, PhD, CISSP-ISSMP[2]

CTO, School of Graduate Studies
Norwich University, Northfield VT

Contents

[1] This paper is a compilation with minor edits of four articles published in *Network World Security Strategies* in June and July 2008.

[2] M. E. Kabay, PhD, CISSP-ISSMP is Program Director of the Master of Science in Information Assurance < http://www.graduate.norwich.edu/infoassurance/ > and CTO of the School of Graduate Studies at Norwich University in Northfield, VT. Mich can be reached by e-mail at < mailto:mekabay@gmail.com >; Web site at < http://www2.norwich.edu/mkabay/index.htm >.

1 A Sound Methodology

As most people realize, all published information about data-security breaches should be examined with critical faculties fully engaged. Studies and statistics about computer crimes consistently suffer from the following methodological problems:

- Limited ascertainment (the crimes may not be detected)
- Restricted reporting (many organizations don't want to report breaches at all and there is no centralized reporting facility to collate the data)
- Non-random samples (it is not possible to generalize from the samples to a wider population because the reports come from self-selected reporting organizations).[3]

The Verizon Business Risk Team recently published a valuable analysis of four years of data on security breaches among their clients entitled "2008 Data Breach Investigations Report."< http://www.verizonbusiness.com/resources/security/databreachreport.pdf > Wade H. Baker, C. David Hylender and J. Andrew Valentine are the authors; their contributors include my old friend and colleague Dr Peter Tippett, MD, PhD, A. Bryan Sartin, Stan S. Kang, Christopher Novak, and members of the Verizon RISK Team.[4]

Brad Reed has pointed out the main findings recently in *Network World* < http://www.networkworld.com/news/2008/061108-vz-breach.html > and the paper itself includes a good executive summary; therefore, in the next few columns, I will elaborate on the implications of specific points from the report.

In this section, I want to draw readers' attention to the methodology of this landmark study. I believe that the study is unique in drawing upon a massive database of more than 500 specific investigations carried out by the Verizon RISK Team over the last four years. As the authors write,

> Furthermore, it contains firsthand information on actual security breaches rather than on network activity, attack signatures, vulnerabilities, public disclosures, and media interpretation that form the basis of most publications in the field. While many reports in the security industry rely on surveys as the primary data collection instrument, this data set is inherently more objective.

Surveys are inherently limited because it is difficult or impossible to determine whether the willingness to participate in the survey is correlated with any particular attributes of the participants; e.g., perhaps those who refuse to participate have worse security than those who participate – or vice versa. We don't know and cannot know based on the survey results.

In contrast, the organizations studied in the Verizon report were clients of the Risk Team (or on an incident response retainer contract) either *before* they had breaches or they were referred to Verizon after the breaches. In either case, the fact that these are known clients increases the reliability of the findings compared with surveys where anonymous respondents can fill in the blanks without verification of their identity. No one is claiming that the sample is a *random* sample that allows

---

3 For more information about these issues, see my paper, "Understanding Computer Crime Studies and Statistics v4."< http://www2.norwich.edu/mkabay/methodology/crime_stats_methods.pdf >

4 Disclaimer: I worked for a company that eventually became the Verizon Business Risk Team many years after I left, but my only involvement today is my continued friendship with several employees who work there.

generalization of the sample results to the universe of all possible corporate victims of security breaches; the authors themselves warn,

> Though challenges such as sampling techniques, response rates, and self-selection are not relevant to the research method used in this study, it cannot be concluded that the findings are therefore unbiased. Perhaps most obvious is that the data set is dependent upon cases which Verizon Business was engaged to investigate. Readers familiar with publicly available statistics on data loss will quickly recognize differences between these sources and the results presented in this report. This has much to do with caseload. For instance, it is simply more likely that an organization will desire a forensic examination following a network intrusion than a lost laptop.

It is refreshing to see a security report written with this degree of statistical awareness.

Most important, the detailed statistics, including causes or methods, numbers of records compromised, types of data involved, time span of events, discovery methods, and estimated costs were based on analysis by trained professionals, not on self-reported, unverified guesswork by anonymous respondents. One of the most serious methodological problems of studies which rely on multiple-choice responses by unknown respondents is that it is difficult to validate the data; presentation of cost-classes, for example, naturally attracts respondents to whichever categories are presented. Checking a box on a form is a lot easier than actually measuring costs or analyzing causes, but the unverified results are of dubious reliability.

Survey-design courses demonstrate many methods for validation of surveys, none of which are ever used in popular security surveys as far as I have seen in over 20 years of study. Examples of internal and external survey-validation techniques include multiple questions in different parts of the survey instrument addressing the same metric using different wording and different scales; repeated administration of the instrument to identifiable individuals to measure intra-respondent variability; and follow-up studies to compare the survey results with data collected independently.[5]

In the next section, I'll look at the implications of a surprising finding: "In a finding that may be surprising to some, most data breaches investigated were caused by external sources."

---

5 For an excellent brief (30-page) introduction to sound survey design, see Dr David S. Walonick's free tutorial.< http://www.statpac.com/surveys/surveys.pdf >

2 Outsider Attacks

In the preceding section, I quoted a surprising finding: "In a finding that may be surprising to some, most data breaches investigated were caused by external sources." Today I want to explore the implications of that finding.

The authors explain their terminology for sources of data breaches:

> Internal threat sources are those originating from within the organization. This encompasses human assets—company executives, employees, and interns as well as other assets such as physical facilities and information systems. Most insiders are trusted to a certain degree and some, IT administrators in particular, have high levels of access and privilege.

The three threat sources used in the study are as follows (quoting with elisions as shown):

- **External**— Intuitively, external threats originate from sources outside the organization. Examples include hackers, organized crime groups, and government entities but also environmental events such as typhoons and earthquakes. Typically, no trust or privilege is implied for external entities.
- **Internal**— Internal threat sources are those originating from within the organization. This encompasses human assets—company executives, employees, and interns as well as other assets such as physical facilities and information systems….
- **Partner**— Partners include any third party sharing a business relationship with the organization. This value chain of partners, vendors, suppliers, contractors, and customers is known as the extended enterprise….

The researchers found that outsiders, not insiders, were responsible for "data compromises" in about three-quarters of the cases studied; "…business partners were involved in 39 percent of the data breaches handled by our investigators. Internal sources accounted for the fewest number of incidents (18 percent), trailing those of external origin by a ratio of four to one." The percentages add up to more than 100 percent because more than one type of source was observed in many breaches.

Speaking personally, I am going to have to rethink my long-held stance – originating in the 1980s – claiming that the bulk of the threats to information systems are internal. I have taught that about half the problems observed in organizations come from errors and omissions, with dishonest and disgruntled employees coming in next and adding up to about three quarters of the cases informally reported by consultants. The Verizon study casts serious doubt on this vague generalization and I will be telling my introductory information assurance students to follow the guidance of my favorite bumper sticker: QUESTION AUTHORITY – in this case, me!

Again, no one is claiming that the results of the Verizon study can be extended to the totality of all security breaches; nevertheless, their results are certainly giving me something to think about. I hope that readers will find the study equally stimulating.

In the next section of this paper, I'll look at the research findings concerning breach size and source.

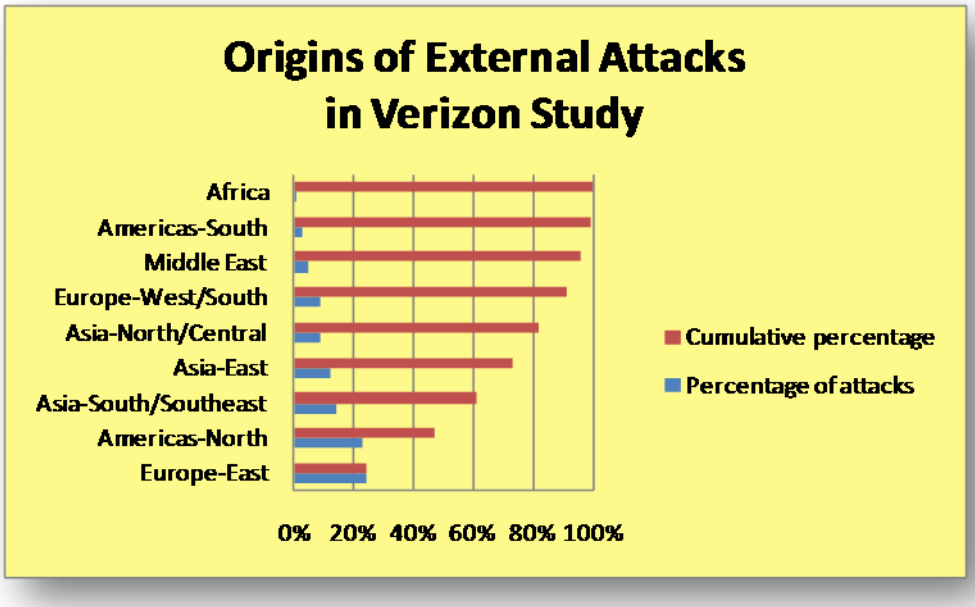
3 Breach Size and Source

The most interesting aspect of the data is that "The median size (as measured in the number of compromised records) for an insider breach exceeded that of an outsider by more than 10 to one. Likewise, incidents involving partners tend to be substantially larger than those caused by external sources." I was pleased to see the authors using the median, not the mean, of the number of records compromised; most of the reports published in our field erroneously use means (arithmetic averages) even though the variables have drastically skewed (asymmetric) frequency distributions that make those averages much less useful than for symmetric distributions.< http://www.npr.org/templates/story/story.php?storyId=5567890 >

When the authors corrected for the number of cases involving external sources, internal sources, and partners, the numbers of records likely to be involved in a breach showed that "partners represent the greatest risk for data compromise, followed closely by insiders." These observations support "the principle that privileged parties are able to do more damage to the organization than outsiders."

Using as much information as they could bring together on the IP addresses of external attacks, the Verizon team found that the geographic distribution of attack origins looked like this (some of these numbers are not shown in the report but were supplied by author Wade Baker for this article):

Zone	%
Europe-East	24%
Americas-North	23%
Asia-South/Southeast	14%
Asia-East	12%
Asia-North/Central (incl. Russia)	9%
Europe-West/South	9%
Middle East	5%
Americas-South	3%
Africa	1%
Europe-North (Scandinavia)	0%
Oceania (Austrialia/NZ)	0%
Americas-Central	0%



So over 80% of the estimated attack-sources are from Eastern Europe, North America, and Asia. These results surprised me, since I have fallen into the habit of thinking of China as the number one source of threats to information security today; I have to correct my impressions and be more careful in my teaching, lecturing and writing.

On the insider front, the analysts found that half the insider attacks involved IT administrators and about 41% involved other non-executive employees. These results are consistent with the long-held view that privileged insiders must be selected with care and consistently monitored as part of any effective security program.

Many breaches in the data set involved breaches mediated through weaknesses in partner systems: "Partner-side information assets and connections were compromised and used by an external entity to attack the victim's systems in 57 percent of breaches involving a business partner. Though not a willing accomplice, the partner's lax security practices—often outside the victim's control—undeniably allow such attacks to take place. Exacerbating this situation, the victim organization frequently lacks measures to provide accountability for partner-facing systems. This contributed to the 21 percent of breaches in which partner involvement was evident but specific persons were not identified."

These findings on attacks via compromised partner systems support the view that it makes sense to insist on external security audits of partner organizations before establishing and while maintain extended business relationships involving electronic data interchange. Consult your attorneys to discuss their views on due diligence in executing fiduciary responsibility to corporate stakeholders.

4 Attack Vectors

In section, I will look at the findings on attack vectors, called "Common Attack Pathways" in the report. The paper provides the following summary data:

Attack Vector	%
Remote Access and Control	42%
Web Application	34%
Internet-Facing System	24%
Physical Access	21%
Wireless Network	9%

The authors comment, "In over 40 percent of the breaches investigated during this study, an attacker gained unauthorized access to the victim via one of the many types of remote access and control software. On many occasions, an account which was intended for use by vendors in order to remotely administer systems was compromised by an external entity. These vendor accounts were then used to illegitimately access enterprise information assets. This scenario is particularly problematic due to the fact that, from the victim's perspective, the attacker appears to be an authorized third party. In many of these cases, the remote access account is configured with default settings, making the attacker's job all too easy."

These findings support the long-established warnings about canonical accounts (i.e., accounts that have the same name and characteristics on all comparable systems). Such accounts are even worse risks when system administrators fail to change the canonical passwords that are often included as part of the installation of specific application or utility software.

One of the interesting counter-intuitive results is the low involvement of wireless networks as an attack vector. "Despite the large amount of media attention given to the supposed weakness of wireless networks, this vector was exploited considerably less than others presented in Figure 17. When wireless infrastructure was the means of entry, it was due to poor configuration and weak encryption rather than a successful attack against an adequately secured WLAN."

The Verizon report is well organized and well written; the language is simple and engaging and never stuffy. The authors make no claims that go beyond the value of their data set and they use reasonable statistical measures to describe their data. I hope that their excellent work will influence others to improve security studies.

This section concludes my look at a few highlights of the Verizon study; I encourage readers to study the findings themselves in more detail.