Improving Cybersecurity in Hospital Information Systems through Anonymization Techniques

Jakub Rapšík
University of Zilina
Žilina, Slovakia
rapsik@stud.uniza.sk

Michal Kvet
University of Zilina
Žilina, Slovakia
Michal.Kvet@uniza.sk

***Abstract*— In an era of increasing cybersecurity threats, protecting sensitive patient information in hospital information systems (HIS) is critical. This paper analyzes and tests various anonymization techniques within a HIS developed for healthcare. Techniques such as generalization, k-anonymity, pseudonymization, and data masking were evaluated for their effectiveness in mitigating data leakage risks while maintaining system performance. The findings highlight the importance of balancing security with operational efficiency, showing that anonymization enhances data privacy but can introduce performance reduction. These results offer a practical approach for securing HIS without compromising service delivery.**

Keywords— Data Anonymization, Hospital Information System, Cybersecurity, Data Security,

I. Introduction

Over the last years the healthcare industry has been quick to digitize due to the increasing necessity to track patients and to better deliver healthcare. At the center of these changes is the HIS, a comprehensive system used to regulate the administrative, financial, and clinical activities of healthcare facilities. The HIS is an integrated information system that includes patient data, clinical charts, diagnostics, medication management and billing. There are, of course, many advantages to digitalizing all these processes, but there are also serious privacy and security issues.

With so much personal information in the HIS, security of confidentiality, integrity, and availability of data is of the utmost importance. Patient records are very personal, with demographic information, medical history, diagnostic results, and plans for treatment. Disclosure of such data to unauthorized persons could lead to serious consequences, including identity theft, financial loss, and damage to patient trust. Also, health care providers must adhere to very strict regulatory frameworks, for example the European Union's General Data Protection Regulation (GDPR), which requires the protection of the data and any anonymization techniques to be appropriate.

One of the main methods used to preserve confidential information in health care is anonymization. Anonymization techniques involve the conversion of identifiable patient data into non-identifiable information to preserve patient privacy yet still allow health care providers to utilize the data for secondary uses such as research and quality improvement. But doing anonymization in HIS is not an easy task either, all because of the delicate balance between data utility and privacy.

This paper will discuss the anonymization of sensitive data in HIS and how it affects system performance. This paper will also cover cyber threats to hospitals such as SQL injection, denial of service (DoS), brute force attacks, phishing and social Engineering attacks. It is these types of threats that give rise to a very strong need for robust security mechanisms, especially as HIS systems are becoming more vulnerable to cyberattacks. Using this approach, an overview of different anonymization methods will be presented, the performance characteristics of each method will be discussed.

II. Cybersecurity Threats In His

As the use of HIS continues to expand, it is crucial that their security is also enhanced. Not only because these systems are subject to hacker attacks due to the nature of the information they handle, medical records, patient data, treatment plans. The following are the major security threats to the confidentiality, integrity and availability of HIS.

A. SQL Injection Attack

SQL injection is a harmful method that enables attackers to alter the SQL query of a web application by injecting malicious SQL code. This manipulation allows attackers to gain unauthorized access to a database, retrieve sensitive data, alter or delete data, and perform administrative tasks. The vulnerability primarily occurs due to inadequate validation of input and the combination of user-supplied data with SQL commands, which allows attackers to execute arbitrary SQL statements. For example, a basic login form on a website could be vulnerable to SQL injection if it fails to properly validate user input. [1]

SQL injection is one of the most common attack in healthcare information systems (HIS), exploiting vulnerabilities in healthcare database software or web interfaces. These attacks can lead to the illegal viewing of patient information, alteration of medical records, or worse, fake billing information. The impact is especially serious with HIS because if data integrity is compromised, it can result in misdiagnosis, improper treatment, or financial fraud. The vulnerability of these systems to SQL injection attacks therefore not only compromises the security of the data, but also directly jeopardizes patient care and hospital operations.

B. Distributed Denial-of-Service (DDoS) Attacks

DoS or DDoS attack is also one of the most common attack. DDoS attack works on the principle of overwhelming the data flow of a website, causing it to stop and then crash. The attacker most often sends fake requests that the server

cannot process. Effective DDoS attacks require a large number of computers functioning in a parallel configuration under the control of the attacker. DDoS attacks are the most accessible attacks on the Internet, primarily because of their availability. It is possible to find many sites that offer DDoS attacks as a paid service.

A distributed denial of service (DDoS) attack poses a significant threat to the security and integrity of patient health data and the overall functionality of healthcare services. The attack can severely disrupt the capacity and performance of the healthcare network, leading to potential harm to patients and hindering the delivery of critical medical services. [2][3]

C. Brute Force Attacks

A brute force attack is a type of cyberattack that tries every possible combination of a password or other phrase that an attacker wants to crack. Although this method is simple, it does not necessarily mean that the password is easy to crack. The time required to crack a given password depends on the complexity of the password or phrase rules defined.

The pattern with 2 lowercase letters and 4 numbers can be cracked in about 20 seconds. However, an improved pattern with length of 8, although cracked in a single attack, took 10 hours and 20 minutes. A more detailed attack covering all possible arrangements of letters and digits separately would take less than one second to complete. An incremented attack up to six characters in length took 2 minutes and 12 seconds and resulted in a large number of cracked passwords. A separate attack focusing on digits only, ranging from seven to twelve characters, took 3 minutes and 18 seconds. [5]

D. Phishing attack

Phishing is a form of cyberattack in which the attacker (phisher) aims to acquire personal information from the victim through different types of phishing attacks. The target of a phishing attack can be an individual user or multiple organizations or institutions. The attacker's primary objective is to obtain personal data, such as login credentials, PIN codes, credit card details, and other information that can be used to commit identity theft or financial fraud.

Based on HIMSS, the top security concern is often phishing. General email phishing is mentioned by 71% of the participants, followed by spear phishing at 67%, phishing/voice phishing at 27%, whaling at 27%, hacking company email at 23%, SMS phishing at 21%, phishing websites at 20%, and social media phishing at 16%.[3][6]

All staff members are eligible to undergo security awareness training in order to mitigate the risk of falling victim to phishing attacks within healthcare institutions. This training should be conducted regularly through various methods such as training sessions, webinars, communication, and reminders, rather than just once or once a year. In addition, multifactor authentication is a mandatory requirement for accessing software and IT services. This means that even if an employee provides their login credentials, attackers will still need to go through an additional, more secure authentication process. Endpoints should be safeguarded with anti-malware software, as some phishing attacks may succeed through endpoint detection and response.[3][7]

E. Social Engineering attacks

Social engineering is a powerful tool for compromising health information systems. Attackers can gather information about employees at medical facilities by observing their online behavior, social media presence, work environment, and casual interactions. This information can then be exploited by attackers to target these individuals.[3]

After acquiring enough details about a specific target, malicious individuals can develop highly customized attacks that are often difficult to identify. Within the healthcare sector, where employees are frequently under significant time constraints, these strategies can be especially impactful. An attacker might masquerade as a trusted coworker, supplier, or government representative, tricking the target into revealing confidential login details or patient information. Because of the urgency often linked with healthcare choices, staff members may be more inclined to disregard security procedures, particularly when responding to what seems to be an urgent or crucial demand.

Attackers commonly use pretexting or also known as Physical Breach Attacks as a method to manipulate targets into revealing sensitive information. For example, they may pose as IT staff in urgent need of access to fix a system issue. In the healthcare industry, where time-sensitive data is prevalent, personnel may be inclined to grant access without thoroughly verifying the authenticity of the request if the scenario seems plausible.[4]

Another threat in healthcare environments is tailgating, where unauthorized individuals gain physical access to restricted areas by closely following an authorized employee. This tactic can lead to unauthorized individuals accessing areas where sensitive information, such as patient records or networked medical devices, are stored.

Social engineering also capitalizes on trust and human error through vishing and phone impersonation, where attackers pose as insurance representatives, government officials, or even patients seeking medical advice. In these cases, healthcare workers may inadvertently share sensitive data due to a misplaced sense of urgency or obligation to assist. Addressing these threats requires regular training programs and awareness-raising campaigns.

III. Data Anonymization Techniques in His

In the context of HIS, safeguarding patient information is paramount. The process of anonymizing data is crucial as it enables the analysis of medical data for research, operational, and administrative purposes without compromising patient privacy. By employing anonymization techniques, HIS can ensure compliance with regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), both of which mandate stringent privacy protection. Through anonymization, personal identifiers are removed, significantly reducing the risk of re-identifying individuals.

A. Anonymization Techniques

The implementation of various anonymization techniques in HIS ensures that data remains valuable for analysis while protecting the privacy of individuals. The most used techniques are data masking, pseudonymization, differential privacy, K-Anonymity

1) Generalization

Generalization is a fundamental technique used in data anonymization, particularly in healthcare systems such as the HIS. This process involves replacing specific, detailed data

with broader, more generalized values, making it more challenging to trace the data back to an individual. For example, instead of displaying an exact age, such as 42 years, the age could be represented as a range, like 40-50 years. Similarly, specific geographical locations, such as postal codes, could be generalized to broader areas, such as cities or regions.

The primary goal of generalization is to preserve the usefulness of the data while reducing the risk of re-identification. For instance, a healthcare dataset might need to maintain some level of detail about patient demographics to be valuable for medical research. Generalization enables this by finding a balance between data anonymity and utility.

One reason why generalization is often considered the easiest anonymization technique to implement is that it primarily involves grouping data into broader categories. This simplicity makes it computationally efficient and easier to integrate into existing data processing systems. For healthcare systems like HIS, which handle large volumes of sensitive patient information, generalization can be an effective first line of defense in protecting privacy while maintaining the functionality of the data for operational and analytical purposes.

However, the trade-off with generalization is that excessive broadening can reduce the data's usefulness, as overly generalized data may lose its relevance or accuracy for decision-making and research. Despite this, in many scenarios, such as for healthcare databases, the ease of implementation and relatively low computational cost make it a popular choice for anonymization.

2) K-Anonymity

K-anonymity technology was introduced by Sweeney in 1998. It mandates that a minimum of k records are indistinguishable based on identifiers in the released data. This prevents attackers from singling out specific individual privacy information, thus safeguarding individual privacy. K-anonymity sets a maximum acceptable risk of information leakage for a user through the parameter k. While K-anonymity offers some protection for individual privacy, it also reduces data availability.[8][9]

K-anonymity often builds on generalization as one of its primary methods. Generalization transforms specific values into broader categories, and K-anonymity then ensures that for each combination of quasi-identifiers (e.g., age, gender, location), there are at least k records sharing the same generalized characteristics. Therefore, K-anonymity can be seen as an enhanced or structured version of generalization because it imposes a more formalized requirement: not just that data is generalized, but that it is generalized sufficiently to prevent re-identification of individuals within a group of size k

In a healthcare information system like HIS, patient data may contain quasi-identifiers such as age, postal code, and gender. Through the application of K-anonymity, the system guarantees that for every combination of these quasi-identifiers, there are at least k patients with identical values. If k=5, this implies that within each group of data records sharing similar quasi-identifiers, each patient would be indistinguishable from at least four others. This serves as a safeguard against potential attackers seeking to re-identify individuals by correlating external information with the dataset.

3) Data Masking

Data masking is a key technique used in data anonymization to protect sensitive information by replacing the original data with modified values. Unlike encryption, which necessitates decryption keys to access the original data, data masking alters the data in a manner that renders it irreversible or usable only within specific confines. This makes it especially effective for scenarios such as testing, development, or sharing datasets, where actual data is unnecessary, but the dataset's structure and consistency are still needed. There are various types of data masking, each serving a specific purpose.

a) Static Data Masking: This method entails concealing data within a dataset stored in a database, creating a permanently modified dataset for use in non-production environments. It is frequently used when duplicating databases for testing or analytical purposes. The original data remains unaltered in the production environment.

b) Dynamic Data Masking: Dynamic masking takes place in real time and modifies the data as it is accessed, unlike static masking. It is a valuable technique for concealing sensitive data from specific users without making changes to the original database. For example, in a healthcare information system, a user could access patient data, but certain sensitive fields such as Social Security numbers or diagnosis codes would be dynamically masked to avoid unnecessary exposure.

c) Deterministic Masking: This form of masking guarantees that identical input values are consistently substituted with the same masked values. This is crucial for maintaining referential integrity across various datasets. For instance, by consistently masking a patient's name across different records, data relationships can be preserved without compromising sensitive information.

d) Random Data Masking: This form of masking, as the name implies, substitutes sensitive data with random values. It is beneficial when the masked information does not have to bear any connection to the original data. For instance, random masking can be applied to patient names for the purpose of anonymized research.

Data masking in healthcare offers a crucial advantage by allowing the sharing and utilization of patient data while safeguarding personal information. This helps healthcare institutions adhere to regulatory standards

4) Pseudonymization

Pseudonymization is a commonly employed data anonymization method that substitutes identifiable information in a dataset with pseudonyms, typically randomly generated values or code names. This process allows for the re-identification of individuals under specific controlled conditions, while also enabling the retrieval of the original data, when necessary, provided that the pseudonym mapping (the key) is securely maintained. Unlike full anonymization, where data is irreversibly altered, pseudonymization retains the potential for data recovery.

The primary pseudonymization techniques employed in healthcare involve identifier replacement and hashing, with or without an additional key, commonly referred to as a "salt." Identifier replacement substitutes data identifying an individual with a unique identifier, such as a monotonic counter, which does not directly disclose the individual's

identity. On the other hand, hashing replaces this data with a distinct cryptographic value computed by a one-way (hash) function, either over the data alone or over the data and an additional key, or "salt." Additional pseudonymization techniques utilized in healthcare encompass tokenization and encryption.[10][11][12]

5) Differential Privacy

Differential privacy offers robust protection against the re-identification of individuals in datasets, making it particularly valuable for safeguarding sensitive data in healthcare information systems (HIS). In contrast to conventional anonymization methods like generalization or k-anonymity, which often involve altering or eliminating data points to preserve privacy, the differential privacy framework aims to mathematically ensure individuals' privacy while still permitting valuable data analysis. The fundamental concept behind differential privacy is to introduce randomness into the results of data queries, thereby preventing attackers from deducing whether a specific individual's information is included in the dataset.

Adding noise to data is a widely used method to protect differential privacy. This technique involves introducing random values, typically drawn from a Laplace probability distribution to each data point in a dataset. By doing so, it helps to safeguard the privacy of individuals. Noise addition is commonly applied to aggregate data, such as the sums or counts of individual-level data. It is important to note that adding noise to individual data could potentially distort crucial information about the individual, so it is primarily reserved for aggregate data.[13]

In the realm of health information systems (HIS), the concept of differential privacy is especially valuable in situations where healthcare data needs to be shared with researchers or public health organizations for analysis without compromising the confidentiality of patient information. For instance, public health data may be shared for the purpose of studying disease trends, treatment effectiveness, or allocation of resources. By employing differential privacy, it is ensured that the shared data retains its statistical integrity for analysis, while thwarting attempts by malicious parties to trace specific data points back to individual patients.

IV. Performance Impact of Anonymization in HIS

In the previous section, we discussed the application of several anonymization techniques within the HIS, which we developed to safeguard patient data and test anonymization methods. Each approach, whether it involves generalization, k-anonymity, pseudonymization, or differential privacy, comes with its own set of strengths and limitations. However, these techniques also bring about varying degrees of computational complexity and impact the overall performance of the system. The following analysis assesses the performance implications of these anonymization techniques based on their implementation and the scale of data processed in our HIS.

In our performance analysis, we evaluated anonymization techniques using a dataset comprising three primary tables from the HIS. These tables include the patient records table, which contains approximately 16,000 entries with detailed personal and medical information, and the patient card and illness tables, which hold additional sensitive data related to medical treatments and diagnoses. We selected these tables for testing due to their reflection of a typical healthcare data environment, where ensuring patient information confidentiality is of utmost importance. The size and complexity of the dataset allowed for a thorough evaluation of the performance impact of different anonymization techniques, especially when applied to large volumes of sensitive healthcare records.

A. Generalization

The concept of generalization, as previously discussed, involves condensing the precision of data by categorizing values into broader groups. In our HIS system, we mainly utilized this method for fields such as age, location, and medical history, where specific details could potentially reveal individuals' identities. For example, instead of specifying an exact age, patients' ages were grouped into broader ranges such as "20-30" or "40-50."

In the HIS, the performance impact of generalization was relatively minimal compared to other techniques. This is due to the straightforward nature of grouping data, which entails replacing specific values with generalized categories and does not significantly increase computational load. The time complexity of generalization in data retrieval is close to O(1), as it does not require additional cryptographic or search operations. However, for extensive datasets, especially when multiple fields are generalized, there may be a slight increase in query processing time to accommodate broader data categories.

In HIS, the performance reduction due to generalization was calculated to be approximately 5-7%, especially in scenarios involving large-scale queries on anonymized fields. However, this trade-off is generally considered acceptable given the ease of implementation and low risk of data integrity violations.

B. K-Anonymity

K-anonymity, an improvement of generalization, ensures that each record cannot be distinguished from at least k-1 other records. In our HIS, k-anonymity was utilized to prevent individual patients from being re-identified based on combinations of quasi-identifiers such as age, gender, and diagnosis.

The performance reduction introduced by k-anonymity was more significant compared to generalization. This is because k-anonymity often requires additional steps, such as iterating through the dataset to ensure that the k-anonymity condition is met for each record. For example, the system may need to repeatedly check if a group of records with similar attributes (such as the same age group and diagnosis) has at least k identical entries. This necessitates frequent scanning of the dataset, resulting in a higher computational cost.

In HIS, the impact of k-anonymity on performance was calculated to be around 10-12%, depending on the size of the dataset and the k value chosen. Larger values of k naturally result in higher computational costs, as more records need to be grouped together to satisfy the anonymity condition. However, this technique significantly enhances privacy protection, making it a valuable option for safeguarding sensitive health information.

C. Pseudonymization

In the HIS, pseudonymization was extensively utilized to safeguard patient identifiers, such as names, social security

numbers, and medical record numbers, by replacing them with unique codes or pseudonyms while storing the original data separately in a secure location. The replacement of each identifier with a unique pseudonym ensured that accessing the pseudonymized data would not enable re-identification of patients without access to the mapping table.

The performance impact of pseudonymization in HIS was considerable, resulting in an estimated 15% decrease in performance. This was mainly due to the additional computational steps needed for creating and managing pseudonyms. Each time data was accessed or updated, the system had to consult a mapping table to convert pseudonyms back to their original values, increasing the time complexity of data access operations. The time complexity for pseudonymization is typically O(N) due to the necessity of traversing the dataset and preventing collisions, ensuring that no two different identifiers are mapped to the same pseudonym.

Despite the performance reduction, pseudonymization is a powerful method for safeguarding sensitive data, especially in scenarios where data must be processed or shared without disclosing individuals' identities.

D. Data Masking

Data masking is a method that involves modifying or concealing data in a manner that renders it unreadable or meaningless to unauthorized users, while retaining its utility for specific processes or testing. In the HIS we developed, data masking was implemented for highly sensitive fields such as patient names, social security numbers, and insurance details. The masked data was only accessible to users with proper authorization, such as hospital administrators or medical professionals.

The data masking process typically entails substituting sensitive data with fictitious yet realistic values. In our HIS, we replaced names with randomly selected names from a repository of names and generated new social security numbers with the same structure as the original ones. This approach ensures that even if unauthorized individuals gain access to the database, they will not be able to extract any useful information.

In terms of performance impact, data masking in HIS caused a moderate level of performance impact with an estimated 8-10% reduction in performance. This is primarily because the masked data must be dynamically generated or retrieved whenever a query involving sensitive fields is executed. The complexity of data masking operations is generally O(N), as the system needs to traverse and mask multiple fields for each record.

Although the computational costs of data masking are higher than simpler techniques such as generalization, it is generally faster than pseudonymization or differential privacy, as it does not necessitate the creation or management of mappings or the addition of statistical noise. The trade-off between performance and privacy is quite balanced, making data masking a widely used technique for securing data that still needs to be accessed in certain situations, such as for testing or reporting.

E. Differential Privacy

Differential privacy represents an advanced anonymization technique that safeguards the privacy of individuals by introducing statistical noise to the data, making it challenging for potential attackers to deduce information about specific individuals. Within the HIS, differential privacy was implemented for aggregated data queries, such as determining the number of patients with a particular condition within a specified time period.

While differential privacy offers robust privacy protections, it does entail a more substantial computational reduction compared to other anonymization techniques. This is because the process involves adding noise to the data in a manner that ensures privacy while maintaining the overall usefulness of the dataset. The amount of noise added must be meticulously calculated to strike a balance between privacy and accuracy, necessitating additional computations.

In HIS, the performance impact of implementing differentiated privacy was calculated to be approximately 20%, mainly due to the complexity of incorporating and managing noise in real-time queries. The more frequent or detailed the queries, the greater the amount of noise that needs to be added, further escalating the computational burden. Nevertheless, differential privacy proves highly effective in scenarios where statistical analysis is required without compromising individual privacy.

V. Discussion

The potential consequences of different methods of anonymization used in the HIS were examined by this publication, along with how they affect performance. The results highlight the need for a careful balance between safeguarding data privacy and ensuring efficient operation of healthcare information systems, especially considering the sensitive nature of health data and the growing number of cybersecurity risks.

A. Benefits of Anonymization

The advantages of anonymization methods are numerous. One important benefit is the improved protection of data privacy. Methods such as generalization, k-anonymity, and pseudonymization notably enhance the privacy of patient data. By effectively concealing sensitive information, these techniques safeguard patient identities from unauthorized access, thereby reducing the risk of data breaches.

Anonymization is crucial in maintaining the confidentiality of patient records, particularly in the face of increasing cyberattacks on healthcare systems. Furthermore, the implementation of anonymization is in line with legal and regulatory frameworks, including GDPR and HIPAA. Compliance not only reduces legal risks but also strengthens public trust in healthcare institutions, which are increasingly responsible for protecting patient information. Adopting robust anonymization strategies can help meet these compliance requirements while ensuring that sensitive data is used responsibly.

Another significant advantage is the facilitation of data sharing. Anonymization enables healthcare organizations to share valuable data for research and analytical purposes without compromising patient confidentiality. This collaborative approach promotes advancements in medical research and enhances healthcare outcomes. By making anonymized datasets accessible for academic research, institutions can contribute to the broader healthcare ecosystem, potentially leading to breakthroughs in treatment and patient care.

Moreover, the adaptability of different anonymization techniques allows for customized solutions based on the specific needs of the HIS. For example, generalization can be used for demographic data to prevent the disclosure of specific patient identities, while k-anonymity can ensure that no individual can be distinguished from at least k-1 others in a dataset. This flexibility ensures optimal functionality across various data types and usage scenarios, meeting diverse operational requirements within the healthcare setting.

B. Limitations of Anonymization

It's important to acknowledge certain limitations. Pseudonymization can lead to a performance reduction, while techniques like generalization may introduce latency. This can impact healthcare operations, especially in emergency scenarios. Anonymization can improve privacy but reduce data utility, making it challenging to derive meaningful insights. Implementing advanced techniques like differential privacy may be complex and demand substantial resources, especially for smaller healthcare facilities. Despite anonymization, there's still a risk of re-identification through advanced data analytics or external datasets.

Another important point is that the success of anonymization techniques depends on the knowledge and training of healthcare personnel. Staff members should be well-informed about the significance of data privacy and the anonymization methods utilized in their systems. Fostering a culture of security and awareness can greatly bolster the efficacy of these measures and minimize the likelihood of human error resulting in data breaches.

VI. CONCLUSION

The study highlights the importance of using strong anonymization methods in HIS to improve data privacy while maintaining operational efficiency. The techniques examined, including generalization, k-anonymity, pseudonymization, data masking, and differential privacy, offer major benefits in protecting patient information but also present challenges, such as performance reduction, potential data utility loss, and implementation complexity.

Future work should focus on refining these methods to minimize their performance impacts and maximize data utility. For instance, exploring hybrid approaches that combine different anonymization techniques could lead to better results in terms of both privacy protection and data usability. Additionally, organizations must regularly update their anonymization strategies to effectively counter emerging threats as data analytics continue to evolve.

Furthermore, it is essential for healthcare institutions to invest in training programs that foster a culture of data privacy awareness among staff. A well-informed workforce can act as the first line of defense against potential data breaches and ensure the effective use of anonymization techniques.

Ultimately, a more effective implementation of anonymization strategies will help protect sensitive patient information, thereby reinforcing the integrity and trustworthiness of healthcare systems. By addressing the limitations identified in this study, healthcare institutions can better navigate the complex landscape of data privacy and security. The insights from this research not only stress the importance of anonymization techniques in healthcare but also pave the way for future advancements that prioritize both privacy and performance.

REFERENCES

[1] N. D. Bobade, V. A. Sinha and S. S. Sherekar, "A diligent survey of SQL injection attacks, detection and evaluation of mitigation techniques," *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal, India, 2024, pp. 1-5. [Online], Available: https://ieeexplore.ieee.org/document/10481914

[2] Latif, Rabia, Abbas, Haider, Latif, Seemab, Masood, Ashraf, EVFDT: An Enhanced Very Fast Decision Tree Algorithm for Detecting Distributed Denial of Service Attack in Cloud-Assisted Wireless Body Area Network, *Mobile Information Systems*, 2015, 260594, 13 pages, 2015. [Online], Available: https://doi.org/10.1155/2015/260594

[3] A. M. Mohamad Al-Aboosi, S. N. Huda Sheikh Abdullah, M. Z. Murah and G. S. AL Dharhani, "Cybersecurity Trends in Health Information Systems," *2022 International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates, 2022, pp. 01-04. [Online], Available: https://ieeexplore.ieee.org/document/9995952

[4] R. Salama, F. Al-Turjman, S. Bhatla and S. P. Yadav, "Social engineering attack types and prevention techniques- A survey," *2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, Ghaziabad, India, 2023, pp. 817-820. [Online], Available: https://ieeexplore.ieee.org/document/10140957

[5] L. Bošnjak, J. Sreš and B. Brumen, "Brute-force and dictionary attack on hashed real-world passwords," *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, 2018, pp. 1161-1166 [Online], Available: https://ieeexplore.ieee.org/document/8400211

[6] G. Saira, S. Arvind, and D. Mike, "Cyber-attacks are a permanent and substantial threat to health systems: Education must reflect that," Digital Health. 2022, vol. 8, 2022, [Online], Available: https://www.researchgate.net/publication/361389978_Cyber-attacks_are_a_permanent_and_substantial_threat_to_health_systems_Education_must_reflect_that

[7] R. Abdillah, Z. Shukur, M. Mohd and T. M. Z. Murah, "Phishing Classification Techniques: A Systematic Literature Review," in *IEEE Access*, vol. 10, pp. 41574-41591, 2022 [Online], Available: https://ieeexplore.ieee.org/document/9755138

[8] Y. Xie, Q. He, D. Zhang and X. Hu, "Medical ethics privacy protection based on combining distributed randomization with K-anonymity," *2015 8th International Congress on Image and Signal Processing (CISP)*, Shenyang, China, 2015, pp. 1577-1582 [Online], Available: https://ieeexplore.ieee.org/document/7408136

[9] S. Vijayrani, A. Tamilarasi and M. Sampoorna, "Analysis of Privacy Preserving k-anonymity Methods and Techniques", *Proceeding of the International Conference on Communication and Computational Intelligence*, pp. 540-545, 2010 [Online], Available: https://ieeexplore.ieee.org/document/5738788

[10] I. Basdekis *et al*., "Pseudonymisation in the context of GDPR-compliant medical research," *2023 19th International Conference on the Design of Reliable Communication Networks (DRCN)*, Vilanova i la Geltru, Spain, 2023, pp. 1-6, [Online], Available: https://ieeexplore.ieee.org/document/10108370

[11] European Union Agency for Cybersecurity, Drogkaris, P. and Bourka, A., *Recommendations on shaping technology according to GDPR provisions – An overview on data pseudonymisation*, Drogkaris, P.(editor) and Bourka, A.(editor), European Network and Information Security Agency, 2018, [online] Available: https://data.europa.eu/doi/10.2824/74954

[12] Deploying Pseudonymisation Techniques The case of the Health Sector, 3 2022, [online] Available: https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques.

[13] R. Subramanian, "Differential Privacy Techniques for Healthcare Data," *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, San Antonio, TX, USA, 2022, pp. 95-100. [Online], Available: https://ieeexplore.ieee.org/document/992303