Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 7105625, 15 pages
https://doi.org/10.1155/2020/7105625

Review Article

Text Data Security and Privacy in the Internet of Things: Threats, Challenges, and Future Directions

Umair Khadam,[1] Muhammad Munwar Iqbal,[1] Meshrif Alruily,[2] Mohammed A. Al Ghamdi,[3] Muhammad Ramzan,[3] and Sultan H. Almotiri[4]

[1]*Department of Computer Science, University of Engineering and Technology, Taxila 47050, Pakistan*
[2]*Faculty of Computer and Information Sciences, Jouf University, Sakaka City, Saudi Arabia*
[3]*Computer Science Department, Umm Al-Qura University, Makkah City, Saudi Arabia*
[4]*Department of Computer Science & IT, University of Sargodha, Sargodha, Pakistan*

Correspondence should be addressed to Muhammad Munwar Iqbal; munwariq@gmail.com

Received 22 November 2019; Accepted 8 January 2020; Published 19 February 2020

Academic Editor: Ghufran Ahmed

In our daily life, Internet-of-Things (IoT) is everywhere and used in many more beneficial functionalities. It is used in our homes, hospitals, fire prevention, and reporting and controlling the environmental changes. Data security is the crucial requirement for IoT since the number of recent technologies in different domains is increasing day by day. Various attempts have been made to cater the user's demands for more security and privacy. However, a huge risk of security and privacy issues can arise among all those benefits. Digital document security and copyright protection are also important issues in IoT because they are distributed, reproduced, and disclosed with extensive use of communication technologies. The content of books, research papers, newspapers, legal documents, and web pages are based on plain text, and the ownership verification and authentication of such documents are essential. In the current domain of the Internet of Things, limited techniques are available for ownership verification and copyright protection. In the said perspective, this study includes the discussion about the approaches of text watermarking, IoT security challenges, IoT device limitations, and future research directions in the area of text watermarking.

1. Introduction

With the rapid development of embedded technology, computer technology, mobile communication network, and the Internet, IoT emerges at a historic moment. The primary feature of IoT is global perception, reliable transfer, and intelligent processing of information. The key is to realize the interaction of information between people and machine or machine and machine. Since its introduction, the IoT has caused major repercussions around the world because many human and material resources have been invested in supporting research, and remarkable results have been achieved. The rapid growth of IoT has brought significant changes to the industry, which is considered the third wave of the global information industry after the computer and the Internet. The Internet of Things is a collection of elements embedded in software, actuators, and electronic components that share and collect data over an Internet connection. IoT devices can be used in many environments that are equipped with sensors and low processing power [1]. The significant difference between the traditional internet and IoT is the absence of human role. IoT devices can create, analyze, and take action on information about an individual's behavior [2]. IoT offers a lot of benefits for humans but facing many issues regarding security and privacy [3].

Current security challenges for IoT that need to be sorted out are presented in Figure 1. This shows that data integrity, security, privacy, automation, updating, a common framework, and encryption capabilities are the main challenges. In IoT, text documents integrity and security issues are exist in a modern digital world [5]. A large number of text documents are generated daily and shared through IoT. Due to

Figure 1: Recent security challenges in IoT [4].

advanced technologies, these documents can be easily copied and redistributed [6]. IoT has unlimited benefits, but on the contrary, illegal use of these documents creates a problem for the original. Nowadays, a number of ways have been used by hackers to infect or access the information. Digital text document protection is a crucial issue for researchers in the modern world [7]. The use of digital libraries, Internet technologies, mobile phones, e-commerce, and iPods are a fast and easy way of broadcasting information [8]. However, the security and privacy of digital content are difficult to handle. In this case, it is necessary to provide protection to digital materials that are traveling over the internet [9, 10].

2. IoT Security Challenges

Currently, 23 billion IoT devices are connected worldwide. By the end of 2020, it will further rise and reach up to 30 billion, and by the end of 2025, it will reach over 60 billion [11]. The security challenges for IoT are mention below.

2.1. Updating. The majority of IoT devices update their software automatically, while other devices had to be updated manually [12]. Some manufacturers only offer updates for a short period of time and then stop it. It is challenging to manage the upgrade of millions of devices that are connected to IoT. All the devices do not support the automatic update and require manual updating, which are time-consuming and lead to security loopholes if any mistake happens [13].

2.2. Automation. As in our daily lives, IoT devices continue to invade and deal with the number of IoT devices. It is challenging to manage an enormous amount of user data. The fact cannot be denied that any single error in an algorithm will bring down the entire infrastructure [14].

2.3. Common Framework. In IoT, there is an absence of a common framework, so all the manufacturers retain privacy and security at their own risk. Once a standard framework is implemented, then the security issue will be resolved [15].

2.4. Security and Privacy Issues. Different IoT devices can share data among various platforms. The IoT devices exchange and gather data for multiple reasons, such as decision-making, better service, and improving efficiency. Thus, it is essential that the endpoint of data shall be secured completely.

2.5. Data Integrity. Billions of IoT devices are interlinked and exchange data on a daily basis. The data integrity is the main issue in the IoT that no one can manipulate data at any point. Digital watermarking and blockchain should be implemented in order to ensure data integrity [16, 17].

3. IoT Device Limitations

There are two main issues IoT devices have: first one is battery capacity, and the second one is computing power [18]. Since some IoT devices are placed in such environments where we cannot charge them or charge is not available, the devices should perform the designed functionality in limited energy, and heavy security instructions may drain with limited power [19]. To mitigate this issue, three possible techniques can be used: first, to minimize the security requirements, and second, to raise the capacity of the battery. That seems impossible because most IoT devices are small in size and designed to be lightweight. However, a large battery has no extra room. The third approach is to harvest energy from natural resources such as heat, light, wind, and vibration, but such techniques are required on hardware upgradations and increase the monetary cost. The IoT devices have limited memory space that cannot store and handle the computational requirements of advanced security algorithms [19]. The IoT devices should be smart and manage all these requirements.

4. IoT Devices Architecture

Internet of Things includes many connected sensors and devices, and every device uses different communication standards and protocols. There are no precisely defined rules and standards for communication. In addition, the applications of the Internet of Things would not be limited and increase from day today. Different IoT devices are produced by different manufacturers even if they perform the same functionality. So, this challenge is related to the nature of the IoT and may lead to a lack of unified standardization.

4.1. IoT Devices Data Storage Issues. Data storage becomes a significant issue, as the amount of data increases rapidly. When the stored information is damaged, it is a challenging task to back up all. There is no assurance that data and information are securely transmitted over the IoT devices. Furthermore, it is a significant challenge for management

companies and data storage to develop tools and standards that handle data provided and security issues.

4.2. Limited Resources of Infrastructure. IoT devices generally have limited memory and low processing capacities. Designing comprehensive security measures in 64 kB to 640 kB memory is a big challenge for software developers and IoT hardware manufacturers. In addition, they must have enough storage for security software to defend against security threats.

4.3. Data Privacy Protection. Anyone can access integrated devices from anywhere with IoT, which affects sensitive data confidentiality and privacy. Therefore, specific standards or rules must be defined to avoid the privacy violation. For example, some IoT devices share data with other devices, and in this case, the data become unsafe. This helps attackers and intruders to breach the security of the IoT system.

4.4. Lack of Skills. Specific skills and expertise are critical factors in the design, development, implementation, and management of security that must be considered. Any of this factor disruption may cause damage to the IoT security system. In addition, the lack of skills and expertise slows down the adoption of IoT technologies [20]. There are very limited people who can adequately handle the IoT system. The number of qualified people who master in IoT techniques is very limited. The benefits of IoT technology and dealing with its challenges depend mostly on individual capabilities.

5. Digital Watermarking

Digital watermarking belongs to information hiding and plays an essential role in copyright protection, ownership verification, and authentication [21]. In digital media, when we talk about information hiding, text watermarking is the least discussed subject. The protection of digital content is a difficult task, especially plain text [22–24]. The information hiding is categorized into steganography, cryptography, and watermarking as shown in Figure 2. A secret message is embedded in digital content without affecting the original text, which authenticates the ownership verification [25, 26].

Researchers have significant challenges that information growth rate is higher, which requires an appropriate technique for watermarking. It is crucial to maintain data integrity while ensuring the confidentiality and availability of information [27]. However, with the practical development of the watermarking application, security issues of watermarking have emerged and achieved significant progress in this field [28].

Many techniques have been proposed in the last two decades, for hiding information in terms of steganography and text watermarking for copyright protection [29], authentication, copy control [30], ownership verification [31–37]. The main contributions of this study are listed as follows:

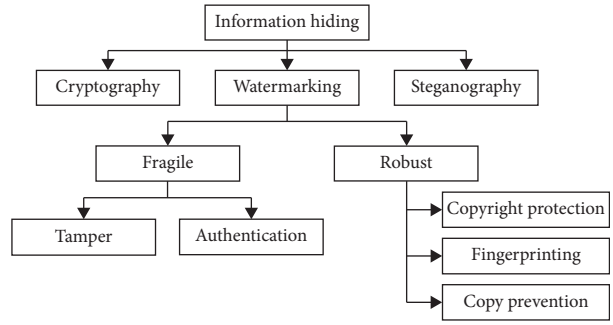

Figure 2: Common information hiding techniques.

(i) We briefly describe the IoT current security and privacy issues and recommendations

(ii) We conduct an extensive investigation about the approaches of text watermarking, IoT security challenges, IoT device limitations, and future research directions

(iii) We summarize the text watermarking approaches/techniques that are used for digital watermarking

(iv) A comparative analysis of previous techniques has been conducted on the basis of robustness, security, capacity, and imperceptibility. Their efficiency evaluated on the basis of set criteria, also identifying the drawbacks of exiting techniques

6. Digital Watermarking and Its Applications

In the real-world, watermarking can be used in a variety of applications that categorized into image, audio, video, and text [38]. Authorized documents, such as websites, certificates, business plans, articles, poems, books, corporate documents, e-mails, and SMS, can be protected through watermarking [39]. The applications of digital watermarking can be used for authentication, copyright protection. Some other application of the watermarking in the text listed below [40–42].

6.1. Authentication. The plain text in articles and newspapers highlighted various problems with authentication. Watermarking is a verification tool to authenticate the integrity of the plain text. To prove the authentication if the watermark (author information) is perceived, then it has genuine document else text has been tempered and cannot be measured. The authentication mechanism can be used for a text document to detect any tampering. If tampering is identified, then the document cannot be considered as original, also for legal purposes, and it is necessary to authenticate text document [43].

6.2. Copyright Protection. Watermarking is also used in copyright protection of digital contents, like e-books, web content, research papers, poetry, and other documents. The author inserts a watermark in the document for copyright, and this watermark is extracted in the future from the given material to prove ownership. Digital watermarking is very helpful to settle the copyright issues in court.

6.3. Tamper Detection. A large number of text documents are available for users to read online, and these documents can be confronted with a series of attacks such as copying, unauthorized access, and redistribution. Tamper detection is one of the digital watermarking applications that can detect and recover the tampered region from the digital contents. Text watermarking is used as a fragile tool against these attacks [34, 44].

6.4. Copy Control. Publishers are looking for more consistent ways to control the copy of their important documents. Likewise, they want their essential documents to be available on the Internet for revenue generation. The watermarking is also applied here to provide access control and stop illegal copying [43].

6.5. Forgery Detection. Text documents reproduction and plagiarism are serious issues, and it is rapidly growing. Text watermarking is applied here to embedding watermark in the original document before publishing online [45]. Almost every private and public organizations deal with text documents on a daily basis, and digital text watermarking application can be applied here to control the forgery detection problem.

Watermarking major applications [46] is shown in Figure 3.

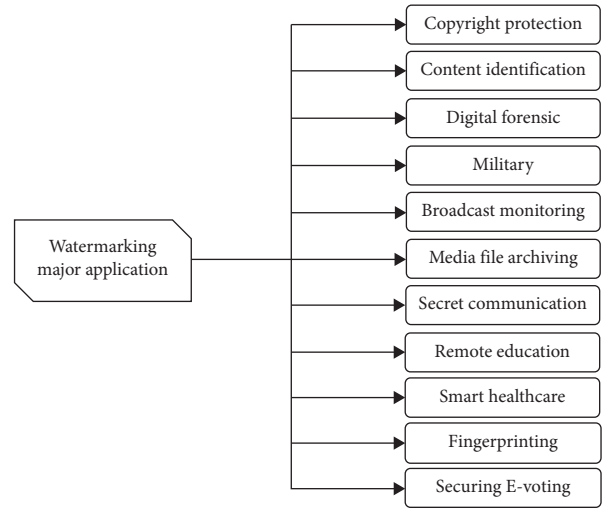

Figure 3: Watermarking major applications.

7. Text Watermarking Evaluation Criteria

The researchers count a lot of parameters while developing novel techniques. However, digital text watermarking evaluation criteria can be classified into security, capacity, robustness, imperceptibility, and computational cost. It is not possible to design such a system of watermarking that can cover all these properties. In the below content, each property of watermarking mentioned above is described [44, 47, 48].

7.1. Robustness. Robustness means that if watermark information is tempered and then it is still survived [49]. The mean of robustness is that it will be almost impossible without a license and without the content that defeat marked a great extent the content is not suitable and reliable [50]. When a technique of watermarking is designed, it is essential to revenue in consideration of the future application and the equivalent number of attacks that are possible. On the bases of watermark distortion rate (WDR) and pattern matching rate (PMR), the robustness of text watermarking is computed. That is formalized from (1) and (2).

$$\mathrm{PMR} = \frac{N_{\mathrm{m}}}{N_{\mathrm{w}}}. \tag{1}$$

$$\mathrm{WDR} = 1 - \frac{N_{\mathrm{m}}}{N_{\mathrm{w}}}, \tag{2}$$

where N_{m} determines the number of patterns matched correctly and N_{w} defines the number of watermark patterns.

7.2. Imperceptibility. The imperceptibly in the primary and fundamental requirement that means the watermark is securely embedded into the document objects. The watermark information could not feel the audience, or the watermark should not affect the original text. The watermarked and original information should be similar, and the content should be perceptually equal [51]. Peak signal–to-noise ratio (PSNR) and similarity percentage (SIM) is used to ensure the imperceptibility using the following equation (3) [52]:

$$\mathrm{PSNR} = 20 \log_{10} \cdot \frac{O_{\mathrm{doc}}\,(\max)}{\mathrm{RMSE}}, \tag{3}$$

where O_{doc} (Max) is the maximum pixel value in the document image, RMSE stands for root–mean-squared error, and it is calculated using the following equation (4):

$$\mathrm{RMSE} = \sqrt{\frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[O_{\mathrm{doc}}\,(\mathrm{i},\mathrm{j}) - W_{\mathrm{doc}}\,(\mathrm{i},\mathrm{j})\right]^2}. \tag{4}$$

The following equation (5) is used to calculate the similarity parentage (SIM):

$$\mathrm{SIM} = \left[1 - \frac{\mathrm{RMSE}}{O_{\mathrm{doc}}\,(\max)} \times 100\right]. \tag{5}$$

7.3. Capacity. The capacity indicates that the maximum bits of watermark information that can be stored in the host document. If a technique can hold large hiding capacity without affecting the visibility, then it is considered. The capacity can be measured using the following equation (6):

$$\mathrm{capacity} = \frac{\text{total no of bits (secret data)}}{\text{total no of cover file data (kB)}} \times 100. \tag{6}$$

7.4. Security. There is another scheme for watermarking which is security. It states that the information of the author (watermark) is hidden from unauthorized users. They do not

have access to detect the watermark. Watermark still exists and the payload still remains coved is the mean of security. Unapproved and unauthorized parties are not capable of identifying the author's information. Security is measured on the bases of the imperceptibility, capacity, and robustness as shown in the following equation (7):

$$\text{security} = [\text{imperceptibility} + \text{capacity} + \text{robustness}]. \tag{7}$$

7.5. Computational Cost. Text watermarking techniques are computationally less complex for small text documents. More computation power is required for text documents that occupy many pages. In general, less complex algorithms are used for systems with limited resources to reduce the cost [44].

8. Watermarking Embedding and Extraction Process

Watermarking is the technique of information hiding that provides ownership verification and copyright protection to text documents against illegal usage [53–55]. Digital watermarking has two steps: the first one is watermark embedding, and the second step is watermarking extraction or verification. In watermark embedding, secret information (watermark) is inserted into the original document without affecting the content of the document. A key can be used to encrypt the secret information for security purposes, and then the same key is applied for decryption. When an illegal attempt happens, then the watermark information is used to verify the original owner of the document. The reverse process of watermark embedding is called watermark extraction. Basically, this process is applied to verify the originality of the document. The architecture of watermarking is presented in Figure 4.

9. Existing Techniques of Text Watermarking

Digital text watermarking arose in 1994 [56, 57] and grew with the passage of time, as the communication and Internet start all over the world. These techniques are based on words and sentences, acronym, synonym, presupposition, syntactic tree, typo error, noun-verb, and text images for German, Persian, French, Spanish, and English languages. The text watermarking techniques and attacks are presented in Figure 5.

An information hiding technique is proposed in [58] that hides information in a binary text document; they use the boundary of characters for information hiding. Five pixels long, 100 pairs of border patterns were defined. There were two different models for each pair, an "A" model and a "D" model, which can be changed into each other when the pair is returned. A bit is embedded in the five-pixel long border by browsing the patterns. Kim et al. suggested a technique based on the classification of words and interword spaces to insert a watermark [59]. All words are classified in the document according to adjacent words and specific text attributes which comprise a segment, and it is further categorized according to the names of the class and the words in the segment. Each segment class contains the same amount of information.

Zhou et al. [60] introduced a method which used a chaotic encrypting algorithm to generate the watermarks, and the host document splits into two blocks using Chinese mathematical expressions. Two different text blocks and keys were generated to calculate the stoke numbers and the Chinese character frequency. When the content of the watermarked text document is modified, results from two blocks of text do not match, and text document result authentication will be false. In [61], a technique is proposed that is based on a particular part of speech (POS) for text zero watermarking. POS is the category of a word which has similar grammatical properties. The chaotic function is used to extract the sequences that are used to develop a watermark without altering cover data, and the imperceptibility problem is also resolved. This method provides excellent security because the order of the selected POS tag is unknown by an attacker.

Meng et al. [62] introduced a technique where sentence entropy is used to calculate the watermark key. Entropy defines as the average expected value of data that a message contains. Through word frequency and important selection, the sentence entropy is calculated, and according to the order of the crucial sentence, the watermark is embedded. Some unknown attacks were also applied to this method, which includes insertion, deletion and synonym substitution to check the robustness of this method, which is good but shows a very low success rate. Jalil et al. [63] suggested a technique that embeds through generating a watermark key. To find the nonvowel character that occurs most frequently, the occurrence of nonvowel ASCII character analyzes first in each partition. The maximum occurrence of nonvowel and author key letters is used for watermark generation. Certification authorities are used to a registered watermark in order to provide security. Extracted watermark accuracy is analyzed through insertion and deletion attacks. In [64], the author proposed a watermarking technique that generates watermark key on the bases of the preposition, double letters, and cover file partition is analyzed through the repeating letter frequency. The key is generated through a count of double letters in a time interval. The conversion of the image into the text is performed to generate the hidden data that is included in the host document. In insertion, deletion, reordering, and other attacks, the proposed method is robust and more secure.

Cheng et al. [65] introduced an algorithm on the strategy of fragments regrouping for watermark embedding. The original watermark is divided into different fragments of order numbers and then embedded in the characters of the document. After deleting and tamper attack when some fragments are deleted or changed, the destroyed fragment is recovered using other correct fragments that are embedded in the phrases. Kim et al. [66] proposed a method that is based on syntactic displacement and morphological division in natural language watermarking for Korean. Syntax-based watermarking is used in this approach, usually, a Korean

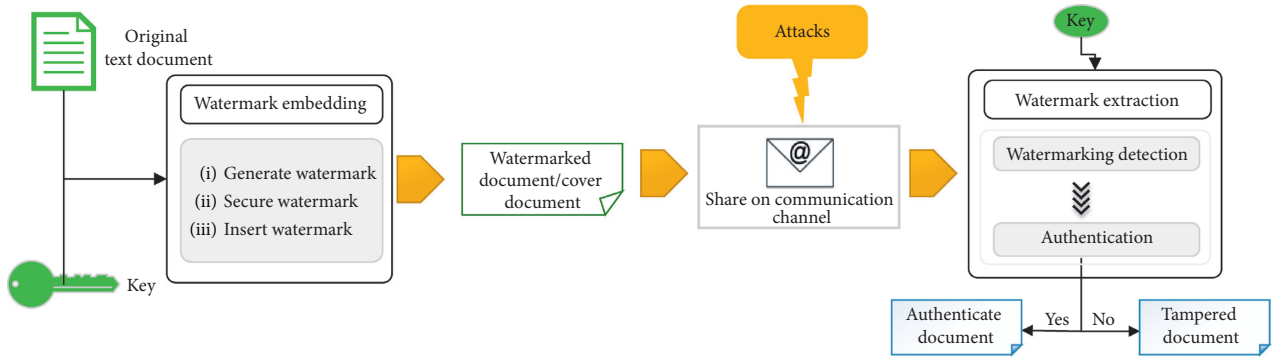

Figure 4: Digital watermarking architecture.

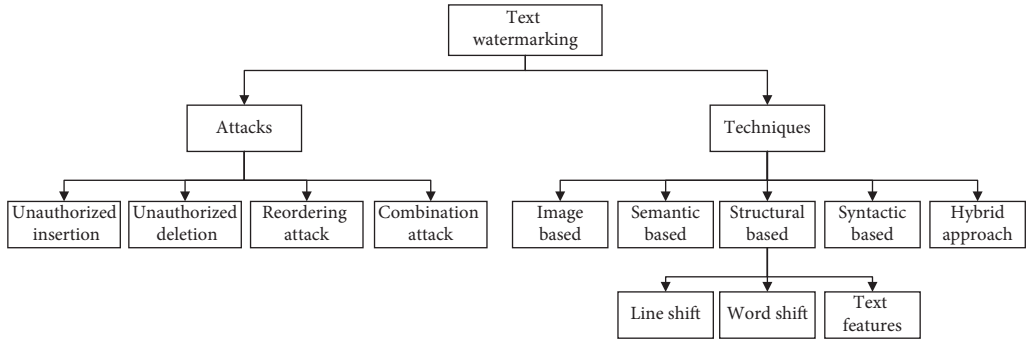

Figure 5: Text watermarking techniques and attacks.

word consists of function morpheme and content morpheme. Through the use of word characteristics, the word is divided into two content morphemes into two new words, which are used for watermark embedding.

In [67], a model based on 3-D using 2-D coordinates of word-level and weights of sentences to construct zero watermarking is introduced. The structure of the 2D word space includes the length and frequency of words, then that the 2-D model is extended into 3-D. Three frequent attacks are tested on the proposed model that are synonym replacement, syntactic transformation, and deleting attack. The test report shows that the proposed method is robust, secure, and useful imperceptibility. Al-Wesabi et al. [68] proposed Markov's model-based approach for watermarking, where the watermark key is generated through cover file probabilistic features. The use of the hidden Markov model information for text watermarking is analyzed and stored in the document for authentication. It offers protection against attacks with a higher percentage of watermark distortion than all attacks. In [69], the author suggests a zero watermark approach that uses the Arabic character's characteristics for embedding the watermark without changing the original text. In an initial phrase, name/number of sura and number of verses are checked, and then from each verse of Holy Quran, key is generated. With this algorithm, a character watermark bit of the word set is inserted. The proposed method built a system to verify the sensitive of the Holy Quran digital text. With this technique, changes in the original text content can be detected and only minimal hardware resources are required.

Alginahi et al. [70] introduced an approach that generates the watermark key by converting the image into text. A duplicated cover file is used for embedding the image logo where it is classified and processed, and using its characteristics watermarking key is generated. The proposed technique offers authorized content manipulation and copyright protection. Through using blind and fragile watermarking approaches, the watermark key is secured. This method produces excellent results after evaluating the computational time of watermark encoding and decoding. Ba-Alwi et al. [71] presented a novel technique based on probabilistic models for ownership verification and tamper detection in English documents. The probabilistic pattern is extracted by using natural language processing based on the Markov model. Each text document content is analyzed in English and extracts the probabilistic characteristics between these contents.

In [72], the authors suggested a technique based on word items and particular attributes of robustness and excellent performance, which can hide information in a Word document. A novel method is proposed to enhance the robustness of the watermark. Watermarking information is divided into 5 groups. After this, it is embedded into the plain text one by one as a group no. An advantage of this method is challenging to extract hidden information because its first encrypted information is divided into several groups and then embedded into word properties. After the experiments, most of the watermarked text is the same, but in two or three lines, some characters are changed which also changes text meaning. The scheme is not very good on the

base of imperceptibility. Chen et al. [73] suggested a semantic technique for embedding watermark information in the text. The watermark information is embedded through the mapping location of each digit. The proposed algorithm does not change the integration of text and format. The author claims that it is robust against watermarking attacks and text format transformation.

Ahvanooey et al. [29] offer a novel text watermarking method for web pages. Structural and syntactic rules are used to embed watermark, which is encoded and converted into zero-width control characters with a binary model classification. Hypertext Markup Language (HTML) is used as a cover file to embed the transparent zero-width watermark. In [74], the author suggests a novel method for embedding information in text, which is based on font code that embeds a watermark into text by disrupting text character glyphs while retaining text content. The glyph recognition method is also presented to restore the information that is embedded in the encrypted document. A new approach is proposed for Arabic text using pseudospace in [75]. The connected letters are isolated with pseudospace to hide watermark bits, which are used to hide watermark bits. In the first method, the watermark is embedded in the punctuation of the Arabic text by inserting a pseudorandom, and in the second method, the pseudospace is added to the standard space, thus increasing the capacity. The proposed method is robust and imperceptible against formatting and tampering attacks. Wen et al. [76] suggested algorithms for Extensible Markup Language (XML) document to hide information. The first method is the eXtensible Stylesheet Language Transformation- (XSLT-) related that is designed with the inclusion of additional codes to provide copyright protection. In the second method, the functional dependency is used for the XML file as a function for zero watermark. The proposed method performs well in alternation attacks, compression attacks, reorganization attacks, and selection attacks. From the study of Hakak et al. [77] in this work, a complete framework is presented with regard to the automatic authentication and distribution of the digital Quran and Hadith verses. The verification process is divided into two phases, security and verification. The watermarking technique in case of the security phase secured the confirmed and tested verse. For verification, the Boyer–Moore algorithm is used for extraction. The efficiency analysis of the existing techniques is presented in Table 1.

10. Attacks in Text Watermarking

Watermark content has specific attacks depending on the application. Some attacks are significant from other attacks. The basic types of attacks are an illegal insertion, illegal updation, illegal deletion, reordering attack, and the mixture of all these attacks. Table 2 presents the analysis of robustness attacks. This includes insertion, deletion, reordering, formatting, copy and paste, and retyping attacks. In the following categories, these attacks are placed [7, 72, 99–104].

10.1. Unauthorized Insertion. When an attacker wants to add false information, then such type of attack occurs, i.e., in the case of legal documents. Each time a dispute concerning the application of copyright occurs, and this type to identify the first recorded content stamp is used.

10.2. Unauthorized Detection. The ability to be detected in some applications is restricted. It is believable that the aptitude of a challenger to quickly identify whether a mark in a particular plant is present endangers the security of the watermarking system.

10.3. Unauthorized Deletion. An attacker can delete some words or sentences from the text to remove the original author's identity. All watermark application required security against illegal deletion. It is crucial to restrict the attacker to remove watermark information. The system is called secure if the watermark is still extracted from the text after applying the attack.

11. Research Challenges and Future Direction

Text watermarking research is at an early stage, although the watermarking process has been extensively studied. There are several significant issues in text watermarking that have remained unresolved. In addition, applications continue to pose new challenges, and many organizations still need to implement text watermarks.

11.1. Information Availability. Information availability means that a user can access information easily and securely. Millions of Internet users around the world generated and shared information on a daily basis, which required protection against illegal usage fully. In the text watermarking context, the availability of information remains constant and prevents any change in the text content. An active system is required to ensure data availability in secure manners, where a user can access information after an independent self-monitoring system.

11.2. Data Integrity. Data integrity is one of the critical aspects of text watermarking, which is related to reliability, usability, relevance, value, and quality. Data consistency and accuracy assurance can be part of integrity [105]. The explosion of the internet allows users to access a vast amount of information, where the integrity of information also required. With the development of internet technologies such as cloud, data can be easily shared through different communication. The main issue is how to ensure the integrity of data over the Internet.

11.3. Originality Protection. It is difficult to identify the originality and quality of data that is available online or come from all sorts of databases that are always well preserved in all cases. The implemented techniques' processing time is still high and lacks imperceptibility. The challenge is how to find the appropriate method that protects the originality of

Table 1: Efficiency analysis chart.

No.	Authors and Years	Parameters					Efficiency analysis	Drawbacks
		Medium	Capacity	Security	Imperceptibility	Robustness		
1	Hamdan and Hamarsheh [26]—2016	Text	Low	High	Medium	High	High robustness and security	The main drawback is the length (capacity) of the cover message
2	Gutub et al. [78]—2010	Text	High	Medium	Low	Low	High capacity	Imperceptibility and robustness are low when applying formatting attacks
3	Kim et al. [59]—2003	Text	High	NA	High	Low	High capacity and imperceptibility	Low robustness when applying distance algorithm spaces between words are deleted
4	Yang and Kot [79]—2004	Text	Low	High	High	Medium	The proposed system has a high capacity, imperceptibility, and capacity	Robustness is down when applying distance algorithms
5	Alginahi et al. [80]—2013	Text	Low	Medium	High	Medium	High imperceptibility	Capacity is low and no robustness against formatting attacks
6	Meng et al. [62]—2010	Text	Medium	Medium	High	High	High imperceptibility and robustness	Low capacity
7	Jalil and Mirza [41]—2010	Image plus text	Low	High	Low	High	High robustness and security	Low imperceptibility and capacity
8	Jaiswal and Patil [81]—2013	Text	High	Low	High	Low	High imperceptibility	Robustness and security are down
9	Cheng et al. [65]—2010	Text	High	Medium	Low	Medium	High capacity	Low imperceptibility and no robustness against reformatting attack
10	Mir [82]—2014	Text	Medium	High	High	Medium	High security and imperceptibility	No robustness against attacks
11	Meng et al. [67]—2011	Text	Low	Medium	High	High	High robustness, security, and imperceptibility	Low capacity
12	Zhang et al. [72]—2010	Text	High	High	Low	Medium	High capacity and security	No robust against copy paste and retyping attacks and low imperceptibility
13	Liu et al. [83]—2015	Text	Low	High	High	High	High robustness, security, and imperceptibility	Low capacity
14	Alginahi et al. [35]—2014	Text	High	Medium	Low	Low	High capacity	Low imperceptibility and no robust against attacks
15	Liang and Iranmanesh [84]—2016	Text	Low	Medium	High	Low	High imperceptibility	No robustness against attacks and low embedding capacity
16	Alotaibi and Elrefaei [75]—2017	Text	High	Medium	High	Medium	High capacity and imperceptibility	Robustness is medium because vulnerable to retyping attacks
17	Yingjie et al. [85]—2017	Text	Low	High	Medium	High	High robustness and security	Low capacity
18	Wen et al. [76]—2018	Text	Low	High	Medium	High	High robustness and security	Low capacity
19	Kuribayashi et al. [86]—2018	Text	High	Medium	Low	High	High robustness and capacity	Low imperceptibility
20	Jalil and Mirza [41]—2010	Image plus text	Low	High	Low	High	High robustness and security	Low imperceptibility and capacity
21	Taha et al. [87]—2018	Text	High	Medium	Medium	Low	High capacity	Not robust against formatting attacks
22	Xiao et al. [74] 2018	Text	Low	High	Medium	High	High robustness and security	Low capacity and only applicable to one font

Table 1: Continued.

No.	Authors and Years	Parameters					Efficiency analysis	Drawbacks
		Medium	Capacity	Security	Imperceptibility	Robustness		
23	Tan et al. [88] 2018	Text	High	High	Low	Medium	High capacity and security	Low imperceptibility

Table 2: Robustness analysis against attacks

Sr. No	Authors	Insertion	Deletion	Reordering	Reformatting	Copy and paste	Retyping
1	Al-Nofaie et al. [89]	✓	✓		✓	✓	
2	Rizzo et al. [90]			✓	✓	✓	
3	Alotaibi et al. [75]	✓			✓	✓	
4	Ahvanooey et al. [29]	✓		✓	✓	✓	✓
5	Alotaibi et al. [91]	✓			✓	✓	
6	Ahvanooey et al. [31]	✓		✓	✓	✓	
7	Mir. [82]		✓				✓
8	Por et al. [36]	✓			✓	✓	
9	Chou et al. [92]	✓			✓	✓	
10	Umair et al. [7]	✓		✓	✓	✓	✓
11	Alginahi et al. [35]	✓	✓		✓	✓	
12	Gutub et al. [78]	✓			✓	✓	
13	Lee et al. [93]	✓		✓	✓	✓	
14	Bender et al. [94]	✓			✓	✓	
15	Lu et al. [95]		✓		✓	✓	✓
16	Mali et al. [24]	✓			✓	✓	✓
17	Halvani et al. [96]	✓		✓	✓	✓	✓
18	Kim et al. [66]		✓		✓	✓	✓
19	Meral et al. [97]		✓	✓	✓	✓	✓
20	Topkara et al. [98]	✓			✓	✓	✓

data and balance between robustness, capacity, and imperceptibility. Most of the prior techniques are either robust or imperceptible or improves the hiding capacity but failed to maintain the balance between all these parameters.

11.4. Sensitive Information Protection. Sensitive information cannot support the smallest change, such as a slight change in a character or word. When we alter confidential information, then the meaning of the text can change, or the original purpose of the text also changed [106]. This case usually involves religious writings, financial documents, government documents, and political documents. Such issues in text watermarking have been addressed with regard to the protection of the religious scriptures of the Arabic text. A lot of studies address the sensitive issue in text watermarking but not to be resolved yet. A precise text watermarking technique is required to resolve the sensitive issue.

11.5. Confidentiality of Information. Confidentiality or secrecy of information means it is not available for unauthorized persons or organizations. Specific measures need to be taken for information protected from unauthorized persons. Specific techniques must be implemented to control the confidentiality of the content in text watermarking. A suitable technique is required for the protection of information confidentiality.

11.6. Cryptography. The embedded data security must be further secured using cryptography, which helps prevent the key and make sure that watermark information is out of reach for an unauthorized user. A lot of methods have been proposed in the past to solve the copyright issues but still needs improvements. A new security framework is necessary for a trusted organization that relies on text watermarking techniques.

11.7. Language Flexibility. The majority of text watermarking techniques is only applicable for certain languages such as English, Arabic and Chinese, which reduce the usability and applicability of the techniques. It is a core challenge for the researcher to identify a suitable and proper text watermarking technique that should be implemented in any type of text language.

11.8. Document Transformation. When a watermarked document is transformed into other formats like Word to PDF and vice versa, there is a risk of losing the watermark information. It is crucial for the researcher to identify a proper text watermarking technique that supports the format transformation.

12. Recommendations

Text documents belong to almost all companies or organizations, such as banks, audit firms, or any public or private organizations. Both electronic and soft copies of sensitive

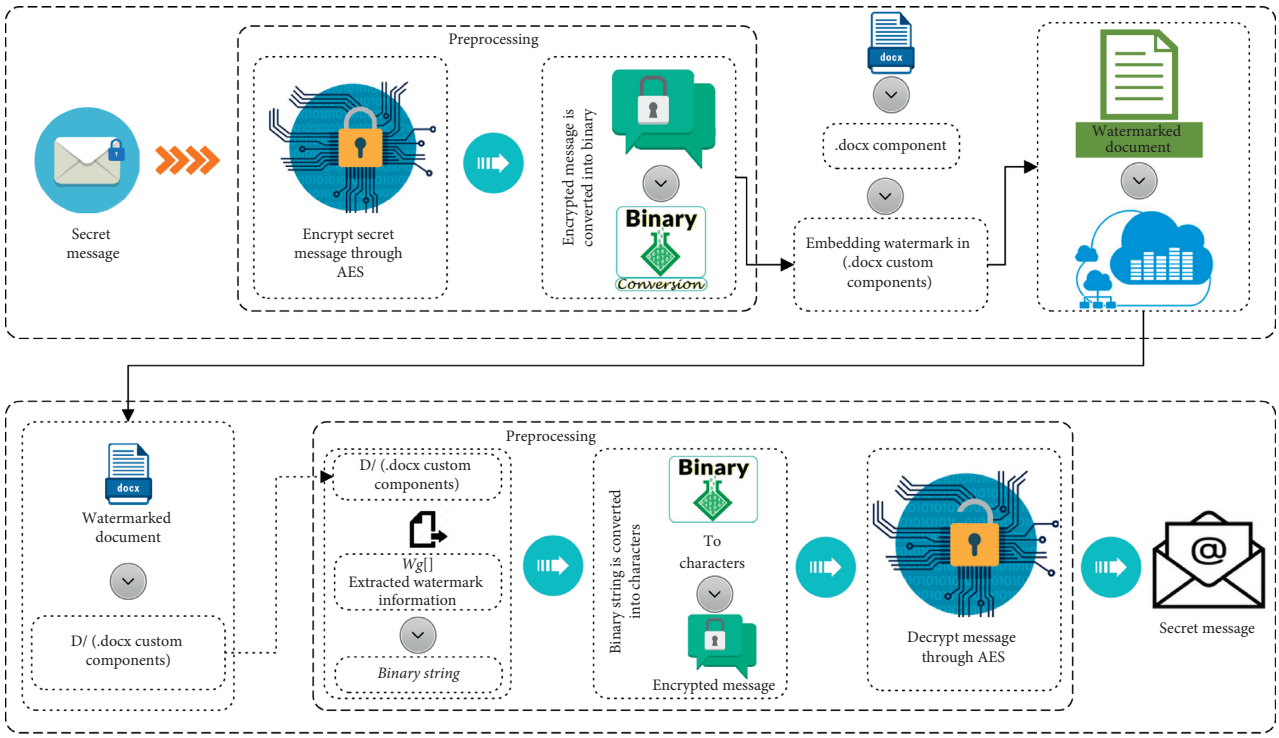

Figure 6: Proposed model for text document security and privacy in IoT.

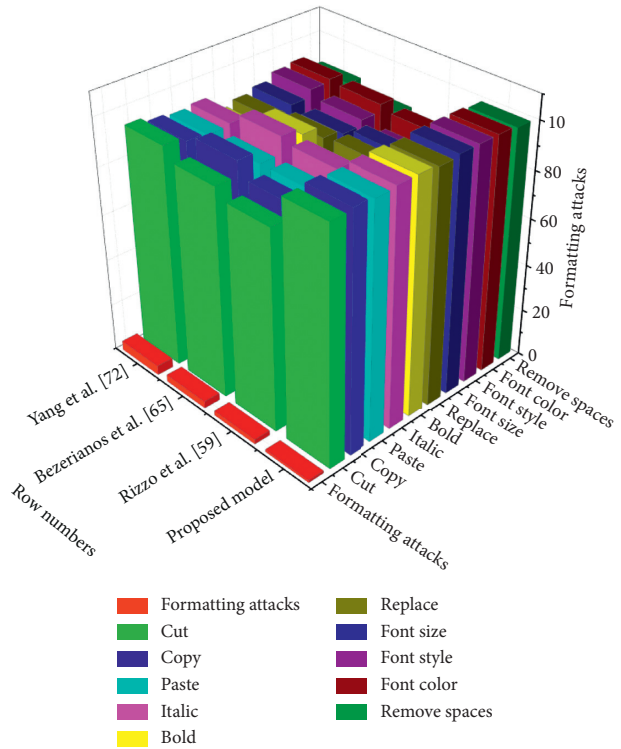

Figure 7: Robustness results of the proposed model with existing techniques.

text documents are processed. Such as soft degrees, birth certificates, legal notes, financial statements, classified reports, and declarations. The challenge is to define a reliable method to authenticate these documents and to guarantee the originality and protection of textual documents by copyright. An appropriate watermark technique is needed that is robust against formatting attacks and improves hiding capacity, imperceptible, and secure. This problem can be solved by a new framework to address the current challenges in text watermarking.

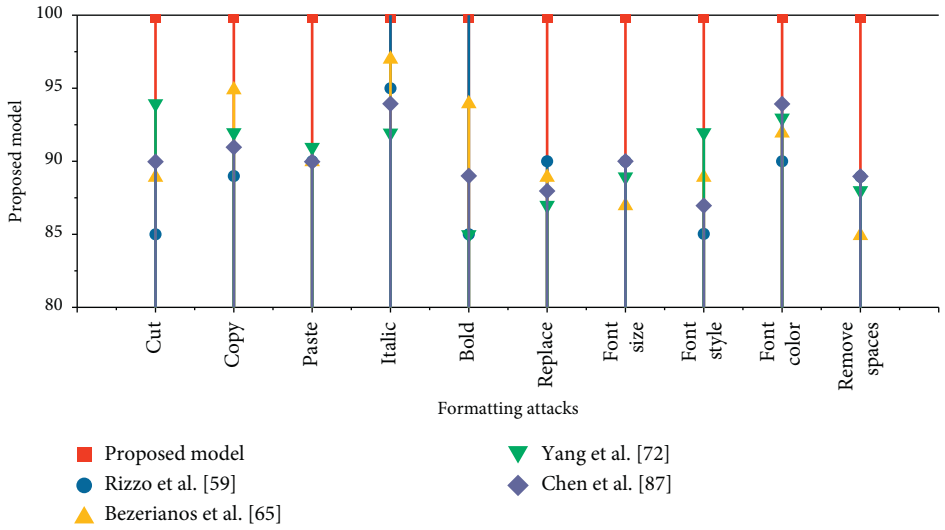

Figure 8: Proposed model and existing technique result against formatting attacks.

12.1. Proposed Model. We proposed a novel framework that overcomes the current challenges of security and privacy in the IoT paradigm based on digital watermarking as shown in Figure 6. The proposed system can provide secure communication of text documents on both local and cloud paradigms. In the proposed framework, the watermark is embedded into the custom properties of a text document. These custom properties are suitable for three reasons. First, they are not referred to with the parts of the primary document. Second, the watermarking process does not change to the original content of the document. Third, it can hide an adequate amount of secret message.

In the proposed model, the secret message (M_S) is given as input, and then in the pre-processing phase, M_S is encrypted through Advanced Encryption Standard (AES). The ASE is a simple encryption technique that is used to secure M_S. The encrypted message (M_E) is converted into a binary string, and then it is divided into n number of groups. The suitable components of the original document D_O are inspected, and M_E groups are embedded into these components. As mentioned above, the custom components are ideal for three reasons, capacity, security, and robustness. M_S has no influence on the original content of the document and does not disturb the imperceptibility. After concealing the secret information, the watermarked document (D_W) is generated that is stored or shared via the local and cloud paradigm.

Through the experimental results, our proposed model achieves excellent results against all the parameters. The proposed method is robust against all formatting attacks and more secure as compared with previous techniques, as revealed in Figure 7. Various kinds of brute force attacks are applied to check the robustness of the watermarked document. These attacks include content and format-based attacks. Figure 8 presents the comparison of the proposed method with [59, 65, 72, 83] against content and format-based attacks, which illustrate that the proposed model is robust against all possible mentioned attacks.

Our system can be applied for copyrights and owner authentication of text documents on both local and cloud computing paradigm. It can also protect the text documents against illegal use.

In addition, through the initial experimental results, we found that the proposed framework is robust, imperceptible, and supports high embedding capacity because the watermark information is stored in document components.

13. Conclusion

In this investigation, we have presented security and privacy issues in IoTs, text watermarking issues, current techniques, attacks, future research direction, and recommendations. We also classified the existing approaches of text watermarking, security and privacy issues. IoT emerging technologies and the latest applications brought new challenges for the researchers to required their attention. We have discussed and summarized the main difficulties for text document protection in IoT. This article deliberated the most common challenges and issues in text watermarking. The text is the most common medium that travels across the internet and needs full protection. Digital text watermarking is more famous for copyright protection and also hides secret information in digital contents. A lot of techniques have been proposed in this field of research, but still a new model that identifies the approaches, requirements, application of text watermarking, and its embedding process is needed. This article deliberated the most common challenges and issues in text watermarking and proposed a novel method. A novel framework for evaluating text watermarking methods is proposed that is easily and readily accessible. It is consulted by the relevant organizations and the research community. The experimental results and analysis prove that the proposed model is robust against content format-based attacks and improves the ability of concealment as compared to the previous techniques. In future research, the main tasks have been marked, and

further investigation in the area of watermarking in IoT is awaited. We also analyzed the other possible attacks in the further, which enhance the robustness and improves the ability of concealment. In future, other Microsoft Word and Excel documents other than special properties will be examined for watermarking. We also investigated the Portable Document Format (PDF) document that is the most popular document format in the world.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Dabbagh and A. Rayes, "Internet of things security and privacy," in *Internet of Things from Hype to Reality*, pp. 211–238, Springer, Berlin, Germany, 2019.

[2] M. A. Habib, M. Ahmad, S. Jabbar, S. H. Ahmed, and J. J. P. C. Rodrigues, "Speeding up the internet of things: LEAIoT: a lightweight encryption algorithm toward low-latency communication for the internet of things," *IEEE Consumer Electronics Magazine*, vol. 7, no. 6, pp. 31–37, 2018.

[3] M. A. Habib, M. Ahmad, S. Jabbar et al., "Security and privacy based access control model for internet of connected vehicles," *Future Generation Computer Systems*, vol. 97, pp. 687–696, 2019.

[4] Colocation America, *Current Security Challenges Facing the Internet of Things*, Colocation America, Los Angeles, CA, USA, 2018.

[5] U. Khadam, M. M. Iqbal, M. A. Azam, S. Khalid, S. Rho, and N. Chilamkurti, "Digital watermarking technique for text document protection using data mining analysis," *IEEE Access*, vol. 7, pp. 64955–64965, 2019.

[6] M. Ahmad, A. Ahmad, S. Jabbar et al., "TCP CUBIC: a transport protocol for improving the performance of TCP in long distance high bandwidth cyber-physical systems," in *Proceedings of the 2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2018.

[7] U. Khadim, "Information hiding in text to improve performance for word document," *International Journal of Technology and Research*, vol. 3, no. 3, p. 50, 2015.

[8] S. D. Lin and Y.-H. Huang, "An integrated watermarking technique with tamper detection and recovery," *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 11, pp. 4309–4316, 2009.

[9] Q. Gu, Q. Han, Q. Gao, and Q. Chen, "A novel adaptive reversible watermarking algorithm based on wavelet lifting scheme," in *Proceedings of the 2009 International Conference on Information Engineering and Computer Science*, December 2009.

[10] I. Ghafir, J. Saleem, M. Hammoudeh et al., "Security threats to critical infrastructure: the human factor," *The Journal of Supercomputing*, vol. 74, no. 10, pp. 4986–5002, 2018.

[11] H.-N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran, "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies," *Enterprise Information Systems*, pp. 1–25, 2019.

[12] C. Alcaraz, *Security and Privacy Trends in the Industrial Internet of Things*, Springer, Berlin, Germany, 2019.

[13] K. R. Sollins, "IoT big data security and privacy versus innovation," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1628–1635, 2019.

[14] M. M. Gaber, A. Aneiba, S. Basurra et al., "Internet of Things and data mining: from applications to techniques and systems," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 3, Article ID e1292, 2019.

[15] H.-J. Kim, H.-J. Chang, H.-J. Suh, and H.-J. Shon, "A study on device security in IoT convergence," in *Proceedings of the 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)*, May 2016.

[16] X. Zheng, Z. Cai, and Y. Li, "Data linkage in smart internet of things systems: a consideration from a privacy perspective," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 55–61, 2018.

[17] M. A. Ferrag, M. Derdour, M. Mukherjee, A. Derhab, L. Maglaras, and H. Janicke, "Blockchain technologies for the internet of things: research issues and challenges," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2188–2204, 2018.

[18] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, 2017.

[19] S. Sicari, A. Rizzardi, L. A. Rizzardi, and A. Coen-Porisini, "Security, privacy and trust in internet of things: the road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015.

[20] L. Mainetti, L. Manco, L. Patrono, I. Sergi, and R. Vergallo, "Web of topics: an iot-aware model-driven designing approach," in *Proceedings of the 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, December 2015.

[21] A. Čolaković and M. Hadžialić, "Internet of Things (IoT): a review of enabling technologies, challenges, and open research issues," *Computer Networks*, vol. 144, pp. 17–39, 2018.

[22] X. Zhou, W. Zhao, Z. Wang, and L. Pan, "Security theory and attack analysis for text watermarking," in *Proceedings of the 2009 International Conference on E-Business and Information System Security*, May 2009.

[23] Z. Jalil and A. M. Mirza, "A Review of Digital Watermarking Techniques for Text documents," in *Proceedings of the 2009 International Conference on Information and Multimedia Technology*, December 2009.

[24] M. L. Mali, N. N. Patil, and J. Patil, "Implementation of text watermarking technique using natural language watermarks," in *Proceedings of the 2013 International Conference on Communication Systems and Network Technologies*, June 2013.

[25] X. Liu, J. Zhang, H. Wang, X. Gong, and X. Cheng, "A novel text watermarking algorithm based on graphic watermarking framework," in *Proceedings of the 2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications*, November 2014.

[26] A. M. Hamdan and A. Hamarsheh, "AH4S: an algorithm of text in text steganography using the structure of omega network," *Security and Communication Networks*, vol. 9, no. 18, pp. 6004–6016, 2016.

[27] N. S. Kamaruddin, A. Kamsin, L. Y. Por, and H. S. Rahman, "A review of text watermarking: theory, methods and applications," *IEEE Access*, vol. 6, pp. 8011–8028, 2018.

[28] M. Barni, I. Cox, T. Kalker, and H. J. Kim, "Digital Watermarking," in *Proceedings of the 4th International Workshop, IWDW 2005*, vol. 3710, Springer, Siena, Italy, September 2005.

[29] M. Taleby Ahvanooey, H. Dana Mazraeh, and S. H. Tabasi, "An innovative technique for web text watermarking

(AITW)," *Information Security Journal: A Global Perspective*, vol. 25, no. 4-6, pp. 191–196, 2016.

[30] A. A. Mohamed, "An improved algorithm for information hiding based on features of Arabic text: a Unicode approach," *Egyptian Informatics Journal*, vol. 15, no. 2, pp. 79–87, 2014.

[31] M. Talebi Ahvanooei and S. H. Tabasi, "A new method for copyright protection in digital text documents by adding hidden unicode characters in Persian/English texts," *International Journal of Current Life Sciences*, vol. 4, no. 8, pp. 4895–4900, 2014.

[32] M. T. Ahvanooei, S. H. Tabasi, and S. Rahmani, "A novel approach for text watermarking in digital documents by zero-width interword distance changes," *DAV International Journal of Science*, vol. 4, no. 3, pp. 550–558, 2015.

[33] Z. Jalil and A. M. Mirza, "A robust zero-watermarking algorithm for copyright protection of text documents," *Journal of the Chinese Institute of Engineers*, vol. 36, no. 2, pp. 180–189, 2013.

[34] M. Bashardoost, M. S. M. Rahim, and N. Hadipour, "A novel zero-watermarking scheme for text document authentication," *Jurnal Teknologi*, vol. 75, no. 4, pp. 49–56, 2015.

[35] Y. M. Alginahi, M. N. Kabir, and O. Tayan, "An enhanced Kashida-based watermarking approach for increased protection in Arabic text-documents based on frequency recurrence of characters," *International Journal of Computer and Electrical Engineering*, vol. 6, no. 5, pp. 381–392, 2014.

[36] L. Y. Tayan, K. Wong, and K. O. Chee, "UniSpaCh: a text-based data hiding method using Unicode space characters," *Journal of Systems and Software*, vol. 85, no. 5, pp. 1075–1082, 2012.

[37] M. Dalla Preda and M. Pasqua, "Software watermarking: a semantics-based approach," *Electronic Notes in Theoretical Computer Science*, vol. 331, pp. 71–85, 2017.

[38] K. Gopalakrishnan, N. Memon, and P. L. Vora, "Protocols for watermark verification," *IEEE MultiMedia*, vol. 8, no. 4, pp. 66–70, 2001.

[39] M. Naseri, S. Heidari, M. Baghfalaki et al., "A new secure quantum watermarking scheme," *Optik*, vol. 139, pp. 77–86, 2017.

[40] M. H. Shirali-Shahreza and M. Shirali-Shahreza, "A new synonym text steganography," in *Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, August 2008.

[41] Z. Jalil and A. M. Mirza, "Text watermarking using combined image-plus-text watermark," in *Proceedings of the 2010 Second International Workshop on Education Technology and Computer Science*, March 2010.

[42] A. M. Alattar and O. M. Alattar, "Watermarking electronic text documents containing justified paragraphs and irregular line spacing," in *Proceedings of the Security, Steganography, and Watermarking of Multimedia Contents VI*, San Jose, CA, USA, January 2004.

[43] R. Petrovic, B. Tehranchi, and J. M. Winograd, "Security of copy-control watermarks," in *Proceedings of the 2007 8th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, September 2007.

[44] M. H. Alkawaz, G. Sulong, T. Saba, A. S. Almazyad, and A. Rehman, "Concise analysis of current text automation and watermarking approaches," *Security and Communication Networks*, vol. 9, no. 18, pp. 6365–6378, 2016.

[45] P. Singh and R. Chadha, "A survey of digital watermarking techniques, applications and attacks," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, no. 9, pp. 165–175, 2013.

[46] C. Kumar, A. K. Singh, and P. Kumar, "A recent survey on image watermarking techniques and its application in e-governance," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3597–3622, 2018.

[47] M. Agarwal, "Text steganographic approaches: a comparison," 2013, https://arxiv.org/abs/1302.2718.

[48] J. Guru and H. Damecha, "Digital watermarking classification: a survey," *International Journal of Computer Science Trends and Technology (IJCST)*, vol. 5, pp. 8–13, 2014.

[49] H. Kabetta and B. Y. Dwiandiyanta, "Information hiding in CSS: a secure scheme text-steganography using public key cryptosystem," 2012, https://arxiv.org/abs/1201.1968.

[50] M. Nazari, A. Sharif, and M. Mollaeefar, "An improved method for digital image fragile watermarking based on chaotic maps," *Multimedia Tools and Applications*, vol. 76, no. 15, pp. 16107–16123, 2017.

[51] M. Shirali-Shahreza, "Text steganography by changing words spelling," in *Proceedings of the 2008 10th International Conference on Advanced Communication Technology*, February 2008.

[52] N. A. S. Al-maweri, "Robust digital text watermarking algorithm based on unicode extended characters," *Indian Journal of Science and Technology*, vol. 9, no. 48, 2016.

[53] L. Robert and T. Shanmugapriya, "A study on digital watermarking techniques," *International Journal of Recent Trends in Engineering*, vol. 1, no. 2, pp. 223–225, 2009.

[54] J.-M. Shieh, D.-C. Lou, and M.-C. Chang, "A semi-blind digital watermarking scheme based on singular value decomposition," *Computer Standards & Interfaces*, vol. 28, no. 4, pp. 428–440, 2006.

[55] L. Gongshen, X. Ding, B. Su, and K. Meng, "A text information hiding algorithm based on alternatives," *Journal of Software*, vol. 8, no. 8, 2013.

[56] J. T. Brassil, S. Low, N. F. Maxemchuk, and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 8, pp. 1495–1504, 1995.

[57] J. Brassil, "Hiding information in document images," in *Proceedings of the Conference Information Sciences and Systems (CISS-95)*, pp. 482–489, Johns Hopkins University, Baltimore, MD, USA, 1995.

[58] Q. G. Mei, E. K. Wong, and N. D. Memon, "Data hiding in binary text documents," in *Proceedings of the Security and Watermarking of Multimedia Contents III. International Society for Optics and Photonics*, San Jose, CA, USA, 2001.

[59] Y.-W. Kim, K.-A. Moon, and I.-S. Oh, "A text watermarking algorithm based on word classification and inter-word space statistics," in *Proceedings of the ICDAR*, Edinburgh, UK, August 2003.

[60] X. Zhou, S. Wang, W. Zhao, and R. Peng, "A semi-fragile watermarking scheme for content authentication of Chinese text documents," in *Proceedings of the 2009 2nd IEEE International Conference on Computer Science and Information Technology*, August 2009.

[61] L. He, "A part-of-speeach tag sequence text zero-watermarking," in *Proceedings of the Second Symposium International Computer Science and Computational Technology (ISCSCT'09)*, Citeseer, Huangshan, China, December 2009.

[62] Y. Meng, T. Guo, Z. Guo, and L. Gao, "Chinese text zero-watermark based on sentence's entropy," in *Proceedings of the 2010 International Conference on Multimedia Technology*, October 2010.

[63] Z. Jalil, H. Aziz, S. B. Shahid, M. Arif, and A. M. Mirza, "A zero text watermarking algorithm based on non-vowel ASCII characters," in *Proceedings of the 2010 International Conference on Educational and Information Technology*, September 2010.

[64] Z. Jalil and A. M. Mirza, "An invisible text watermarking algorithm using image watermark," in *Innovations in Computing Sciences and Software Engineering*, pp. 147–152, Springer, Berlin, Germany, 2010.

[65] W. Cheng, H. Feng, and C. Yang, "A robust text digital watermarking algorithm based on fragments regrouping strategy," in *Proceedings of the 2010 IEEE International Conference on Information Theory and Information Security*, December 2010.

[66] M.-Y. Kim, O. R. Zaiane, and R. Goebel, "Natural language watermarking based on syntactic displacement and morphological division," in *Proceedings of the 2010 IEEE 34th Annual Computer Software and Applications Conference Workshops*, July 2010.

[67] Y. Meng, L. Gao, X. Wang, and T. Gao, "Chinese text zero-watermark based on space model," in *Proceedings of the 2011 3rd International Workshop on Intelligent Systems and Applications*, May 2011.

[68] F. N. Al-Wesabi, A. Z. Alshakaf, and K. U. Vasantrao, "A zero text watermarking algorithm based on the probabilistic weights for content authentication of text documents," in *IJCA Proceedings on National Conference on Recent Trends in Computing (NCRTC)*, pp. 26–31, Foundation of Computer Science, New York, USA, May 2012.

[69] Y. M. Alginahi, O. Tayan, and M. N. Kabir, "A zero-watermarking verification approach for Quranic verses in online text documents," in *Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, December 2013.

[70] Y. M. Alginahi, O. Tayan, and M. N. Kabir, "An adaptive zero-watermarking approach for authentication and protection of sensitive text documents," in *Proceedings of the International Conference on Advances in Computer and Information Technology—ACIT 2013*, Kuala Lumpur, Malaysia, May 2013.

[71] F. M. Ba-Alwi, M. M. Ghilan, and F. N. Al-Wesabi, "Content authentication of English text via internet using zero watermarking technique and Markov model," *International Journal of Applied Information Systems (IJAIS)*, vol. 7, no. 1, pp. 25–36, 2014.

[72] Y. Zhang, H. Qin, and T. Kong, "A novel robust text watermarking for word document," in *Proceedings of the 2010 3rd International Congress on Image and Signal Processing*, October 2010.

[73] J. Chen, J. Yang, J. Ma, and J. Lu, "Text watermarking algorithm based on semantic role labeling," in *Proceedings of the 2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)*, July 2016.

[74] C. Xiao, C. Zhang, and C. Zheng, "FontCode: embedding information in text documents using glyph perturbation," 2017, https://arxiv.org/abs/1707.09418.

[75] R. A. Alotaibi and L. A. Elrefaei, "Improved capacity Arabic text watermarking methods based on open word space," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 2, pp. 236–248, 2017.

[76] Q. Wen, Y. Wang, and P. Li, "Two Zero-Watermark methods for XML documents," *Journal of Real-Time Image Processing*, vol. 14, no. 1, pp. 183–192, 2018.

[77] S. Hakak, A. Kamsin, J. Veri, R. Ritonga, and T. Herawan, "A framework for authentication of digital Quran," in *Information Systems Design and Intelligent Applications*, pp. 752–764, Springer, Berlin, Germany, 2018.

[78] A. A.-A. Gutub, F. Al-Haidari, K. M. Al-Kahsah, and J. Hamodi, "E-Text watermarking: Utilizing "Kashida" extensions in Arabic language electronic writing," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 1, pp. 48–55, 2010.

[79] H. Yang and A. C. Kot, "Text document authentication by integrating inter character and word spaces watermarking," in *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763)*, June 2004.

[80] Y. M. Alginahi, M. N. Kabir, and O. Tayan, "An enhanced Kashida-based watermarking approach for Arabic text-documents," in *Proceedings of the International Conference on Electronics, Computer and Computation (ICECCO)*, November 2013.

[81] R. J. Jaiswal and N. N. Patil, "Implementation of a new technique for web document protection using unicode," in *Proceedings of the 2013 International Conference on Information Communication and Embedded Systems (ICICES)*, Febuary 2013.

[82] N. Mir, "Copyright for web content using invisible text watermarking," *Computers in Human Behavior*, vol. 30, pp. 648–653, 2014.

[83] Y. Liu, Y. Zhu, and G. Xin, "A zero-watermarking algorithm based on merging features of sentences for Chinese text," *Journal of the Chinese Institute of Engineers*, vol. 38, no. 3, pp. 391–398, 2015.

[84] O. W. Liang and V. Iranmanesh, "Information hiding using whitespace technique in microsoft word," in *Proceedings of the 2016 22nd International Conference on Virtual System & Multimedia (VSMM)*, October 2016.

[85] M. Yingjie, L. Huiran, S. Tong, and T. Xiaoyu, "A zero-watermarking scheme for prose writings," in *Proceedings of the 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, October 2017.

[86] M. Kuribayashi, T. Fukushima, and N. Funabiki, *Data Hiding for Text Document in PDF File*, Springer International Publishing, Cham, Switzerland, 2018.

[87] A. Taha, A. S. Hammad, and M. M. Selim, "A high capacity algorithm for information hiding in Arabic text," *Journal of King Saud University-Computer and Information Sciences*, 2018.

[88] L. Tan, K. Hu, X. Zhou, R. Chen, and W. Jiang, "Print-scan invariant text image watermarking for hardcopy document authentication," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13189–13211, 2018.

[89] S. Al-Nofaie, M. Fattani, and A. A.-A. Gutub, "Capacity improved Arabic text steganography technique utilizing 'kashida'with whitespaces," in *Proceedings of the 3rd International Conference on Mathematical Sciences and Computer Engineering (ICMSCE 2016)*, Lankawi, Malaysia, February 2016.

[90] S. G. Rizzo, F. Bertini, D. Montesi, and C. Stomeo, "Text watermarking in social media," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in*

Social Networks Analysis and Mining-ASONAM '17, July 2017.

[91] R. A. Alotaibi and L. A. Elrefaei, "Utilizing word space with pointed and un-pointed letters for Arabic text watermarking," in *Proceedings of the 2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim)*, April 2016.

[92] Y.-C. Chou, C.-Y. Huang, and H.-C. Liao, "A reversible data hiding scheme using cartesian product for HTML file," in *Proceedings of the 2012 Sixth International Conference on Genetic and Evolutionary Computing*, August 2012.

[93] I.-S. Lee and W.-H. Tsai, "Secret communication through web pages using special space codes in HTML files," *International Journal of Applied Science and Engineering*, vol. 6, no. 2, pp. 141–149, 2008.

[94] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3.4, pp. 313–336, 1996.

[95] H. Lu, G. P. Ma, D. Y. Fang, and X. L. Gui, "Resilient natural language watermarking based on pragmatics," in *Proceedings of the 2009 IEEE Youth Conference on Information, Computing and Telecommunication*, September 2009.

[96] O. Halvani, M. Steinebach, P. Wolf, and R. Zimmermann, "Natural language watermarking for German texts," in *Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security*, June 2013.

[97] H. M. Meral, B. Sankur, A. Sumru Özsoy, T. Güngör, and E. Sevinç, "Natural language watermarking via morphosyntactic alterations," *Computer Speech & Language*, vol. 23, no. 1, pp. 107–125, 2009.

[98] U. Topkara, M. Topkara, and M. J. Atallah, "The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions," in *Proceedings of the 8th Workshop on Multimedia and Security-MM&Sec '06*, September 2006.

[99] K. Bennett, *Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text*, Purdue University, West Lafayette, IN, USA, 2004.

[100] D. Huang and H. Yan, "Interword distance changes represented by sine waves for watermarking text images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1237–1245, 2001.

[101] M. Khairullah, "A novel text steganography system using font color of the invisible characters in microsoft word documents," in *Proceedings of the 2009 Second International Conference on Computer and Electrical Engineering*, December 2009.

[102] L. Guoyuan and W. Guohui, "A new information hiding method based on word 2007," in *Proceedings of the IEEE 2nd International Conference on Software Engineering and Service Science*, July 2011.

[103] H. H. Nasereddin, "Digital watermarking a technology overview," *International Journal of Research and Reviews in Applied Sciences*, vol. 6, no. 1, pp. 89–93, 2011.

[104] M. M. Iqbal, U. Khadam, K. J. Han, J. Han, and S. Jabbar, "A robust digital watermarking algorithm for text document copyright protection based on feature coding," in *Proceedings of the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, June 2019.

[105] J. E. Boritz, "IS practitioners' views on core concepts of information integrity," *International Journal of Accounting Information Systems*, vol. 6, no. 4, pp. 260–279, 2005.

[106] S. Hakak, A. Kamsin, O. Tayan, M. Y. Idna Idris, and G. Amin Gilkar, "Approaches for preserving content integrity of sensitive online Arabic content: a survey and research challenges," *Information Processing & Management*, vol. 56, no. 2, pp. 367–380, 2017.