Стратегия за данни и връзката със сигурността на приложенията

OWASP Sofia

Svetlana Videnova

18/10/2024

План

Какво е стратегия за данни
Примерна архитектура
Паралел със сигурността на приложения
Казуси

Какво е стратегия за данни

Може ли да има дата стратегия без да споменем сигурността?

Кога се използва

Какво включва

Защо е важна

От кого се създава

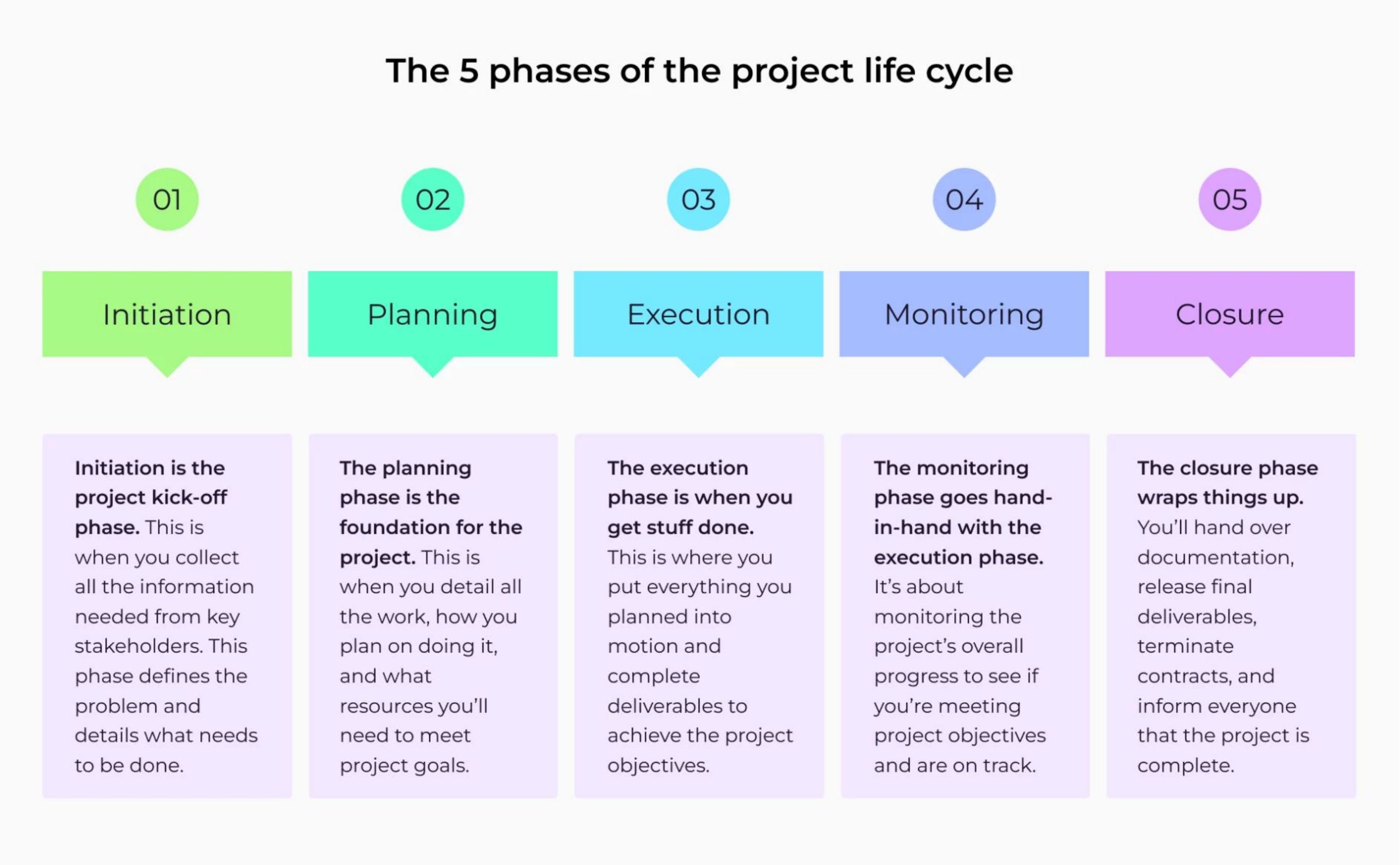
От кой се използва

Примери

Какво е стратегия за данни

Стратегията за данни е цялостен план, който определя как организацията обработва и използва данните си за постигане на своите бизнес цели.

Тази стратегия очертава начините, по които организацията събира, осигурява, анализира и управлява данни през своя жизнен цикъл, като гарантира, че данните се третират като ценен актив. Стратегията свързва дейностите, свързани с данните директно с бизнес целите, помага на организациите да взимат решения за управление на данни, да подобрят оперативните ефективни и да създават нови възможности. Идентифицира план за действие в случай на инциденти, свързани със сигурността или етични проблеми



The 5 phases of the project life cycle

01

Initiation

Initiation is the project kick-off phase. This is when you collect all the information needed from key stakeholders. This phase defines the problem and details what needs to be done.

02

Planning

The planning phase is the foundation for the project. This is when you detail all the work, how you plan on doing it, and what resources you'll need to meet project goals.

03

Execution

The execution phase is when you get stuff done. This is where you put everything you planned into motion and complete deliverables to achieve the project objectives.

04

Monitoring

The monitoring phase goes hand-in-hand with the execution phase. It's about monitoring the project's overall progress to see if you're meeting project objectives and are on track.

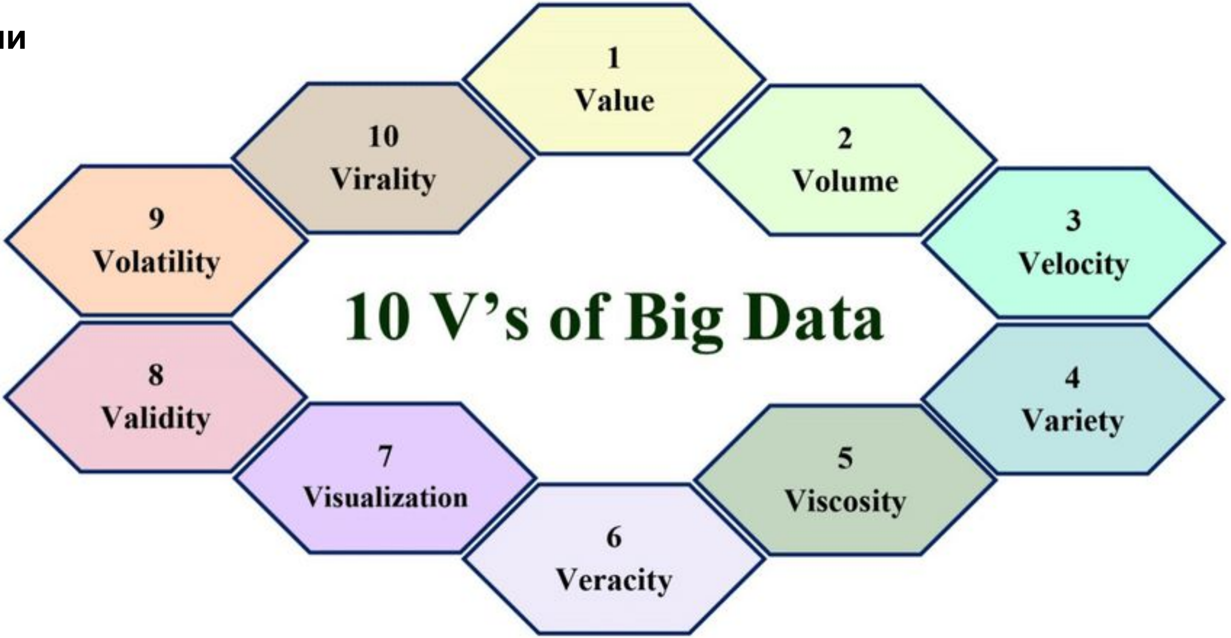
05

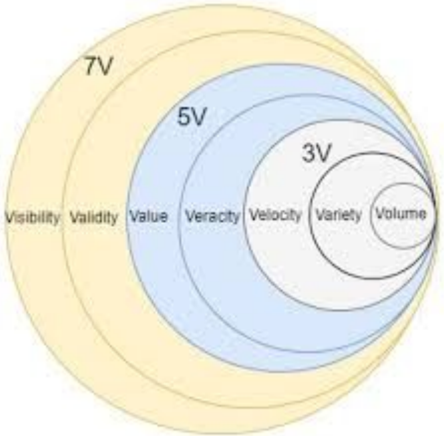
Closure

The closure phase wraps things up. You'll hand over documentation, release final deliverables, terminate contracts, and inform everyone that the project is complete.

Ключови аспекти на стратегия за данни

1. Управление на данни
2. Архитектура на данни
3. Събиране и интегриране на данни
4. Съхранение и управление на данни
5. Анализ на данни и прозрения
6. Сигурност и поверителност на данните
7. Монетизиране на данни
8. Култура на данни и обучение на екипите





John Mashey

Какво включва

Big Data vs Data traditionnelle

1. Volume de données :
- Données traditionnelles : Volumes gérables par des systèmes de bases de données classiques.
- Big Data : Volumes massifs dépassant les capacités des systèmes traditionnels, de l'ordre des téraoctets, pétaoctets ou plus.
2. Variété des données :
- Données traditionnelles : Principalement structurées dans des formats prédéfinis.
- Big Data : Mélange de données structurées, semi-structurées et non structurées provenant de sources diverses.
3. Vélocité :
- Données traditionnelles : Mises à jour périodiques, traitement par lots.
- Big Data : Flux de données en temps réel ou quasi-réel, nécessitant un traitement rapide.
4. Infrastructure :
- Données traditionnelles : Bases de données relationnelles sur des serveurs centralisés.
- Big Data : Systèmes distribués, cloud computing, nouvelles technologies comme Hadoop.

prédéfinis.

- Big Data : Analyses prédictives et prescriptives, utilisation d'algorithmes d'apprentissage automatique.

6. Objectif :
- Données traditionnelles : Répondre à des questions spécifiques, rapports standardisés.
- Big Data : Découvrir de nouvelles corrélations, tendances et insights inattendus.
7. Complexité :
- Données traditionnelles : Gestion relativement simple avec des outils standards.
- Big Data : Nécessite des compétences avancées et des outils spécialisés pour l'extraction, le traitement et l'analyse.

En résumé, la Big Data se distingue par sa capacité à traiter des volumes massifs de données variées à grande vitesse, ouvrant de nouvelles possibilités d'analyse et de prise de décision que les approches traditionnelles ne permettent pas.

Синхрон на инициативите за данни с бизнес целите

Оптимизира ресурсите: Помага да се възползвате максимално от технологиите, таланта и самите данни.

Подобрява взимането на решения: точни, навременни анализи.

Намалява риска: Намалява рисковете, свързани със сигурността на данните, поверителността и съответствието.

Насърчава иновациите: Отключва възможности за иновации чрез разкриване на анализи и активиране на нови продукти или услуги, управлявани от данни.

Защо е важна

От всеки свързан с проекта - колкото повече гласове се чуят толкова по-ефикасна ще е

Data Leadership & Strategy

- Chief Data Officer (CDO)
- Chief Analytics Officer (CAO)
- Head of Data Science
- Data Strategy Consultant

Data Governance & Compliance

- Data Governance Manager
- Data Steward
- Compliance Officer (Data Privacy/Protection Officer)
- Data Risk Manager

Data Engineering & Architecture

- Data Engineer
- Data Architect
- ETL Developer
- Big Data Engineer
- Cloud Data Architect
- Database Administrator (DBA)

Data Science & Analytics

- Data Scientist
- Machine Learning Engineer
- AI Specialist
- Data Analyst
- Business Intelligence Analyst (BI Analyst)
- Quantitative Analyst (Quant)

Software Development & Application

- Data Product Manager
- Software Engineer (Data Applications)
- API Developer
- DevOps Engineer (for Data)

Data Visualization & Reporting

- Data Visualization Specialist
- BI Developer
- Dashboard Designer

Business & Stakeholder Roles

- Business Analyst
- Product Owner (Data Products)
- Stakeholders (e.g., department heads, executives)

Data Operations & Management

- Data Operations Manager
- Data Quality Analyst
- Master Data Manager
- Metadata Manager
- Data Wrangler

Security & Privacy

- Data Security Officer
- Cybersecurity Analyst (Data Focus)
- Data Privacy Analyst

Project Management & Support

- Data Project Manager
- Scrum Master (for Data Projects)
- Data Analyst Intern
- Data Consultant

Specialized Roles

- Natural Language Processing (NLP) Engineer
- Geospatial Data Analyst
- Data Ethics Specialist

От кого се създава

От всеки свързан с проекта или даже извън него

Data Leadership & Strategy

- Chief Data Officer (CDO)
- Chief Analytics Officer (CAO)
- Head of Data Science
- Data Strategy Consultant

Data Governance & Compliance

- Data Governance Manager
- Data Steward
- Compliance Officer (Data Privacy/Protection Officer)
- Data Risk Manager

Data Engineering & Architecture

- Data Engineer
- Data Architect
- ETL Developer
- Big Data Engineer
- Cloud Data Architect
- Database Administrator (DBA)

Data Science & Analytics

- Data Scientist
- Machine Learning Engineer
- AI Specialist
- Data Analyst
- Business Intelligence Analyst (BI Analyst)
- Quantitative Analyst (Quant)

Software Development & Application

- Data Product Manager
- Software Engineer (Data Applications)
- API Developer
- DevOps Engineer (for Data)

Data Visualization & Reporting

- Data Visualization Specialist
- BI Developer
- Dashboard Designer

Business & Stakeholder Roles

- Business Analyst
- Product Owner (Data Products)
- Stakeholders (e.g., department heads, executives)

Data Operations & Management

- Data Operations Manager
- Data Quality Analyst
- Master Data Manager
- Metadata Manager
- Data Wrangler

Security & Privacy

- Data Security Officer
- Cybersecurity Analyst (Data Focus)
- Data Privacy Analyst

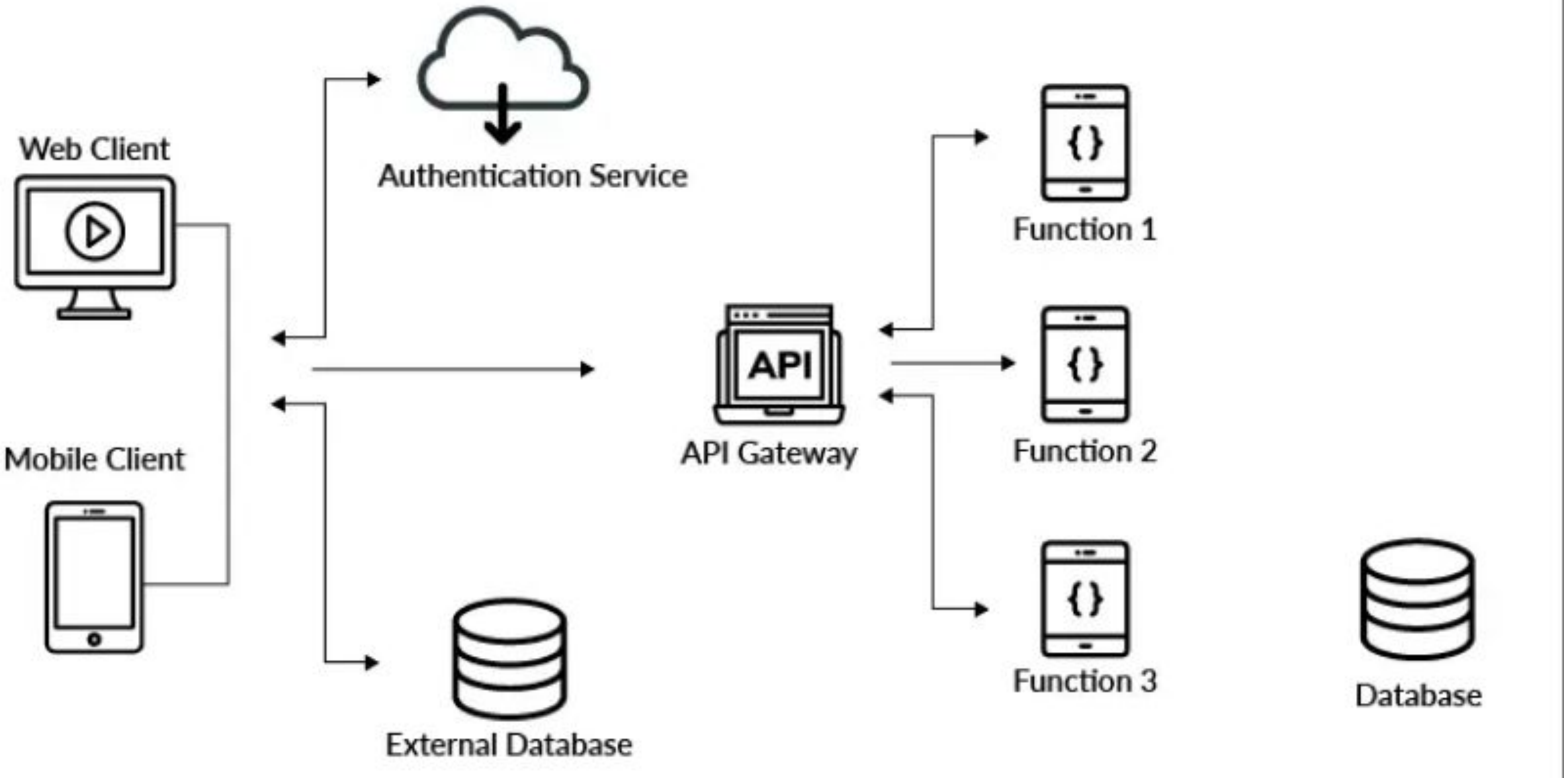
Project Management & Support

- Data Project Manager
- Scrum Master (for Data Projects)
- Data Analyst Intern
- Data Consultant

Specialized Roles

- Natural Language Processing (NLP) Engineer
- Geospatial Data Analyst
- Data Ethics Specialist

От кой се използва



Типична архитектура на приложение

Заплахи
и
План на действие

Силни механизми за криптиране и контрол на достъпа.

Стабилни политики и обучение за управление на данни.

Редовни одити на сигурността и оценки на уязвимостта.

Многофакторно удостоверяване (MFA) за достъп до чувствителни данни.

Техники за маскиране на данни, анонимизиране и псевдонимизиране.

Непрекъснато наблюдение за подозрителна дейност и вътрешни заплахи.

Съответствие с разпоредбите за поверителност на данните и най-добрите практики.

Несигурно архивиране на данни

Злоупотреба с данни или свръхсъбиране

Неупълномощен достъп PPI

Нарушения на поверителността на данните GDPR

Неадекватно управление на данните
Липса на ясни политики, роли и отговорности при управлението и защитата на данните, което води до лошо качество на данните

Липса на криптиране на данни

Фишинг и социално инженерство

Злоупотреба с данни или свръхсъбиране

Загуба на данни
Случайно изтриване или повреда на данни

Shadow IT

Изтичане на данни
Нежелано излагане на данни чрез несигурни системи, приложения или API, което води до достъп до чувствителна информация или споделяне с неупълномощени трети страни.

Вътрешни заплахи
Служители или партньори, които злоупотребяват с достъпа си до данни

Лошо качество на данните

Неправилното управление на задържането и изтриването на данни може да доведе до запазване на чувствителни данни по-дълго от необходимото, което увеличава риска от излагане или нарушение.

Манипулиране на данни
Умишлено или неумишлено манипулиране на данни което води до неточен анализ и вземане на решения въз основа на неверни данни.

Несигурна интеграция на данни
Несигурните връзки между системите по време на обмен на данни, особено в мулти-облачни или хибридни среди

Рискове на трети страни

Проблеми със суверенитета на данните
Трансгранични трансфери на данни, които могат да нарушат местните закони за пребиваване на данни или да доведат до правни усложнения в случай на нормативно неспазване.

Ransomware атаки

Липса на анонимизиране на данните

Сигурност на приложения

- Офлайн приложение
- Пен тест
- Инфосек
- Одит / оценка на сигурността

Стартегия за данни

- Етика на данните
- Идентификация и защита за нарушени данни
- План на действие
- Планиране и намаляване на риска

OWASP principles

- The principle of Least Privilege and Separation of Duties
- The principle of Defense-in-Depth
- The principle of Zero Trust
- The principle of Security-in-the-Open

Интеграция в стратегията за данни

- Управление на данни (Data Governance)
- Архитектура
- Анализи

За да се гарантира, че данните са едновременно полезни и защитени.

Шифроване, контрол на достъпа, маскиране на данни, сигурно подреждане/боравене/изхвърляне на данни.

Uber Data Breach (2016)

Problem: Uber's breach exposed the personal information of 57 million users and drivers. The breach was concealed for over a year, and attackers gained access through compromised credentials for cloud storage.

How a Strong Data Strategy Could Have Helped:

- **Cloud Security**: A well-thought-out data strategy would include cloud security measures like multi-factor authentication (MFA) and stronger encryption for sensitive data stored in the cloud.
- **Data Encryption and Masking**: A data strategy focused on securing personally identifiable information (PII) could have reduced the breach's severity by encrypting or masking user and driver data.
- **Incident Response Plan**: A data governance framework with a well-defined incident response plan would ensure timely disclosure and faster recovery.

Takeaway: By embedding security in every aspect of its data strategy, Uber could have mitigated both the breach’s impact and the subsequent damage caused by the delayed response.

Facebook–Cambridge Analytica Scandal (2018)

Problem: Cambridge Analytica harvested the personal data of 87 million Facebook users without consent. The data was used to influence political campaigns and elections, raising concerns over data privacy and security.

How a Strong Data Strategy Could Have Helped:

- **Data Privacy by Design**: A well-structured data strategy emphasizing privacy by design would have ensured that user consent and privacy settings were embedded into Facebook’s data-sharing practices.
- **Data Governance and Auditing**: Strong data governance policies could have included regular audits of third-party access to user data, identifying and addressing misuse much earlier.
- **Access Control and Permissions**: Facebook’s data strategy should have implemented tighter control over how external developers could access and use data, ensuring that data was only used for authorized purposes.

Takeaway: The scandal highlighted how a failure in data governance and third-party oversight can lead to massive breaches of user trust, damaging a company’s reputation and leading to regulatory consequences.

Sony Pictures Hack (2014)

Problem: A cyberattack on Sony Pictures exposed sensitive employee information, internal emails, and unreleased films. The attack was allegedly linked to geopolitical motives related to the release of a controversial movie, "The Interview."

How a Strong Data Strategy Could Have Helped:

- **Data Segmentation**: A good data strategy would involve segmenting critical data and sensitive internal communications from less sensitive data, making it harder for attackers to access everything in a single breach.
- **Incident Response and Recovery**: A well-defined incident response plan, integrated into the data governance framework, would have enabled faster response and recovery, potentially minimizing the damage.
- **Data Retention Policies**: The breach revealed embarrassing old emails and documents, suggesting the lack of proper data retention and deletion policies. A strong data strategy could have enforced regular data purging to reduce the amount of sensitive information at risk.

Takeaway: The Sony hack demonstrates the importance of having a proactive data strategy that incorporates strong segmentation, retention, and response policies, especially in industries handling highly sensitive information.

Yahoo Data Breaches (2013 and 2014)

Problem: Yahoo suffered two of the largest data breaches in history, exposing the information of 3 billion accounts. The breach included names, email addresses, and hashed passwords, with poor incident management leading to delays in disclosure.

How a Strong Data Strategy Could Have Helped:

- **Proactive Incident Detection**: A robust data strategy that prioritized real-time monitoring and incident detection could have identified the breach sooner, reducing exposure.
- **Stronger Password Encryption**: Yahoo's data strategy should have focused on using stronger encryption techniques for passwords, rather than outdated hashing methods like MD5, which are easily compromised.
- **Breach Disclosure Protocols**: A comprehensive data governance framework would have included clear guidelines for the prompt and transparent disclosure of breaches, minimizing reputational damage.

Takeaway: Yahoo's failure to protect and respond adequately highlights the need for organizations to integrate advanced encryption, monitoring, and breach response mechanisms into their data strategy.

LinkedIn Data Breach (2021)

Incident: A hacker scraped data from 700 million LinkedIn profiles, which included names, email addresses, phone numbers, and professional information. This data was made available for sale on the dark web.

Application Security Issue:
LinkedIn's application lacked sufficient protection against **API scraping**. The breach was not due to unauthorized access to internal systems but to the abuse of publicly available data through LinkedIn's API. While the data scraped wasn't private in nature, the lack of rate limiting or detection of abnormal API usage led to a large-scale data breach.

Link to Data:
Though the data accessed wasn't confidential, LinkedIn still faced scrutiny because sensitive information like emails and phone numbers were exposed, potentially leading to phishing attacks and identity theft.

Takeaway: Application security needs to address not only preventing unauthorized access but also protecting data that could be misused if aggregated at scale. API security measures such as rate limiting, user activity monitoring, and behavior analysis could have mitigated this breach.

MyFitnessPal Data Breach (2018)

Incident: MyFitnessPal, owned by Under Armour, experienced a data breach exposing 150 million user accounts. The attackers obtained usernames, email addresses, and hashed passwords.

Application Security Issue:
The breach was linked to a **vulnerability in password hashing algorithms** used by the MyFitnessPal application. Though the passwords were hashed, they were hashed using the outdated MD5 algorithm, which is vulnerable to brute-force attacks. Attackers could easily crack many of the passwords by exploiting the weak hashing method.

Link to Data:
Weak password hashing compromised the security of user data, making it easier for attackers to reverse the hashed passwords. Despite the initial layer of protection, the choice of algorithm was insufficient to protect sensitive information.

Takeaway: Strong hashing algorithms like bcrypt or Argon2 are essential for securely storing passwords. This example highlights the critical need to regularly update and assess the cryptographic standards used in applications to protect user data.

Slack OAuth Token Incident (2015)

Incident: Slack suffered a breach where attackers accessed a database containing OAuth tokens used for integration with external applications. These tokens allowed access to users’ Slack accounts and the data within them.

Application Security Issue:
The breach was caused by **insecure handling of OAuth tokens**, which are used by third-party applications to authenticate users. Attackers gained access to these tokens and were able to use them to access private messages, files, and other sensitive data within Slack accounts.

Link to Data:
This incident demonstrated how insecure application token management can lead to unauthorized access to user data. The exposure of OAuth tokens gave attackers a direct path to sensitive information stored within the Slack application.

Takeaway: Secure handling of authentication tokens and the use of short-lived tokens or multi-factor authentication (MFA) can help reduce the risk of unauthorized access. Application security is critical to protecting data linked to third-party integrations and authentication mechanisms.

Въпроси

Благодаря

Svetlana (Lana) Videnova
lana.videnova@gmail.com
https://www.linkedin.com/in/lana-videnova/
https://videnova.com/