



# City Research Online

## City St George's, University of London

**Citation:** Tahir, S., Tahir, H., Tahir, R., Rajarajan, M. & Abbas, H. (2022). Water Is a Viable Data Storage Medium: A Security and Privacy Viewpoint. *Electronics*, 11(5), 818. doi: 10.3390/electronics11050818

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.





**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/28024/>

**Link to published version:** <https://doi.org/10.3390/electronics11050818>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

## Article

# Water Is a Viable Data Storage Medium: A Security and Privacy Viewpoint

Shahzaib Tahir <sup>1,\*</sup>, Hasan Tahir <sup>2,†</sup>, Ruhma Tahir <sup>3,†</sup>, Muttukrishnan Rajarajan <sup>4,\*</sup> and Haider Abbas <sup>1,†</sup>

- <sup>1</sup> Department of Information Security, College of Signals, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; dr.h.abbas@ieee.org
- <sup>2</sup> Department of Computing, School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; hasan.tahir@seecs.edu.pk
- <sup>3</sup> School of Computer Science and Electronics Engineering, University of Essex, Colchester CO3 3SQ, UK; rtahir@essex.ac.uk
- <sup>4</sup> School of Mathematics, Computer Science and Engineering, University of London, London EC1V 0HB, UK
- \* Correspondence: shahzaib.tahir@mcs.edu.pk (S.T.); r.muttukrishnan@city.ac.uk (M.R.)
- † These authors contributed equally to this work.

**Abstract:** The security of IoT devices is a major concern that needs to be addressed for their wide adoption. Users are constantly seeking devices that are faster and capable of holding large amounts of data securely. It is purported that water has memory of its own and the ability to retain memory of the substances that are dissolved into it, even after being substantially and serially diluted. It was also observed in the lab setting that the microscopic pattern of water obtained from the same vessel by different people is unique but can easily distinguish those individuals if the same experiment is executed repeatedly. Furthermore, extensive research is already underway that explores the storage of data on water and liquids. This leads to the requirement of taking the security and privacy concerns related to the storage of data on water into consideration, especially when the real-time collection of data related to water through the IoT devices is of interest. Otherwise, the water memory aspect may lead to leakage of the data and, consequently, the data owners identity. Therefore, this article for the first time highlights the security and privacy implications related to water memory and discusses the possible countermeasures to effectively handle these potential threats. This article also presents a framework to securely store sensitive data on water. The proof-of-concept prototype is implemented and tested over a real-world dataset to analyze the feasibility of the proposed framework. The performance analysis yields that the proposed framework can be deployed once data storage on water is widely used.

**Keywords:** water memory; security; privacy; distinguishability; searchable encryption; DNA



**Citation:** Tahir, S.; Tahir, H.; Tahir, R.; Rajarajan, M.; Abbas, H. Water Is a Viable Data Storage Medium: A Security and Privacy Viewpoint. *Electronics* **2022**, *11*, 818. <https://doi.org/10.3390/electronics11050818>

Academic Editor: Dimitra I. Kaklamani

Received: 12 January 2022

Accepted: 3 March 2022

Published: 5 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Today's world has become a global village where people rely on automation and technology to perform routine tasks. Although the reliance on technology has numerous benefits, including on-demand availability of resources and data, the existing data storage devices cannot cope with the data storage needs that are exponentially increasing. Furthermore, it is estimated that the world's data will hit 175 zettabytes by 2025 [1,2]. Smart devices are generating colossal amounts of data, which are processed and analyzed to make smarter decisions, provide necessary support and optimize functionalities.

The internet of things (IoT) is a technological revolution incorporating the latest technologies in both hardware and software. The IoT has already had a noteworthy impact on multiple societal verticals, such as healthcare [3], smart cities [4], manufacturing [5], home automation [6], communications [7], agriculture [8], etc. The IoT owes its success primarily to innovations in the field of communications and sensing technologies. The IoT ecosystem is positioned to make an even stronger impact as innovators and device

manufacturers conceive new designs and application areas. The world is constantly looking for IoT devices that are smarter, incorporate a wider range of functionality, possess higher resources and are able to inter-operate with one another. While all these factors are very much needed, little emphasis is placed on security and privacy concerns associated with IoT devices. To place IoT devices in every home and office, designers will need to provide security as a fundamental design element. Adversaries are able to take advantage of limited or poor security implementation, thus compromising the security and privacy of individuals [9]. Owing to their pervasive nature, IoT devices store and process large quantities of data. Data are frequently stored, using conventional magnetic storage of semiconductor-based chip storage. Sophisticated technologies are now equally accessible to adversaries, which makes it possible for them to capture device data and communications. To thwart attacks against storage mediums, a novel form of data storage is needed. This leads to a compelling case to have alternate methods of data storage that is able to meet growing demands for data storage with the provision of security. To overcome this problem, the use of liquid and water deoxyribonucleic acid (DNA) to store data has been studied [10].

It is purported that water has the ability to retain memory of the antibodies that come in contact with it, and thus the term “Water Memory” or “Memory of Water” evolved [11]. The study of homeopathy [12] and magnetized water is also based on the belief that water changes its molecular structure, with its ability to remember interactions, and retains this new structure in the future when exposed to external substances, materials or objects. While storing data on liquid or water, an attacker can exploit the property of water memory, leading to security and privacy concerns. Data security in this context means limiting unauthorized access to the data stored on the water, thus dealing with data confidentiality. Privacy refers to safeguarding the user identity or the identity of the individual who has come in contact with water. By default, the security and privacy concerns remain unexplored in the literature.

### *Contributions*

The following contributions are made through this research:

- This article for the first time sheds light on the potential security and privacy risks associated with data storage on the DNA of liquid and water. The importance of this study stems from the fact that if the concerns remain unaddressed, they could lead to data misuse and data theft.
- The countermeasures are also proposed in this article to effectively mitigate the possibility of attacks. A framework is presented to securely store sensitive data on water and presents strategies to counter the security and privacy risks.
- A proof-of-concept prototype is implemented and tested over a real-world dataset to analyze the performance of the proposed framework. The performance results demonstrate the efficiency of the proposed framework, and hence, it can be used in a real-world environment.

The paper is organized as follows: Section 2 highlights the water memory beliefs by explaining the concept of water memory. It also cites the articles in favor of and against the concept of water memory. Section 3 for the first time explains the security and privacy implications associated with data storage on water and water memory. Section 4 presents a novel framework for achieving data security and privacy. Section 5 presents the proof-of-concept prototype and highlights the complexity of the system by testing it over a real-world dataset. The conclusion and future work are drawn toward the end in Section 6.

## **2. Water Memory Beliefs**

Before exploring the evidence of water memory, it is important to define the term “water” in this context. Then we discuss the research studies that are in favor of this concept, followed by the studies that are against the evidence of water memory. This article is primarily based on the proposition of water having memory and brings to light many concerns.

### 2.1. Water in the Context of “Water Memory”

Water is a mineral solvent and has the capacity to dissolve numerous fluids and substances. Real, pure water, free from all foreign molecular species, cannot be produced. This also holds true for distilled and deionized water. Pure water is generally known as neutral water, i.e., pure water has a pH value of exactly 7 (where a pH value between 0 and 7 means the water is acidic; a pH value between 7 and 14 describes alkaline water). Therefore, water is considered to be a very complex structure, especially when used in the context of water memory and data storage on water. Water in the context of water memory is assumed to be a solution of various, varying materials and substances that are dissolved/suspended in the water [13]. The water used in the experiment presented in the next section is tap water.

### 2.2. Evidence in Favor of “Water Memory”

This section discusses the experiments that were conducted to demonstrate the ability of water to store information. This discussion lays the foundation and will help us highlight the security and privacy concerns that are introduced by storing data on water.

#### 2.2.1. Water Exposed to Individuals

Kröplin and Henschel in [14,15] analyzed water droplets under a darkfield microscope with the intention to identify whether water samples can distinguish between the people carrying out the experiment. Figure 1 demonstrates the experiment and the outcome. The experiment was performed between four participants. The experiment is conducted in the same room, and the participants are placed 1.5 m apart to avoid interactions. Every individual uses a one-way syringe to place 4 small drops of tap water on the microscope slide. Once the slides have dried, they are photographed using a dark field microscope and a camera, where the dark field microscope is set to magnification factor 40. The experimental results illustrate how each row corresponds to the experiment carried out by a particular individual. A pattern can be observed within a single sample (row). However, intra-sample (columns) patterns are not similar. Therefore, under this experiment, water has the ability to store individual specific information.

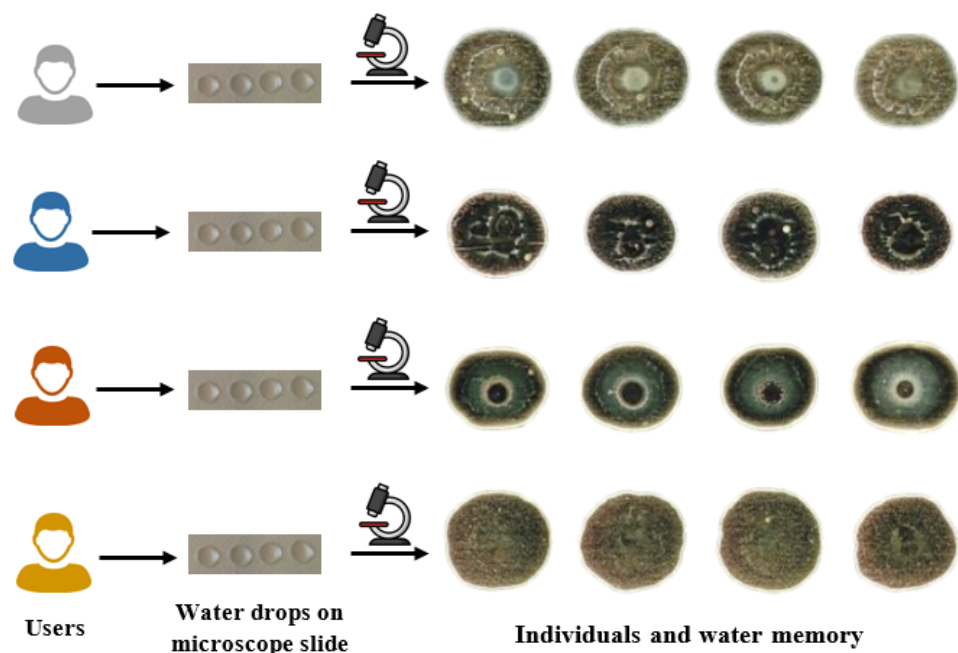


Figure 1. Experiment—water drops on a microscope slide, individuals and water memory.

### 2.2.2. Water Exposed to “Things”

This experiment was conducted by Masaru Emoto [16]. The researcher, through several experiments, demonstrated that water, when exposed to different things, such as music, written words, visual images and photographs, has the ability to absorb, hold and re-transmit human feelings. The author took a drop of water and froze it at  $-20\text{ }^{\circ}\text{C}$ , then analyzed the crystal structure under a microscope. The experimental results are indeed spectacular (as shown in the Figure 2). It was observed that water is a sensitive medium and the crystal structure is affected when water is exposed to harsh words, specific and concentrated thoughts, such as “I will kill you” and “You fool”, or heavy metal music. The pattern is mainly dull and asymmetric in the case of negative exposure. Similarly, if loving words, light music or beautiful pictures are directed toward water, the crystal structure shows a colorful, symmetric snowflake pattern. This experiment clearly indicates that water has the ability to store and re-transmit human feelings.



Figure 2. Experiment—hidden messages.

### 2.3. Counter-Evidence for “Water Memory”

Although the concept of water memory sounds interesting and groundbreaking, it gave rise to much controversy, and the idea was not widely accepted by researchers. The reason behind this idea failing to convince the community was that it was thought that the experiment was against our knowledge on the physics of water, and other teams failed to reproduce the same experiment. Readers are referred to [17] for insight on the counter evidences. The same paper [17] gave a new direction as to the controversy and explained why the experiments could not reproduce the results: “that the outcomes of these experiments were related to quantum-like interference of the cognitive states of the experimenter”. Therefore, a quantum-like probabilistic model would allow to describe how the experiments were carried out.

As we begin to use water as a data storage medium, it is necessary to give importance to even the least possibility of water memory, as it could cause severe security and privacy leakages. Therefore, we consider the aspect where data has the ability to retain memory.

### 2.4. Water and Liquid as a Data Storage Media

It is believed that the liquid hard drive will have the ability to store a terabyte of data in a tablespoon full of liquid [18]. In fact, when liquid and water are examined minutely, there are microscopic particles suspended within them. These microscopic particles can replicate a solid hard drive’s platters by encoding the data in binary and storing it on the DNA strands. DNA computing offers the following advantages [19]:

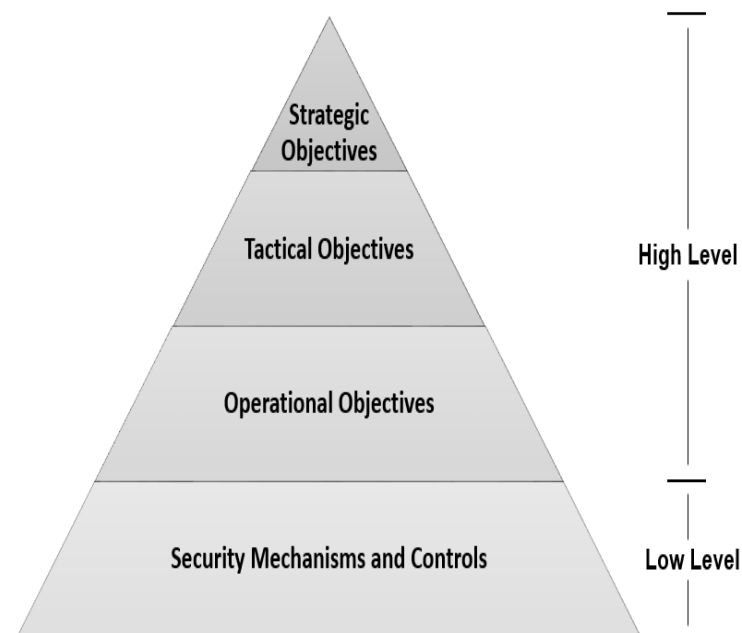
- **Increased performance rate** allows 1026 simultaneous operations per second, enabling the performance of the data storage on the DNA strands to increase at an exponential rate.

- **Low power consumption**, as the increased performance rate allows to operate 10,000 times the speed of today's supercomputers while consuming fewer resources.
- **Incredibly lightweight** due to efficient and lightweight biochemical operations.
- **Increased data capacity** as the DNA has the ability to store 2.2 exabytes per gram, i.e., a single cubic centimeter of DNA holds more information than a trillion CDs.
- **Imperishable storage** as DNA strands have the ability to retain data and remain in a stable state under controlled conditions.

Although data DNA computing and data storage on DNA have many advantages, the security and privacy concerns are yet to be explored in relation to water memory. The meaning of security and privacy in the context of water memory is already discussed in Section 1.

### 3. Data Security and Privacy Implications

This section focuses on the security and privacy implications related to data storage on water. Figure 3 highlights the hierarchy of controls deployed to achieve data security and privacy within an organization. Considering data storage on water, the objectives and controls are broadly categorized into a low level and high level. Low-level areas refer to possible threats whose mitigation strategy may require the modification/tuning of the organization's architecture at the system level. This would help to integrate data storage on water, and achieve data security and privacy. High-level areas are generally those that add value to the security and privacy mechanisms that were put in place at the policy and governance level (low level). Table 1 extends these areas and highlights the possible concerns while storing data on water/liquid and presents some recommendations. The below discussion extends the recommendations further by exploring the strategies that can be adopted to achieve security and privacy, and to contain the risks.



**Figure 3.** Information security controls.

**Table 1.** Possible security and privacy concerns while storing data on water.

Areas	Recommendations
Low Level	
Data Protection	<ul style="list-style-type: none"> <li>• Ensure appropriate methods exist for the protection of data at rest, motion and in use.</li> <li>• Provision methods for data sanitization.</li> </ul>
Identity and Access Management	<ul style="list-style-type: none"> <li>• Ensure security mechanisms for authentication, authorization, access control are implemented and adequate for the business processes and systems.</li> </ul>
Compliance	<ul style="list-style-type: none"> <li>• Understand the laws and regulations applicable, mainly related to data, its security and privacy such as GDPR and HIPAA.</li> <li>• In case an organization is providing water-based data storage-as-a-service be aware of the contract terms mentioned in the service level agreements (SLA)</li> </ul>
High Level	
Governance	<ul style="list-style-type: none"> <li>• Have Standardized policies and procedures</li> <li>• Distribute Management</li> <li>• Define clear roles and responsibilities</li> <li>• Put audit mechanisms in place</li> </ul>
Trust	<ul style="list-style-type: none"> <li>• Deliver confidence to stakeholders by permitting security and privacy controls.</li> <li>• Create a monitoring method that supports decision making, performance monitoring, compliance, value delivery.</li> <li>• Establish a risk management protocol that is adaptive to changing risk landscape.</li> </ul>
Architecture	<ul style="list-style-type: none"> <li>• Establish an understanding of the system and implemented security controls.</li> <li>• Determine the impact of the security controls on the system end to end.</li> </ul>
Availability	<ul style="list-style-type: none"> <li>• Ensure business continuity in the event of loss through availability, backup, data and disaster recovery.</li> </ul>
Data Lock-in	<ul style="list-style-type: none"> <li>• Create methods that reduce data lock-in and promote platform sharing and data portability by limiting vendor specific behavior.</li> </ul>

### 3.1. Low-Level Areas

#### 3.1.1. Data Protection

Data protection refers to implementing security controls for the protection of data owners and data consumers. Ultimately, the level of data protection is dictated by the organization or the standards with which it complies. To be effective, data protection has to be provided by the design. This means the protection by default privacy must be ensured and that it is proactive in nature. The data protection mechanisms should be fully functional and service provision should be end-to-end, while respecting user privacy at the same time. Modern data protection laws also dictate the need for placing privacy controls in the hands of the data owners and not the data consumers. The data protection can be widely achieved through deploying cryptographic mechanisms.

#### 3.1.2. Identity and Access Management

Identity and access management relates to the mechanisms that entail identity proofing and authentication mechanisms. These mechanisms in place allow only an authorized person to access sensitive data stored on a storage medium. As discussed in Section 2, water has its own memory; therefore, it is essential that only an authorized person should be able to access the water and the data stored on it. While storing sensitive data on water, it is very important to have identity and access management in place that governs the user provisioning, maintenance and protection of the sensitive data outsourced to an organization. The traditional data structure, such as access control lists and access control matrix, are able to conform to water-based data storage.

### 3.1.3. Compliance

Before an organization shifts toward storing data on water, it is important to be aware of regulations that apply to the target sector. For example, an organization may require regulation compliance that includes GDPR, FISMA, HIPAA, NIS Directive, etc. For organizations interested in using water as a data storage medium, regulation compliance is crucial, as the organizations may be handling sensitive and private data. For such organizations to be compliant, the understanding of standards helps achieve the regulatory compliance. The standards mainly belong to the family of ISO-27000 that helps manage the security of the assets, or PCI-SSC that standardizes the payment card industry. To fully benefit from data storage on water, an organization should understand the domain-specific regulations and tune/design a water-based data storage system such that the regulatory requirements are met. This would help an organization to effectively manage the compliance risk and with the compliance auditing. The regulatory compliance is also reflected in the service level agreements and contracts between the client, and the client may consider himself/herself to be protected against data leakage and data theft.

## 3.2. High-Level Areas

### 3.2.1. Governance

Governance in terms of security refers to having a set of tools, personnel and business processes in place that achieve the security goals for meeting the organization's needs. Information security governance also defines an organization's structure, roles and responsibilities, performance measurement, defined tasks and oversight mechanisms. An organization planning to use water as a data storage medium must focus on governance, as it adds another layer of accountability for attaining transparency and traceability. The predefined policies, standards and procedures help mitigate the risks and help achieve confidentiality, integrity and availability. Such an organization may rely on frameworks, such as ISO 38500, balanced scorecard, COBIT, systems security engineering capability maturity model (SSE-CMM), etc. [20].

### 3.2.2. Trust

To fully benefit from the services that an organization has to offer requires the customers to have full trust of the organization. Trust is generally subjective in nature and depends upon the type and sensitivity of the data that a user is willing to store. If an organization is planning to provide data storage as a service on water, they would need to have trust mechanisms and service level agreements in place. Although outsourcing has many benefits, it gives rise to the risk associated with insider access and debatable data ownership, which could lead to loss of data control and eventually reduced trust. Cryptographic mechanisms may be deployed to achieve security and privacy. Additionally, threat-monitoring and risk-management protocols should be deployed to effectively reduce the potential threats.

### 3.2.3. Architecture

It is believed that prevention is better than cure. For an organization willing to use water as a data storage medium, it is important to understand the architecture. This mainly refers to the architecture of the software and hardware used to deliver the services of data storage on water to the clients. The architecture is very important, as the clients are generally located at different geographic locations and rely on virtual machines or hypervisors to access the remote resources. By introducing a remote location for data storage, the attack surface may be extended. Thus, a clear understanding of the architecture is required, and controls, such as virtual firewalls, intrusion detection and prevention systems, may be deployed.

### 3.2.4. Availability

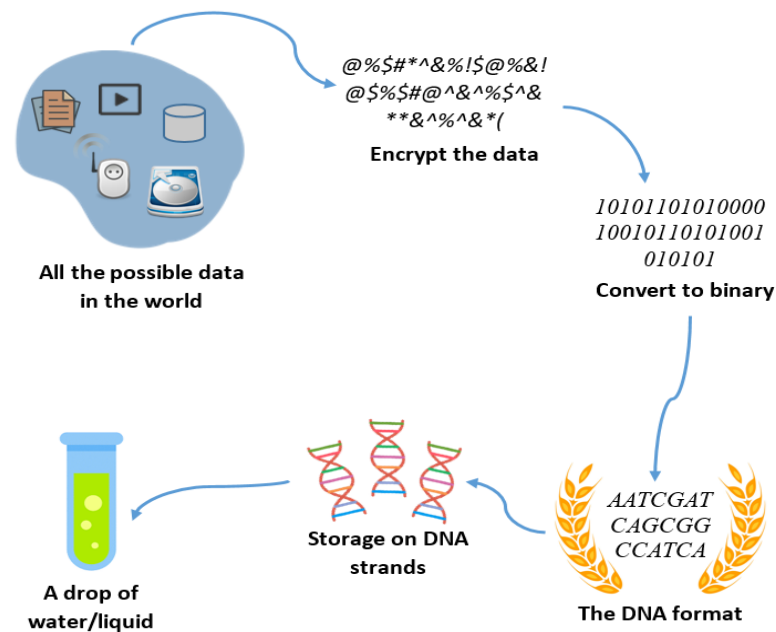
As water-based storage is considered a space-efficient data storage method, enterprises will be willing to use it as a suitable method for off-site and on-site backup. The offering of availability through water-based storage requires that in the event of a disaster, enterprises are able to restore functionality and switch between primary and backup sites with minimal downtime. Other requirements when considering water-based backup is the cost efficiency compared to the benefits brought to the business. The ability to perform an automated switchover and ensuring high system performance post switching is also mandatory.

### 3.2.5. Data Lock-In

Data lock-in [21] refers to the inability of being able to move/migrate data between different storage data structures and the cloud. The movement of data from one storage to another is a necessary requirement when one is switching vendors, performing data backup or is involved in any migration-related activity. The inability to move data means that users are locked into the current data storage method and cannot migrate to and from water-based storage. The prevention of lock-in will require services and APIs that assist in the process of migration. Additionally, proprietary data types and data structures need to be avoided since they are not readily understood by other vendors. This brings to light the need for the standardization of water-based storage data structures.

## 4. Proposed Framework to Achieve Security and Privacy over Sensitive Data Stored on Water

With the help of a scenario, a framework is presented to explain one of the possible routes that may be taken to outsource sensitive data to an organization that uses water-based data storage. This framework concentrates on low-level threats (discussed in the previous section) and presents mitigation strategies. Suppose Bob outsources his sensitive data but he does not trust the data storage service provider. Direct (unencrypted) data storage on water can lead to unauthorized data access and misuse of the data. The general data protection regulations (GDPR) also insist that people should have full control over their data. Therefore, to achieve data protection, avoid data misuse and comply with the GDPR requirements, a method is to encrypt and then store the data on the DNA strands. Figure 4 presents a holistic view of the flow of events that would take place to store the encrypted data on a drop of water or liquid. Firstly, all the data that need to be stored on the water medium are encrypted, and the corresponding hexadecimal values are obtained. The hexadecimal values are converted to binary. The binary values are represented as a sequence of nucleotide bases (As, Ts, Cs, and Gs) and stored as DNA strands that are contained in molecules of water or any other liquid. Later on, the DNA strands may be accessed to retrieve the data stored on it. DNA cryptography is a mathematical and a systematic approach that achieves storage of encrypted data on a DNA sequence.



**Figure 4.** Encrypted data storage on a drop of water/liquid.

#### 4.1. DNA Cryptography

Cryptography is a technique that performs mathematical operations on the data aimed to achieve information security, such as data confidentiality, data integrity and non-repudiation. DNA cryptography is the technique of hiding the data that are stored in DNA strands and sequences. In simpler terms, DNA cryptography relies on the conversion of each alphabet into a different combination of four bases. This is an active area of research, and over the years, different methods to achieve DNA cryptography have been proposed [22,23].

The storage of encrypted data on DNA strands seems to be a feasible solution but gives rise the problem of searching and sifting through the large amounts of encrypted data stored in the DNA sequences. To handle this significant problem, a naive approach would be to decrypt the data and then search over it. However, this approach results in an increase in the computation overhead. Therefore, a mechanism is required that allows on-the-fly searching over the encrypted data. Therefore, searchable encryption is proposed.

#### 4.2. Searchable Encryption on DNA Strands

Searchable encryption [24,25] is a technique widely being explored in the area of cloud computing, where large amounts of data are stored in the cloud. Searchable encryption allows a user to generate a search query (also called a trapdoor) and can delegate the search across the peers involved in a network. A searchable encryption schemes offers the following properties:

- Only an authorized person is allowed to generate a meaningful trapdoor/search query that can be used to perform the search.
- Only an authorized person is able to decrypt the data.

Searchable encryption over a DNA sequence [26] is an emerging area of research, and very little research has been conducted to address this problem. In the context of data storage on DNA strands and water, searchable encryption can be used to achieve confidentiality of the data [27]. Searchable encryption will limit the possibility of data theft and data misuse. Furthermore, modern searchable encryption schemes also preserve the privacy of the data owner and the user performing the search by generating probabilistic trapdoors. A probabilistic trapdoor is a unique search query generated for the same keyword searched repeatedly. Therefore, an attacker maintaining a history of the past

searches cannot uncover the underlying text or the keywords that have been searched. A comparison of different searchable encryption techniques is presented in [28].

#### 4.3. Authentication, Authorization and Access Control

DNA cryptography and searchable encryption primarily achieve confidentiality of the data; however, the interaction of a user with water and water memory itself can lead to serious privacy concerns that need to be addressed separately. Only authenticated entities should be able to access the data. This also ensures non-repudiation and accountability based on secure credentials associated with the individuals. There are several methods to achieve privacy of the data source to restrict unauthorized access to the water/liquid. This framework proposes to integrate technologies specifically designed for enforcing authentication and access control. The framework utilizes the services of an application encryption (AE) server [29] to govern all the access control policies that would be involved in the framework. The AE is able to manage security objects, such as X.509 certificates, symmetric and asymmetric encryption keys and tokens, as well as providing attribute-based access control (ABAC) services based on the XACML standard. For authentication, token-based authentication can be used, which authenticates users with their usernames and passwords, and obtains a time-limited cryptographically secure token upon successful authentication. The token can be used for further authentication for a session of limited duration. Access control is provided through a XACML policy decision point (PDP), which is a decision engine that evaluates user or administrator defined access control policies to provide fine-grained access control to the available resources. An advantage of this authentication and access control approach is that the users can share their tokens with some trusted entities not only for a limited time, but also a limited set of resources, without having to share their usernames, passwords or other sensitive security credentials. This approach is also useful for security evaluation and auditing purposes.

Therefore, the amalgamation of the above-mentioned technologies requires the tuning of the system on low level and helps achieve data protection, identity and access management, and compliance. An organization can build on top of this framework and achieve high-level security objectives that include governance, trust, architecture, availability and data lock-in.

### 5. Performance Metrics

To demonstrate the feasibility of the proposed framework, a proof-of-concept prototype was developed and tested over a real-world dataset. Before discussing the complexity of the proposed framework, the system specifications and dataset description are presented:

#### 5.1. System Specifications

The workstation used for the development and testing is an 11th Generation Intel(R) Core (TM) i5-1135G7 @ 2.40 GHz with 12 GB RAM. The implementation was done in Python and the Homomorphic Encryption library used is HElib version 2.1.0 [30] over the Ubuntu operating system.

#### 5.2. Dataset Description

The prototype was tested over a real-world dataset of large genome databases. The dataset is available at [31] and was already used in DNA-based searchable encryption techniques [17]. The dataset contains 10,000 binary DNA sequences. Each sequence is of 2158 bits long.

#### 5.3. Computational Performance

To demonstrate the performance of the prototype, all the phases of the proposed framework were implemented. However, in future work we will extend the proof of concept to include the last phase, i.e., data storage on water. Figure 5 extends Figure 4

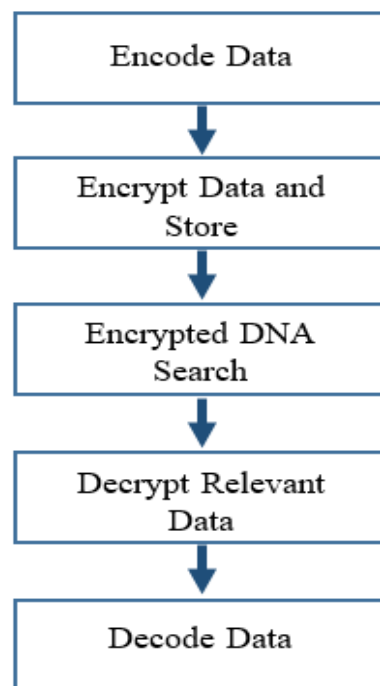
and highlights the processes that are involved. All the phases are discussed, and the performance results are presented below:

### 5.3.1. Encode Data

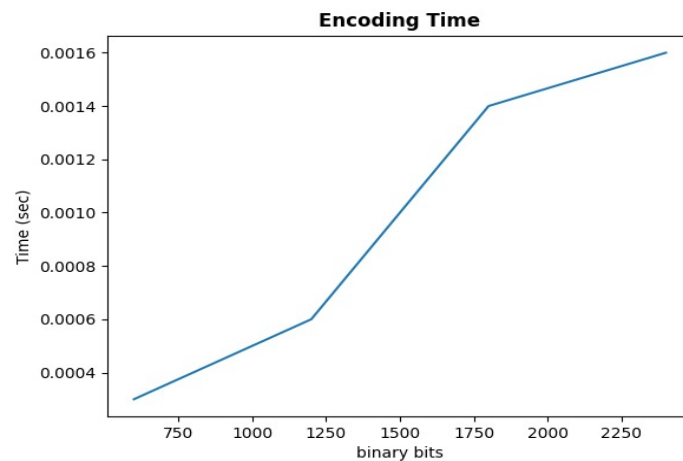
Encoding the data is the process where the data are firstly converted into binary and then in their ASCII representation. In the dataset, each tuple is already converted into bits, so we convert every octet into its decimal representation and then we derive the corresponding character from there. This helps us reduce the number of data segments that were to be encrypted and also reduces the storage overhead. The time required for encoding the bit stream is presented in the Figure 6. For a stream of 2300 bits, the encoding requires a total of 0.0016 s.

### 5.3.2. Encrypt Data and Store

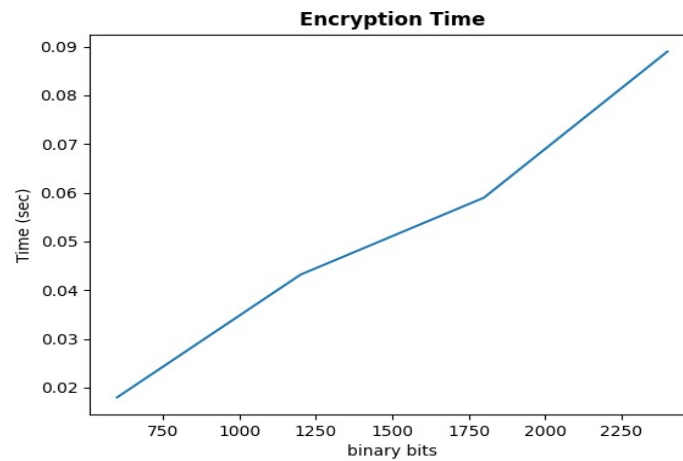
The security is reliant on the encryption of the data. For this purpose, we used fully homomorphic encryption, HeLib. The proposed framework is not dependent upon a particular variant of homomorphic encryption, so we also refer readers to [32] that presents a comparison of the existing schemes. The authors also discussed the key sizes and different security parameters. We used HELib version 2.1.0. The encryption time is represented in Figure 7. It can be seen that for 2300 bits, after encoding, the encryption time is 0.09 bits.



**Figure 5.** Process diagram to securely store the data.



**Figure 6.** Computational complexity for encoding.



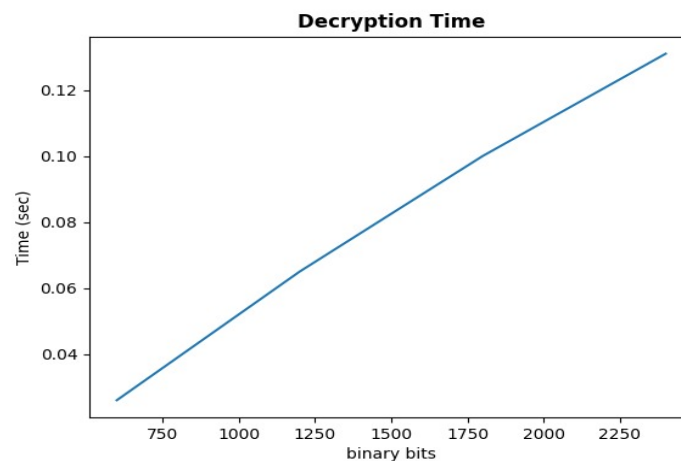
**Figure 7.** Computational complexity for encryption.

### 5.3.3. Encrypted DNA Search

Search over the DNA strand is the most important aspect of the proposed architecture, as it becomes a challenge to identify the required data segments once the data are encrypted. The time required to search over 2300 bits homomorphically encrypted is 206.57 s.

### 5.3.4. Decrypt Relevant Data

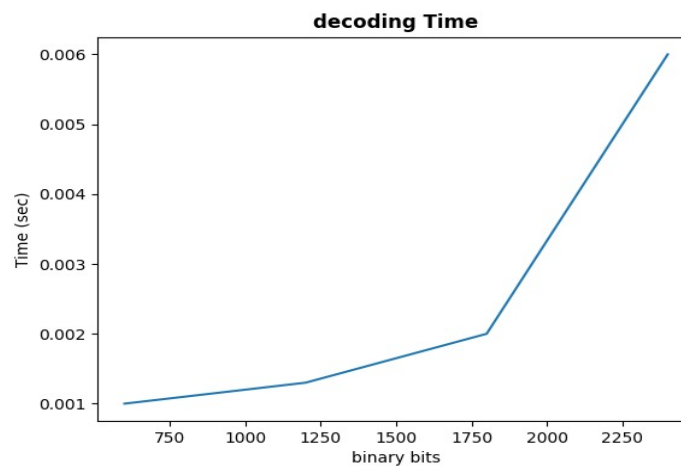
The data segments identified as a result of the search need to be decrypted so that the underlying plaintext can be recovered. As shown in Figure 8, the decryption time for 2300 bits is 0.014 s.



**Figure 8.** Computational complexity for decryption.

### 5.3.5. Decode Data

Upon the successful search, once the data are decrypted, they need to be decoded to give the original binary representation so that the data are aligned with the original text that was stored. The time required for this operation is presented in the Figure 9. For 2300 bits, the decoding time is 0.006 s.



**Figure 9.** Computational complexity for decoding.

## 6. Conclusions and Future Work

The popularity of IoT devices has placed them in a variety of environments and settings. Users are always looking for devices that are faster and increasingly intelligent. IoT devices sense data, store them and possibly make decisions on the data at hand, thus the demand for increased secure data storage. Water is known to be most vital compound for man's well being. Recently, this essential compound has further risen to prominence as researchers explore its use in liquid hard drives to store data on DNA strands. Although the liquid hard drive has the ability to store terabytes of data in a tablespoon of water/liquid, the storage of sensitive data in the water and liquid could lead to a lack of data confidentiality. Furthermore, it was claimed that water has memory that can lead to potential security and privacy breaches. This article for the first time highlighted the security and privacy concerns related to data storage on water and the effects of water memory. This study also proposed a framework and mitigation strategies to deter attackers from launching successful attacks on data stored on water and DNA strands. A proof-of-concept prototype was implemented and tested over a real-world dataset. The results demonstrated the practicality of the framework. In future, we propose to extend

this proposition to a fully functional testbed and demonstrate the capability in a real-world environment.

**Author Contributions:** Conceptualization, S.T. and M.R.; methodology, S.T. and H.T.; validation, H.T., R.T. and H.A.; formal analysis, H.T. and H.A.; data curation, H.T. and R.T.; writing—original draft preparation, S.T. and R.T.; writing—review and editing, H.T., R.T. and H.A.; supervision, M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This work was done by the Information Security and Privacy Lab, NUST supported by the National Centre for Cyber Security, Pakistan, under the project titled “Privacy Preserving Search over Sensitive Data Stored in the Cloud”.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Coughlin, T. 175 Zettabytes By 2025. Forbes. 2018. Available online: <https://www.forbes.com/sites/tomcoughlin/2018/> (accessed on 6 January 2022).
2. Morgan, S. The World Will Store 200 Zettabytes of Data by 2025. Cybercrime Magazine. 2020. Available online: <https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/> (accessed on 6 January 2022).
3. Islam, S.M.; Lloret, J.; Zikria, Y.B. Internet of Things (IoT)-based wireless health: Enabling technologies and applications. *Electronics* **2021**, *10*, 148. [CrossRef]
4. Bellini, P.; Nesi, P.; Pantaleo, G. IoT-Enabled Smart Cities: A Review of Concepts, Frameworks and Key Technologies. *Appl. Sci.* **2022**, *12*, 1607. [CrossRef]
5. Kalsoom, T.; Ahmed, S.; Rafi-ul-Shan, P.M.; Azmat, M.; Akhtar, P.; Pervez, Z.; Imran, M.A.; Ur-Rehman, M. Impact of IoT on Manufacturing Industry 4.0: A New Triangular Systematic Review. *Sustainability* **2021**, *13*, 12506. [CrossRef]
6. Stolojescu-Crisan, C.; Crisan, C.; Butunoi, B.P. An IoT-Based Smart Home Automation System. *Sensors* **2021**, *21*, 3784. [CrossRef] [PubMed]
7. Kumar, S.; Tiwari, P.; Zymbler, M. Internet of Things is a revolutionary approach for future technology enhancement: A review. *J. Big Data* **2019**, *6*, 111. [CrossRef]
8. Ayaz, M.; Ammad-Uddin, M.; Sharif, Z.; Mansour, A.; Aggoune, E.-H.M. Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk. *IEEE Access* **2019**, *7*, 129551–129583. [CrossRef]
9. Dhanjani, N. *Abusing the Internet of Things: Blackouts, Freakouts, and Stakeouts*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015.
10. Church, G.M.; Gao, Y.; Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* **2012**, *337*, 1628. [CrossRef] [PubMed]
11. Davenas, E.; Beauvais, F.; Amara, J.; Oberbaum, M.; Robinzon, B.; Miadonnai, A.; Tedeschi, A.; Pomeranz, B.; Fortner, P.; Belon, P.; et al. Human basophil degranulation triggered by very dilute antiserum against IgE. *Nature* **1988**, *333*, 816–818. [CrossRef] [PubMed]
12. Teixeira, J. Can water possibly have a memory? A sceptical view. *Homeopathy* **2007**, *96*, 158–162. [CrossRef] [PubMed]
13. Chaplin, M.F. The Memory of Water: An Overview. *Homeopathy* **2007**, *96*, 143–150. [CrossRef] [PubMed]
14. Kröplin, B.; Henschel, R.C. *Water and Its Memory: New Astonishing Insights in Water Research*; GutesBuch Verlag UG: Cham, Switzerland, 2017.
15. The World in a Drop- the Ability of Water Functioning as Memory and Mirror. World in a Drop. 2021. Available online: [https://www.weltimtropfen.de/index\\_english.html](https://www.weltimtropfen.de/index_english.html) (accessed on 6 January 2022).
16. Emoto, M. *The Hidden Messages in Water*; Simon and Schuster: Hillsboro, OR, USA, 2011.
17. Beauvais, F. “Memory of Water” without Water: The Logic of Disputed Experiments. *Axiomathes* **2014**, *24*, 275–290. [CrossRef]
18. Whitwam, R. The Liquid Hard Drive That Could Store a Terabyte of Data in a Tablespoon of Fluid. ExtremeTech. 2014. Available online: <https://www.extremetech.com/extreme/186797-the-liquid-hard-drive-that-could-store-a-terabyte-of-data-in-a-tablespoon-of-fluid> (accessed on 6 January 2022).
19. El-Moursy, A.E.; Elmogy, M.; Atwan, A. Dna-based cryptography: Motivation, progress, challenges, and future. *J. Softw. Eng. Intell. Syst.* **2018**, *3*, 67–82.
20. Ahuja, S.; Chan, Y.E. IT security governance: A framework based on ISO 38500. In *Proceedings of the CONF-IRM 27*; Association for Information Systems: Atlanta, GA, USA, 2015; pp. 1–14.
21. Raj, M.; Tahir, S.; Khan, F.; Tahir, H.; Zulkifl, Z. A Novel Fog-based Framework for Preventing Cloud Lock-in while Enabling Searchable Encryption. In *Proceedings of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, Islamabad, Pakistan, 20–21 May 2021; pp. 1–6.
22. UbaidurRahman, N.H.; Balamurugan, C.; Mariappan, R. A novel DNA computing based encryption and decryption algorithm. *Procedia Comput. Sci.* **2015**, *46*, 463–475. [CrossRef]

23. Zhang, Y.; Wang, F.; Chao, J.; Xie, M.; Liu, H.; Pan, M.; Kopperger, E.; Liu, X.; Li, X.; et al. DNA origami cryptography for secure communication. *Nat. Commun.* **2019**, *10*, 5469. [[CrossRef](#)] [[PubMed](#)]
24. Mei, L.; Xu, C.; Xu, L.; Yu, X.; Zuo, C. Verifiable Identity-Based Encryption with Keyword Search for IoT from Lattice. *Comput. Mater. Contin.* **2021**, *68*, 2299–2314. [[CrossRef](#)]
25. Tahir, S.; Ruj, S.; Rahulamathavan, Y.; Rajarajan, M.; Glackin, C. A new secure and lightweight searchable encryption scheme over encrypted cloud data. *IEEE Trans. Emerg. Top. Comput.* **2017**, *7*, 530–544. [[CrossRef](#)]
26. Ray, I.G.; Rahulamathavan, Y.; Rajarajan, M. A New Lightweight Symmetric Searchable Encryption Scheme for String Identification. *IEEE Trans. Cloud Comput.* **2020**, *8*, 672–684.
27. Leontiadis, I.; Li, M. Storage Efficient Substring Searchable Symmetric Encryption. In *Proceedings of the 6th International Workshop on Security in Cloud Computing (SCC '18)*; Association for Computing Machinery: New York, NY, USA, 2018; pp. 3–13.
28. Bösch, C.; Hartel, P.; Jonker, W.; Peter, A. A survey of provably secure searchable encryption. *ACM Comput. Surv. (CSUR)* **2014**, *47*, 1–51. [[CrossRef](#)]
29. Tahir, S.; Rajarajan, M.; Sajjad, A. A ranked searchable encryption scheme for encrypted data hosted on the Public Cloud. In *Proceedings of the 2017 International Conference on Information Networking (ICOIN)*, Da Nang, Vietnam, 11–13 January 2017; pp. 242–247.
30. Halevi, S.; Shoup, V. Design and Implementation of HELib: A Homomorphic Encryption Library. Cryptology ePrint Archive. 2020. Available online: <https://www.shoup.net/papers/helib-design.pdf> (accessed on 6 January 2022).
31. Available online: <https://github.com/iskana/pbwt-sec/> (accessed on 6 January 2022)
32. Acar, A.; Aksu, H.; Uluagac, A.S.; Conti, M. A Survey on Homomorphic Encryption Schemes: Theory and Implementation. *ACM Comput. Surv.* **2018**, *51*, 79. [[CrossRef](#)]