Global Data Leakage Report, 2016

© InfoWatch Analytical Center, 2017
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of contents</td>
<td>2</td>
</tr>
<tr>
<td>In figures</td>
<td>3</td>
</tr>
<tr>
<td>Summary</td>
<td>4</td>
</tr>
<tr>
<td>Methodology</td>
<td>5</td>
</tr>
<tr>
<td>Research results</td>
<td>7</td>
</tr>
<tr>
<td>Leak channels</td>
<td>12</td>
</tr>
<tr>
<td>Industry-specific map</td>
<td>15</td>
</tr>
<tr>
<td>Regional specifics</td>
<td>20</td>
</tr>
<tr>
<td>Conclusion and findings</td>
<td>21</td>
</tr>
<tr>
<td>Leakage monitoring on InfoWatch’s website</td>
<td>22</td>
</tr>
<tr>
<td>Glossary</td>
<td>23</td>
</tr>
</tbody>
</table>
In figures

✓ In 2016, InfoWatch Analytical Center registered 1,556 data leaks revealed (in the media and other sources) worldwide, which is 3.4% more than in 2015.

✓ External attacks were behind 38% of the data leaks, with a share of externally triggered leaks increasing by six percentage points (p.p.) year over year.

✓ 93% of the leaks jeopardized personal data and payment details, with a total of 3.1+ billion records being compromised over the reporting period (three times more than in 2015).

✓ 44 mega leaks were registered in 2016, each resulting in the loss of no less than 10M personal data items. Mega leaks accounted for 94.6% of all compromised records.

✓ Inside companies, employees were responsible for 36% of the leaks, while executives caused 2.2% of the cases.

✓ In 2016, Russia came in second, recording 213 data leaks from domestic companies and government agencies (+80% YoY).
Summary

This is an annual data leakage report prepared by InfoWatch Analytical Center.

In 2016, we saw a new trend of political hacking or hacktivism, with intruders stealing dozens of millions of data belonging to Philippine, Mexican, Turkish, and US voters. The recent U.S. President election campaign was suffering from regular scandals related to data leaks, which might have even tipped the election.

External attackers compromised data of Facebook, Foursquare, GitHub, iCloud, LinkedIn, MySpace, Snapchat, Telegram, Tumblr, and Twitter users. Indeed, there was hardly a single online service, users of which avoided external and internal attacks in 2016.

Any prominent company will definitely grab the headlines if there is even the slightest hint on leakage due to its fault. This year, such breaking news hit Alibaba, AliExpress, Amazon, American Express, Apple, AT&T, Baidu, BMW, Cisco, Credit Suisse, Dell, Deutsche Telekom, Experian, Google, Huawei, owners of Hyatt and Marriott hotels, KFC, Microsoft, Nokia, Oracle, T-Mobile, Toyota, Uber, Valve, Vodafone, Walmart, and Yahoo.

In 2016, the leakage scale and related financial damage were truly immense. According to FireEye, FIN6 hackers alone compromised 20 million credit cards and thus grasped more than $400 million.

Attacks were also aimed at government agencies, regional administrations, ministries, and even law enforcement authorities and police offices. Data leaks were recorded in the U.S., including the Department of State, Olympic Committee, Internal Revenue Service, and election campaign offices of both Donald Trump and Hillary Clinton.

The growing number of leaks and ever-increasing volumes of compromised data shape a global trend that has nothing to do with regional specifics, being fueled by new opportunities of data handling in the digital world, such as digitalization of services, e-commerce, web money, and digital intellectual property.

The more opportunities we get, the more tempting digital data looks for hackers. Global leakage analysis shows a data compromising threatscape of today: which data channel is the most vulnerable now and why, which industry is the most attractive for intruders, and which hackers are more dangerous (external or internal).

The report authors are confident that the research results will be of interest to information and economic security experts, journalists, company owners and executives of organizations using confidential information (trade, bank, and tax secrets), or other valuable information assets.
Methodology

The report is powered by the InfoWatch Analytical Center's own database that has been nurtured by its experts since 2004. The database contains public information on data leaks from commercial and non-commercial (public, municipal) organizations and government agencies as a result of malicious or negligent actions by employees or other persons. InfoWatch leak database consists of several thousand registered incidents.

Before entering the database, each leak is classified according to several criteria, such as organization size, scope of activity (industry), amount of damage, leak type (by intent), leak channel, types of leaked data, and attack vector.

Moreover, incidents are classified by nature of intruder's actions. In addition to 'ordinary' leaks, the report also covers the following 'qualified' leaks:

— when an authorized officer tampers with payment details and non-public information
— when an employee abuses its access privileges to get data he/she does not need for the job.

According to the authors, the research covers maximum 1% of all estimated leaks due to high latency of incidents associated with compromised data. However, InfoWatch selected the leak classification criteria in such a way that each group of categories could contain sufficient or excessive number of elements (actual data leaks).

---

1 Information about data leaks published by government agencies, mass media, bloggers, as well as Internet forums, and other open sources.

2 Information (data) leak is a loss of control over the information (data) as a result of an external attack or acts by a person who has an authorized access to information or by a person who gained an unauthorized access to such information.

3 Data leaks are divided into intentional (malicious) and unintentional (accidental) depending on whether or not a person is guilty of actions that resulted in a data leak. The terms "intentional" and "malicious" ("unintentional" and "accidental") are equivalents and are used as synonyms herein.

4 Leaks are classified by source (person responsible). Along with malicious insiders, the classification includes external intruders.

5 InfoWatch Analytical Center ranks companies by size depending on the known or estimated number of personal computers (PCs): small companies with up to 50 PCs, medium ones with 50 to 500 PCs, and large ones with over 500 PCs.

6 Information about damage and a number of records compromised is obtained from media publications.

7 "Leak channel" means a certain scenario when cumulative actions by a corporate information system user in relation to hardware or software services result in the loss of control over information or a breach of confidentiality. Leak channels are determined only for those leaks caused by actions of an internal offender.

8 "Attack vector" means a type of intruder's actions behind data leakage, including intruders who attack company's web resources and information infrastructure from the outside and insiders who obtain unauthorized access to resources and misuse confidential information, etc.
This approach to survey fielding allows having theoretical sample, with the findings and
trends identified in the sample being representative for the entire assembly.

When preparing diagrams (breakdowns), we excluded from the sample those leaks, which
were not defined according to the breakdown criteria. For example, breakdown by attack
vector (external threats, malicious insider’s activities) does not contain leaks with a non-defined
vector. The same is true for breakdowns by person responsible, intent, and other criteria.

The report authors deliberately excluded both too large data leaks (over 10 million personal
data records) and too small ones (less than 100 records) from the industry-specific map and
diagrams provided in the 'Industry-specific Map' section in order to avoid any
misrepresentations. The use of a limited sample for diagrams in the said section is expressly
specified.

In addition, the sample did not include breach of confidentiality, other cybersecurity incidents
(such as DDoS attacks) that did not result in data leaks, or leaks from an unclear data source
(when it is impossible to determine an organization which owned the compromised data).

The authors did not have a goal of either finding the exact number of data leaks, or
evaluating actual or potential damage caused to organizations. The report is focused on
identifying trends and pace in the global, industry-specific, and regional data leakage maps.
Research results

In 2016, InfoWatch Analytical Center registered 1,556 confidential data leaks (see Fig. 1) compromising over 3.1 billion personal data records, such as social security numbers, bank card details, and other critical information.

![Figure 1. Number of data leaks and volume of personal data compromised, 2011 - 2016](image)

In 2016, leak growth rate slowed down (Fig. 2) coming to 3.4% vs. 7.9% in 2015.

![Figure 2. Number of registered data leaks, 2006 - 2016](image)

In 2016, for the first time during all reporting years, we witnessed a three-fold increase in both the volume of data compromised by leaks and the number of personal data records compromised per leak (see Fig. 3).
Soaring volumes of compromised data do not result from just one or two big cases; otherwise we could call it a sporadic growth. In fact, we registered 79 leaks, each compromising over 1 million records.

Thus, the leakage landscape changed dramatically in 2016, opening the era of mass data compromising as ever-increasing number and maturity of data leaks are currently making the difference.

In 2016, we registered 540 (38.2%) data leaks by external intruders and 873 (61.8%) by internal offenders (see Fig. 4).

^9"Attack vector" means a type of intruder’s actions behind data leakage, including hackers who attack company’s web resources and information infrastructure from the outside and insiders who obtain unauthorized access to classified resources and confidential information, etc.
In 2016, the share of data leaks pushed by external attacks increased by over 6 p.p. (from 32% in 2015), thus compromising 2.53 billion personal data records or 80% of the total records compromised this year and leading to 34 of 44 registered mega leaks

**databreaches.net**: External hackers stole personal and financial information of 203 million customers of Experian, including names, dates of birth, addresses, phone numbers, and other data, and put the database dump for sale at BTC 0.8082 ($640 at the current exchange rate).

Data leaks caused by external attacks are notable for greater volume of compromised data, with an average of 4.69 and 0.56 million personal data records being compromised per external attack and insider's intent or negligence, respectively. As a rule, external attackers try to take out all they can reach.

**Motherboard**: A hacker has downloaded over 200 GB of data from a Department of Justice (DoJ) server, including names, email addresses and phone numbers of 20,000 FBI employees and 10,000 Department of Homeland Security (DHS) employees. The data was obtained, the hacker told Motherboard, by first compromising the email account of a DoJ employee. The hacker then called FBI and told an operator he was a new user and did not understand how to access DoJ infrastructure.

It is worth noting that external leaks provoke an interesting 'information effect'. Any organization suffering from a big leak has to make a public statement. However, PR teams are not always able to address the challenge with flying colors, and some leaks may unhorse CEOs and government officials.

**Thesmokinggun.com**: A hacker under the nickname Guccifer 2.0 stole personal data of democratic members of the U.S. House of Representatives and leaked it as an Excel file via Twitter. Previously, he/she had also taken credit for hacking the network of the Democratic National Committee (DNC) and exposing approximately 20,000 emails on WikiLeaks, with DNC’s chairperson Debbie Wasserman Schultz supposedly having to resign after that.

However, internal leaks are hardly less destructive than external; just consequences for a victim are different.

While external leaks are similar to carpet bombing, internal leaks remind of spot bombing where mission-critical information is at stake and financial damage is almost unlimited, sometimes reaching the total value of victim’s business.

**bloomberg.com**: UK online gaming operators 888 Holdings Plc and Rank Group Plc called off the acquisition of William Hill, another major UK bookmaker, worth $4.1 billion due to a data leak.

---

10 “Mega leaks” are leaks compromising 10+ million personal data records.
Specific nature of internal leaks draws attention to privileged users, such as top managers, system administrators, and other employees, including information security officers themselves, who have almost unlimited data access. Such high-ranking employees are hard to control, whereas consequences of their mistakes or malicious actions can be easily compared to natural disasters.

*Hitsniffer, a web analytics company, declared suspension of its operations due to customer base theft by its former employee.*

In 2016, current (33.9%) and former (2.1%) employees caused 36% of the data leaks. In more than 2% of the cases, company executives (top managers, heads of departments and business units) and system administrators were at fault, while 6% of the leaks were caused by contractor’s personnel who was authorized to access sensitive data (see Fig. 5).

*Figure 5. Leaks by source (person responsible), 2016*
Personal data and payment details accounted for similar shares in the leakage breakdown by information type as in the previous years (93%) (see Fig. 6).

![Figure 6. Leaks by data type, 2016](image)

In 2016, the share of leaks followed by data fraud (bank fraud) went down by 3.3 p. p. to 7%.

![Figure 7. Incidents by nature, 2016](image)

10% of the incidents are classified as violations associated with unauthorized access to information (access rights abuse, manipulation of data that personnel does not need for the job).

**Conclusion**

The triple growth in the volume of compromised data indicates ever-increasing value of digital information. While the growth in the number of leaks is difficult to forecast (it may even stop), the volume of compromised data is most likely to keep increasing every year.
Leak channels

In 2016, there were fewer leaks through 'Equipment Loss' (-4.8 p.p.), 'E-mail' (-1 p.p.), and 'Paper Documents' (-7 p.p.), while leaks through removable media and mobile devices remained the same as in 2015. However, the 'Network' channel share increased by 11.6 p.p. (see Fig. 8).

For unintentional leaks by employees, the most significant channels were e-mail (23.7%), paper documents (16.2%), removable media (6.7%), and equipment theft/loss (5.3%), with 47.1% of all recorded unintentional leaks going through the network channel.

Figure 8.Leaks by channel, 2015 – 2016

Figure 9. Accidental and intentional leaks, 2016
Malicious leaks prefer the network channel, with more than 90% of the cases involving unauthorized information transfer or disclosure via the Internet (including web services, e-mail, and other online resources).

Figure 10. Personal data leaks by channel, 2016

The network channel prevails in both the number of leaks and volume of compromised personal data, accounting for more than 66% of such incidents (see Fig. 10).

Internal offenders chose a variety of scenarios such as uploading confidential information to Box, OneDrive, and other clouds, as well as using free webmail.

External break-ins are less varied. As a rule, hackers are not familiar with corporate data structure and do not know where to find the most valuable information and how it looks like, and thus they steal anything of value:

- **gazeta.ru**: Upon hacking of the Philippine Election Commission’s website, LulzSec Philippines stole and posted online personal data of 55 million Philippine citizens, including names, dates of birth, passport details, and even fingerprints. Previously, Anonymous Philippines had hacked the same website allegedly to attract public attention to possible violations during the election.

Network became the most popular channel for both malicious and unintentional leaks, being used to transmit mission-critical data and huge volumes of compromised information.
**RIA Novosti:** French naval contractor DCNS has suffered a hacker attack resulting in the leak of 22,400 pages outlining secret details of six Scorpene submarines that DCNS designed for the Indian Navy, which raised doubts about the security of the company's other project for building 12 submarines ordered by Australian Navy. The data leaked to The Australian newspaper and, possibly, third parties included the descriptions of communications, combat control, navigation, torpedo firing systems, and underwater sensors.

Offenders use mobile devices, removable media, e-mail, and paper documents to obtain data less and less. Advanced intruders know that the cutting-edge control tools can intercept the transfer of confidential information through these channels, and so they don't risk being caught.

**Conclusion**

The fact that network is now the main channel for accidental and intentional leaks shows its growing importance for business. An abundance of web-based communication services and annually increasing number of human errors expand a share of accidental leaks when distributing, posting data over the Internet, etc.

On the other hand, offenders less often use knowingly controlled data transfer channels, such as e-mail and instant messengers. In this context, network remains a channel where control and protection tools surpass the capabilities of offenders.
Industry-specific map

Leak breakdown by organization type has not changed dramatically since 2015 (see Fig. 11).

The largest number of registered leaks came from the healthcare sector (25.8%), while the smallest share is attributed to the manufacturing and transportation sectors (3.9%). High-tech companies, especially large online services and web stores, accounted for 73.6% of all data records compromised in 2016, leaving 11.9% to retailers, hotels, and restaurants and 9.9% to government and municipal authorities (see Fig. 12).

Figure 11. Leaks by organization type, 2015 – 2016

Figure 12. Number of leaks and compromised personal data volume by industry, 2016
These diagrams show actual leakage landscape and the volume of compromised data by industry, while it is more important to understand what sectors are the magnet for cybercriminals.

The attractiveness of industry directly depends on data liquidity in this sector\(^1\). Industry may be less attractive if hackers believe its data protection is sophisticated enough. To put it simply, here is a formula:

\[
\text{Number of intentional leaks} \leftarrow \frac{\text{Data liquidity}}{\text{Alleged information security maturity}}
\]

The number of intentional leaks in a particular industry shows how attractive such industry is. Having seeing how malicious leaks of one type are broken down by industry, we identify the most attractive and, therefore, the most vulnerable segments.

In 2016, retail and transportation companies once again were the most attractive. However, this year they were joined by the finance sector.

securitylab.ru: Hackers stole personal data of almost six million customers of Three UK, a British largest telecom provider, including names, dates of birth, phone numbers, and addresses.

More than a half of the leaks compromising personal data were of malicious nature in the above industries (see Fig. 13).

\(^{11}\) The easier stolen information is to monetize, the more attractive the segment is.
If we look at these industries in terms of attack vectors, then we'll see how attractive each of them is for both external intruders and insiders (see Fig. 14).

The diagram shows that high-tech and transportation companies and banks were the most frequent victims of data leaks caused by external offenders, with a small share of personal data stolen by insiders.

At the same time, healthcare institutions and banks more often suffered from insiders' malicious actions, mainly because such information is very easy to sell.

**databreaches.net**: The health insurer database with 9.3M entries has been hacked and put for sale at the price of BTC 750, which is almost $500,000.00. The data compromised included addresses, social security and mobile phone numbers.

The industry-specific map shows a more comprehensive picture of leaks (see Fig. 15).
A bubble size represents a total volume of compromised records by all segment companies (in millions of personal data records), while vertical axis shows a number of leaks by industry. The map is divided into three diagrams depending on the size of an affected company (small, medium, and large).

Figure 15. Industry-specific personal data leakage map, million, 2016

Industry-specific leakage includes personal data leaks where the exact number of compromised data is known. However, the compromised data volume calculated for the industry excludes mega leaks, i.e. incidents with over 10 million records being compromised.
High-tech companies (online services, digital service providers, and mobile operators) account for a major share of compromised personal data and pioneer the use of digital technologies for data processing and storage. Here, the largest volume of compromised data is just an inevitable evil, which accompanies any advanced data management approach.

In 2016, medium (up to 500 PCs) and large (500+ PCs) companies were almost equal in both the number of leaks and compromised data volumes (see Fig. 16).

![Number of leaks and Data volume](image)

**Figure 16. Leaks by organization size, 2016**

Large companies accounted for slightly more leaks but lost less data per leak than medium-sized businesses, probably, due to insufficient cybersecurity budgets in SMB.

**Conclusion**

The industry-specific data leakage map is not likely to drastically change in the near future, as both factors affecting leakage growth by industry (i.e. data liquidity and security) are very stable.

For a long time, banks and insurance companies have been considered to process the most liquid data, and people still believe that some industries are better protected than the others. As a result, our idea and understanding of today’s most attractive and thus most vulnerable industries do not change much.

Having this understanding and actual leakage landscape, we are positive about more frequent and powerful external attacks to be initiated against hi-tech, healthcare, medical insurance, and finance organizations in the future.
Regional specifics

In terms of geography, the US traditionally ranked first with 838 leakage incidents or 57% of the total volume of leaks in 2016. Russia has retained the second position with 213 leaks, almost twice as many as in 2015.

![Graph showing leaks by country, 2016](image)

*Figure 17. Leaks by country, 2016*

The research authors have already noted that, despite slight local differences, a modern global data leakage landscape is almost the same for all countries where information is handled electronically. Differences between regions and countries are mostly due to peculiar mindset, attitude to data leakage, evaluation of its implications, possible damage, and threats. Thus, more and more leakage-related publications appear in the media of Indonesia, Vietnam, and India, which have never been included in our sample before.

*economictimes.indiatimes.com:* The payment details of 3.2 million customers in 19 Indian banks were compromised during the largest nation-wide leakage, journalists say.

If people enjoy e-government and other online services, in a quick and convenient manner, where electronically handled personal data is now used instead of paper documents, then their countries are exposed to the risk of such data misuse. For example, an identity theft has already become a U.S. common crime, which is committed not by qualified hackers but assistant nurses, waiters, and police officers—regular folks who simply want to make some extra cash with other people’s data.

The situation is slightly different in less digital regions: even though they suffer from much fewer leaks, the scale and nature of attacks can be compared with the most sophisticated leaks hitting Western countries.
Conclusion and findings

Over the last two-three years, we witnessed a dramatic change in the leakage landscape. Today's compromised data volume is boosted by external leaks alone (over 94% of all cases) and thus tripled in 2016.

However, the above case studies show that even a single leaked document can jeopardize the existence and business of the company. C-level managers and other privileged users are usually at fault here, with a share of such leaks significantly increasing in 2016.

The network channel prevails when it comes to intentional leaks, while also increasing its share in accidental leaks every year. Such trend is mainly due to the evolution of communication services opening new ways to process, store, and transfer information electronically. In addition, offenders stopped using controlled data channels (e-mail and removable media) long time ago.

Remarkably, medium and large companies have changed their shares in the leakage landscape. Although mid-sized business used to be a piece of cake for offenders, and currently it is still more vulnerable to outside attacks and, generally, is not as good in combating modern cybersecurity threats, we do not see any noticeable gap between medium and large companies anymore.

Apparently, medium business started to pay more attention to the security of their own information and customer data.

However, despite all positive trends, the general conclusion is discouraging as we are entering the era of mass data compromising. To paraphrase a well-known expression, it is not about which company can be hacked but when it happens. Therefore, business owners and cybersecurity officers are to realize what should be protected as the most valuable asset, against whom and how, because protection of everything from all threats is hardly possible anymore.
Leakage monitoring on InfoWatch's website

InfoWatch Analytical Center regularly posts data leakage reports on its website, as well as the most notorious incidents commented by InfoWatch experts.

Follow the leakage news, new reports, analytical and popular articles via our channels:

- Emailing
- Facebook
- Twitter

InfoWatch Analytical Center
http://www.infowatch.com/analytics
Glossary

**Cybersecurity incidents** in this research mean restricted information compromising through leaks and/or destructive actions by company employees.

Data leak is a loss of control over the information (data) as a result of an external attack or actions by a person who has an authorized access to information or by a person who gained an unauthorized access to such information.

**Destructive actions by employees** mean staff actions resulted in restricted information compromising, such as: use of confidential information for personal needs associated with fraud; illegal access to information (access rights abuse).

**Confidential information** in this context means information which can be accessed by clearly identified and limited set of persons subject to its non-disclosure to third parties without consent of the information owner. In this report, Confidential Information also includes personal data.

**Intentional/Accidental leaks. Intentional leaks** mean information leakages when a user, who works with information, could foresee negative implications of his or her actions, knew about their illegal nature, was warned about liability, and acted from mercenary motives and for personal benefit. This results in a risk of losing control over the information and/or committing a confidentiality breach. In this case, it does not matter whether user's actions actually led to negative consequences or corporate losses.

**Accidental leaks** mean information leakages when a user neither foresees negative implications of his/her actions, nor acts for personal benefit. In this case, it does not matter whether user's actions actually led to negative consequences or corporate losses. The terms "intentional" and "malicious" ("unintentional" and "accidental") are equivalents and are used as synonyms herein.

Attack vector means a type of intruder's actions behind data leakage, including intruders who attack company's web resources and information infrastructure from the outside (External Attackers) and insiders who obtain unauthorized access to classified resources and misuse confidential information, etc. (Internal Offenders).

**Data Channel** means a scenario resulting in the loss of control over information or a breach of confidentiality. Currently, we distinguish eight independent channels:

- Theft/loss of equipment (server, data storage system, laptop, PC), with the information being compromised during the equipment maintenance or due to its loss
- Mobile devices where data leakage occurs because of unauthorized use or theft of a mobile device (smartphone, tablet), which is used within the BYOD (Bring-Your-Own-Device) paradigm.
- Removable media where CDs and flash drives are lost/stolen.
- Network where data is leaked via a browser (sending data to personal email, filling in browser forms); unauthorized use of intranet resources, FTPs, and cloud services; and unauthorized information posting on a website.
- E-mail where data is leaked via corporate e-mail.
✓ Paper documents which can cause a data leakage in the case of improper storage, utilization or printing (confidential information printing out, stealing or taking out).

✓ Instant messengers (data leakage via voice, chat, and video communication).

✓ Non-defined where a media publication about an incident does not reveal a specific leak channel.