Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.049215

ARTICLE

# Privacy-Preserving Information Fusion Technique for Device to Server-Enabled Communication in the Internet of Things: A Hybrid Approach

**Amal Al-Rasheed[1], Rahim Khan[2,3,*], Tahani Alsaed[4], Mahwish Kundi[2,5], Mohamad Hanif Md. Saad[6] and Mahidur R. Sarker[7,8]**

[1]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

[2]Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, 23200, Pakistan

[3]Faculty of Computing and Informatics, University Malaysia Sabah, Kota Kinabalu, Sabah, 88400, Malaysia

[4]Applied College, Taibah University, Madinah, 41477, Saudi Arabia

[5]Maynooth International Engineering College, Maynooth University, Co Kildare, W23 X021, Irland

[6]Department of Mechanical Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, 43600, Malaysia

[7]Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Selangor, 43600, Malaysia

[8]University of Design, Innovation and Technology (UDIT), Av. Alfonso XIII, 97, Madrid, 28016, Spain

*Corresponding Author: Rahim Khan. Email: rahimkhan@awkum.edu.pk, mahidursarker@ukm.edu.my

Received: 30 December 2023 Accepted: 27 March 2024 Published: 18 July 2024

## ABSTRACT

Due to the overwhelming characteristics of the Internet of Things (IoT) and its adoption in approximately every aspect of our lives, the concept of individual devices' privacy has gained prominent attention from both customers, i.e., people, and industries as wearable devices collect sensitive information about patients (both admitted and outdoor) in smart healthcare infrastructures. In addition to privacy, outliers or noise are among the crucial issues, which are directly correlated with IoT infrastructures, as most member devices are resource-limited and could generate or transmit false data that is required to be refined before processing, i.e., transmitting. Therefore, the development of privacy-preserving information fusion techniques is highly encouraged, especially those designed for smart IoT-enabled domains. In this paper, we are going to present an effective hybrid approach that can refine raw data values captured by the respective member device before transmission while preserving its privacy through the utilization of the differential privacy technique in IoT infrastructures. Sliding window, i.e., $\delta_i$ based dynamic programming methodology, is implemented at the device level to ensure precise and accurate detection of outliers or noisy data, and refine it prior to activation of the respective transmission activity. Additionally, an appropriate privacy budget has been selected, which is enough to ensure the privacy of every individual module, i.e., a wearable device such as a smartwatch attached to the patient's body. In contrast, the end module, i.e., the server in this case, can extract important information with approximately the maximum level of accuracy. Moreover, refined data has been processed by adding an appropriate nose through the Laplace mechanism to make it useless or meaningless for the adversary modules in the IoT. The proposed hybrid approach is trusted from both the device's privacy and the integrity of the transmitted information perspectives. Simulation and analytical results have proved that the proposed privacy-preserving information fusion technique for wearable devices is an ideal solution for

resource-constrained infrastructures such as IoT and the Internet of Medical Things, where both device privacy and information integrity are important. Finally, the proposed hybrid approach is proven against well-known intruder attacks, especially those related to the privacy of the respective device in IoT infrastructures.

**KEYWORDS**
Internet of things; information fusion; differential privacy; dynamic programming; Laplace function

## 1 Introduction

The Internet of Things (IoT) is defined as a networking infrastructure where smart physical devices, appliances, vehicles, and other objects (preferably physical, are embedded with appropriate sensors, actuators, transceivers, and software) for the timely collection of useful information about a particular phenomenon either through active or passive interaction and sharing it with a centralized module, i.e., a server or edge [1]. These devices or smart objects could range from very simple devices, such as thermostats in smart homes to wearable smart watches, to very complex industrial transportation and machinery, whose sole purpose is to minimize human efforts while carrying out different activities with an equal level of precision and accuracy, along with improving productivity. However, due to their resource-constrained nature, i.e., limited processing and communication power, the data values captured by these devices are highly susceptible to outliers or noise. It occurs either due to malfunctioning of the respective deployed module, i.e., the physical device, or through interference during transmission [2]. Secondly, duplicate data values are quite common in IoT infrastructure, as it is highly likely for neighboring devices to capture similar data. Therefore, captured data values by the respective IoT network should be passed through a sophisticated refinement process. In this process, outliers and duplicate data values are removed or replaced with approximate, accurate values while preserving the overall integrity of the respective data. In addition to the outliers or duplicate data values, a more crucial issue, which has a direct correlation with the adoption of IoT in different application domains, is the privacy of the respective source device that is required to be ensured, as every individual in an IoT-enabled smart environment, i.e., a smart home or hospital, does not want to reveal his or her identity while providing respective information to the centralized module [3]. Thus, smart devices could be a great addition to homes, hospitals, and offices, but their associated risks, as these are highly vulnerable to both security and privacy attacks from intruders, should be kept in mind as well. Therefore, newly developed fusion methodologies should incorporate appropriate measures to ensure data integrity through sophisticated noise-evading mechanisms. Additionally, it should preserve the privacy of the underlined smart device in the IoT through the integration of smart security parameters. device in IoT through the integration of smart security parameters Data fusion and aggregation are among the de facto standards to ensure the integrity of the captured data, i.e., both from outliers and duplicate perspectives, through the respective smart devices in the IoT environment [4]. These approaches could be adopted both locally, i.e., on the deployed smart device, such as a temperature sensor, or centrally on the respective server or edge modules. However, in both cases, these mechanisms are bound to ensure the maximum possible integrity, precision, and accuracy, along with the minimum possible information loss ratio, irrespective of the application domains [5]. In the multiple sensor domain, an influential value, i.e., high density, is reduced by the approximate outcome of the respective fusion model, where functions are classified as directionally or simply monotones [6].

Likewise, an effective feature extraction model has been integrated with a deep learning-enabled multi-model scheme to ensure precise and accurate diagnosis of faults, especially from two different signals related to time [7]. Furthermore, a device vulnerability-aware fusion approach has been reported to make sure that the integrity of the refined data is consistent with the captured raw data in resource-constrained networks [8]. An effective fusion approach was developed by Fitzgerald et al. [9] for minimizing the ratio of duplicate data values through two different models, i.e., *1-K* and *n-K*, where the former was applicable for multi-sensor scenarios whereas the latter is useful when actuators along with appropriate sensors are required. For the human activity recognition model, a smart fusion approach has been developed where wearable devices are bound to carry out processing on the visual data in the multiple head-enabled systems [10]. Similarly, false readings have been rectified through a lightweight fusion scheme developed by Singh et al. [11]. Although these fusion approaches are smart enough to refine captured data values with the maximum possible integrity level of the refined data, the information loss ratio and complexity are far beyond the acceptance bounds. Moreover, these schemes are specifically designed for information fusion and do not care about the privacy of the respective smart devices.

The privacy of a smart device could be preserved through a sophisticated privacy mechanism that must ensure that information has sufficient details, enabling the intended receiver to extract valuable information while preserving the privacy of the respective smart device in the IoT domain. Federated learning has been introduced to guarantee the privacy of the respective source devices from adversaries in IoT infrastructures [12]. It is a distributed machine learning methodology where data captured by respective devices is transmitted in weighted form instead of actual data values. Moreover, certain ratios of noise could be incorporated into real data before the aggregation process [13]. Likewise, federated learning and blockchain-based architecture have been developed to safeguard the privacy of patients in the smart healthcare environment of the Internet of Medical Things (IoMTs) [14]. The fusion of federated learning and blockchain has been developed for privacy preservation in IoT infrastructure [15]. Additionally, Monero and deniable ring signatures, along with licensing technology, were integrated to form a sophisticated privacy-preserving model for resource-constrained networks [16]. Differential privacy is among the most prominent approaches that are developed to provide the expected level of privacy to the respective source device, especially in the open communication module of resource-limited networking infrastructures [17]. In differential privacy, the expected output function is not directly correlated to the presence or absence of individual records, as depicted in Fig. 1. An interesting feature of differential privacy is that the privacy of the device is guaranteed irrespective of the adversary's knowledge while carrying out the attack. Although these approaches have resolved security and privacy issues, particularly those linked with the direct communication of device and server modules over a shared medium, fusion-enabled strategies have been completely neglected by the research domain. Information fusion is not only important to improve the accuracy and precision ratio of the various decisions taken by the respective server or edge modules, but also equally important for the reliability of data as well. Secondly, existing approaches have not considered device location and propagation ratio, which are quite crucial in the development of a reliable and consistent privacy-preserving communication approach, especially in the IoT networking infrastructure. Thirdly, differential privacy is very handy if incorporated into the security protocol as it guarantees the safety of the concerned device even if the adversary has sufficient knowledge. Therefore, a hybrid algorithm with embedded fusion and device privacy needs to be developed which incorporates exceptional capabilities of information fusion and differential privacy, especially in the IoT.

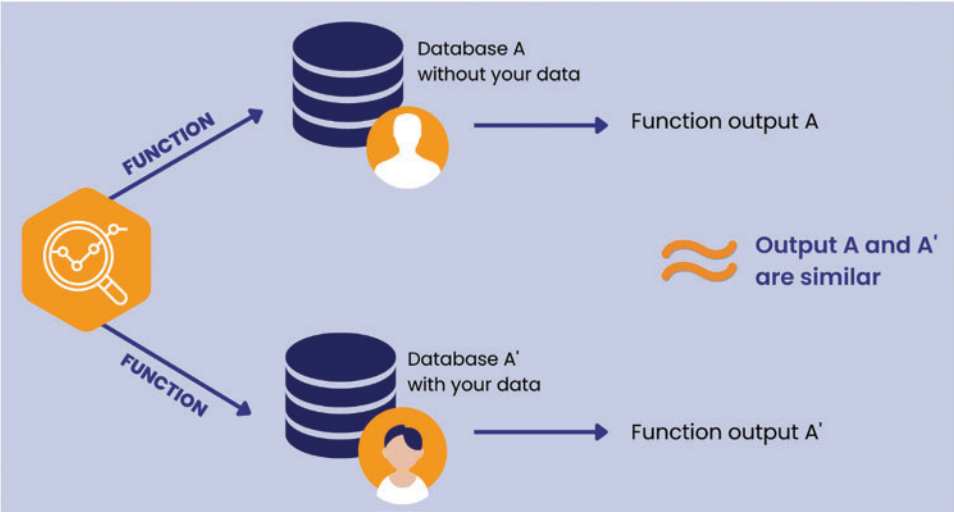


**Figure 1:** Differential privacy is not affected by presence OR absence of a record

In this paper, a privacy-preserving information fusion-enabled model has been developed to perform refinement of raw data values, which are captured through smart devices, in IoT infrastructures. This model consists of two broad steps, i.e., (i) data fusion, which is carried out by the respective device, i.e., locally, to refine captured data through a sophisticated procedure that is sliding window-based dynamic programming methodology that is highly pruned against outliers or noisy data; (ii) secondly, to preserve the privacy of the respective smart device, i.e., a mobile phone or smartwatch, through differential privacy, where noise is added to actual raw data before transmission to make it meaningless for the intruders or adversaries in the IoT. The main contributions of this paper are given below:

1. A dynamic programming-enabled data fusion approach that is carried out locally on every smart device in the IoT ensures that refined data is more trustworthy than raw data as it is noise-and duplicate-free (where applicable).
2. The concept of a sliding window is integrated with a traditional dynamic programming algorithm to minimize its processing and response time, additionally, making it applicable for real-time data sets.
3. Differential privacy has been adopted to ensure the privacy of the respective smart device or source module by adding sufficient noise to actual data before its transmission in the IoT; and
4. Mechanism for sharing refined data with a centralized module without revealing the identity of the source device.

The remaining manuscript is organized as follows: In Section 2, a short but comprehensive review of the most relevant literature is presented, where different issues with the existing schemes are identified. In the next section, i.e., Section 3, a detailed description of the proposed methodologies, both information fusion and device privacy, is provided with sufficient detail along with supported mathematical models. In Section 4, the performance of the proposed hybrid information fusion and privacy-preserving scheme is presented both graphically and in textual format. Finally, concluding remarks along with future research directives are described in Section 5.

## 2 Literature Review

Noisy data and privacy of devices, that is both source and destination, are among the crucial issues, which form the basis for the adoption of IoT infrastructure in different application domains. In literature, these issues were addressed separately, i.e., sophisticated fusion approaches were developed to minimize the expected ratio of the noisy data whereas federated learning, particularly differential privacy, and blockchain technology were developed to preserve the device's privacy. In this section, a detailed description of those techniques where either fusion or device privacy issues have been addressed especially for the IoT domain. A brief overview of the existing approaches is presented in the following subsections.

### *2.1 Information Fusion and Aggregation*

Information fusion and aggregation are used to refine raw data that is captured through various sensing devices and transmitted via a shared medium to the respective server [18,19]. These approaches could resolve both duplicates and outliers locally, i.e., at the respective source devices, or centrally at the concern server or edge module. In scenarios where multiple devices are active, an extended version of the function, i.e., pre-aggregation, was developed where the influencing factor of certain attributes or values, preferably those with maximum value is reduced through the separation of functions into monotone and others. These measures were taken to attain the expected outcome from the model [6]. Likewise, effective feature extraction and deep learning-enabled multi-models have been integrated to ensure the precise and accurate diagnosis of faults, especially from two different signals related to the time [7]. Furthermore, a device vulnerability-aware fusion approach has been reported to make sure that the integrity of the refined data is consistent with the captured raw data in resource-constraint networks. An effective fusion approach was developed by Fitzgerald et al. [9] for minimizing the ratio of duplicate data values through two different models, i.e., *1-K and n-K*. The former was used for multi-sensor scenarios. In contrast, the latter is useful when actuators along with appropriate sensors are required. For the human activity recognition model, a smart fusion approach has been developed where wearable devices process the visual data in multiple head-enabled systems [10]. Similarly, false readings have been rectified through a lightweight fusion scheme developed by Singh et al. [11]. Furthermore, a smart methodology has been reported to refine vibration signals especially those collected through the respective wearable devices. Secondly, Bayesian fusion models particularly those with built-in interval estimators are applied directly to active channels in the IoT domain [20]. Entropy and fusion methodologies have increased the accuracy level of the underlying prediction system, as false readings directly affect the performance of an autonomous system. Additionally, this system completes the fusion process before the identification of injuries and illness classification [21]. Pradhan et al. [22] have developed a methodology of truncated bits where unimportant bits are dropped automatically to control the computational tasks of the underlined system. A power-effective fusion methodology has been developed by Xiao et al. [23] to resolve duplicates and outliers' issues autonomously. Moreover, this approach has guaranteed that captured readings are rectified without compromising the overall integrity level of the data set. Apart from this, the least mean square especially with embedded multi-directional features was utilized to minimize the approximate ratio of the duplicate data values in IoT [24]. It is a weighted-enabled approach, which is constantly updated to find optimal values to produce a more precise way of handling both duplicates and noise values. A hybrid model of cluster and power-aware techniques was developed where both optimal path identification and aggregation issues were resolved through fuzzy logic and capuchin algorithms [25]. Although these approaches have addressed both fusion and aggregation issues, these are either

application-specific or very hard to implement in other domains. Secondly, these approaches have not considered the security and privacy issues of the communicating devices in the IoT.

### *2.2 Privacy Preserving Fusion Approaches*

Device or data privacy is a critical issue that is directly correlated with the IoT infrastructures as every device must share its captured data, i.e., sensitive, and private, via un-secure wireless media, which is highly susceptible to numerous adversary attacks and, thus, could lead to its possible leakage [26]. Therefore, privacy-preserving methodologies have been adopted in different application domains of the IoT to ensure that the privacy of the respective wearable device, i.e., smartphone or watch, is preserved, even if data is transmitted via open communication channels. Anonymity is a common technique used to preserve the privacy of the source device in numerous IoT networks. In this technique, heuristic algorithms are utilized to divide available records into groups and then every record is replaced with a central value, thus hiding important details from the adversary [27]. Furthermore, an attack on communication medium and background knowledge is prevented through an enhanced version of k-anonymity, i.e., L-diversity. Similarly, a k-anonymity and $\theta$ sensitive hybrid methodology has been introduced where diversity levels are separated through a $\theta$-oriented threshold value and append certain noise to preserve the privacy of both data and device [28]. Similarly, differential privacy has been introduced for the effective preservation of data by applying certain functions to append noise and make transmitted data meaningless to the adversary or intruder module. Differential privacy is a committed approach where it is almost impossible for the intruder to extract meaningful information irrespective of how much knowledge and processing power, he/she has [29]. Fan et al. [30] have presented an enhanced version of the traditional differential privacy by adding noise through the Laplace mechanism where the information loss ratio is sufficiently reduced. Similarly, adopted of Kalman's filtering methodology has been reported to enhance privacy levels while keeping the information loss ratio to its minimum possible threshold [31]. Federated learning has been introduced to guarantee the privacy of the respective source devices from adversaries in IoT infrastructures [11]. It is a distributed machine learning methodology where data captured by a respective device is transmitted in weighted form instead of actual data values. Moreover, a certain ratio of noise could be incorporated into real data before the aggregation process [12]. Likewise, federated learning and blockchain-based architecture have been developed to safeguard the privacy of patients in the smart healthcare environment of IoMTs [13]. The fusion of federated learning and blockchain has been developed for privacy preservation in IoT infrastructure [14]. Additionally, Monero and deniable ring signatures along with licensing technology were integrated to form a sophisticated privacy-preserving model for the resource's constraint networks [15]. Furthermore, a lightweight and privacy-preserving technique was developed by Zhang et al. [31] to extract a noise-free data set. A structure-oriented and efficient routing-based technique was introduced and applied to the next-generation wireless sensor networks [32]. Although, both information fusion and privacy issues are addressed separately quite well in an effective manner, however, a hybrid methodology is required to be developed, which ensures both the privacy and accuracy of the data in the IoT domain.

## 3 Proposed Hybrid Information Fusion and Privacy-Preserving Scheme

Generally, in IoT and other resource-constrained networking infrastructures, devices, i.e., smart watches or mobile phones, are deployed preferably near the underlined phenomenon, which may be passive or active. As these devices are resource constraints, i.e., limited computation, storage, and transmission power, therefore, it is highly likely that captured data has a certain ratio of false readings, which should be refined before its transmission to the respective module, i.e., server or cloud in this

case. Secondly, as refined data is transmitted over open communication media an intruder could intercept it and the privacy of the source device could be compromised. Therefore, privacy-preserving approaches should be utilized to keep the anonymity of the source device for both intruders irrespective of how knowledgeable he/she is. A detailed discussion of the proposed information fusion approach has been provided in the following subsection, which is followed by an explanation of the proposed differential privacy-based approach.

### *3.1 Dynamic Programming and Sliding Window-Based Information Fusion: Device Level*

In IoT infrastructures, every smart device $C_i$ is deployed in a closed neighborhood of the respective phenomenon, which is required to be monitored 24/7, and captures data values simultaneously after a predefined time interval. These captured readings are shared with the centralized module, i.e., server $S_j$ in this case, through a shared communication medium where important decisions, preferably informed, are made based on these readings. Thus, the accuracy and preciseness of the underlined decision support system implemented on the respective server or edge module have a direct correlation to these readings, which are captured through the respective smart devices. As these devices $C_i$ are resources limited and the probability of false ready is high, therefore, captured data of every device is required to be processed, i.e., refined, either locally by the respective device or centrally by the concerned server module. Local data fusion is very fruitful as not only is it reliable, but it saves considerable resources through a strict communication or transmission policy, i.e., only accurate and precise data is transmitted. To address this issue, we have presented a sliding window and dynamic programming-based fusion approach, which could be implemented directly on smart devices without compromising performance measures. In this approach, every device is bound to capture data consecutively and share it with the respective server module in blocks instead of individual reading and $C_{\mathrm{i}}$ must ensure that every reading in a particular block should be within the predefined bounds. These bounds are the defined threshold values or expected range of the underlined sensor module, i.e., upper, and lower. For example, the upper and lower bounds for a temperature sensor would be approximately 10 and 60 $^0$C depending on the region where IoT is deployed. Thus, every device must ensure that every value in a particular block should fall within these ranges. If a value is exceeded or less than these bounds, then it could be an outlier and, thus, required to be refined before the activation of the communication process. Let's say we have two vectors $X = x1, x2, x3....xn$ and $Y = y1, y2, y3....ym$ represents the upper and lower bound of a particular smart device in IoT, then vector $Z = z1, z2, z3....zk$ is accurate if every value of $Z$ *f* alls within these bounds as described Eq. (1).

$$\forall_{i\,=\,1,\,2,\ldots,k} Z_k \in [X_n,\ Y_m] \tag{1}$$

It is important to note that if all member values of a particular block lie within the defined bounds, i.e., upper, and lower, then the whole block is assumed as accurate and is ready to be transmitted to the concerned destination server or edge module. Alternatively, if any value, say $Z_2$ of a particular block, does not fall within these ranges or bounds, then surely, it is an outlier and should be refined, which is sorted out by replacing this outlier with an average value of the whole block except outliers as given in the Eq. (2).

$$z_2 = Avg\left(\frac{\sum_1^k Z_k}{k-1}\right) \tag{2}$$

For example, a simple scenario of noisy data values, i.e., temperature sensor readings, in a particular block are represented in Fig. 2 where the upper and lower bounds for our country are

approximately −10 and 60, respectively. After completion of the respective refinement process, the concerned block of data, i.e., $Z_k$, is ready to transmit via wireless media to the concerned destination module. If this procedure is repeatedly applied to every block of raw data before the communication, then the accuracy ratio of various decisions made by the respective server could be approximately 97% as the transmission medium is not ideal and has a certain ratio of interference. Moreover, the information loss ratio in this case is very rare and could be considered as zero, cos only false readings are replaced with average values.

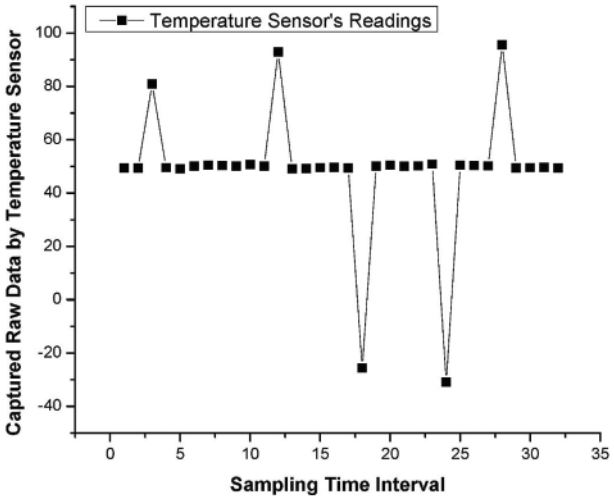


**Figure 2:** Generic overview of outliers or noisy data values in a particular block

Although noisy data or outliers are rectified through repeated application of the procedure, duplicate data values persist as these are accurate values in a particular block. The proposed scheme compares every captured reading with existing accurate values in the block as given in the Eq. (3).

$$\forall_{i\,=\,1,\,2,\ldots,n}[Z_i\} \notin Z_{1,\,2,\,3,\ldots,n-1} \tag{3}$$

where the most recent value captured through the respective device is represented by $Z_i$.

$$d\,(Z_k,\ X_i) = \sqrt{\frac{\sum_{i\,=\,1}^{n}(Z_k - X_i)^2}{n}} \tag{4}$$

The proposed refinement process becomes more effective if we integrate the sliding window-based concept with the traditional dynamic programming approach. In the proposed setup, every block of data could be a possible window of size $\delta$, where a dynamic programming approach is used to find similarity indexes of the currently captured block with the previously transmitted block. If we assume that the variable Length is used to represent the length of the Longest common subsequence between the two blocks, i.e., current, and previously transmitted, then its calculation is based Eq. (5).

$$Length_{i,\,j,\,\delta} = \begin{cases} i = 0 & l_i = 0\ OR\ l_j = 0 \\ Lenghth_{i-1,\,j-1} + 1, & if\ l_i\,l_j \geq 0\ and\ X_i = Y_j \\ Max\left[Lenghth_{i-1,\,j},\ Lenghth_{i,\,j-1}\right], & if\ l_i\,l_j < 0\ and\ X_i \neq Y_j \end{cases} \tag{5}$$

Similarity indexes of two blocks are computed using a sliding window-based dynamic programming algorithm by filling out the indexes table with values representing several duplicates in the current

window. For example, a value four in the indexes table represents that four occurrences of a particular value exist. As soon as the indexes table is computed, the next step is to find the longest common subsequence, which is duplicate data values and needs to be discarded to save valuable resources as these values have already been transmitted in the previous block.

### *3.2 Smart Device's Privacy Preservation through Differential Privacy in IoT*

As soon as a particular block of data has been accurately refined through the proposed sliding window-based dynamic programming algorithm and now it is ready to be shared with the respective server or edge module in IoT. However, before the transmission, the privacy of the source device must be preserved, which is ensured through the effective utilization of differential privacy. For this purpose, the very first step is finding the privacy budget, which describes the actual amount of noise required to be added to the underlined block of data to ensure the expected level of privacy as depicted in Fig. 3.

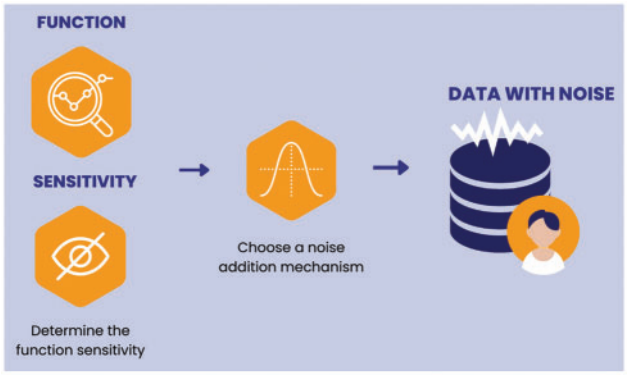


**Figure 3:** Expected ratio of noise required to be added through Laplace function

Two functions or blocks, say $X$ and $Y$, are assumed as neighbors if these blocks or functions differ only in a single individual where all possible outputs are represented by $S$ as given in the Eq. (6).

$$\frac{P_r[F(X) \in S]}{P_r[F(Y) \in S]} < e^{\varepsilon} \tag{6}$$

where $F$ represents a randomized function, a function having multiple outputs under the same input. Moreover, a probability distribution that describes its out does not represent point distribution. It is important to note that out of function $F$ could be the same with or without data of an individual, thus, enough randomness should be incorporated in function $F$ such that it is difficult for an adversary to differentiate between $X$ or $Y$ based on the output of $F$. Thus if an intruder could not decide which of the input, i.e., from $X$ and $Y$, has been used, then surely, he/she cannot determine whether actual data was embedded in the message input or not. Finally, $\epsilon$ is the budget parameter, which can tune the privacy level of the input and its value is inversely proportional to the privacy, i.e., a smaller value of $\epsilon$ means the highest possible privacy and vice versa. We have used the Laplace mechanism for addition noise to the input, i.e., block of data in this, to ensure the privacy of the source module as given in Eq. (7).

$$F(X) = f(x) + Lap\left(\frac{S}{\varepsilon}\right) \tag{7}$$

where *Lap(s)* and *s* are respective representations of sampling and sensitivity, respectively. The probability density function of the Laplace function is given in the Eq. (8).

$$f\left(\frac{x}{u}, b\right) = \frac{1}{2b}\exp\left(-\frac{x-u}{b}\right) \tag{8}$$

where $\mu$ and $b$ represent location parameter and diversity, respectively. Likewise, the Laplace distribution function is very easy to integrate (if we are interested in separating two symmetric cases, which is quite common in the IoT infrastructure). It is due to the effective utilization of the absolute value function. The cumulative distribution function of a random variable $X$ is that $X$ is bounded to take values that are less than or equal to $x$ in any domain. Thus, the cumulative distribution function of the Laplace function is given by Eq. (9). Additionally, it is highly recommended to note that Eq. (9) is applicable only if we are interested in separating approximately symmetric cases in the real environment.

$$F(x) = \int_{-\infty}^{\infty} f(u)du = \begin{cases} \frac{1}{2}\exp\left(\frac{\infty-u}{b}\right) & if\ x < u \\ 1 - \frac{1}{2}\exp\left(-\frac{\infty-u}{b}\right) & if\ x \geq u \end{cases} \quad = \frac{1}{2} + \frac{1}{2}sgn(x-u)(1-exp) - \frac{x-u}{b}\Big)\Big) \tag{9}$$

In the next step, the Laplace mechanism is applied to the function F to add sufficient noise to the actual data values in such a way that the privacy of the respective source device is preserved irrespective of how much knowledge and computational powers the adversary has. This integrated function will convert respective blocks of data to a form secure version that is deferentially private. As soon as this deferentially private version of the underlined refined block is generated, then it could be transmitted via an insecure channel to the intended destination module, i.e., server or edge in this case. Even though, if this deferentially private version of the block is intercepted by the intruder, then he/she will not be able to extract any useful information from it, i.e., this version of data is meaningless to the intruder and, thus, the privacy of the respective source device, $C_i$, is preserved.

The proposed hybrid algorithm, which is fusion-oriented and privacy-preserving, is presented below as Algorithm 1, where m and n are used to describe block lengths. Additionally, X and Y are the input variables along with sliding window control parameter is $\delta$. Finally, two functions are used in the proposed algorithm that is the computation of the similarity indexes table and the length of the longest common subsequence that is used to present how similar two captured data sets are.

**Algorithm 1:** Sliding Window-Based Dynamic Programming Algorithm for IoT

**Input:** Blocks of Data $A_n$ & $B_m$ Captured through Smart Device $C_i$
**Output:** Refined Block of Data (Noise and Duplicate Free)
$Z_k \leftarrow (A_n, B_m)$;
m: Length of Block A;
n: Length of Block B;
$\delta \leftarrow \omega_i$;
X [1, 2, . . . , m, 1, 2, . . . , n];
Y [0, 1, 2, . . . , m, 0, 1, 2, . . . , n];
Function-LCSS (m, n, $\delta$);
Function-Length (Y, X, i, j, $\delta$);

(Continued)

**Algorithm 1 (continued)**

```
Remove Duplicates from A_n or B_m;
Refined Block under δ A_n and B_m;
LCSS (block (A_δi, B_δi ){
    for I = 0...m do
        Lengthi, 0 = 0
    end for
    for j = 0...n do
        Length0, j = 0
    end for
    for i = 1...m do
        for j = 1...n do
            if A_i = B_j then
                X_i,j,δ = X_i−1,j−1,δ +1;
                Y_i,j = "↖";
            else
                if Y_i,j−1,δ ≤ Y_i−1,j,δ then
                    X_i,j,δ = X_i−1,j,δ;
                    Y_i,j,δ = "↑";
                else
                    X_i,j,δ = X_i,j−1,δ;
                    Y_i,j,δ = "←";
                end if
            end if
        end for
end for
}
Length (Arg1, Arg2, Arg3, Arg4) {
    if i = 0 & j = 0 & δ then
        return 0;
    end if
    if Y_i,j,δ == ↖ then
        LCSS (Y, X, i – 1, j – 1, δ);
        print Xi;
    else
        if Y_i,j,δ == ↑ then
            LCSS (Y, X, i – 1, j, δ);
        else
            print-LCSS (Y, X, i, j – 1, δ);
        end if
    end if
    }
return Location Aware Noise
```

## 4 Performance Evaluation of the Proposed Hybrid Approach

This section thoroughly describes numerous simulation results on different evaluation metrics to double-check the performance of the proposed privacy-preserving information fusion approach in comparison to the existing state-of-the-art approaches. Therefore, these mechanisms were implemented in an open-source simulator, that is *NS2*, where every scheme was deeply analyzed through effective performance evaluation metrics such as the ratio of fused data (both duplicate and outliers), Privacy level, Vulnerability, Computation, and transmission time, lifetime, and the ratio of the transmitted packets along with congestion control. Initially, a random deployment methodology for devices was adopted to make it consistent with the real-time IoT infrastructures. Secondly, every device was deployed such that it fell within the coverage area of at least one server module. The coverage area for both member devices and server modules was assumed to be 450, which is the standard coverage area of the Xbee transceiver attached to the Wasp-Mote Board from Libelium. Moreover, other specific parameters are provided in Table 1 along with a brief but comprehensive description. Moreover, every server in IoT does not need to have an equal number of member devices as far as random deployment is concerned. Lastly, we have selected on-board battery power according to the available standards such as 1150, 2300, 6600, and 13,000 mAh.

**Table 1:** Simulation setup for the IoT

| Parameters | Approximate values |
|---|---|
| IoT's covered area | 1000 m × 1000 m |
| Deployed devices $C_i$ | Approximately (96, 192, 286, 384) |
| Server $S_j$ | 3, 9, 18, 27 |
| Edge | 1 |
| Battery power ($E_i$) | 1150, 2300, 6600, 13,000 mAh |
| Residual energy ($E_r$) | $E_i$–$E_{cons}$ |
| Power consumed on packet transmission ($P_{Tx}$) | 91.4 mW |
| Power consumed on packet receiving ($P_{Rx}$) | 59.1 mW |
| Transceiver's coverage area ($T_r$) | 450 m |
| Length of beacon $S_j$ | 70 to 100 bytes |
| Back-off timer $S_j$ | Random |
| Signal to noise ratio (SNR) p | 10 dB |
| Channel delay | 10 ms |
| Consumption of power in idle mode | 1.27 mW |
| Consumption of power in sleep mode | 15.4 μW |
| Transceiver's power cost ($T_i$) | 1 mW |
| Power threshold (Reception) | 1024 bits |
| Packet size | 128 bytes |
| Distance among devices and server | 400 m |
| Sampling rate | 10 s |
| Typologies | Static and random |

### *4.1 Accuracy and Precision Ratio Metric: Proposed Information Fusion Algorithm*

Figures accuracy and precision ratio are among the crucial measures used to evaluate the performance of the newly developed fusion approaches and their effectiveness in increasing the accuracy level of the underlined decision support system. Hence, the proposed sliding window and dynamic programming-enabled fusion approach has been tested thoroughly especially in the context of data accuracy along with field-proven algorithms preferably under similar conditions and infrastructures. During the simulation setup and experiments, we have observed that the proposed fusion approach has achieved a higher accuracy level than its counterpart algorithms as depicted in Fig. 4. Enhanced accuracy level of the proposed fusion approach has a direct correlation with the concept of the sliding window, i.e., $\delta$, which not only enables the respective device to carry out the refinement process on limited data, i.e., block but bounds it to work on the most recently captured readings. Finally, as the proposed fusion approach is applied locally on every device, thus, valuable resources are saved along with reducing or controlling traffic.

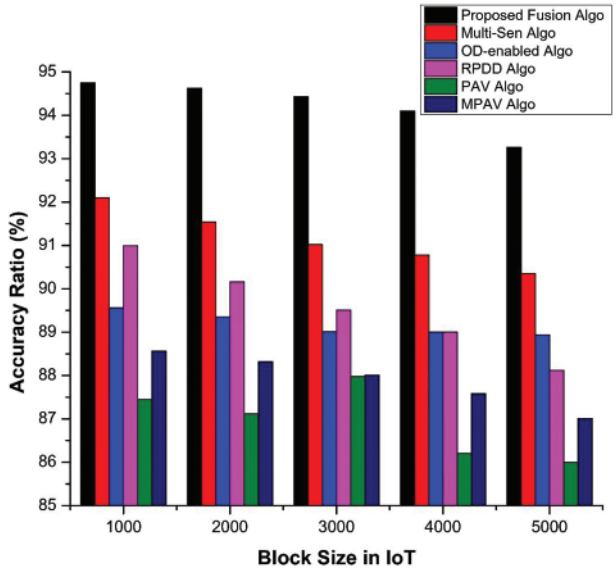


**Figure 4:** Precision and accuracy ratio of the proposed information fusion and existing techniques

### *4.2 Refinement Ratio: Device-Level Information Fusion*

A crucial issue linked with both duplicate and outlier data values is the wastage of valuable resources such as bandwidth, battery, and transceiver if both are tackled properly and in time. Although global data fusion is fruitful in terms of the overall refinement ratio, it compromises the efficient utilization of the available resources, which is crucial in resource-limited domains such as IoT. Therefore, local dynamic programming and a sliding window-based local data fusion approach have been implemented, which performs exceptionally well than other counterpart algorithms. Fig. 5 shows the effectiveness of the proposed fusion approach as it has a maximum possible refinement ratio than existing approaches, i.e., outliers' detection ("OD"), Multiple Sensor-based, multiple pattern anomaly value ("MPAV"), "PAV" and rare pattern drift detection ("RPDD"), under similar technological and environmental conditions. Additionally, the performance of the proposed fusion algorithm is not correlated to the size of the data set whereas the performance of other algorithms is affected by a

slight increase in data set size. The reason behind this non-correlation is its utilization of a fixed block preferably of most recently captured values.

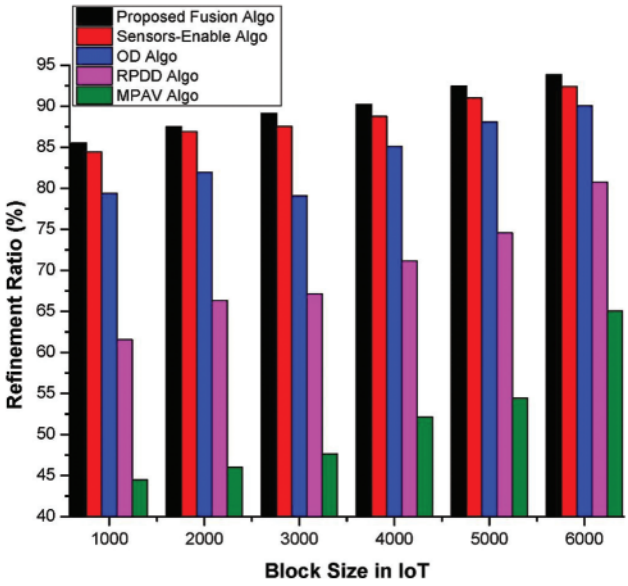


**Figure 5:** Refinement ratio of the proposed information fusion and existing techniques

### *4.3 Processing OR Computational Overhead*

Another crucial performance testing measure is processing or computational time overhead, which is the approximate time required to complete the fusion process either locally or globally. An algorithm, i.e., fusion, with approximately the lowest computational overhead and maximum accuracy ratio is assumed as an ideal solution for IoT, but achieving both goals simultaneously is very hard. However, the proposed information fusion algorithm has performed exceptionally well than existing algorithms as shown in Fig. 6. Firstly, the minimum processing or computational overhead of the proposed information fusion approach is due to the sliding window concept, which enables a device to operate only on the most recently captured readings and within a specific domain, i.e., block. Secondly, as the proposed approach is implemented on the device level, therefore, it must operate the captured reading of a single device. Moreover, an interesting feature of the proposed information fusion approach is its fixed computational overhead as it has to operate a fixed size block, which makes an ideal solution for the resource's constraint networks such as IoT.

#### *4.3.1 Information Loss Ratio*

Information loss is among the critical issues associated with both information fusion and aggregation mechanisms in IoT. However, this ratio could be reduced through appropriate measures instead of removing a record, it should be replaced with a more precise and accurate record. Secondly, this refinement process should be subjected to careful utilization of those readings, captured in closed time intervals, i.e., if a value is identified as noise and it is captured around 10:00 AM, then readings collected in the closed quarter such as 9:59 AM and 10:01 AM should be considered. The proposed information fusion approach has the minimum possible information loss ratio as it bounds the refinement process within a particular block and window. Therefore, the information loss ratio in

the proposed information fusion approach falls within the acceptable range of existing approaches as shown in Fig. 7.

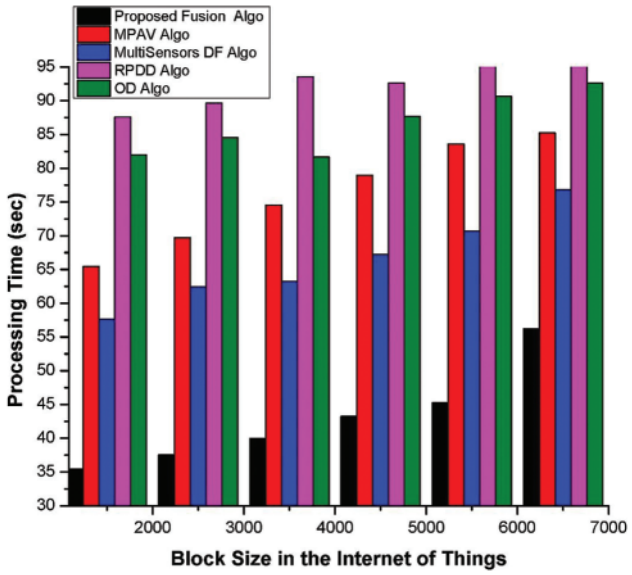


**Figure 6:** Computational time overhead of the proposed information fusion algorithm and existing techniques

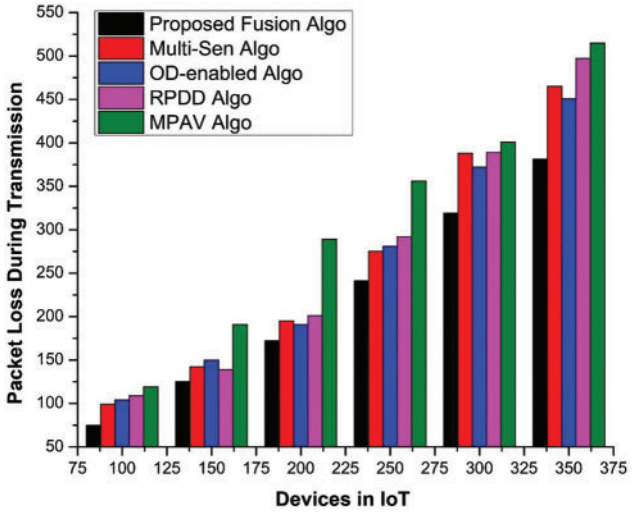


**Figure 7:** Information loss ratio of the proposed fusion and existing approaches

### *4.4 Device's Privacy Preservation in IoT*

Approximate entropy is an indication of the overall complexity ratio of the underlined data set, i.e., a slightly bigger value of entropy concludes that more information is available whereas a smaller value less information. Simulation results have been checked for different values $\epsilon$, i.e., $\epsilon = 0.2$, $\epsilon = 0.3$, $\epsilon = 0.4$, $\epsilon = 0.5$, and $\epsilon = 0.6$, which shows that appropriate noise, preferably random, has been

added to the original data value, which makes it hard for the adversary module to get any clue about information being transmitted. Thus, not only data has been communicated securely, but the privacy of the respective source device is preserved. Additionally, as depicted in Fig. 8, data with embedded noise through the differential privacy is quite different from the original data streams, which are captured by the respective device. Apart from this, differential privacy-enabled communication approaches are pruned against well-known intruder attacks such as man in the middle or masquerading.

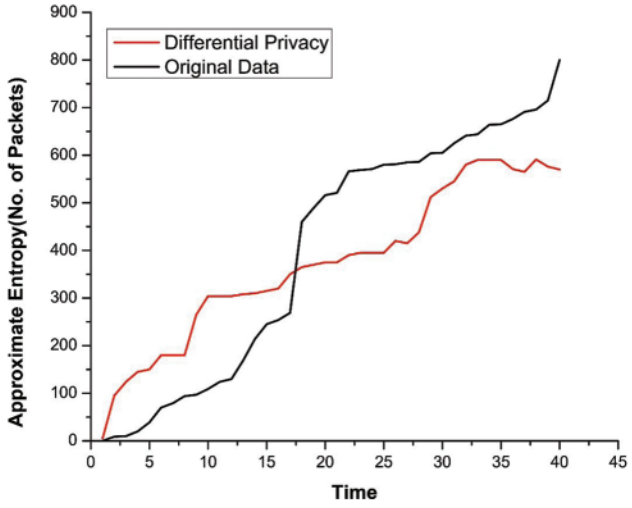


**Figure 8:** Approximate entropy in differential privacy where the value of is u is less than 1

### *4.5 Well-Known Security Attacks*

The proposed privacy-preserving approach is pruned against well-known intruder attacks as sufficient noise has been embedded through the Laplace function, which makes actual data meaningful for intruders or adversaries in the IoT. If one of the adversaries can intercept this message, which is encrypted using differential privacy with the Laplace function, then he/she will not be able to extract any meaningful information if he/she tries to apply head and trial or another sophisticated hacking function. Thus, the proposed approach is to prune against masquerading attacks. Additionally, it is pruned against man in a middle attack, which is launched by adversaries to pretend itself as someone else who is hijacking an ongoing communication session where every message from the source and destination module is intercepted by the intruder and updated copies of captured message is shared. In this case, both source and destination are not aware that their communication has been hijacked and the contents of every message are compromised. However, the proposed scheme appends sufficient noise to the actual data, which is based on budget, i.e., $\epsilon$, and, thus, makes it impossible for the adversary to understand the contents of the underlined messages. Likewise, it is pruned against perfect forward and backward secrecy, phishing, device, and server impersonation attacks as shown in Table 2.

**Table 2:** Simulation setup for the IoT

| Security attack | Traditional | DES | AES-128 | AES-256 | EC | Proposed |
|---|---|---|---|---|---|---|
| User impersonation | No | Yes | Yes | Yes | Yes | Yes |

(Continued)

**Table 2 (continued)**

| Security attack | Traditional | DES | AES-128 | AES-256 | EC | Proposed |
|---|---|---|---|---|---|---|
| Device impersonation | No | Yes | Yes | Yes | Yes | Yes |
| Server impersonation | Yes | Yes | No | No | No | Yes |
| Eavesdropping | No | Yes | Yes | Yes | Yes | Yes |
| Perfect forward and backward | Yes | No | Yes | Yes | Yes | Yes |
| Man in the middle | No | No | Yes | Yes | Yes | Yes |

## 5 Conclusion and Future Work

Internet of Things (IoT) has been widely used in different application domains where member devices could collect information that is extremely private and sensitive. Therefore, appropriate security and privacy measures are required to be adopted to make these networking infrastructures trustworthy for both individual persons and organizations. In addition to the privacy and security of data and source modules, noisy data or outliers are another crucial aspect of IoT networks, and they could affect the accuracy and precision ratio of those systems deployed in the actual working environment of controlling and monitoring. Similarly, duplicate data values should be handled without compromising the accuracy and precision ratio of the IoT. In this paper, we have presented a privacy-preserving information fusion approach to ensure noise-free communication between the device and server module while preserving the privacy of both devices simultaneously. A sliding window-based dynamic programming approach is presented to guarantee timely detection and correction of the noisy data especially at the device level. Moreover, refined data values are secured through a sophisticated methodology, i.e., differential privacy, which not only safeguards the data but also preserves the privacy of the source module. Simulation results have concluded that the proposed scheme is an ideal solution as it has an embedded noise detection and correction facility along with strong privacy.

The proposed privacy-preserving information fusion approach is likely to be extended for IoT infrastructures where the communication party, i.e., devices $C_i$ and servers $S_j$ could be mobile.

**Acknowledgement:** We are thankful to Ministry of Higher Education of Malaysia under the Research Grant LRGS/1/2019/UKM-UKM/5/2 and Princess Nourah bint Abdulrahman University for financing this researcher through Supporting Project Number (PNURSP2024R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Funding Statement:** Ministry of Higher Education of Malaysia under the Research Grant LRGS/1/2019/UKM-UKM/5/2 and Princess Nourah bint Abdulrahman University for financing this researcher through Supporting Project Number (PNURSP2024R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Author Contributions:** The authors confirm their contribution to the paper as follows: study conception and design: Amal Al-Rasheed, Rahim Khan, Mahwish Kundi; data collection: Amal Al-Rasheed, Mahwish Kundi, Tahani Alsaed, Mohamad Hanif Md.; analysis and interpretation of results: Rahim khan, Tahani Alsaed, Mahidur R. Sarker, Mohamad Hanif Md.; draft manuscript preparation: Amal Al-Rasheed, Rahim Khan, Tahani Alsaed, Mahidur R. Sarker, Mohamad Hanif Md. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** This article does not involve data availability and this section is not applicable.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References


[1] W. Ding, X. Jing, Z. Yan, and L. Yang, "A survey on data fusion in the Internet of things: Towards secure and privacy-preserving fusion," *Inf. Fusion*, vol. 51, pp. 129–144, 2019.

[2] J. A. M. Calero *et al.*, "Bindi: Affective Internet of things to combat gender-based violence," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21174–21193, 2022. doi: 10.1109/JIOT.2022.3177256.

[3] H. Lin, S. Garg, J. Hu, X. Wang, M. J. Piran and M. S. Hossain, "Data fusion and transfer learning empowered granular trust evaluation for Internet of Things," *Inf. Fusion*, vol. 78, pp. 149–157, 2022.

[4] M. H. Shirvani and M. Masdari, "A survey study on trust-based security in Internet of Things: Challenges and issues," *Internet of Things*, vol. 21, pp. 100640, 2023.

[5] Y. Liu, H. Wang, M. Peng, J. Guan, J. Xu and Y. Wang, "DeePGA: A privacy-preserving data aggregation game in crowdsensing via deep reinforcement learning," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4113–4127, 2019.

[6] G. Beliakov, S. James, A. Kolesárová, and R. Mesiar, "Cardinality-limiting extended pre-aggregation functions," *Inf. Fusion*, vol. 76, pp. 66–74, 2021. doi: 10.1016/j.inffus.2021.05.004.

[7] C. Che, H. Wang, X. Ni, and R. Lin, "Hybrid multimodal fusion with deep learning for rolling bearing fault diagnosis," *Measurement*, vol. 173, pp. 108655, 2021. doi: 10.1016/j.measurement.2020.108655.

[8] R. Khan, M. Zakarya, Z. Tan, M. Usman, M. A. Jan and M. Khan, "PFARS: Enhancing throughput and lifetime of heterogeneous WSNs through power-aware fusion, aggregation, and routing scheme," *Int. J. Commun. Syst.*, vol. 32, no. 18, pp. e4144, 2019. doi: 10.1002/dac.4144.

[9] E. Fitzgerald, M. Pióro, and A. Tomaszwski, "Energy-optimal data aggregation and dissemination for the Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 955–969, 2018. doi: 10.1109/JIOT.2018.2803792.

[10] M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Multi-level feature fusion for multimodal human activity recognition in internet of healthcare things," *Inf. Fusion*, vol. 94, pp. 17–31, 2023. doi: 10.1016/j.inffus.2023.01.015.

[11] S. Singh and D. Kumar, "Energy-efficient secure data fusion scheme for IoT based healthcare system," *Future Gener. Comput. Syst.*, vol. 143, pp. 15–29, 2023. doi: 10.1016/j.future.2022.12.040.

[12] Y. Shen, S. Shen, Q. Li, H. Zhou, Z. Wu and Y. Qu, "Evolutionary privacy-preserving learning strategies for edge-based IoT data sharing schemes," *Digit. Commun. Netw.*, vol. 9, no. 4, pp. 906–919, 2023. doi: 10.1016/j.dcan.2022.05.004.

[13] K. Wei *et al.*, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3454–3469, 2020. doi: 10.1109/TIFS.2020.2988575.

[14] S. Singh, S. Rathore, O. Alfarraj, A. Tolba, and B. Yoon, "A framework for privacy-preservation of IoT healthcare data using federated learning and blockchain technology," *Future Gener. Comput. Syst.*, vol. 129, pp. 380–388, 2022. doi: 10.1016/j.future.2021.11.028.

[15] S. K. Singh, L. T. Yang, and J. H. Park, "FusionFedBlock: Fusion of blockchain and federated learning to preserve privacy in Industry 5.0," *Inf. Fusion*, vol. 90, pp. 233–240, 2023. doi: 10.1016/j.inffus.2022.09.027.

[16] T. Li, H. Wang, D. He, and J. Yu, "Blockchain-based privacy-preserving and rewarding private data sharing for IoT," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 15138–15149, 2022. doi: 10.1109/JIOT.2022.3147925.

[17] B. Jiang, J. Li, G. Yue, and H. Song, "Differential privacy for industrial internet of things: Opportunities, applications, and challenges," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10430–10451, 2021. doi: 10.1109/JIOT.2021.3057419.

[18] J. Song, Y. Liu, J. Shao, and C. Tang, "A dynamic membership data aggregation (DMDA) protocol for smart grid," *IEEE Syst. J.*, vol. 14, no. 1, pp. 900–908, 2019. doi: 10.1109/JSYST.2019.2912415.
[19] A. Saleem *et al.*, "FESDA: Fog-enabled secure data aggregation in smart grid IoT network," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6132–6142, 2019. doi: 10.1109/JIOT.2019.2957314.
[20] C. Brüser, J. M. Kortelainen, S. Winter, M. Tenhunen, J. Pärkkä and S. Leonhardt, "Improvement of force-sensor-based heart rate estimation using multichannel data fusion," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 227–235, 2014. doi: 10.1109/JBHI.2014.2311582.
[21] J. Fang *et al.*, "Data fusion in forecasting medical demands based on spectrum of post-earthquake diseases," *J. Ind. Inf. Integr.*, vol. 24, pp. 100235, 2021. doi: 10.1016/j.jii.2021.100235.
[22] S. Pradhan, E. Sinha, and K. Sharma, "Data fusion by truncation in wireless sensor network," in *Adv. Comput. Commun. Paradigms: Proc. Int. Conf. ICACCP 2017*, Singapore, 2017, vol. 1, pp. 544–551.
[23] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *IPSN 2005. Fourth Int. Symp. Inform. Process. Sens. Netw., UCLA*, Los Angeles, CA, USA, Apr. 2005, pp. 63–70.
[24] N. Mahesh and S. Vijayachitra, "Hierarchical autoregressive bidirectional least-mean-square algorithm for data aggregation in WSN based IoT network," *Adv. Eng. Softw.*, vol. 173, pp. 103275, 2022. doi: 10.1016/j.advengsoft.2022.103275.
[25] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," *Peer Peer Netw. Appl.*, vol. 16, no. 1, pp. 189–209, 2023. doi: 10.1007/s12083-022-01388-3.
[26] F. Loukil, C. Ghedira-Guegan, K. Boukadi, A. N. Benharkat, and E. Benkhelifa, "Data privacy based on IoT device behavior control using blockchain," *ACM Trans. Internet Technol. (TOIT)*, vol. 21, no. 1, pp. 1–20, 2021. doi: 10.1145/3434776.
[27] A. Solanas, A. Martínez-Ballesté, and J. Domingo-Ferrer, "V-MDAV: A multivariate microaggregation with variable group size," in *17th COMPSTAT Symp. IASC*, Rome, Italy, 2006, pp. 917–925.
[28] G. Wu, X. Chen, Z. Gao, H. Zhang, S. Yu and S. Shen, "Privacy-preserving offloading scheme in multi-access mobile edge computing based on MADRL," *J. Parallel Distr. Comput.*, vol. 183, pp. 104775, 2024. doi: 10.1016/j.jpdc.2023.104775.
[29] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proc. 2010 ACM SIGMOD Int. Conf. Manage. Data*, New York, NY, USA, Jun. 2010, pp. 735–746.
[30] L. Fan, L. Xiong, and V. Sunderam, "Differentially private multi-dimensional time series release for traffic monitoring," in *Data Appl. Secur. Priv. XXVII: 27th Annu. IFIP WG 11.3 Conf., DBSec*, New York, NY, USA, Jul. 15–17, 2013, pp. 33–48.
[31] J. Zhang, Y. Zhao, J. Wu, and B. Chen, "LVPDA: A lightweight and verifiable privacy-preserving data aggregation scheme for edge-enabled IoT," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4016–4027, 2020. doi: 10.1109/JIOT.2020.2978286.
[32] K. Haseeb, N. Islam, T. Saba, A. Rehman, and Z. Mehmood, "LSDAR: A light-weight structure based data aggregation routing protocol with secure internet of things integrated next-generation sensor networks," *Sustain. Cities Soc.*, vol. 54, pp. 101995, 2020. doi: 10.1016/j.scs.2019.101995.