A Study on Global Data Leaks in 2018

Conducted by InfoWatch Analytics Center

InfoWatch Analytics Center
https://www.infowatch.com/analytics

Table of Contents

In Figures

- ✓ In 2018, InfoWatch Analytics Center registered **2,263** data leaks reported in the media and other sources worldwide, which is **6%** more than in 2017.
- ✓ **63.5%** of the leaks were caused by insiders. The share of intentional and accidental leaks caused by authorized users (insiders) went up by 3% compared to 2017.
- ✓ External attacks were behind the breach of **4.1 billion** records, which is **59.9%** of all records leaked in 2018.
- ✓ Personal and payment information make up **86%** of all leaks. In total, more than **7.3 billion** data records were leaked in the reporting period.
- ✓ In 2018 there were **47** mega leaks, each exposing at least **10** million personal data records. The mega leaks harvested **97.5%** of all leaked records. Compared to 2017, the number of mega leaks increased by 20%.
- ✓ Employees were accountable for **53%** of the leaks, while executives caused **3.2%** of the cases.
- ✓ 2018 is the first reporting period to show an almost double decrease in the volume of leaked data and a significant, almost double reduction in the volume of records per leak compared to 2017.

Summary

This is an annual report on confidential information leaks prepared by InfoWatch Analytics Center.

2018 saw a series of notorious leaks. When reporting data leaks, the media are increasingly focusing not only on the 'household' names of affected companies, but also on the specific figures, such as the volume of leaked data and the financial losses.

In 2018, the biggest damage caused by a single incident was $534 million as reported by Japanese cryptocurrency exchange Coincheck that had the data of its customers' digital wallets leaked.

The UK fined Facebook £500,000 for user data misuse. The Italian Competition Authority gave Facebook two fines for a total of €10 million, and Ireland may add to that a staggering amount of approximately $1.6 billion.

Uber will pay $148 million, which is the biggest data leak payout in history, to settle claims related to the breach of personal data of 57 million users and drivers, with 25 million of them being US citizens.

> *polit.ru: Phone numbers and other personal data of several British cabinet members and MPs leaked through a mobile app flaw. The breach is being investigated. The developer may get up to £20 million in fines.*

In 2018 data leak payouts awarded in courts were also at a record high. A court in Texas, USA, ordered insurance provider Amrock to pay $740 million to their rival HouseCanary for trade secret theft. A Connecticut jury awarded a patient over $853,000 in a judgment against the Avery Center for Obstetrics and Gynecology of Westport for releasing her medical records.

The EU General Data Protection Regulation came in the spring of 2018 with sanctions for breach of provisions on the protection of EU citizens' personal data. First fines have already been issued. A company in Austria was fined €4,800 for installing a CCTV camera in front of the establishment that also recorded a large part of the sidewalk. The first GDPR sanction in Germany fined social network Knuddels.de for leakage of data of 1.8 million users. It was discovered that the network did not take the necessary precautions to protect user personal data and stored passwords in plain text, so hackers were able to steal user data from the company's servers. Knuddles will pay €20,000 for the breach. In Portugal, the Central Hospital Barreiro Montijo (CHBM) was issued a fine of €400,000 for patient data access infringements.

Among businesses, massive data leaks were reported for software developer Veeam (440 million records), marketing firm Exactis (340 million records), logistics service provider SF Express (300 million records), sales engagement startup Apollo (200 million records), tech company VNG (around 163 million records) and Under Armour app (150 million records).

Some data leaks have a direct effect on business, as illustrated by Google+. After data of 52 million users were leaked, Google accelerated closure of Google+ social network by four months.

The study compares the general pattern of leaks with respect to various factors on a YoY basis. This allows identifying trends in individual parameters, to make reasonable assumptions about the factors of the greater picture and to predict trends in individual parameters and the general pattern for the next year.

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

Methodology

The report is powered by the InfoWatch Analytical Center's proprietary database updated and maintained by its experts since 2004. The database aggregates publicly available cases[1] of data leaks[2] which hit business, non-profit (public, municipal) organizations and public bodies and resulted from intentional or negligent actions[3] of employees or other parties[4]. The InfoWatch leak database comprises several thousand registered incidents.

Each leak being logged into the database is classified according to several criteria, such as organization size[5], the field of activity (industry), damage size[6], leak type (by intent), leak channel[7], type of leaked data and attack vector[8].

Incidents are also classified by the nature of the violator's actions. Along with ordinary leaks, the authors also point out 'qualified' leaks when either officers abuse their authorized data access to payment information, insider information, etc., or employees get data beyond a need-to-know basis and their access rights.

The said criteria are expressed as numbers to describe the general pattern of data leaks for a year, to compare parameters from different years, to identify any unobvious dependencies to pinpoint factors of the greater picture and predict short-term and long-term trends in data leaks and data protection.

The authors believe the study covers a maximum of 1% of all assumed leaks due to the extremely concealed nature of incidents involving data compromising. However, InfoWatch selected leak classification criteria in such a way that each category group contained sufficient or excessive number of items (actual data leaks). This approach to determining the size of the studied field allows theoretical sampling, with the findings and trends identified in the sample being representative of the general pattern.

When preparing charts (breakdowns), we excluded leaks that remained Non-defined[9] according to the key breakdown criteria from the sample.

In the 'Industry Map' section, we also deliberately excluded both leaks with an inadequately large volume of personal data leaked (over 10 million records) and tiny incidents (below 100 records) from the industry-specific map and charts to avoid any misrepresentations. The use of a limited sample for charts is expressly specified.

In addition, the sampling does not include confidentiality breaches and other information security incidents (such as DDoS attacks) not followed by any data leaks or leaks from an unclear data source (where the owner of leaked data cannot be identified).

The authors did not have a goal of either finding the exact number of data leaks or estimating actual or potential financial damage. The report is aimed at identifying trends and pace in the global, industry-specific and regional data leakage landscape.

[1] Data leaks reported by government agencies, mass media, bloggers as well as message boards and other open sources.
[2] Information (data) leak means losing control over information (data) due to an external intrusion (attack), access abuse or unauthorized access.
[3] Data leaks are divided into intentional (malicious) and unintentional (accidental) depending on whether or not there is an intent to cause a data leak. The terms 'intentional/malicious' and 'unintentional/accidental' are equal and used as synonyms here.
[4] Leaks are classified by the source, with both malicious insiders and external intruders included.
[5] InfoWatch Analytical Center ranks organizations by size depending on the known or estimated number of personal computers (PCs) installed: small companies with up to 50 PCs; medium-size, with 50 to 500 PCs; and large, with over 500 PCs.
[6] Information about damage and the number of leaked records is obtained from mass media articles.
[7] Leak channel means a certain scenario (actions of a corporate information system user in relation to hardware or software services) that results in the loss of control over information or a breach of its confidentiality. Leak channels are determined only for leaks caused by insiders.
[8] Attack vector means the direction of an attack, including intruders who attack corporate web assets and IT infrastructure from the outside to compromise data and insiders who obtain unauthorized access to restricted resources, misuse confidential information, etc.
[9] For example, the breakdown by attack vector (external attacks and insider threats) does not contain leaks with a non-defined vector. The same goes for breakdowns by source, intent and other criteria.

Research Findings

In 2018, InfoWatch Analytics Center registered 2,263 data leaks (see Figure 1). In total, more than 7.28 billion personal data records were leaked, such as social security numbers, payment information, and other critical data.

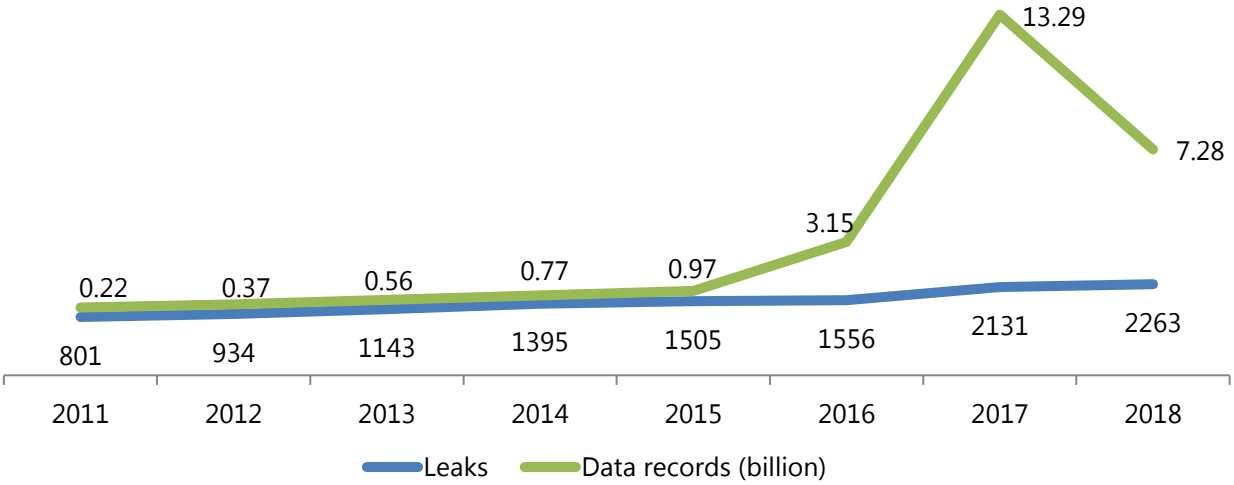
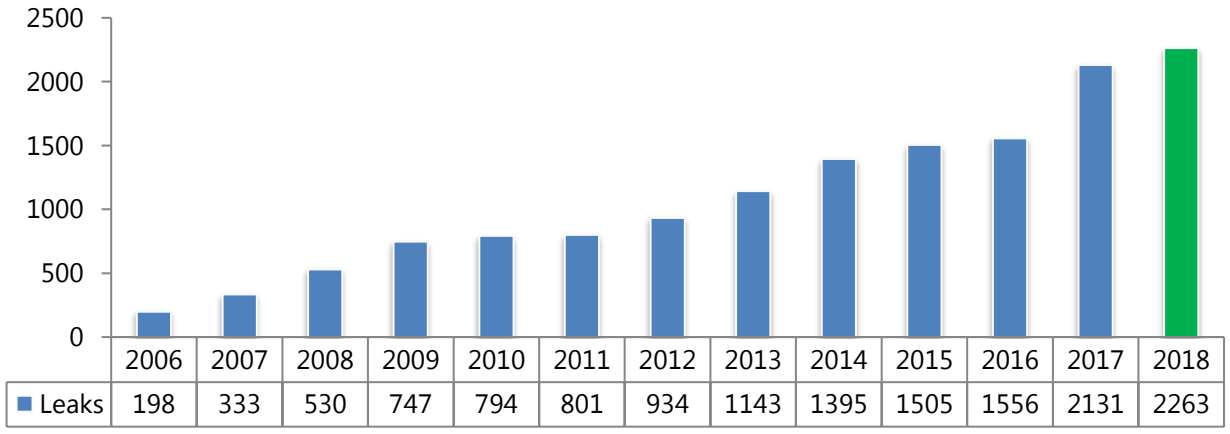


Figure 1. Number of leaks and volume of leaked personal data records in 2011-2018

In 2018, there was a moderate increase in the number of leaks. In 2017, the number of data leaks increased by 36.9% compared to the previous year. In 2018, there were only 6.1% more data leaks.



	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Leaks	198	333	530	747	794	801	934	1143	1395	1505	1556	2131	2263

Figure 2. Number of registered leaks, 2006-2018

In 2018, we registered 81 major leaks (which is 4 leaks less than in 2017). Each of the major leaks exposed over one million data records. In 47 cases, there were more than ten million records compromised, and the number of such 'mega leaks' increased by 20% compared to 2017. Mega leaks exposed 7.1 billion records, which is 97.5% of all records leaked in 2018.

It was the first reporting period to show an almost double (by 45.2%) decrease in the volume of leaked data and an equally significant, almost double (by 48.5%) reduction in 'leak capacity' (i.e. a number of records per leak). In 2018 the average number of records per leak was 3.22 million records (see Figure 3).

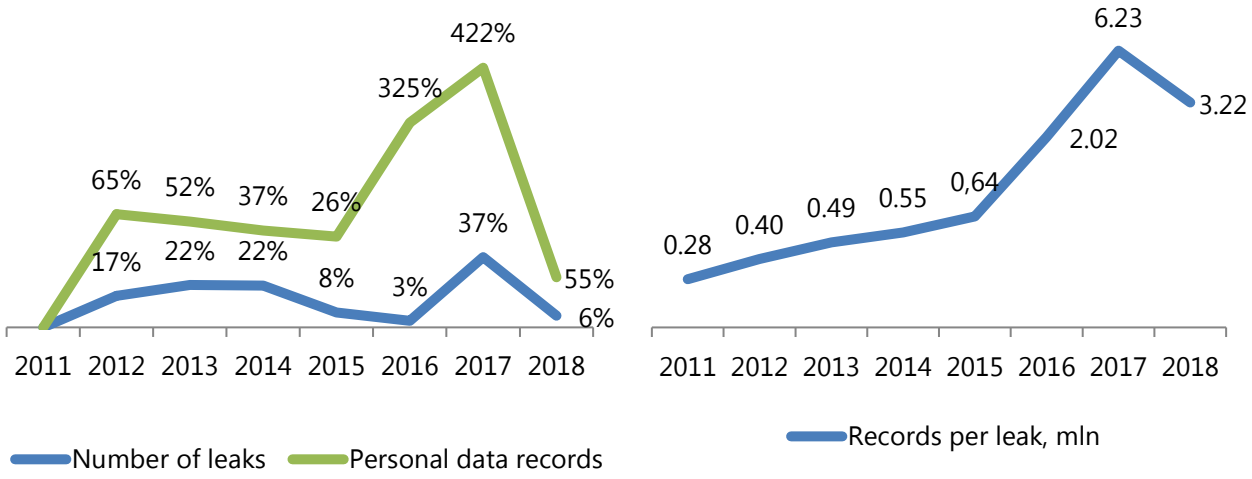


Figure 3. Trends in data leaks and volume of records. Personal data records per leak. 2011-2018

In 2018 intruders were behind 802 (36.5%) registered data leaks, and 1,393 (63.5%) data leaks were caused by insiders (see Figure 4.)

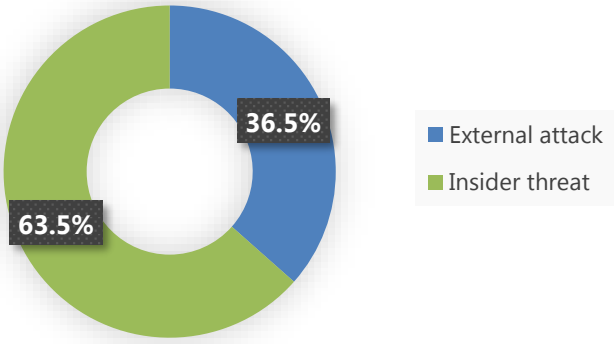


Figure 4. Breakdown by attack vector[10] in 2018

An average external leak exposes 5.15 million records, while an internal leak affects an average of 1.98 million records.

External leaks, therefore, are potentially the most damaging. External attacks were behind leakage of 4.1 billion records, which is 59.9% of all records leaked in 2018. 18 out of 47 registered mega leaks were caused by external attacks.[11]

> *tribuneindia.com: perpetrators used WhatsApp to sell the access key to Aadhaar identification number database for 500 rupees (less than $8). The Aadhaar database is considered the largest repository of resident information. It stores names, addresses, phones, e-mails, and other personal data as well as biometric information of more than a billion Indian residents.*

[10] Attack vector means the direction of an attack, including intruders who attack corporate web assets and IT infrastructure from the outside to compromise data and insiders' actions or failure to act. Leaks caused by insiders can be accidental or intentional, while external attacks can only be intentional.

[11] Mega leaks are leaks resulting in the loss of over 10 million personal data records.

Insider leaks caused by employee error and misconfigured corporate network resources are the closest to external leaks. As a result, huge volumes of information are made publicly available.

> bleepingcomputer.com: *A misconfigured web server exposed taxpayer ID data of 120 million Brazilian taxpayers. The data were leaked because someone had renamed the default index.html file to index.html_bkp. Files with CPF numbers, personal data, military information, phones and addresses of Brazilians were made publicly available.*

In 2018, current (50.5%) or former (2%) employees were the perpetrators in 53% of all cases. In more than 4% of the cases executives (top managers, heads of departments and divisions) and system administrators, (the so-called 'privileged' users) were behind data leaks (see Figure 5).

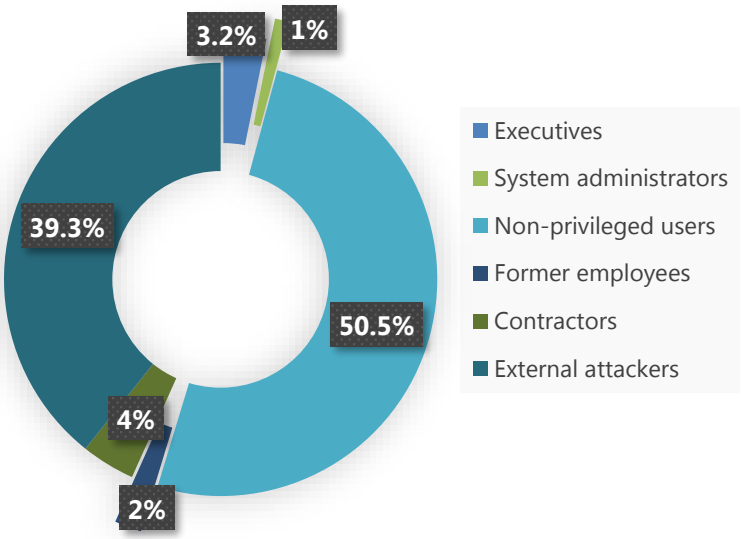


Figure 5. Breakdown by source (the responsible person) in 2018

Contractors whose personnel had authorized access to restricted data caused another 4% of the leaks.

> wired.co.uk: *The Brewhouse & Kitchen pub chain left personal data of 17,000 customers available to anyone through the fault of the contracted Wi-Fi provider. The leaked data included names, dates of birth, e-mails and phone numbers of the patrons.*

The relatively high share of executives and system administrators in the breakdown by source is largely due to their virtually unlimited access privileges to corporate information resources. This is in itself a push towards the misuse of corporate restricted data. The last thing that triggers an unscrupulous employee is the relative ease of cashing in on the stolen data. The most obvious way is to sell the stolen data to the nearest business rival of their employer.

In 2018, we registered a significant number of leaks that followed the above scenario. The fact that in the past year the media widely covered not only the leaks themselves, but also major lawsuits initiated by companies against their employees and against companies that bought the stolen data is telling, at least, of the willingness of businesses to protect their data assets both via technology and via legal tools.

> 5newsonline.com: *A Walmart employee who had worked as a buyer was found guilty of stealing trade secrets. It is reported that shortly before he left his position he collected all possible information about the goods sold by the retailer network, including prices for items bought and sold at Walmart, trade margins and contact details of suppliers. He brought this information to his next employer which is a Walmart vendor, but also a direct competition to Walmart in several product types because it sells similar products online.*

We should also mention that in the East the theft of trade secrets or state secrets (it is difficult to draw the line between the two in some cases) can cost a high-ranking employee their life.

> *scmp.com: Sun Bo, deputy head of China Shipbuilding Industry Corporation, was expelled from the Communist Party of China and sacked for "serious violations of party discipline and causing great damage to the national interest". At least three sources familiar with the case said investigators were looking into allegations that he had passed on confidential information about China's aircraft carrier to foreign intelligence agents. It is unclear what level of confidential information Sun may have given to foreign intelligence agents, but the sources said he could face the death penalty. The punishment depends on the importance of the leaked information.*

The share of leaks of personal and payment information in the breakdown by data type has not changed significantly compared to 2017 and amounted to 86.5%. Trade and state secret leaks were 8.1% and 5.4% respectively (see Figure 6).

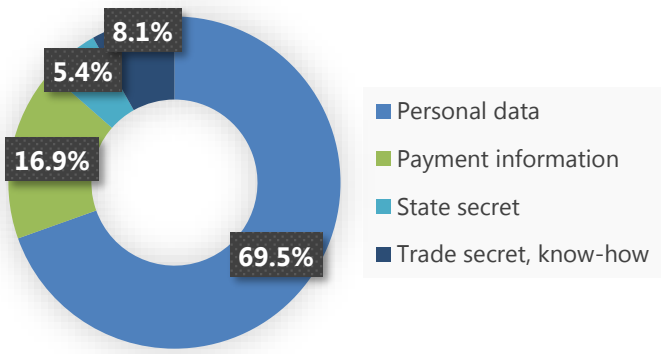


Figure 6. Breakdown by data type in 2018

It is worth mentioning that in the reporting period the difference between leaks of personal and payment information and leaks of other data types was manifest. The consequences of a leaked trade secret are usually easy to predict. The company will suffer direct and indirect financial losses which sometimes amount to dozens or even hundreds of millions of dollars.

However, the personal data situation is radically different. Businesses do not create such data — they only process it. Leaked personal data are not destroyed, and in most cases, there is hardly any actual damage to the processor. Yet, there lies the trick: any major personal data leak results in totally unpredictable losses for the business due to the outflow of customers, falling share prices and fines by regulators for non-compliance.

> *nytimes.com: Personal data of 87 million Facebook users were collected for research purposes and then handed over to Cambridge Analytica, the company that worked on the US President Trump's campaign. After the media learned about the incident, the social network came under unprecedented pressure from the public. Cambridge Analytica management was suspended. In a week, Facebook's share prices on the New York Stock Exchange were knocked down from $177.01 to $159.39. The network's market cap fell by more than $58 billion. Facebook lost a number of major advertisers, among which were Commerzbank and Mozilla.*

In 2018 unskilled (regular) data leaks, i.e. leaks that do not involve further use of leaked data for fraud or personal gain and leaks that are not related to abuse of access rights, made up 83.9% of all incidents. The share of leaks that involved further use of leaked data for fraud (usually bank fraud) is 8.5% (see Figure 7).

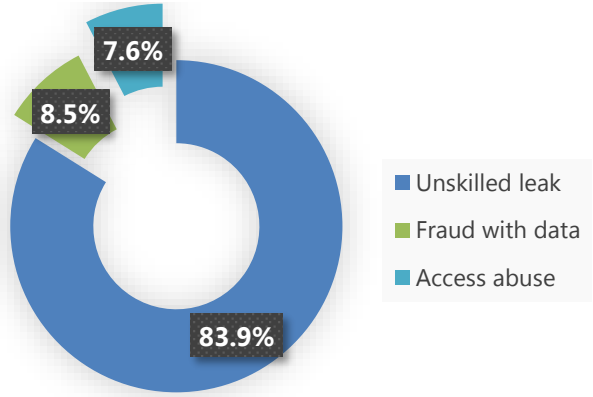


Figure 7. Breakdown by nature in 2018

7.6% of the incidents are registered as involving unauthorized access (abuse of access rights, use of information beyond the need-to-know basis).

> *theepochtimes.com: US prosecution accused a Chinese national of stealing intellectual property from his employer, a multinational energy company. Before quitting his job, Tan Hongjin, 35, downloaded hundreds of confidential files about cell phones and lithium-based battery systems from his employer. The total value of products connected to the information that Tan is claimed to have stolen is estimated at about $1.4-1.8 billion. Tan told a co-worker that he planned to return to China. During its investigation, the FBI found on Tan's laptop an employment agreement from a Chinese company that produces lithium-ion battery materials.*

In breakdown by region for 2018, the USA traditionally took first place (956 cases and 42% of all leaks). Russia came second with 270 leaks. The UK closes the top three, with 124 data leaks in 2018.

Conclusion

In 2018, we observed a significant decline in the growth of the number of personal data leaks and a decrease in the total volume of leaked data as well as in the 'leak capacity'. We can assume that businesses and public bodies that process personal data have decided to bear down on personal data protection violators. For instance, this is evidenced by a decrease in the share of leaks caused by external attacks. External attacks are known to be more sensitive to technical controls. In this respect, the efforts of information security services aimed at countering external threats have not been in vain.

On the other hand, we observed an increase in the number of leaks that exposed over a million records at a time, which suggests both that violators have a considerable interest in user data and that information security is far from perfection.

Personal data, as the easiest to understand and measure the type of data, gradually give way to trade and other secrets. Leaks of these types of data are increasingly becoming the focus of media attention. At the same time, leaks of trade secrets and know-how are not new, they have always been there. However, at the present stage of security systems evolution, perhaps, there are enough capabilities to record the fact of leakage of trade secrets and collect the evidence necessary for the subsequent trial. It stands to reason that in 2018 we registered a lot of reports about companies that pressed charges against their former employees for sabotage, industrial espionage or theft of onfidential information and were successful. This is due in no small part to behavioral analysis techniques that allow determining the weakest link in the team.

Leak Channels

In 2018, the share of leaks via e-mail decreased most significantly (by 5.3% compared to 2017.) There were less leaks due to theft or loss of equipment (0.8% less) and via mobile devices (0.2% less.) We saw an increased share of leaks via printed documents (by 2.8%), network (by 2.3%) and instant messengers (by 1.2%).

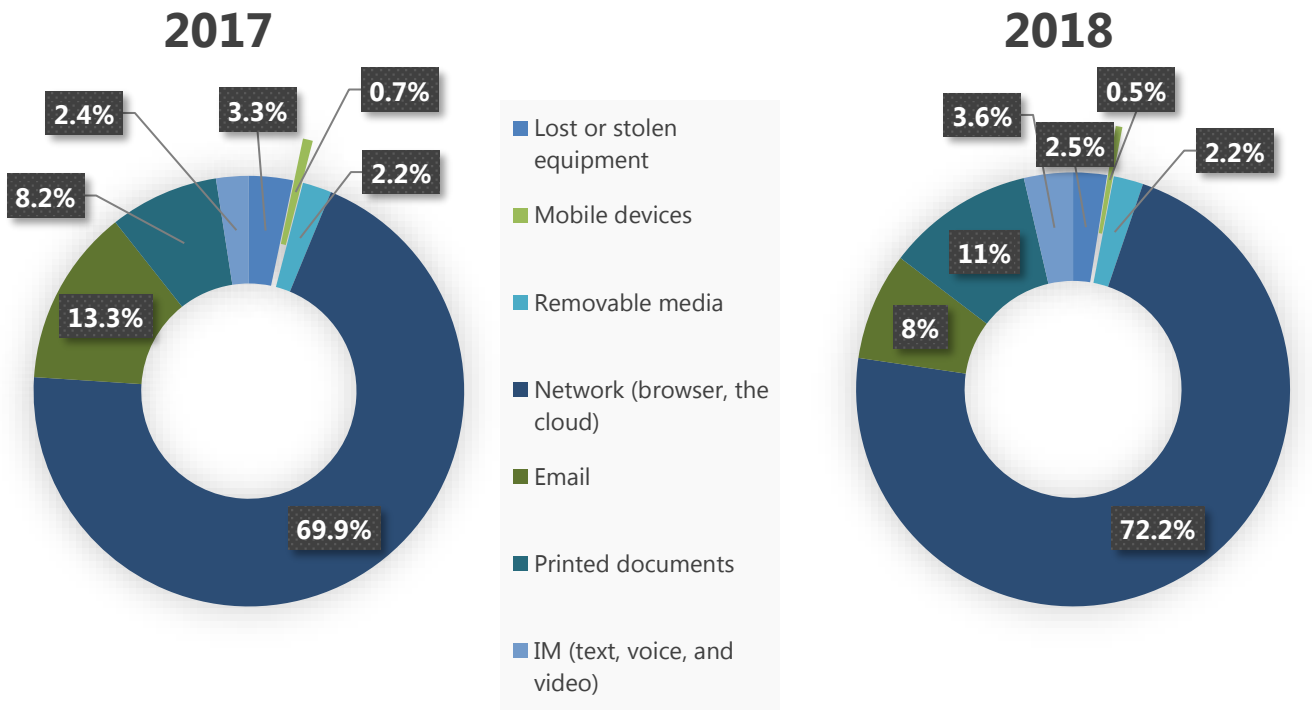


Figure 8. Breakdown by channel in 2017-2018

In the breakdown by channel, the largest proportion of accidental leaks caused by employee error is via printed documents (17.1%), email (15.9%), due to theft or loss of equipment (4.6%), via messengers (voice and text, 4.5%) and removable media (2.7%). 55% of all registered accidental leaks were attributed to the network channel.

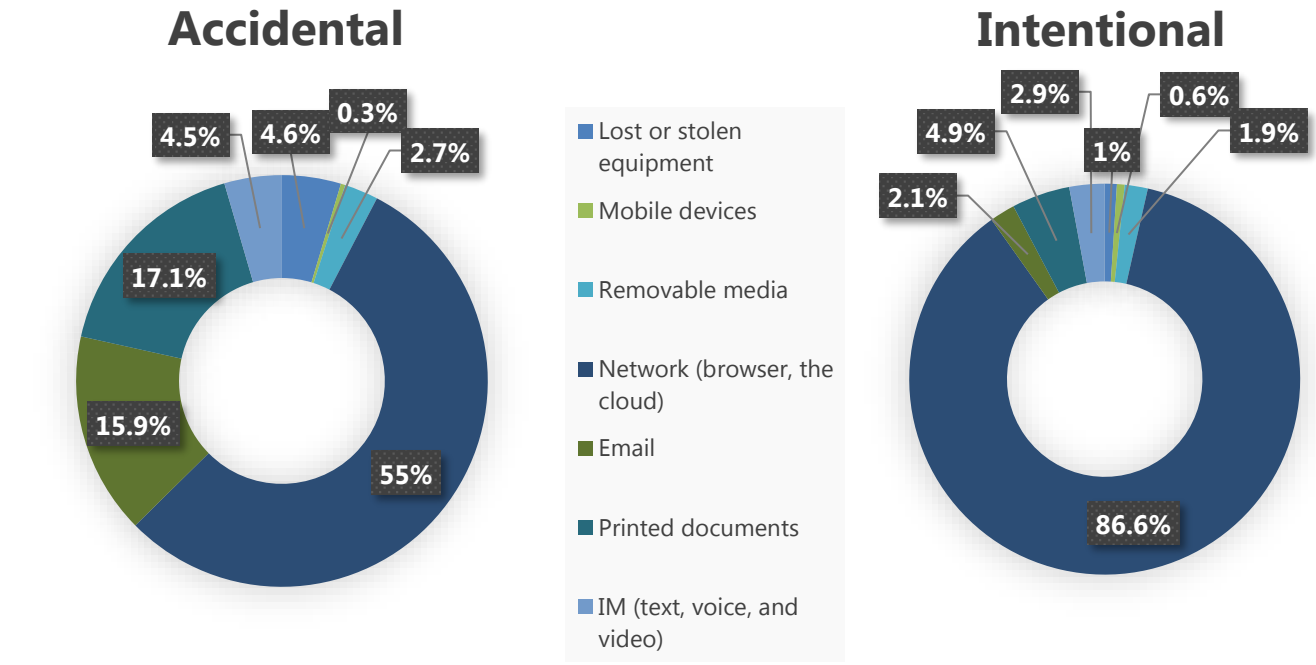


Figure 9. Distribution of accidental and intentional leaks, 2018

The high share of data leaks through paper documents seems puzzling only at first. In day-to-day work, businesses and public bodies around the world continue to use paper in a major part of their document flow. At the same time, the rules on the disposal of such documents are not always observed.

> *abc.net.au: Hundreds of classified government files were found in fireproof filing cabinets purchased at a second-hand shop. Files reveal the inner workings of five separate governments of Australia and span nearly a decade. Most files are systematized; some are classified 'top secret' or are not intended for non-Australian eyes even after the mandatory 20-year storage period. Among other things, the files shine light on the activities of ex-Prime Minister Tony Abbott.*

Intentional leaks are most often associated with the network channel. Almost 90% of intentional leaks are related to illegal transfer or disclosure of information via the Internet (including web services, e-mail, and other online resources.) The share of intentional leaks through paper documents is 4.9%, which is quite high, with 2.1% of intentional leaks occurring through e-mail and 1.9% — through removable media.

> *ico.org.uk: In a British court, a former school employee was fined £700 for unlawful obtaining of other people's personal data. Darren Harrison used a USB stick to upload huge amounts of personal data of students to the server. It was revealed that all information was obtained by him at his previous jobs. During the course of the investigation, Harrison provided no valid explanation as to how the information had appeared on his USB stick. Appearing before the court, Harrison pleaded guilty.*

A relatively new phenomenon in the context of leak channels is leakage of data through mobile app flaws. Formally, such cases can be classified as network leaks or mobile device leaks. However, we should consider that the processor that orders a third-party mobile app has no technical capabilities to protect such data. Of course, the potential liability can be redistributed by legal tools. However, the impact on the reputation of the company that orders the app will be considerable in any case.

yorkmix.com: Personal data belonging to users of the app developed for residents of York, England, by order of the council, were breached. The app had been actively downloaded by residents and was intended to improve the city's environment. The breach affected around 5,900 users. The council refused to use the app and asked the residents to delete it from their devices.

The network channel is leading both in the terms of the total volume of leaked personal data and the number of leaks. Most (>60%) leaks of personal data occur via the network (see Figure 10).

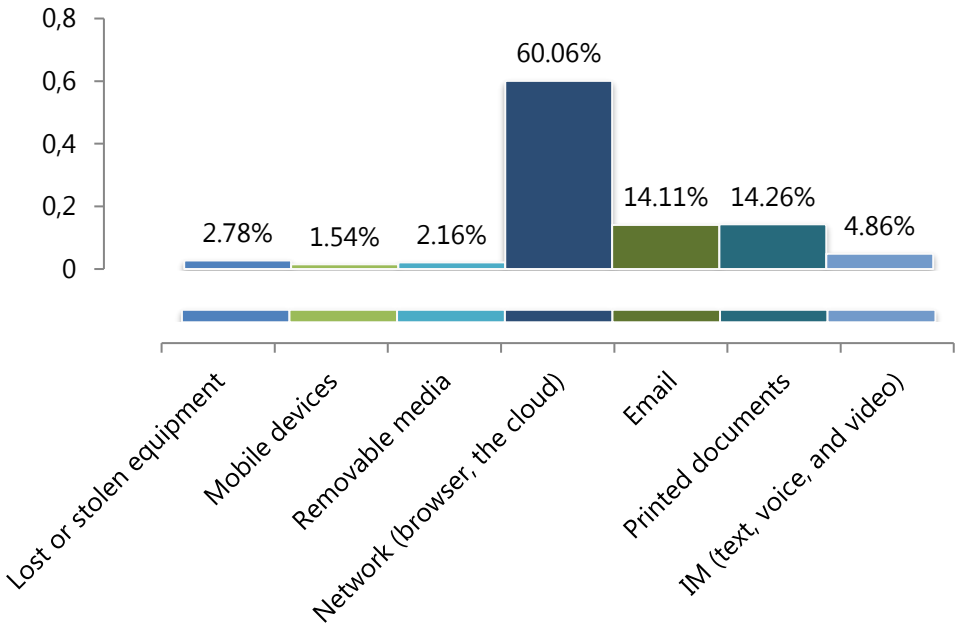


Figure 10. Breakdown by channel, 2018

Dealing with insiders, companies have faced a variety of scenarios. Sensitive data were stored on Box, OneDrive and other cloud storage, leaked through free e-mail accounts (webmail), etc.

bleepingcomputer.com: A former engineer who worked for a US Navy contractor was found guilty of theft of classified information. The man, 35, copied more than 5,000 files to his personal cloud shortly before he left his position. He sent some of the documents to his e-mail. The court found the defendant guilty of 13 counts of data theft. The man faces a maximum of ten years of imprisonment for each guilty verdict.

Scenarios of external attacks are less varied. Hackers, as a rule, are not familiar enough with the structure of the corporate data, such as what is the most valuable information and where it is stored, so they take whatever may be valuable.

time.com: MARRIOTT International reported a massive customer data leak. The company says hackers may have accessed the personal data of 500 million guests. About 327 million stolen records included passport details, e-mail, and postal addresses. Hackers were also able to steal some of the encrypted payment data records. Marriott does not rule out the possibility that this information has been decrypted. The news of the leak caused a stir in the market — Marriott shares price fell by 5.6%.

Network should be recognized as the most 'popular' leak channel both for accidental and intentional leaks. A characteristic of network leaks is the highly critical nature of data and huge volumes of leaked records.

The small share of intentional leaks via mobile devices, removable media, email, and paper documents is due to experienced attackers resorting to these channels less and less. A 'skilled' attacker understands that modern control systems allow to intercept transmission of confidential information via these channels and will not take the risk.

As for instant messengers, there are certain regional differences that are reflected in the global leak pattern. For example, Southeast Asian countries traditionally have more leaks via mobile devices and messengers than the rest of the world, because users prefer mobile devices to desktop PCs both for informal and business communication.

Conclusion

The breakdown by channel only slightly changes every year, which allows us to assume that the factors at its basis are stable. The most important and perhaps the only phenomenon that is relevant for predicting the pattern of leaks in the near future is the steady decline in the number of intentional leaks across all channels except the network.

As we have already mentioned, this is due to the relatively high computer skills of attackers. They know or suspect that there are information security systems and how they work. On the other hand, information security systems are more adapted to the detection of leaks via network and e-mail, but appear less effective if, for example, data are leaked through voice messages.

Besides, the share of intentional leaks not detected by security systems for mobile devices, messengers, voice messages and other 'new' channels seems to be significantly higher than the corresponding share of 'latent' leaks, for example, for the network channel. It seems that security systems are not as effective with intentional leaks as in the case of accidental leaks. The obvious conclusion is that in the next few years information security systems must not be oriented towards control of information only; they also should control the user who accesses the confidential information. It might be that the 'analysis' of user actions, a technology that could detect data leaks by indirect evidence and take into account the human factor, will, if combined with the traditional control system, prove more effective in terms of countering intentional leaks.

Industry Map

Compared to 2017, the breakdown of data leaks by organization type has not changed significantly (see Figure 11).

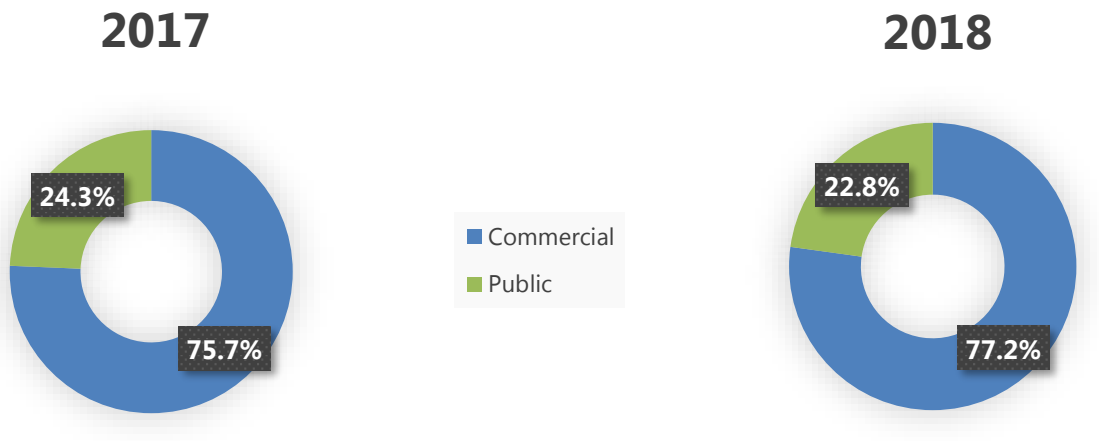


Figure 11. Breakdown by organization type in 2017-2018

Data leaks from high-tech companies and medical organizations were most common, with few leaks from the HoReCa, manufacturing and transport sectors (see Figure 12.)

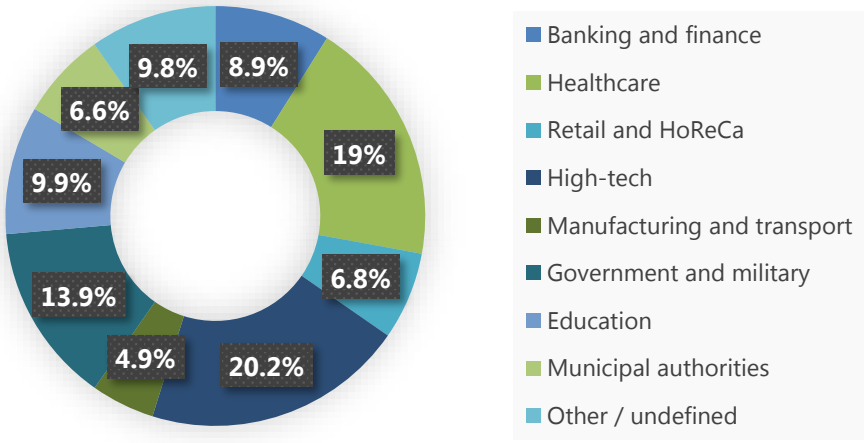


Figure 12. Breakdown by industry, 2018

These charts illustrate the actual pattern and give a general idea of data leaks in various industries. It is more important to find out which sectors are currently the most attractive for attackers.

The 'attractiveness' of an industry is directly related to how easy it is to cash in on the data processed by such companies. In other words, the easier it is to convert the stolen information into money, the more attractive the sector is. For attackers, information security maturity is in inverse relation to the attractiveness of an industry. The number of intentional leaks in a particular industry can be considered the attractiveness indicator. The breakdown of

intentional leaks of the same type of data by industry reveals the most attractive sectors (they are the most vulnerable, too.) This is illustrated in the formula:

$$\text{The share of intentional leaks} \leftarrow \frac{\text{Easy} - \text{to} - \text{sell data}}{\text{Perception of data protection level}}$$

In 2018, the most attractive industries were banking and insurance (shown together as 'Banks and financing'), manufacturing, retail and high-tech companies. More than half of the leaks of personal data were intentional in these industries (see Figure 13.)

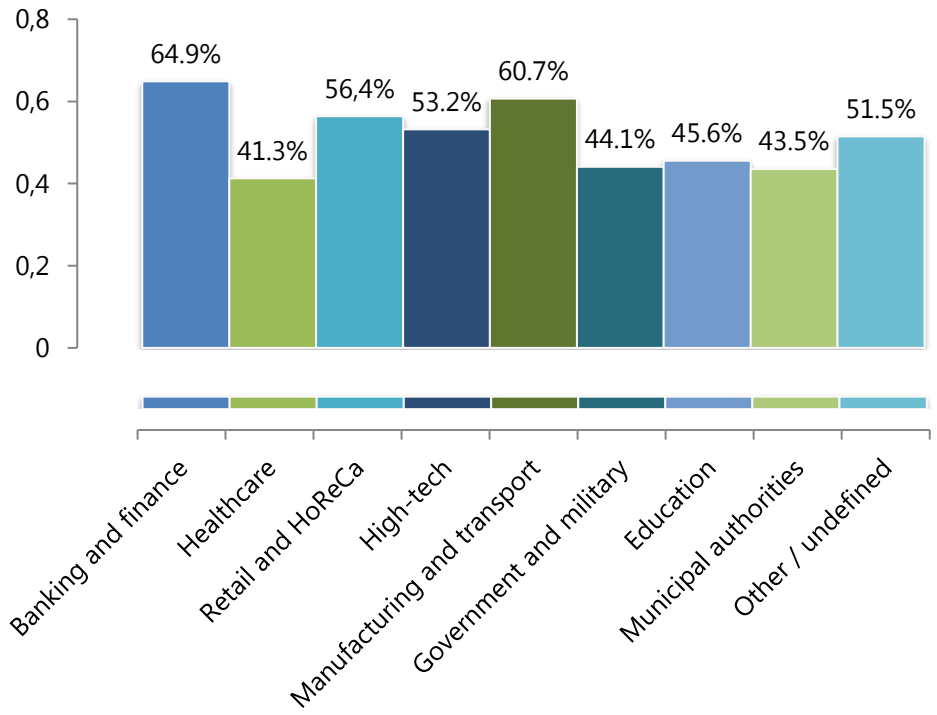


Figure 13. The share of intentional leaks in all personal data leaks by industry, 2018

For a comparative histogram, we take personal data as a protected object common for all industries to clearly show the breakdown of industries by 'attractiveness'. Of course, with respect to banks and financing institutions, the most 'attractive' type of data is different — it is financial information. For high-tech companies and manufacturers, know-how is the most valuable information asset.

> *pcmag.com: The chief executive and eight employees from South Korean company Toptec were charged with stealing Samsung technologies. The investigators claim the employees of Toptec conspired with two managers of an unnamed Chinese company to sell the secrets of Samsung's "3D Lamination" display technology to China for $13.8 million. The 3D Lamination technology was designed with the support of Toptec and is now used in Samsung's flagship phones. Samsung spent six years and around 150 billion won ($134 million) to develop the technology.*

If we rearrange the breakdown by attractiveness depending on the attack vector, we get a clear idea of the attractiveness of a particular sector for external attackers and insiders (see Figure 14.)

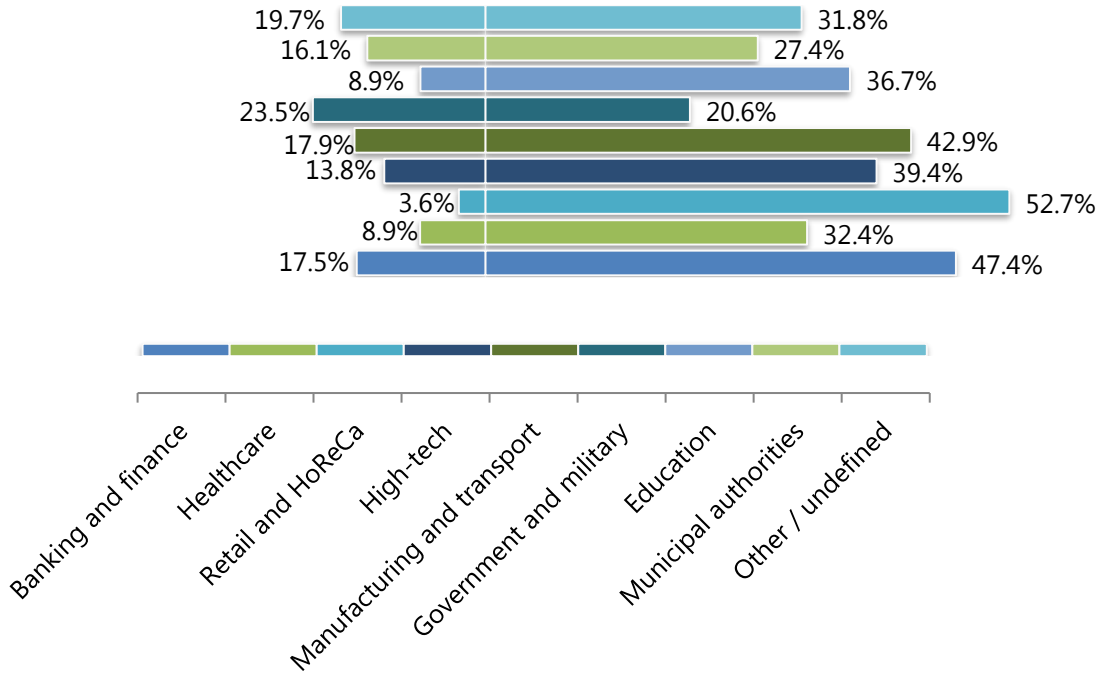


Figure 14. The share of intentional leaks caused by insiders (left) and intruders (right) against the total personal data leaks by industry, 2018

Government and military, manufacturing and transport companies, banks, insurance companies, and financial institutions are leading with respect to insider threats. This is explained by easy-to-sell data processed in these sectors. The ease to sell and potential income make attackers ignore the risk. The evidence from practice shows that attackers who want to steal trade secrets or bank information are not stopped even by the prospect of huge fines and long prison terms.

HoReCa, high-tech and financial institutions attract external attackers most often.

See Figure 15 for a more comprehensive industry map. The size of a bubble reflects the total volume of leaked records in millions (across all companies in the sector), and the vertical position of a bubble shows the number of leaks in the sector [12]. Depending on the size of the affected company, the map is broken down into three charts — small, medium and large companies.

[12] The number of leaks in the industry includes personal data leaks when the precise volume of leaked data is known. The volume of leaked data in the industry is calculated without taking into account mega leaks which leaked over 10 million records.

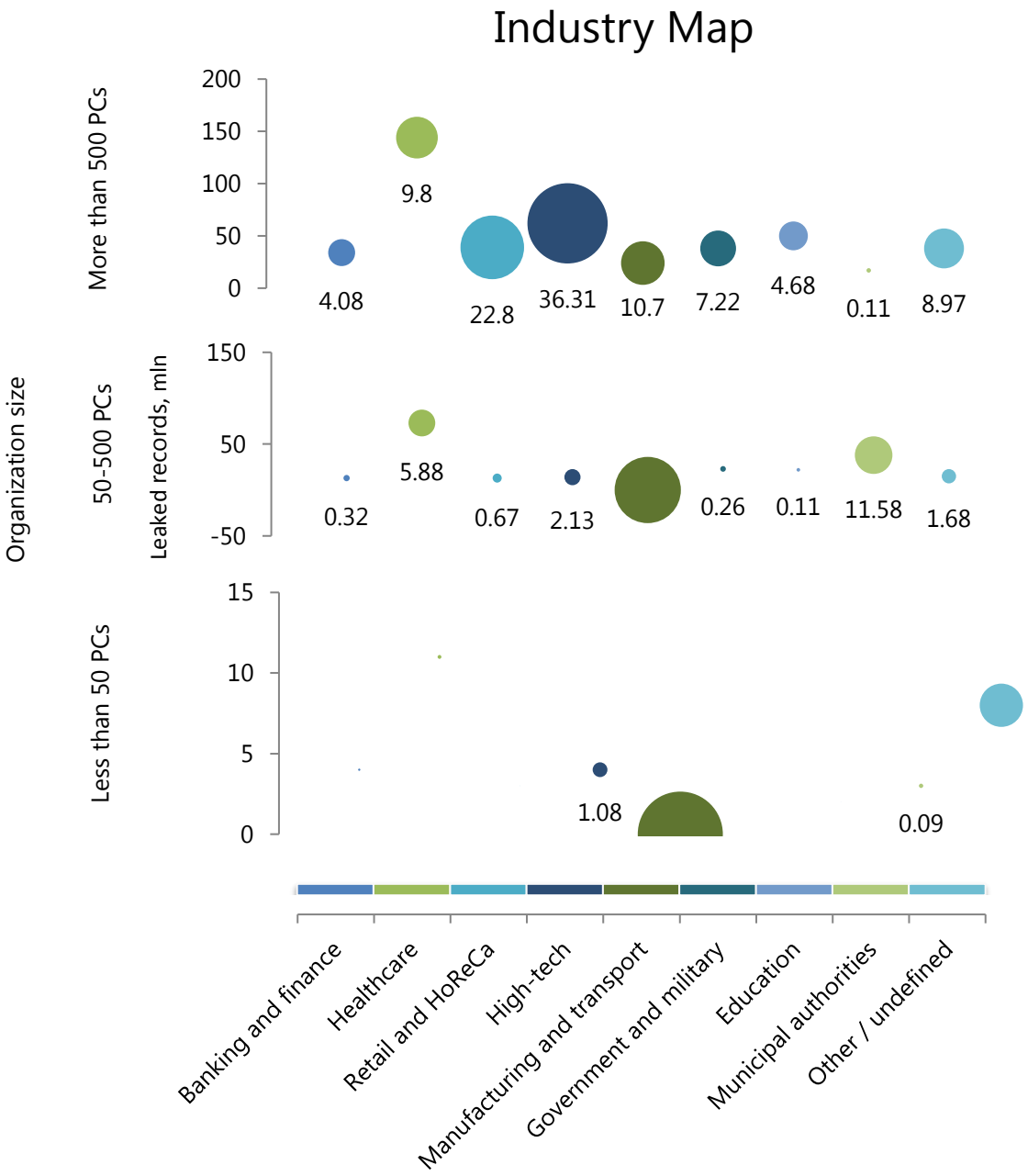


Figure 15. Personal data leak industry map, million of records, 2018

High-tech companies (web-services, digital service providers, mobile operators) have the biggest volume of leaked personal data. These companies can be called pioneers in the use of digital data processing and storage technologies. Customers of major retailers and financial institutions were less affected. The volume of personal data leaked from major manufacturers and government was unexpectedly large.

Conclusion

The trends in the industry map are mainly determined by two factors — how easy it is to sell data and how well data are protected — both being relatively unstable. Changes in the most important indicators of the industry

pattern — the share of intentional leaks in the sector, the share of intentional leaks caused by external attackers and insiders — are associated with how attackers assess the value of data processed in the sector and the cybercrime infrastructure that arouse to allow easy sales of stolen data (either a web service database with millions of personal data records or manufacturing know-how).

There is also a relatively insignificant volume of data leaked from small businesses. In the real world, the issue of major data leaks is totally irrelevant to such companies. Because of the minor volume of data processed, small businesses are almost always of no interest to external attackers, except for small high-tech companies. However, it may occur that, as security improves in large companies, attackers will be forced to turn their attention to small businesses.

Final Thoughts

The changes in the overall pattern of leaks in 2018 can be called insignificant. The factors that form the global pattern of data leaks are stable, which means that figures are stable too. This way, the identified trends are very likely to remain relevant for the next few years.

It should be noted that the growth of the number of leaks and volume of records seems to be slowing down. We conducted a comparative analysis of the pattern of leaks in 2017-2018 and can tentatively assume that this is associated with increased attention paid by government and businesses that process large amounts of data to leaks of personal and payment information.

The state represented by regulators prepares legal acts and with each such legal act increases liability of companies that fail to protect data. Companies are therefore forced to invest in data security, both for their 'own' information, such as trade secrets and know-hows, and for 'outsider' personal data that are not owned but have to be stored by companies.

From the data for the past year, it is clear that different types of data require different approaches to their protection. Until recently, companies wanted to invest in the protection of 'intellectual property', intuitively classifying it as 'theirs'. Yet it was very difficult to justify the need to protect personal data, i.e. the data that the company does not own, but only processes. Attempts to assess the potential damage from the loss of 'outsider' personal data failed against the healthy pragmatism of personal data processors — why protect what the company has not invested in the creation of. The risk of loss of reputation as a result of a personal data leak was either ignored or accepted as unavoidable by businesses.

In 2018, it seems that there is no longer a lack of motivation for processors — huge fines have been introduced for personal data breaches. Now major companies and public bodies that process personal data of their customers and citizens do have something to lose. But the most important conclusion that can be drawn is that each type of data requires a special approach to protection.

The same is true for data channels. Security methods used for the network channel are virtually ineffective when it comes to the 'new' channels, such as mobile devices and messengers. As part of the overall approach to data security, it is necessary to develop individual approaches for each channel, both at the level of technology and at the level of policies and corporate procedures.

The difference between internal and external leaks that we spoke of remains. They even can be viewed not as two different types of leaks, but as two fundamentally different incidents that have more differences than similarities. Let us mention only the most obvious fact. If, to counter external threats, the 'only' thing security tech and services need is every minute readiness and up-to-date knowledge, then countering insider threats, quite otherwise, involves, apart from readiness and knowledge, considerable effort in managing both information (protected objects) and employees, analyzing and controlling employees' behavior and detecting anomalies in technical and information systems.

For the rest, the general pattern of leaks in 2018 was hardly surprising. The capacity of hacker attacks (the average number of leaked records per incident) fell by more than a third in 2018. Still, it is too early to say that there is a tipping point in the fight against external threats. The total number of incidents has not decreased, and hackers still regularly get into huge databases.

In 2018, banking, insurance, manufacturing, trade and high-tech companies proved to be the most attractive for attackers. We have registered an increase in the share of leaks caused by insiders and an increase of leaks through the network channel.

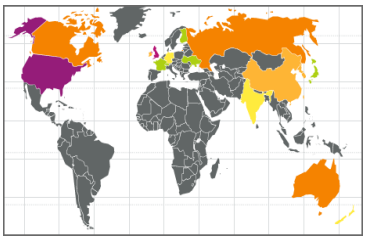
Earlier we mentioned the fact that the owner of information is not always prepared to assess the value of protected object and estimate the losses as one of the major issues related to dealing with consequences of data leaks. Now, with the advent of the global trend of increasing the fines for personal data leaks, with the growing number of guilty verdicts in cases on theft of trade secrets, we may assume that the issue of assessing the value of data is now irrelevant.

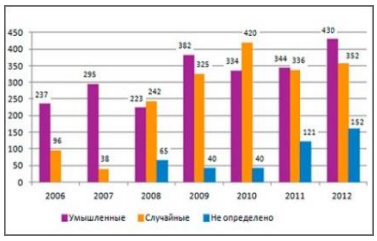
Until recently, one could often hear the cliché saying 'Who owns the information, he owns the world' from information security companies. In modern world, owning the information makes sense, because the information circulating inside a company is acquiring actual value right before our eyes. The value of trade information is determined by the market, while the value of personal data is determined by the potential fines. This means that investment in data security can already be evaluated in terms of efficiency, which seems to be the major accomplishment in 2018.

Data Breach Monitoring on the InfoWatch Website

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

In addition, the website contains data leakage statistics for past years, available in the form of dynamic diagrams.





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

- Email
- Facebook
- Twitter

InfoWatch Analytics Center
https://www.infowatch.com/analytics

Glossary

Information security incidents *in this research mean cases of compromising confidential information as a result of data leaks and/or destructive actions by employees.*

Data leak *means losing control over information due to external intrusion (attack), access abuse, or unauthorized access.*

Destructive actions by employees *mean personnel actions that resulted in the compromising of confidential information, including the use of confidential information for personal needs associated with fraud; illegal access to information (abuse of access rights).*

Confidential information *in this context means information which can be accessed by a limited number of expressly identified persons subject to its non-disclosure to third parties without the consent of an information owner. In this report, the term "confidential information" also includes personal data.*

Intentional/Accidental Leaks. *Intentional leaks mean an information leakage 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 for personal gain or benefit. This results in a risk of losing control over information and/or committing a confidentiality breach. In this case, it does not matter whether such 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 such user's actions actually led to negative consequences or corporate losses. The terms "intentional/malicious" and "unintentional/accidental" are equal and used as synonyms herein.

Attack vector *means a classification criterion of intruder's actions behind data leakage, including intruders who attack company's web assets and IT infrastructure from the outside to compromise data, and insiders who obtain unauthorized access to classified resources and misuse confidential information, etc.*

Data channel *means a scenario which results in the loss of control over information and a breach of its confidentiality. Currently, we identify eight separate leak channels:*

- ✓ *Theft/loss of equipment (server, data storage, laptop, desktop), with information being compromised during maintenance or due to the loss of such equipment*
- ✓ *Mobile devices where data leakage occurs because of unauthorized use or theft of a mobile device (smartphone, tablet) when used as part of BYOD paradigm*
- ✓ *Removable media loss/theft (CDs, flash drives)*
- ✓ *A 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*
- ✓ *Email, with data being leaked via corporate email*
- ✓ *Paper documents which can cause a data leakage if stored or utilized improperly (with confidential information printed, stolen, or taken out)*
- ✓ *Instant messengers (data leakage via voice, chat, and video communications)*
- ✓ *'Non-defined' is a category used when incident details appearing in mass media do not allow for the leak channel identification.*