AmiEs-2023

Enhancing Health Data Privacy through Anonymization and Security Techniques

Marios Vardalachakis[1], Haridimos Kondylakis[1], Manolis Tampouratzis[1], Nikolaos Papadakis[1]

[1]Department of Electrical and Computer Engineering (ECE), Hellenic Mediterranean University (HMU)

Heraklion, GR 71004 Greece

(mtp237@edu.hmu.gr, kondylak@hmu.gr, tampouratzis@hmu.gr, npapadak@hmu.gr)

International Symposium on
Ambient Intelligence and Embedded Systems

27 - 30 September, 2023
Sitia, Crete, Greece

Agenda

- Introduction
- Privacy Concerns with Health Data
- *ShinyAnonymizer:* An innovative Approach for Anonymizing Health Data
- Analyzing Encryption and Hashing Techniques
- Recommendation and Best Practices
- Conclusion

Introduction

- ❑ The online storage and utilization of enormous quantities of personal health information have brought about via an age of innovation in healthcare which has altered both medical practices and academic studies.
- ❑ The large database of clinical information, featuring everything from illnesses to genetic info, offers an opportunity to significantly boost medical treatment, however it additionally has significant issues regarding privacy.
- ❑ Recent advances in technology have raised the value of solid health data privacy to the forefront, preserving data security and confidentiality.
- ❑ In addition to basic confidentiality, health data privacy requires ethical values, promotes reliability, and encourages equal access to healthcare for everyone.

Privacy Concerns with Health Data

According to the extremely complex aspect of the information that are involved, the constantly evolving technological surroundings, with the extensive variety of medical systems, ensuring confidentiality of health data is filled with privacy concerns. Some of the privacy concerns are:

•**Data proliferation**: The fast technological advancement of healthcare records and the growth of mobile devices generates an enormous amount of medical information. It is a huge problem to control and secure this data among various devices and platforms while preserving privacy.

•**Data Linkage and Re-identification**: When data is de-identified, it can always able to re-identify someone through combining it with additional data gathered from various sources or by applying modern steps. It is challenging to maintain information's value while offering complete anonymity.

•**Third-Party Sharing:** Health data may be provided with third parties, such insurance providers, scientists, and technological companies. Managing how these entities utilize and secure the data can prove difficult especially if the data moves between borders where various privacy laws applied.

•**Emerging Technologies:** Significant concerns regarding privacy are presented by the utilization of modern technologies like AI and machine learning in the health care sector. Algorithms created with health data may accidentally reveal confidential data about certain individuals.

Role of Anonymization, Hashing, and Data Encryption in HealthCare

Aspects	Anonymization	Hashing	Encryption
Role	Preserving privacy	Ensuring data integrity	Protecting data confidentiality
Function	Removing or altering identifiers	Creating fixed-length hash values	Transforming data into unreadable ciphertext
Use Cases	Research with privacy concerns	Data integrity verification	Secure data transmission and storage
Reversibility	Partially reversible	One-way function; non-reversible	Reversible with proper decryption
Security	Privacy enhancement; risk of re-identification	Data integrity protection; resistant to tampering	Controlled access; safeguarding from breaches
Examples	De-identified research datasets	Verifying file authenticity	Secure communication channels; encrypted files
Challenges	Balancing utility and privacy; re-ID risk	Potential for hash collisions; weak algorithms	Key management; potential performance impact
Ethical Focus	Privacy preservation; data utility for research	Data integrity and authenticity	Data confidentiality and security

ShinyAnonymizer: An innovative Approach for Anonymizing Health Data

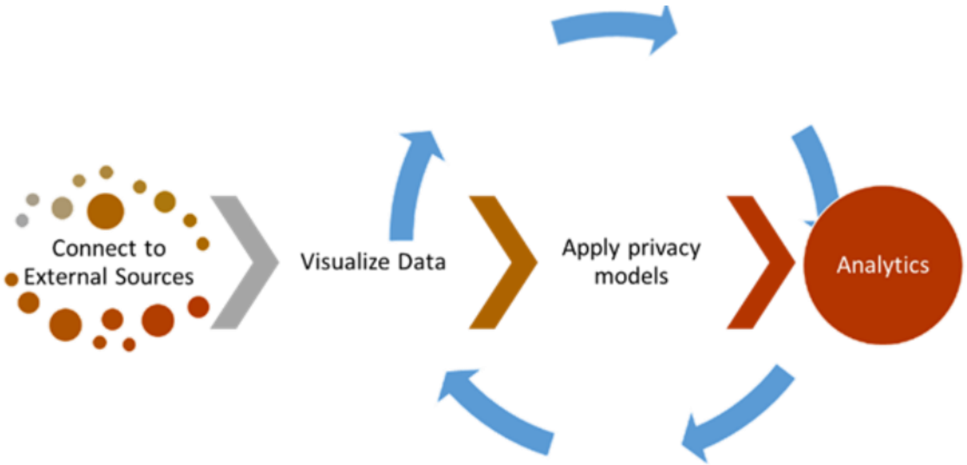
The need for user-friendly anonymization tools has become increasingly evident as organizations and individuals grapple with the complexities of data privacy and compliance. Here are some key points highlighting the necessity of such tools:

•**Rising Privacy Concerns**: In an era of heightened awareness about personal data privacy, individuals and organizations are under pressure to protect sensitive information. User-friendly anonymization tools empower both data controllers and individuals to safeguard their data without requiring extensive technical expertise.

•**Regulatory Mandates:** Stringent data protection regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) impose strict requirements on handling personal information. User-friendly tools enable compliance by simplifying the anonymization process and reducing the risk of non-compliance penalties.

•**Diverse Data Handlers**: Not all users working with data are data scientists or cybersecurity experts. Researchers, analysts, and administrators who lack in-depth technical knowledge also need to perform data anonymization. User-friendly tools cater to this diverse audience, enabling a wider range of users to effectively Anonymize data.

Intro to ShinyAnonymizer



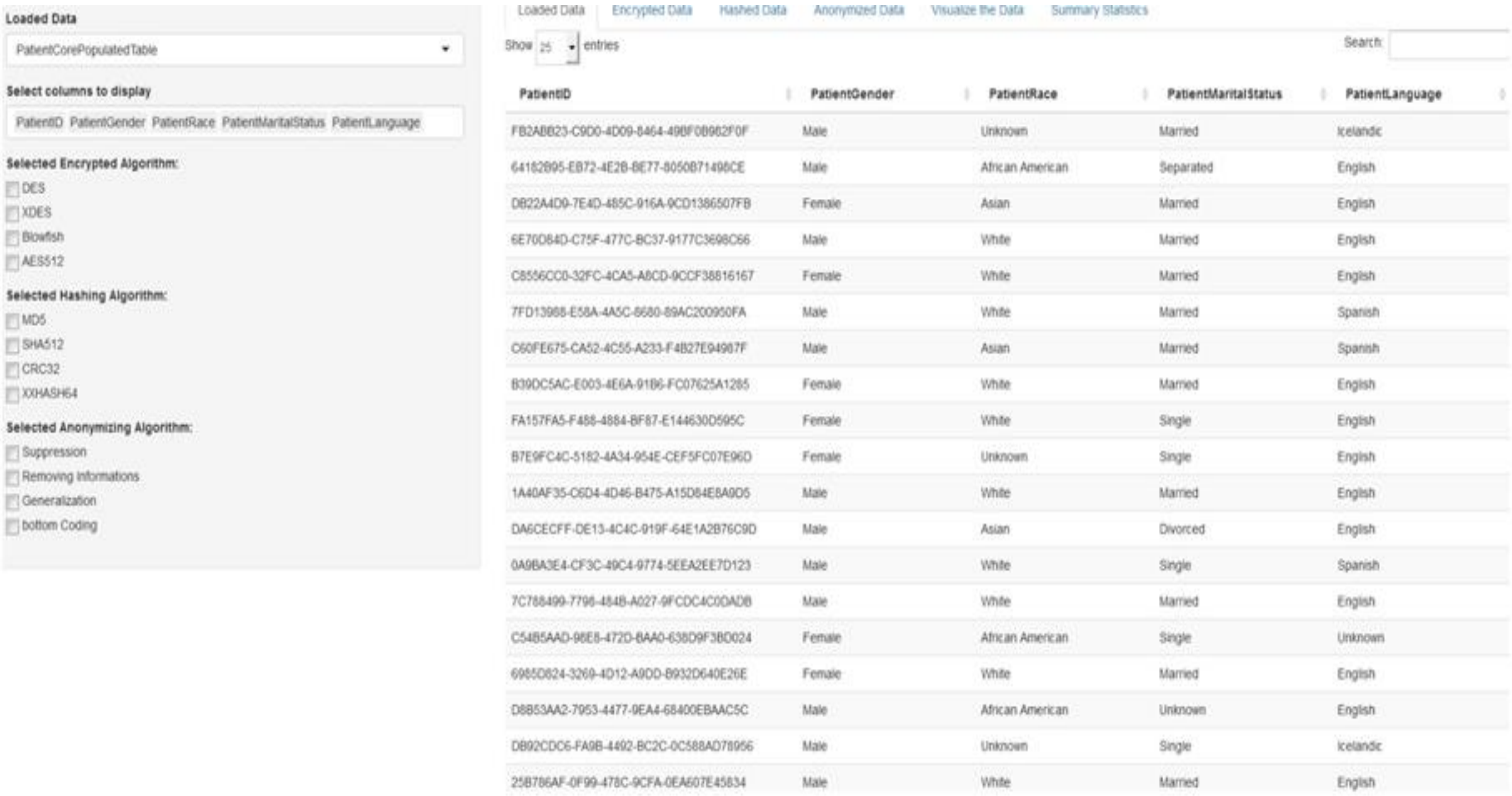
ShinyAnonymizer is a cutting-edge data privacy solution designed to address the critical need for safeguarding sensitive information while maintaining data utility. It is a tool enabling non-expert users to combine of state of the art privacy models :

- **Anonymization** privacy models
- **Hashing** privacy models (MD5,SHA512,CRC32,XXHASH64)
- **Data Encryption** (DES,X-DES,BLOWFISH,AES512)

Multiple data analysis visualization paradigms and statistics.

- Pie charts, bar charts, area charts, histograms and scatter plots

The ShinyAnonymizer Tool Interface



Loaded Data

PatientCorePopulatedTable

Select columns to display

PatientID PatientGender PatientRace PatientMaritalStatus PatientLanguage

Selected Encrypted Algorithm:

- □ DES
- □ XDES
- □ Blowfish
- □ AES512

Selected Hashing Algorithm:

- □ MD5
- □ SHA512
- □ CRC32
- □ XXHASH64

Selected Anonymizing Algorithm:

- □ Suppression
- □ Removing informations
- □ Generalization
- □ bottom Coding

Show 25 entries

Search:

PatientID	PatientGender	PatientRace	PatientMaritalStatus	PatientLanguage
FB2ABB23-C9D0-4D09-8464-49BF0B982F0F	Male	Unknown	Married	Icelandic
64182B95-EB72-4E2B-BE77-8050B71498CE	Male	African American	Separated	English
DB22A4D9-7E4D-485C-916A-9CD1386507FB	Female	Asian	Married	English
6E70D84D-C75F-477C-BC37-9177C3698C66	Male	White	Married	English
C8556CC0-32FC-4CA5-A8CD-9CCF38816167	Female	White	Married	English
7FD13988-E58A-4A5C-8680-89AC200950FA	Male	White	Married	Spanish
C60FE675-CA52-4C55-A233-F4B27E94987F	Male	Asian	Married	Spanish
B39DC5AC-E003-4E6A-91B6-FC07625A1285	Female	White	Married	English
FA157FA5-F488-4884-BF87-E144630D595C	Female	White	Single	English
B7E9FC4C-5182-4A34-954E-CEF5FC07E96D	Female	Unknown	Single	English
1A40AF35-C6D4-4D46-B475-A15D84E8A9D5	Male	White	Married	English
DA6CECFF-DE13-4C4C-919F-64E1A2B76C9D	Male	Asian	Divorced	English
0A9BA3E4-CF3C-49C4-9774-5EEA2EE7D123	Male	White	Single	Spanish
7C788499-7798-484B-A027-9FCDC4C0DADB	Male	White	Married	English
C54B5AAD-98E8-472D-8AA0-638D9F3BD024	Female	African American	Single	Unknown
6985D824-3269-4D12-A9DD-B932D640E26E	Female	White	Married	English
D8B53AA2-7953-4477-9EA4-68400EBAAC5C	Male	African American	Unknown	English
DB92CDC6-FA9B-4492-BC2C-0C588AD78956	Male	Unknown	Single	Icelandic
25B786AF-0F99-478C-9CFA-0EA607E45834	Male	White	Married	English

Key Features and Benefits

Key Features	Benefits
Data Anonymization	Regulatory Compliance
Flexible Techniques	Data Privacy
Preservation of utility	Risk Mitigation
Automated Processing	Ethical Data Use
Customization	Research and Analysis
Audit trails	Customer Trust
Integration Capabilities	Efficiency
User-friendly-Interface	Scalability
	Transparency
	Adaptability

Access To State-of-the-Art Anonymization Techniques

ShinyAnonymizer stands out by providing access to state-of-the-art anonymization techniques that ensure the effective transformation of sensitive data while preserving its usability. The tool offers a range of advanced methods such as masking, suppression, generalization, and perturbation, allowing organizations to tailor their anonymization approach to the specific nature of their data.

Benefits :

- ✓ **Customized Privacy Measures**: ShinyAnonymizer enables organizations to select the most appropriate anonymization technique for each type of data, ensuring that privacy measures are aligned with the sensitivity of the information.
- ✓ **Enhanced Data Utility:** By utilizing cutting-edge anonymization techniques, ShinyAnonymizer ensures that the Anonymized data retains its value for analysis, research, and decision-making. This maintains the utility of the data without compromising privacy.
- ✓ **Adaptation to Regulations:** The availability of diverse anonymization techniques allows organizations to adapt their data processing practices to various regulatory requirements, ensuring compliance with data protection laws such as GDPR, HIPAA, and more.
- ✓ **Mitigation of Re-Identification Risks**: Advanced anonymization techniques reduce the risk of re-identification, ensuring that even if external data is combined, the original identities of individuals cannot be easily uncovered.

Integration of Data Analysis Visualization Paradigms

ShinyAnonymizer goes beyond traditional anonymization solutions by seamlessly integrating advanced data analysis and visualization paradigms. This unique feature allows organizations to not only Anonymize sensitive data but also gain valuable insights through interactive visualizations, enhancing the overall utility of Anonymized datasets.

Benefits

Data Exploration: ShinyAnonymizer integration of data analysis and visualization paradigms enables users to explore Anonymized data interactively. This empowers data analysts, researchers, and decision-makers to uncover patterns and trends without compromising individuals' privacy.

Real-Time Insights: The integration of visualizations within the anonymization process facilitates real-time insights into the Anonymized data, supporting faster and more informed decision-making.

Communication of Findings: Visualizations enhance the communication of findings derived from Anonymized data, making it easier for teams to convey insights to non-technical audiences.

Data-Driven Decisions: By offering dynamic visual representations of Anonymized data, ShinyAnonymizer empowers organizations to make data-driven decisions while upholding privacy standards

Exploring Encryption And Hashing Techniques

ShinyAnonymizer takes data security to the next level by offering a comprehensive exploration of encryption and hashing techniques. These techniques not only ensure data privacy during the anonymization process but also bolster the protection of sensitive information against unauthorized access and breaches.

Benefits:

Strong Data Protection: ShinyAnonymizer integration of encryption and hashing techniques adds an extra layer of security to the anonymization process. This safeguards the data not only from re-identification but also from potential breaches.

Confidentiality: Encryption techniques render the data unreadable to unauthorized users, ensuring that even if the Anonymized data is intercepted, it remains indecipherable and confidential.

Integrity Verification: Hashing techniques allow for data integrity verification, enabling users to confirm that the Anonymized data has not been tampered with during the anonymization process.

Selective Access: Encryption and hashing can be used to control access to specific data subsets. This provides organizations with the ability to share only the necessary information while keeping the rest of the data protected.

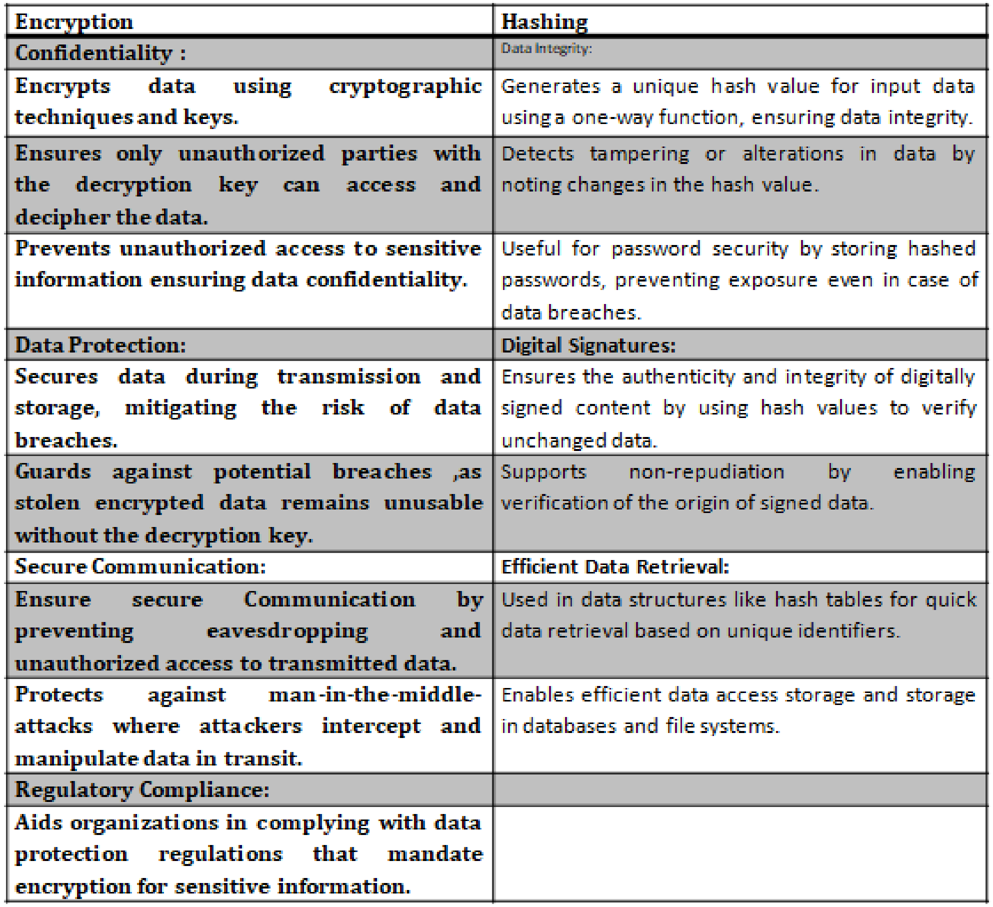
Enhanced Trust: By incorporating encryption and hashing techniques, ShinyAnonymizer builds trust among stakeholders by demonstrating a commitment to robust data security practices.

Compliance with Regulations: The utilization of encryption and hashing aligns with regulatory requirements for securing sensitive information, contributing to compliance with data protection laws and standards.

The role of Encryption and Hashing in Data Security

- Encryption and hashing are fundamental techniques in the realm of data security, playing critical roles in protecting sensitive information from unauthorized access, breaches, and tampering.

- These techniques serve as key components in safeguarding data privacy and maintaining the integrity of digital assets. Here's an overview of the roles of encryption and hashing in data security.

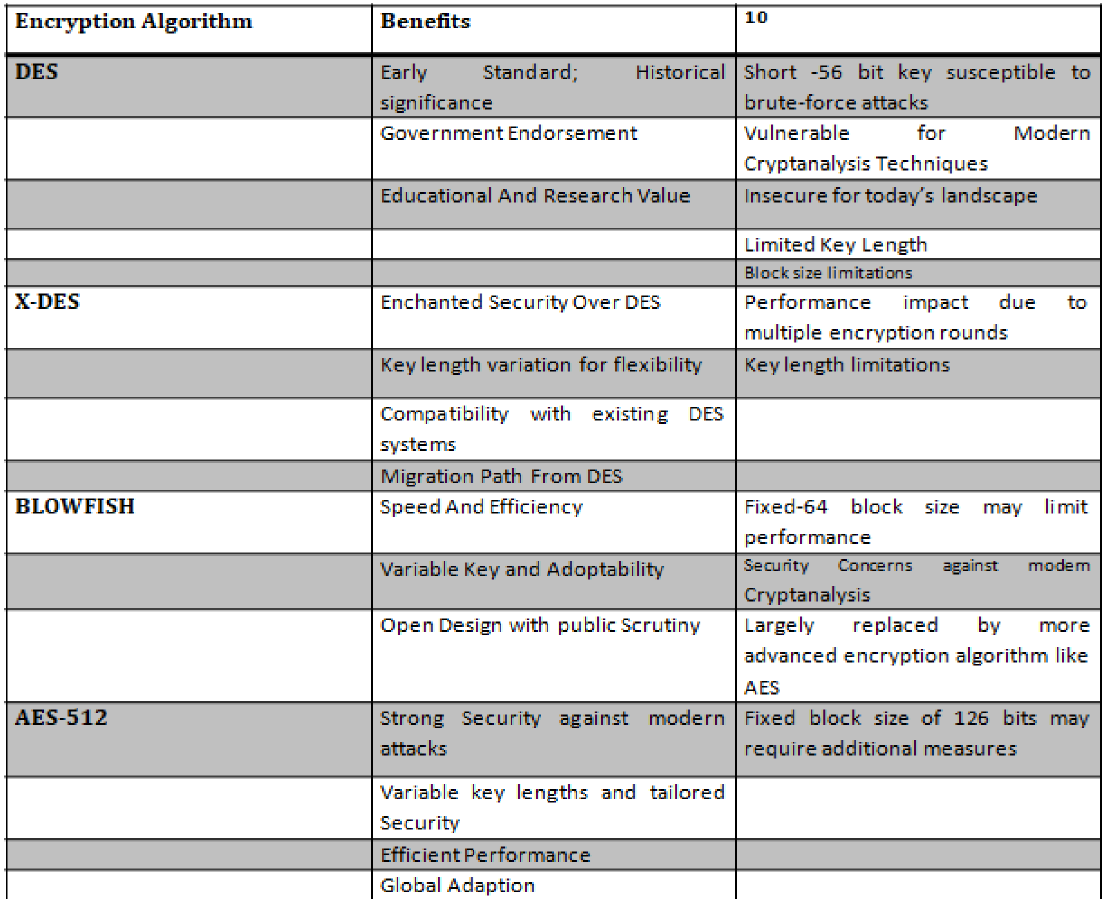


Encryption	Hashing
Confidentiality :	Data Integrity:
Encrypts data using cryptographic techniques and keys.	Generates a unique hash value for input data using a one-way function, ensuring data integrity.
Ensures only unauthorized parties with the decryption key can access and decipher the data.	Detects tampering or alterations in data by noting changes in the hash value.
Prevents unauthorized access to sensitive information ensuring data confidentiality.	Useful for password security by storing hashed passwords, preventing exposure even in case of data breaches.
Data Protection:	**Digital Signatures:**
Secures data during transmission and storage, mitigating the risk of data breaches.	Ensures the authenticity and integrity of digitally signed content by using hash values to verify unchanged data.
Guards against potential breaches ,as stolen encrypted data remains unusable without the decryption key.	Supports non-repudiation by enabling verification of the origin of signed data.
Secure Communication:	**Efficient Data Retrieval:**
Ensure secure Communication by preventing eavesdropping and unauthorized access to transmitted data.	Used in data structures like hash tables for quick data retrieval based on unique identifiers.
Protects against man-in-the-middle-attacks where attackers intercept and manipulate data in transit.	Enables efficient data access storage and storage in databases and file systems.
Regulatory Compliance:	
Aids organizations in complying with data protection regulations that mandate encryption for sensitive information.	

Encryption Techniques: Benefits and Drawbacks

- Throughout the Anonymization procedure strong encryption techniques are employed to enhance data security and privacy in ShinyAnonymizer. Various algorithms for encryption maintain essential data safe against illegal access and attacks.

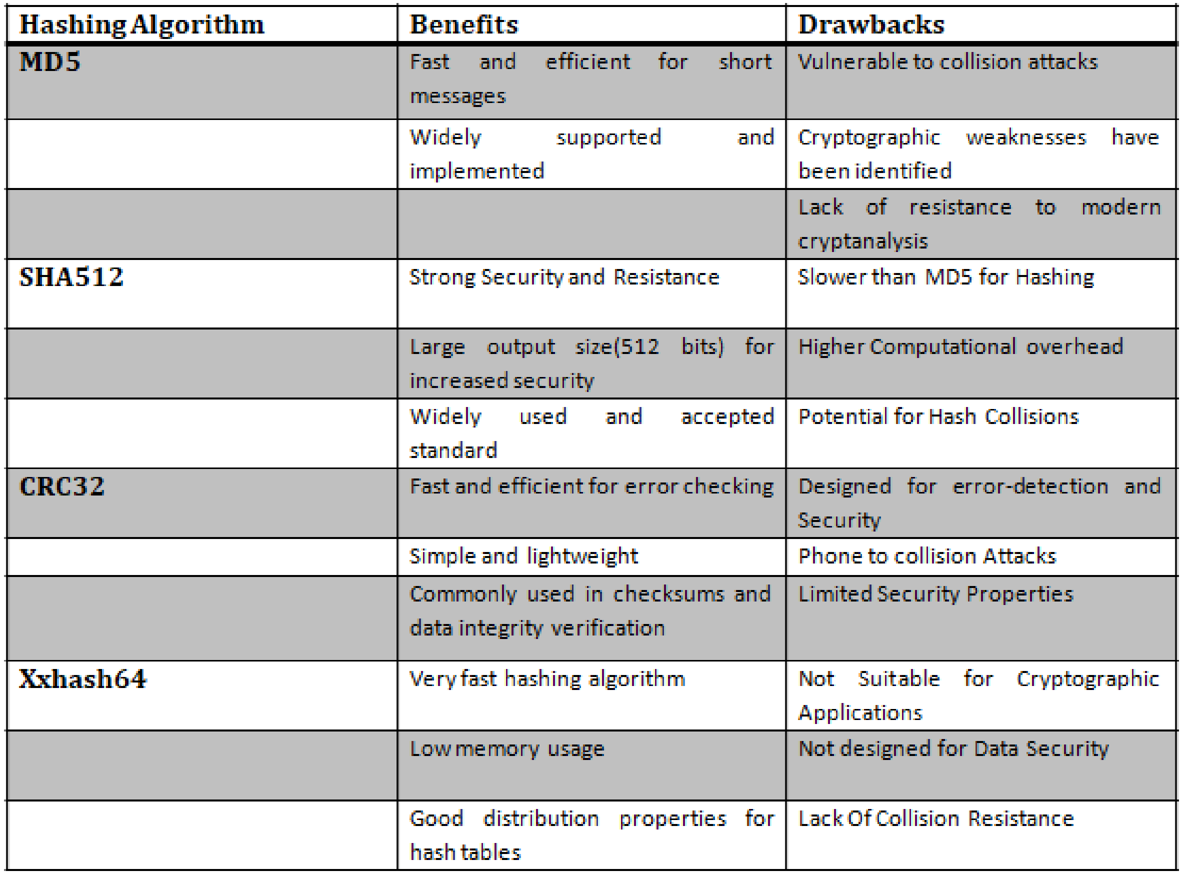
- ShinyAnonymizer also employs symmetric encryption to secure data, using a single secret key for both encryption and decryption. This technique is well-suited for protecting data privacy within the anonymization process.



Encryption Algorithm	Benefits	10
DES	Early Standard; Historical significance	Short -56 bit key susceptible to brute-force attacks
	Government Endorsement	Vulnerable for Modern Cryptanalysis Techniques
	Educational And Research Value	Insecure for today's landscape
		Limited Key Length
		Block size limitations
X-DES	Enchanted Security Over DES	Performance impact due to multiple encryption rounds
	Key length variation for flexibility	Key length limitations
	Compatibility with existing DES systems	
	Migration Path From DES	
BLOWFISH	Speed And Efficiency	Fixed-64 block size may limit performance
	Variable Key and Adoptability	Security Concerns against modem Cryptanalysis
	Open Design with public Scrutiny	Largely replaced by more advanced encryption algorithm like AES
AES-512	Strong Security against modern attacks	Fixed block size of 126 bits may require additional measures
	Variable key lengths and tailored Security	
	Efficient Performance	
	Global Adaption	

Hashing Techniques: Benefits and Drawbacks

- *ShinyAnonymizer* integrates advanced Hashing techniques to verify the integrity of data and enable fast data retrieval.



Hashing Algorithm	Benefits	Drawbacks
MD5	Fast and efficient for short messages	Vulnerable to collision attacks
	Widely supported and implemented	Cryptographic weaknesses have been identified
		Lack of resistance to modern cryptanalysis
SHA512	Strong Security and Resistance	Slower than MD5 for Hashing
	Large output size(512 bits) for increased security	Higher Computational overhead
	Widely used and accepted standard	Potential for Hash Collisions
CRC32	Fast and efficient for error checking	Designed for error-detection and Security
	Simple and lightweight	Phone to collision Attacks
	Commonly used in checksums and data integrity verification	Limited Security Properties
Xxhash64	Very fast hashing algorithm	Not Suitable for Cryptographic Applications
	Low memory usage	Not designed for Data Security
	Good distribution properties for hash tables	Lack Of Collision Resistance

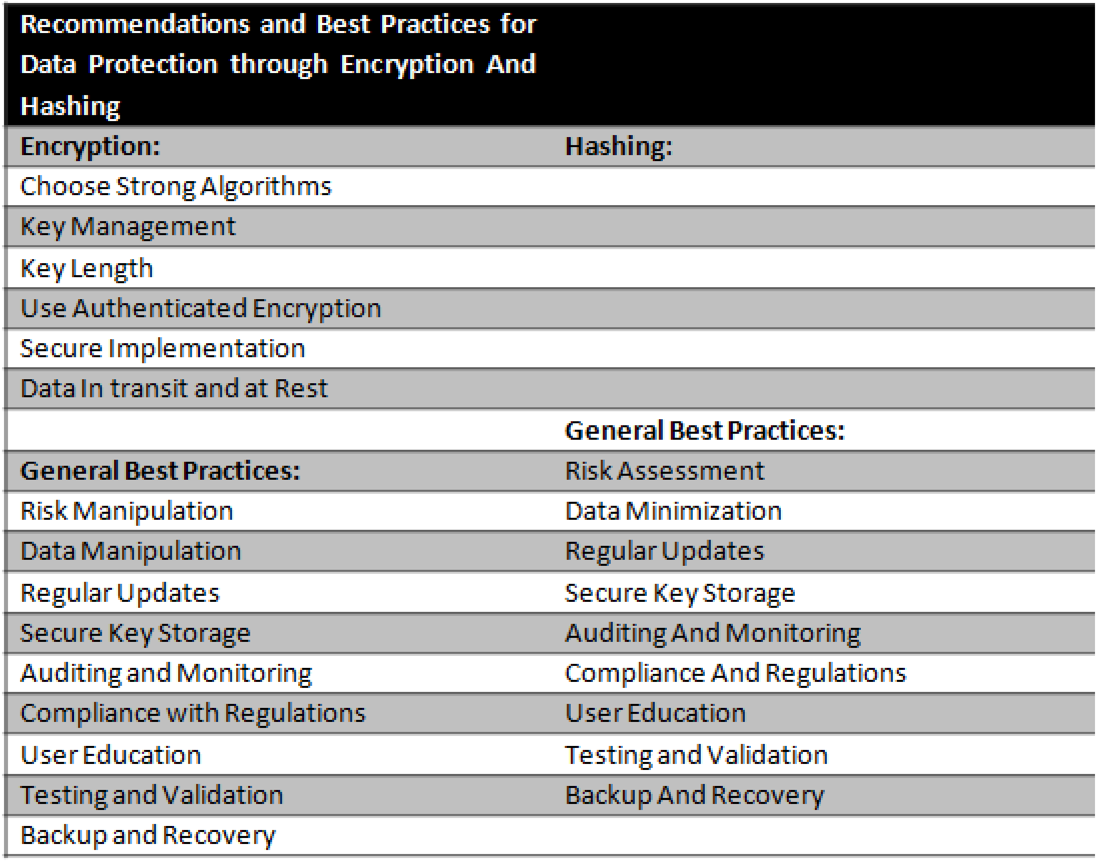
The growing Concern For Personal Data Confidentiality

The dramatic technological advancement of the world, where data is created, obtained, and exchanged on a scale that is unparalleled, is the root of increasing worries for privacy and confidentiality. The risk of fraud, illicit access, and data leaks involving highly personal data are ultimately contributes to all this fear. The concern is mainly brought on with a number of aspects:

- **Data breaches** have revealed thousands of people's info, such names, addresses, credit card numbers, and occasionally health records. Several notable data breaches included big companies, governments, and organizations. The general trust in data security processes gets harmed by such events.
- **Cybercrime:** The increasing number of technological crimes that include phishing, identity theft, and attacks using ransomware highlights how accessible sensitive data is to attackers along with other criminals who seek to profit or damage operations.
- **Persistent Gathering of Information:** In order to enhance their products, services, and promotions, organizations obtain an enormous amount of personal information, notably in the IT and advertisement sectors. Problems concerning the use and protection of the data could come up as an outcome of this gathering.
- **Lack of Controlling:** Individuals usually have little involvement about the way their data is obtained, used, or released. Fear regarding data privacy along with a sense of powerlessness could arise from this absence of visibility.

Recommendations and Best Practices for Data Protection through Encryption and Hashing

•It is essential to comply with identified guidelines and utilize suggested steps when using hashing and encryption algorithms for protecting information.



Recommendations and Best Practices for Data Protection through Encryption And Hashing	
Encryption:	**Hashing:**
Choose Strong Algorithms	
Key Management	
Key Length	
Use Authenticated Encryption	
Secure Implementation	
Data In transit and at Rest	
	General Best Practices:
General Best Practices:	Risk Assessment
Risk Manipulation	Data Minimization
Data Manipulation	Regular Updates
Regular Updates	Secure Key Storage
Secure Key Storage	Auditing And Monitoring
Auditing and Monitoring	Compliance And Regulations
Compliance with Regulations	User Education
User Education	Testing and Validation
Testing and Validation	Backup And Recovery
Backup and Recovery	

Conclusions

- In the ever-evolving landscape of data security, the paramount importance of encryption and hashing cannot be overstated. These two distinct yet synergistic techniques serve as the cornerstones of robust data protection strategies, collectively fortifying the confidentiality, integrity, and authenticity of sensitive information.
- Encryption, with its ability to render data unreadable to unauthorized parties, acts as a safeguard against potential breaches and unauthorized access.
- By transforming plaintext into an incomprehensible format, encryption ensures that only those possessing the decryption key can unveil the original content. This process not only bolsters data confidentiality but also helps organizations meet stringent regulatory requirements, securing their legal standing and reputation.
- Even the slightest modification to the input data generates a distinct hash value, alerting to tampering attempts. Hashing plays a vital role in ensuring the reliability of digital transactions, authenticating digital signatures, and enhancing password security.
- Combining, encryption and hashing within data protection strategies results in a multi-faceted defense mechanism. Encryption safeguards data at various stages, encompassing storage, transmission, and processing. Concurrently, hashing verifies the integrity of data, acting as a sentinel against tampering and unauthorized alterations. This harmonious fusion of techniques creates a fortified fortress of security, contributing to the establishment of trust among stakeholders and the mitigation of potential risks.

Bibliography

[1] Vardalachakis, Marios, et al. “ShinyAnonymizer: A Tool for Anonymizing Health Data”, 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE), pp. 325-332 (2019).

[2] Vardalachakis M., Kondylakis H., Tampouratzis M., Papadakis N. “Anonymization, Hashing and Data Encryption Techniques: A Comparative Case Study” 3rd International Conference on Mathematics and Computers in Science and Engineering (MACISE 2023), Ierapetra, Crete, Greece, 25-27 August 2023.

[3] Olatunji, Iyiola E., et al. "A review of anonymization for healthcare data." Big data (2022).

[4] Majeed, A., & Lee, S. (2020). Anonymization techniques for privacy-preserving data publishing: A comprehensive survey. IEEE Access, 9, 8512-8545.

[5] Abidalrahman, M., Jararweh, Y., Tawalbeh, L. (2011) AES512: 512-bit Advanced Encryption Standard Algorithm Design and Evaluation. Information Assurance and Security (IAS).

[6] Boland, T., & Fisher, G. “Selection of hashing algorithms”. NIST Technical Papers (June 2000).

Enhancing Health Data Privacy through Anonymization and Security Techniques

Marios Vardalachakis[1], Haridimos Kondylakis[1], Manolis Tampouratzis[1], Nikolaos Papadakis[1]

[1]Department of Electrical and Computer Engineering (ECE), Hellenic Mediterranean University (HMU)

Heraklion, GR 71004 Greece

(mtp237@edu.hmu.gr, kondylak@hmu.gr, tampouratzis@hmu.gr, npapadak@hmu.gr)

Thank you for your Attention!!

Questions ??