imperva

WHITEPAPER

More Lessons Learned from Analyzing 100 Data Breaches

Table of Contents

Introduction

As part of the Imperva Threat Research Team's ongoing efforts to monitor and report on the current database threat landscape, we studied and analyzed over 100 of the largest and most well-known data breaches to date.

To get the most value from the analysis of these breaches, Imperva's Threat Research Team extracted each initial breach, the target of each breach, the system that was compromised by the breach, and what records were stolen.

Before we start with the lessons learned, it is important to understand that because of their sensitive nature, database security breaches often go unreported. Also, in many cases, there is not enough data available to provide a comprehensive report.

The data we present here has been compiled from all the available information sources; the web, breach reports, hackers' forums, analysis of stolen database dumps, and information gathered from deploying our own honeypots. We present the specific types of data that were stolen, the root causes of the breaches, and aim to provide some insight into how this data has changed over the last year.

We also profile and characterize the types of attackers organizations face today and provide a flow diagram to help you understand the complicated ecosystem of data breaches themselves.

Our analysis of these data breaches reveals a bleak and troubling reality. Not only has there been a greater number and higher frequency of breaches over the last decade, but an increasing amount of stolen data is being exposed and sold on the dark web. This stolen dark web data is being sold to black hat hackers and used in extortion attempts, and as fuel to create "phishing" and other social engineering campaigns, which in turn leads to more data breaches.

The ecosystem of data breaches is varied and complicated. For example, many breaches are the result of poor security practices like using unprotected publicly accessible services (Microsoft, Advanced Info Service) or weak authentication. Others are caused by unauthorized access of third-party services (Cambridge Analytica) or supply chain attacks (SolarWinds, Accellion). Some of them are turning a small-scale breach into a colossal disaster - like storing database passwords in clear text. Others are caused by numerous additional security practice failures, which we will explore throughout this report.

The current threat landscape

As we will see when we examine the types of data that cybercriminals steal, we can ascertain what they think are the highest value targets. Once they have a plan, they will try different businesses within the scope of where they see the most revenue potential, and try to breach them first. If you hold high volumes of personal data like names, addresses, Social Security numbers, buying histories, financial information, or medical records, you are already at high risk of being attacked, and as hackers get more sophisticated, that risk rises.

As more enterprises move their workflows and data into cloud-managed infrastructures, these new technologies create new data security challenges. Each cloud-managed environment has its own unique infrastructure and APIs. While the infrastructures themselves are purpose-built for security, protecting your data within the environments is your responsibility. Most traditional data security practices do not translate easily, or at all, to cloud cloud-managed environments, leaving enterprises with extensive gaps in the visibility of their most sensitive data.

To combat this, many organizations have decided to ingest all data - on-premises and cloud-based - into a SIEM tool like Splunk, or they turn logs over to their Security Operations Center (SOC) team and hope they can distinguish behavior that violates security policy and remediate it before it becomes an actual data breach. In the case of Splunk, the indexing of enormous volumes of data increases data security and analysis costs prohibitively. In cases where SOC teams are getting raw data, they are simply shutting off potential alerts because there are too many false positives. In both cases, the noise created by forcing the SOC to manage such volumes of data makes it much easier for the stealthy attacker to get what they want, sooner or later.

All these developments lead to more challenges for enterprises and more opportunities for hackers.

Every organization is a hybrid

Today, every organization is a hybrid. Even if your organization has not migrated one piece of data to the cloud, the chances are high that a vendor or business partner has. If you share data with any of these other entities, their security concerns become yours.

If you have a digital presence at all, you are using applications that are developed using microservices and open-source code, and these applications work together due to APIs that have been written to optimize the digital experience. In all of these instances, you have no control over the security of these pieces of code and, despite your best efforts, you are still at risk.

The lure of cloud-managed environments to drive low-cost innovation is not going to dissipate. More and more of your sensitive data is going to be hosted in DBaaS and Data Lakes, and sticking with traditional data security strategies will force your security to depend on manual processes that aren’t effective in your new hybrid environment.

Complexity brings fragmentation in data security management

The obvious answer to stopping data breaches is to adopt new solutions, but even that creates its own set of challenges. In her blog post, Why the Search for Best-Of-Breed Tooling is Causing Issues for Security Teams, Imperva Technical Product Marketing Manager Kelsey Winiarski reported, "the average enterprise deploys 45 cybersecurity-related tools on its networks. The widespread use of too many tools may contribute to an inability to detect and defend against active attacks. Enterprises that deploy over 50 tools ranked themselves 8% lower in their ability to detect threats and 7% lower in their defensive capabilities than organizations employing fewer toolsets."

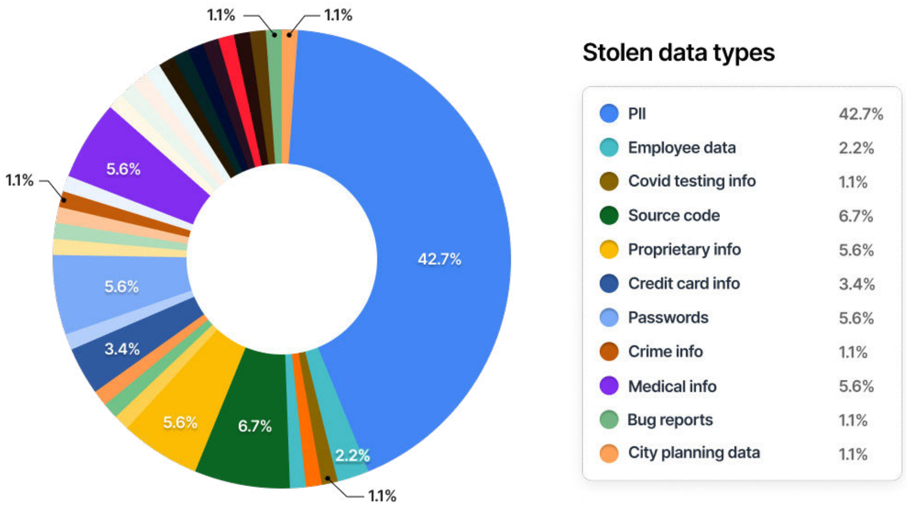
In the end, this proliferation of tools makes a bad problem worse. Kelsey wrote, "most tools can only see one part of an attack, making it difficult to create alerts that tell the whole story. For a tool to be both effective and accurate, solutions need to have full visibility over the network that they're protecting. In most organizations, it's common to see security tools like XDR, UEBA, DLP, and WAF. Each of these serves a specific purpose, and while they each do their job exceptionally well, they lack integration with other solutions deployed in the same environment. As more companies move to the cloud, on-premise DLP tools do not have insight into data that is stored in SaaS applications. In the event of an attack, multiple alerts could be generated from different tools with little indication that these alerts are connected. While many tools offer APIs to further enrich their solutions, each API is not built on the same standards, often requiring custom code to integrate products fully. In the end, analysts are left to piece together information from multiple sources to understand an attack."

A survey of security professionals revealed that 44% spend more than four hours daily on alerts. These alerts often tell just enough information to start an investigation but lack the necessary detail to act accordingly solely based on that alert. Triaging thousands of vague alerts wastes time and resources; senior analysts are pulled in to assist junior analysts in investigating incidents. In the end, it's a critical time sink that almost no organization can afford.

There is a way forward, however. This report explains what data is stolen and from where. It gives a clear idea who the bad actors are that take it, how they take it, and why. Armed with this information, you can build the strategy you need to avoid being a soft target for cybercriminals that want to breach your data.

Types of data most frequently stolen

Imperva's Threat Research Team gathered the data in the chart below over a 12-month period, starting in July 2021.



Hackers can only steal data that is available for them to take. As you can see, Personally Identifiable Information (PII) is the data type of choice for cyber criminals. Credit card information and passwords, which made up 26% combined of stolen data (the last time we collected this data) accounted for just 9.4% in the current period. Why might that be?

It might be that more organizations are using basic security tactics such as Multi-factor Authentication (MFA). Multi-factor authentication employs a process that cross-verifies privileged users with two different forms of identification, usually knowledge of an email address and proof of ownership of a mobile phone. Used on top of the regular username/password verification, MFA makes it much harder for outside cyber attackers to gain the access required to breach data.

On passwords, more people have returned to brick and mortar work environments where perimeter security is harder. The decrease may also be the result of better phishing training and more responsible password management than was being practiced before.

Of all these data types, PII is in the long term, the most valuable. As cyber criminals compile more PII about people from the dark web, they can engage in hard to stop and full-on identity theft - which is where the big fraud money is. Credit cards and passwords can be changed the second there is a breach. When PII is stolen, it could be years before it is weaponized by hackers.

The widespread theft of Personal Data is a strong indication that many organizations are not putting enough protection into place to secure it. Some of this is due to the fact Personal Data is regularly transferred between systems, people, and suppliers to perform common business tasks. As regulations governing data privacy get tougher, it will be critical for organizations to discover, identify, and classify Personal Data across their data estate. Only when you know where the Personal Data is hosted and what applications and users are using it, will you be able to extend the security controls that protect it.

Profiles and tendencies of the four main types of cyber attackers

Knowing what cyber attackers are doing and want can help you better understand the database threat landscape and create a more effective application and data-centric security strategy. To start, it is helpful to characterize the entities from whom you need to protect your assets. Here we'll profile the four principal types of cyber attackers – one originating inside the organization and three from the outside – so that you can gain some insight into their methods and motivations, and use what you know about them to thwart their attacks.

Cyber attackers are commonly split into two groups: **"Inside Threats"** and **"Outside Threats."** The first threat type, originating inside the organization, is generally activated when employees leave data exposed, either maliciously or by mistake/oversight. For malicious inside cyber attackers, the motive is usually money, often accompanied by a dislike for the company. The malicious insider usually has access to assets or credentials and is less suspicious than an outside threat. There are best practices that organizations can engage in to mitigate malicious insider threat risk: The most obvious, and also the easiest thing to do, is to make sure your employees are not doing stupid things like sharing passwords – either internally, or worse externally – and not properly logging out of environments that contain sensitive data when they are done working. Second, it is important for security teams to constantly monitor user permissions and privilege levels to access sensitive data. In short, if a user's job does not require access to specific sensitive data, they shouldn't have it. It is also critical for security teams to have the necessary visibility into the data estate to know what constitutes normal data use. For example, if an internal user who has never accessed a sensitive data source before suddenly starts downloading a lot of sensitive data, security teams should be made aware automatically. You can get a more comprehensive accounting of how to mitigate insider threats here. Attack methods used by outside cyber attackers depend on their motivation, as these three threat profiles show.

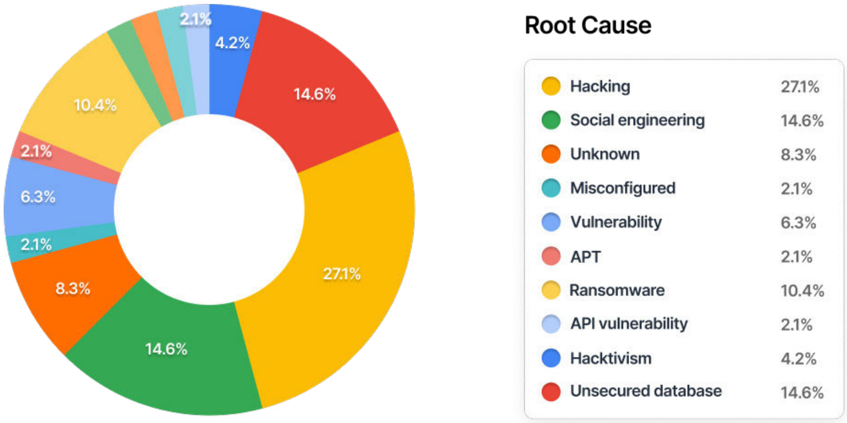
The **Hit and Run attacker** identifies an opportunity – a vulnerability, publicly open database, or something else – takes what they can, and leaves. This kind of attacker won't search for other databases, penetrate the organization's network, or try to execute exotic exploits, etc. They will only take what they can, easily, and sell it to the highest bidder. Many organizations make it easy for Hit and Run attackers to steal data. While most security teams do their best to mitigate the exploitation of newly-discovered vulnerabilities, some DBAs and DevOps people are migrating operations and workloads to publicly open services in the cloud that security teams do not account for. If left sitting out there unsecured, this data can fall easy prey to Hit and Run attackers. If you are using publicly open services, even if only for search and analysis, ensure they are visible, configured properly, and that security updates and patches are current.

The **Curious attacker** usually sets out with a purpose, but has enough interest to look around a little bit, but not too much. They are still focused on their original purpose, malware deployment, data exfiltration, etc.

The **Resident attacker** is the most dangerous type. The Resident will penetrate the organization's network and stay for months, sometimes years. They use keyloggers, sniffers, and other methods to steal credentials and compromise databases, using "Slow & Low" and other methods to stay undetected.

Root causes of data breaches

Looking at the root causes of data breaches over the last 12 months, two data points stick out because they are fairly straightforward to mitigate - Unsecured database at 14.6% and social engineering at 14.6%. A publicly open service increases the risk of a breach to happen, but in most cases, this is not a failure of security practices; it is rather a complete absence of a security posture. For example, when there is a failure to secure non-production environments where company data may be used to perform testing. There are solutions available today that enable organizations to ensure they have visibility into all of their databases so they can apply their security policies to them and prevent breaches.



Social engineering can be very sophisticated, but it is also preventable. The first step is to use the principle of least privilege to ensure that the fewest number of people in the organization even have the authority to access sensitive data. Beyond that, use training to get people to stop clicking things they shouldn't click and from giving away password information. The decision to be beaten by social engineering is yours.

One data point that we expect will rise over time is the number of breaches caused by misconfigured applications. This is a particular problem with cloud-managed infrastructure where configuring for security requires some degree of expertise.

Obviously, with hacking and hacktivism accounting for nearly one in three data breaches, it is critical for organizations to find solutions that make them harder targets for even the most sophisticated hacking. Organizations need to look at things like bad bot protection solutions to help mitigate these causes.

Common mistakes that lead to breaches and how to avoid them

First, a bit more about insider threats versus outside threats. Insider threats present a different sort of danger profile than outsider threats, which are more of a "smash and grab" type of attacker. There are overall strategies that organizations can create that mitigate both of these threat types. Consider what it would take to create a Zero Trust Security Model in your organization. There is no silver bullet here. Even the most comprehensive and effective solutions require organizations to do dozens of things better than before to affect actual data-centric security and stop making the mistakes that cause them to be easy targets.

Managing insider threats

The malicious insider: Someone who maliciously and intentionally abuses legitimate credentials, typically to steal information for financial or personal incentives. An individual who holds a grudge against a former employer, or an opportunistic employee who sells secret information to a competitor. They are familiar with the security policies and procedures of an organization, as well as its vulnerabilities.

- **How to deal with them:** Protect your critical assets better. Form a comprehensive understanding of your critical assets. Ask questions such as: What critical assets do we possess? Can we prioritize our assets? And, what do we understand about the current state of each asset?

The careless insider: An innocent pawn who unknowingly exposes the system to outside threats. The most common type of insider threat is the result of mistakes, such as leaving a device exposed or falling victim to a scam. For example, an employee who intends no harm may click on an insecure link, infecting the system with malware.

- **How to deal with them:** Set and enforce policies. Clearly document organizational policies so you can enforce them and prevent misunderstandings. Everyone in the organization should be familiar with security procedures and should understand their rights in relation to intellectual property (IP) so they don't share privileged content that they have created.

The Mole: An imposter who is technically an outsider but has managed to gain insider access to a privileged network. This is someone from outside the organization who poses as an employee or partner.

- **How to deal with them:** Increase visibility into your data. Deploy solutions to keep track of employee actions and correlate information from multiple data sources. For example, you can use deception technology to lure a malicious insider or imposter and gain visibility into their actions.

The bottom line on insider threats

The risks these data breach threat actors present can be mitigated by a combination of a zero trust security model, the principle of least permission, employee training, and a solution designed and trained through machine learning to identify malicious behavior without burying it in thousands of false positives.

Don’t forget to look at the data, too. Know what your sensitive data is, where it is, and who can locate it. This can only be accomplished through complete visibility into all data stores. Create a robust data access permission policy, and keep up with it as employees and vendors come and go.

The six most common oversights that enable data breaches

1. **No Multi-factor Authentication (MFA).** There is no good reason why an organization is not using an MFA. The more often an outsider is unsuccessful in using credentials gained through criminal activity, the less likely they are to be a threat to your sensitive data.
2. **Limited visibility into all data repositories.** From a single dashboard, your solution should deliver a broad range of data security capabilities, including data discovery, classification, monitoring, access control, risk analytics, compliance management, security automation, threat detection, and audit reporting.
3. **No passwords policy.** Organizations need to train employees on the importance of not duplicating passwords, password management, and not sharing passwords with colleagues, partners or vendors.
4. **Misconfigured data infrastructures.** Every cloud-managed infrastructure is unique and requires a specific skill set to manage properly. Being able to see all cloud-managed data repositories through a single dashboard eliminates the need to maintain configurations for data visibility.
5. **Limited vulnerability protection.** A zero-day vulnerability in a popular piece of code can cause security issues for tens of thousands of organizations. Runtime protection secures your applications from vulnerabilities without leaving your application exposed to potential exploitation.
6. **Not learning from past data breaches.** Use machine learning (ML) to do a rigorous analysis of anomalous behavior to root out malicious activity. ML algorithms can baseline typical access privileged users, send alerts on deviations from that behavior, and keep profiles of how past insiders have breached data.

The data-centric approach to data security

Historically, organizations first focus on their perimeter and endpoints. Strong network and endpoint security, combined with vulnerability management lifecycle toolsets and a mature security operations center, were thought to be the keys to reducing overall risk. However, with insider events occurring more often than external ones, a shift in focus to a data-centric security strategy is clearly warranted.

But what does that really mean? How is a data-centric security strategy different from other approaches? It simply focuses on the lifecycle of the data for which you are responsible. It's about knowing - Where is it? Who is accessing it? How much and how frequently? From where? Why? The more you know about how people should be interacting with your data, the easier it is to detect threats, regardless of their source. Here are the main pillars of a data-centric approach to data security.

- **Gain complete, automatic visibility into all data stores.** You cannot protect what you cannot see. You need 100% visibility so you can see how entities are interacting with your data. When you can see everything from a single dashboard, it is dramatically easier to see something anomalous, no matter if it's an insider or outsider threat.
- **When it comes to your data, know what normal is.** Recognizing normal enables you to configure your security solution to call out only things that are not normal so you can orchestrate a response and remediate efficiently.
- **Don't overwhelm the SOC.** This is what happens when you have poor insight into what constitutes normal interaction with your data. Too many instances seem like potential problems and your SOC wastes a ton of effort investigating false positives - or just as bad: does nothing with the alerts.
- **Plain-language, actionable reporting.** Direct, automated alerts drive effective responses if the information your SOAR gets is clear.
- **Go beyond the platform.** Most security teams use too many tools to protect data. A data security platform may integrate some and not others, and the results are often lacking. A Data Security Fabric integrates seamlessly with the most effective data security tools to provide unprecedented visibility and protection.
- **Train your people.** Data-centric security requires people in the organization to be committed stewards of preventing breaches. Ensure people don't use unsecured public cloud services, follow password policies, and learn to recognize phishing and other social engineering scams.

Imperva Data Security Fabric

How does Imperva Data Security Fabric help customers secure their data? Since we do so much research on cyber threats - not just breaches, but also Bad Bot attack vectors, the evolution of DDoS attacks, and zero-day vulnerabilities. We know a lot more than most about how to avoid these incidents.

Data security today is so much more than logging activity to create audit trails for compliance reporting. Virtually every enterprise that has fallen victim to a high-profile data breach complied. Data security must address the challenges that compliance reporting does not - such as data misuse by insiders. Data security must also address the new reality of the hybrid environment that we discussed earlier. Finally, data security must be automated - there is no bandwidth nowadays to tie so many disparate pieces together to create a security posture without it.

Imperva Data Security Fabric can protect data that traditional data security approaches cannot. The Imperva Data Security Fabric is flexible, scalable, and robust enough to enable any organization to gain visibility into its entire data repository - all structured, semi-structured, or unstructured data - and automatically apply security policies to prevent data breaches.

The Data Breach Matrix

We created a flow diagram to help you better understand the complicated ecosystem of data breaches. It describes the flow of a breach throughout the kill chain steps, from the initial access until the exfiltration.

Initial Access	Discovery / Reconn	Credential Access	Lateral Movement	Privilege Escalation	Persistance	Exfiltration
Application Vulnerability	Network Scan	Leaked / Exposed Credentials	DBLinks	Dynamic SQL	Application Backdoor	SQLi
Exposed Sevice	Network Sniffing	Brute-Force	RCE	User Defined Functions	Create DB User	Dump
Third Party (Supply Chain)		Configuration Files	DB Command Execution		Create OS User	Slow & Low
Malicious Insider		Malware	Read / Write Files			Copy Backups
Phishing						Theft / Lost
						Read / Write Files
						DNS Data Exfiltration

Most data breaches are actually a series of failure points rather than one specific activity that results in a breach. Looking at attacks this way helps us understand the timeline of events, from the initial breach to the actual exfiltration of records.

Databases may be used as a target point for “Data Exfiltration”, or as a penetration point into the organization’s network to get a foothold, deploy malware, etc. Those are two different scenarios, but there is an overlap and common areas in between.

Consider the scenario of Information Disclosure of credentials within a vulnerable web application together with a publicly open database server.

A database with stored credentials that are not encrypted or poorly encrypted can be a fertile jumping point to attack other databases in the organization and can lead to enormous breaches where dozens more databases and the sensitive data within them can be compromised.

The stories are spread from public Github repositories to private S3 buckets, from RCE to Ransom, or from simple and old Web Vulnerability to a massive crypto mining botnet.

Every step in the kill chain can be achieved by different techniques represented as blocks inside each kill chain column.

Summary and recommendations

Traditionally, we've believed the assets within the network perimeter were protected. But as services are digitally transformed, the boundaries of the perimeter have blurred. **The security of an organization is only as strong as the weakest link in the security chain.** Often, the "walls" that protect databases have cracks that allow attackers to put their hands on sensitive data. In many cases, better architecture and cross-organization security practices would do the trick, but those practices are not easy to implement and control. We suggest that organizations implement security for the databases they manage, not just the applications and networks that surround them.

In most of the analyzed breaches, the lack of in-depth security stands out as the main reason. Organizations can reduce the attack surface by securing their database environments. Since a significant number of the attacks target web applications, separating the database server from the application server can make a big difference. Together with a dilution of excessive privileges from key users and strong authentication mechanisms, these practices in combination can help you avoid a data breach.

Protecting your organization's data is a never-ending process; you must always work toward optimizing your security architecture, policies, and practices, both for your assets and employees.

Continuously performing "Discovery and Classification" to locate sensitive information and find security holes is a great way to stay on top of your organization's security posture and eliminate bad practices inside the database environment.

Together with implementing security products like Web Application Firewall **(WAF)** and Imperva Data Security Fabric and "Database Risk Analytics" **(DRA)** it is possible to protect against most of the scenarios presented here.

In the meantime, here are things you should do right now to get on track to mitigate data breach risks:

- **Do not take insider threats lightly.** As we have mentioned, not only are insider threats more prevalent than most of us think, but they are also the most dangerous because they are so difficult to detect and they generally have access to much more sensitive data. Build a multi-faceted strategy for managing them. Use a Zero Trust Model, make the "least privilege principle" your policy, make it someone's job to review privilege levels frequently, and keep tight control. Know what normal behavior is, so you are prepared for anomalous activity.
- **Remember that PII is the real prize for hackers.** As the Imperva Threat Research Team showed us, sensitive personal data is most often stolen because it has the greatest long-term value. It is absolutely critical to have visibility into your sensitive data so you can see what privileged users are doing with it and detect anomalies.
- **Attackers are more sophisticated.** Every attack vector that threatens your data becomes more difficult to detect and stop every day. Your solution must stay on top of the latest hacker behavior. Make sure your solution is scalable, flexible, and continuously building features that match this sophistication.
- **Don't depend on perimeter security.** As threats become more sophisticated and more people are working remotely long-term, shift your focus from securing "outside in" to securing "inside out." The perimeter has and will continue to disintegrate. By making your security strategy data-centric, you are following data through its lifecycle and finding out in real time what users are doing with it. Continuous deep analysis of how privileged users are interacting with your data repositories will tip you off to more damaging data breaches than only paying attention to the perimeter.
- **Find and manage your dark and shadow data.** Many organizations create copious volumes of dark data in the regular course of business, and much of it is sensitive data. You have to lock that down. As for shadow data, you may never be able to bring your DevOps teams to heel on creating shadow data, but you can get a solution that enables you to detect where the data lives so you can secure it.