Received October 29, 2017, accepted December 2, 2017, date of publication December 27, 2017, date of current version March 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2787422

Cognitive Privacy Middleware for Deep Learning Mashup in Environmental IoT

AHMED M. ELMISERY[1], MIRELA SERTOVIC[2], AND BRIJ B. GUPTA[3], (Member, IEEE)

[1]Department of Electronics Engineering, Universidad Tecnica Federico Santa Maria, Valparaiso 1680, Chile
[2]Faculty of Humanities and Social Sciences, University of Zagreb, 10000 Zagreb, Croatia
[3]National Institute of Technology Kurukshetra, Kurukshetra 136119, India

Corresponding author: Ahmed M. Elmisery (ahmedmisery@gmail.com)

This work was partially supported in part by the Dirección General de Investigación, Innovación y Postgrado through the Security in Cyber-Physical Systems for Power Grids Project, Federico Santa María Technical University, Chile, under Grant UTFSM-DGIP PI.L.17.15, in part by the Advanced Center for Electrical and Electronic Engineering CONICYT-Basal Project under Grant FB0008, and in part by the Microsoft Azure for Research under Grant 0518798.

ABSTRACT Data mashup is a Web technology that combines information from multiple sources into a single Web application. Mashup applications support new services, such as environmental monitoring. The different organizations utilize data mashup services to merge data sets from the different Internet of Multimedia Things (IoMT) context-based services in order to leverage the performance of their data analytics. However, mashup, different data sets from multiple sources, is a privacy hazard as it might reveal citizens specific behaviors in different regions. In this paper, we present our efforts to build a cognitive-based middleware for private data mashup (CMPM) to serve a centralized environmental monitoring service. The proposed middleware is equipped with concealment mechanisms to preserve the privacy of the merged data sets from multiple IoMT networks involved in the mashup application. In addition, we presented an IoT-enabled data mashup service, where the multimedia data are collected from the various IoMT platforms, and then fed into an environmental deep learning service in order to detect interesting patterns in hazardous areas. The viable features within each region were extracted using a multiresolution wavelet transform, and then fed into a discriminative classifier to extract various patterns. We also provide a scenario for IoMT-enabled data mashup service and experimentation results.

INDEX TERMS IoT networks, cloud computing, environmental monitoring, smart cities, big data mashup, multimedia data.

I. INTRODUCTION

Environmental hazards of natural origin involve large extensions of land such as earthquakes, tsunamis, volcano eruptions, landslides and forest fires are common in countries like Chile and produce emergency scenarios where roads are often saturated or damaged and power supplies are down, disrupting connectivity. These hazards can easily affect a large number of people and isolate them from their surrounding environment. While information and storage capabilities are becoming virtually limitless, in such situations, accessing the right information at the right time by the right organization is a crucial requirement to take proper decisions and to publish highly relevant information to the affected communities and helpers in charge of handling the emergency situations [1]. Decision makers usually require access to highly accurate information servers and data application to estimate the number of affected citizens in a certain region and the best available ways to support them.

Environmental monitoring is one of the areas, which attracts public concern. The advance of cloud computing and Internet of things reshaped the manner in which the sensed information is being managed and accessed. The advances in sensor technologies have accelerated the emergence of environmental sensing service. These new services grasp the significance of new techniques in order to understand the complexities and relations in the collected sensed information. Particularly, it utilizes portable sensing devices to extend the sensing range, and cloud-computing environments to analyse the big amount of data collected by various Internet of multimedia things (IoMT) networks in a productive form. Various kinds of sensors are being deployed in the environment as the physical foundation for most of the environmental sensing services. It is highly desirable to link the sensed data with external data collected from different services in order to increase the accuracy of the predictions [2] In regions with environmental hazards, a large number of citizens makes intensive observations about these regions using their mobile phone during their daily activities. This massive data is expected to be generated from different sources and published on various Internet of multimedia things (IoMT) context-based services such as Facebook®, Waze® and Foursquare®. In such situation, it is beneficial to include such data in the decision-making process of environmental monitoring services. In this context, Data Mash-up services appear as a promising tool to accumulates this data and manage in an appropriate way. Data mashup [3] is a web technology that combines information from multiple sources into a single web application for specific task or request. Mashup technology was first introduced in [4] and since then it creates a new horizon for service providers to integrate their data to deliver highly customizable services to their customers [3]. Data mashup can be used to merge datasets from external IoMT context-based services to leverage the monitoring service from different perspectives like providing more precise predictions and performance, and alleviating cold start problems [5] for new environmental monitoring services. Due to that, Providers of the next generation environmental monitoring services keen to gain accurate data mash-up services for their systems. However, privacy is an essential concern for the application of mashup in IoMT-enabled environmental monitoring, as the generated insights obviously require the integration of different behavioural and neighbouring environment data of citizens and from multiple IoMT context-based services. This might reveal private citizens' behaviours that were not available before the data mashup. A serious privacy breach can occur if the same citizen is registered on multiple sites, so adversaries can try to deanonymize the citizen's identity by correlating the information contained in the mashuped data with other information obtained from external public databases. These breaches prevent IoMT context based services to reveal raw behavioural data of the citizen to each other or to the mashup service. Moreover, divulgence citizens' data represent infringement against personal privacy laws that might be applied in some countries where these sites operate. As a result, if the citizens know their raw data are revealed to other parties, they will absolutely distrust this site. According to surveys results in [6], [7] the users might leave a service provider because of privacy concerns.

We believe that environmental cognition services can be enriched by extensive data collection infrastructures of IoMT-enabled data mashup services especially in the domain of urban environmental monitoring. IoMT mashup techniques can be used to merge datasets from external IoMT networks to leverage the functionalities of environmental deep learning service from different perspectives like providing more precise predictions and computation performance, improving the reliability toward citizens, minimizing the impacts of environmental hazards on affected citizens, and providing an early response in cases when the event is inevitable. Due to that, Providers of the next generation environmental cognition services keen to utilize IoMT-enabled data mashup services for their systems. Effective multimedia mining is an essential requirement for the IoMT-enabled data mashup services, since, the extracted patterns obviously requires the integration of different multimedia contents generated from multiple IoMT networks. These multimedia contents may contain random noise, which complicates the pattern discovery process. A serious decline in accuracy occurs when the noisy data is present in the pile of contents that will be processed through the data mashup techniques. Handling this noisy data is a real challenge since it is hard to be distinguished from an abnormal data, it could prevent the environmental deep learning service from fully embracing the useful data extracted from the mashup service. Managing this problem will enable the IoMT- enabled data mashup services to execute different recognition methods for identifying the abnormal objects in an effective manner.

In this work, we proposed Cognitive -based middleware for private data mashup (CMPM) that bear in mind privacy issues related to mashup multiple datasets from IoMT context-based services for environmental monitoring purposes. We focus on stages related to datasets collection and processing and omit all aspects related to environmental monitoring, mainly because these stages are critical with regard to privacy as they involve different entities. We present two cognitive concealment algorithms to protect citizens' privacy and preserve the aggregates in the mashuped datasets in order to maximize usability and attain accurate insights. Using these algorithms, each party involved in the mashup is given a complete control on the privacy of its dataset. In the rest of this paper, we will generically refer to behavioural and neighbouring environment data as Items. Section II describes some related work. In section III we introduce IoMT-enabled data mashup network scenario landing our CMPM. In section IV introduces the proposed cognitive concealment algorithms used in our CMPM. In section V introduces the proposed anomaly detection solution used within the environmental cognition service. Section VI describes some experiments and results based on concealment algorithms for IoT context-based services. Finally, Section VII includes conclusions and future work.

II. RELATED WORK

In practice, end-users have shown an increasing privacy concern when they share their behavioural and location data, especially when this data is shared with untrusted parties [8]. This happens due to the following reasons: First, the behavioural and location data collected by the end-users are personal by nature, e.g., the end-users might decline to reveal their physical daily activities, along with the location and time where they perform such activities. Second, despite the apparently benign nature of collected data. This data can be realized to deduce the private data of end-users that have not intentionally shared. For example, private personal information can be inferred from the brainwave data of users wearing popular brainwear wireless EEG headsets [9], such as the digits of PIN numbers, ATM card data, location of residence and other sensitive data. Third, the independent attribute of data collection amplifies the consideration and need for a privacy-respecting technique when handling the gathered data, since the data can be collected by malignant third parties at any time or location without an explicit consent from the end-users. Finally, the ambiguity of third parties' practices and their obligations in the issues related to data breaches due to cyber-attacks or insider attacks. For example, a security researcher succeeded in detecting severe vulnerabilities in drug infusion pumps that allow an attacker to change the amount of injected drug to a fatal dose that should harm the users [10]. The manufacturers of the affected brands failed to patch the security lapses in their products deployments and sued the researcher. Cases similar to this, hinder the wide acceptance of various monitoring services. Hence, this is a crucial need to preserve the privacy of sensitive data of end-users. Based on the results of a recent survey, increasing demand for privacy protection has been a major concern for the end- users who offering their data to untrusted third parties in order to receive any value-added services [8]. The end-users insist that they need full control over the data collection process and cannot tolerate that their data is stored in a remote location and accessible to different external parties.

For the review of related work, we identify two fundamental research categories were identified: First privacy preserving systems are discussed. Finally, vision-based environmental monitoring systems are shortly surveyed.

A. PRIVACY PRESERVING SYSTEMS

The majority of the literature addresses the problem of privacy on third-party services [11]–[16]; Due to it is a potential source of leakage of personally identifiable Information. However, a few works have studied the privacy for mashup services [17]. The work in [3] discussed a private data mashup system, where the authors formalize the problem as achieving a k-anonymity on the integrated data without revealing detailed information about this process or disclosing data from one party to another. In [18] it is proposed a theoretical framework to preserve the privacy of customers and the commercial interests of merchants. Their system is a hybrid recommender that uses secure two-party protocols with public key infrastructure to achieve the desired goals. In [19] and [20] it is suggested another method for privacy preserving on centralized services by adding uncertainty to the data, using a randomized perturbation technique while attempting to make sure that necessary statistical aggregates don't get disturbed much. Hence, the server has no knowledge about true values of individual data for each user. They demonstrate that this method does not decrease essentially the obtained accuracy of the results. But recent research work [21], [22] pointed out that these techniques don't provide levels of privacy as it was previously thought. In [22] it is Pointed out that arbitrary randomization is not safe because it is easy to breach the privacy protection it offers. They proposed a random matrix based spectral filtering techniques to recover the original data from perturbed data. Their experiments revealed that in many cases random perturbation techniques preserve very little privacy.

B. DEEP VISION-BASED ENVIRONMENTAL MONITORING SYSTEMS

Three main research studies based on machine vision were performed in order to estimate monitor Environmental hazards. Martinez-de Dios *et al.* [23], [24] proposed a method, which computes a 3D perception model of forest fires from multispectral complementary views including an aerial one. Infrared cameras in mid-infrared spectral window and in the far infrared windows are used with visible cameras. A statistical sensor fusion approach using Kalman filtering is employed to merge measurements from different sensors in order to obtain an overall estimation. Telemetry sensors, GPS data and artificial beacons or natural marks (such as a tree or a fire fighter truck) are necessary for the calibration procedure. The position of the fire front, the rate of spread and the maximum height of the flames are estimated. Experiments were carried out on lands of up to 2.5 hectares. In [25], a method was proposed in which 3D points are computed from fire feature points matched using stereoscopic images. From these points, the geometrical characteristics of a fire front like its position on the ground, its shape and its surface are estimated. This method does not need reference marks on the working field of view. The use of several stereovision systems allows obtaining a complete 3D form of the fire front and the estimation of its volume, but the technique presented in [26] is only at the laboratory scale. In another work, the use of NIR stereovision systems was introduced to obtain fire measurement even in the presence of smoke [27]. Experiments were carried out indoors and outdoors on platforms of a maximum size of about 0.5 hectare. Verstockt *et al.* [28], [29] have developed a method using a series of cameras distributed around the fire to compute a 3D model of fires and smokes. This framework merges the single-view detection results of the multiple cameras by homographic projection onto multiple horizontal and vertical planes, which slice the scene. The crossing of these slices creates a 3D grid of virtual sensor points. The location of the fire, its size and its direction of propagation are estimated with precision. This procedure is limited to fire fronts not larger than 2 × 4 m2. One of the most important aspects in the extraction of fire characteristics is the detection and extraction of the fire region. The robustness of the measures is correlated with the efficiency of the segmentation technique. This task is very challenging when conducted in outdoor unstructured environment. The majority of the work in wildland fire segmentation is conducted in the visible spectrum. Little work was conducted in the NIR and other infrared spectrums. In the visible spectrum, different colour spaces such as RGB, YCbCr, CIE $L*a*b*$, YUV, HSI are used and it is observed that no colour system seems to be more effective than another to characterize the fire. Concerning the segmentation methods, it is difficult to determine the efficiency of each of them in the specific case of wildland fires because of the lack of a benchmark of these methods on a standard database of wildfire images. However, comparisons of different methods on a representative dataset revealed that the methods of Phillips *et al.* [30], Lucile *et al.* [31] and Collumeau *et al.* [32] Outperform other methods. More recently, machine learning based techniques [33] were developed that have shown an increased performance in the fire front detection and segmentation. A benchmarking of different fire detection and fire segmentation is given in [33]. In the thermal infrared spectrum, the segmentation becomes easier but the thermal infrared cameras are very expensive and have low resolutions compared to their visible counterparts. The use of NIR cameras seems to be promising. Ideally, a hybrid system, which combines visible and infrared spectrums, would perform better in urban fire detection and segmentation.

Regarding the unitization of IoT in environmental monitoring systems. Yuan *et al.* [34] give an overview of past work dealing with the use of aerial vehicles in the context of forest fires. The majority of this work was dealing with the collection of information and an aerial viewof the fire propagation in order to help in firefighting [35], [36]. Fire detection using aerial vehicles was also conducted in other research [37], [38]. A pioneering work using low-cost aerial vehicles with on-board visible and infrared cameras in close range fire detection experiments was conducted in [39]–[42]. Nominating Internet of multimedia things for fire detection which combine insights from ground-based and airborne sensors along with multimodal cooperative vision analytics can permit a better segmentation, detection, and monitoring of urban forest fires. Additionally, the utilization Internet of multimedia things allows collaborative modes in building complementary three-dimensional views, which will enable the extraction of 3D geometrical characteristics of fires at a larger scale.

A methodology for data mashup service for IOMT enabled collaborative monitoring were proposed in [17]. The authors consider the scenario where the IoMT-enabled data mashup (MDMS) integrates datasets from multiple IoMT networks for the environmental cognition service; figure (1) illustrates the architecture supported in this work. The proposed architecture hosts an intelligent middleware for private data mashup (DIMPM), which enables connectivity to diverse IoT devices via varied sensing technologies. In doing so, the functionalities of the proposed architecture support a cloud based infrastructure for environmental cognition services. The cloud environment promotes a service-oriented approach to big data management, providing a deep learning layer for analyzing the merged data. The architecture follows a layered approach, where the bottom layer is the Environmental IoT devices, while the highest layer is the environmental cognition service.

The data mashup process can be summarized as follows;

- The environmental deep learning service sends a query to the IoMT- enabled data mashup service to gather information related to a specific region to leverage its predictions and performance.

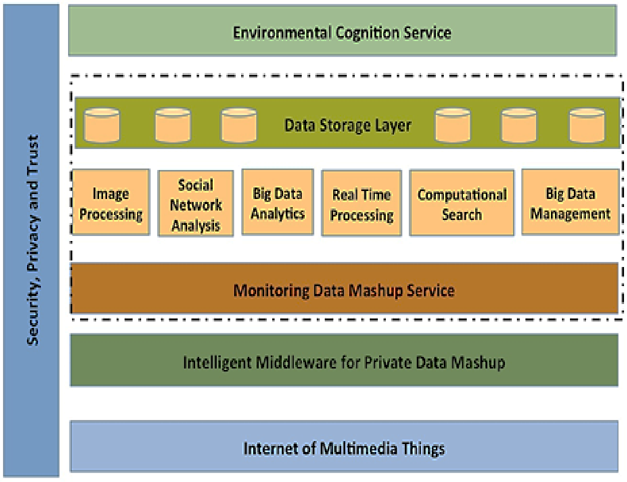
Security, Privacy and Trust
Environmental Cognition Service
Data Storage Layer
Image Processing
Social Network Analysis
Big Data Analytics
Real Time Processing
Computational Search
Big Data Management
Monitoring Data Mashup Service
Intelligent Middleware for Private Data Mashup
Internet of Multimedia Things

FIGURE 1. IoMT-enabled data mashup with Third Party Environmental Cognition Service.

- At the IoMT - enabled data mashup:
  - The coordinator agent at search in its cache to determine the providers which could satisfy this query, then it transforms[∘] the query into appropriate sub-queries languages suitable for each provider's database.
  - The manager agent unit sends each sub-query to the candidate IoT providers to incite them about the data mashup process.
- Based on prior agreement between the mashup provider and data providers, the providers who agree to offer purpose specific datasets to the mashup process will:
  - Forward the sub-query to its manager agent within the intelligent middleware for private data mashup.
  - The manager agent rewrites the sub-query considering the privacy preferences for its host and produces a modified sub-query for the data that can be published. This step allows the manager agent to audit all issued sub-queries and prevent ones that can extract sensitive information.
  - The resulting dataset is concealed to hide real data using the appropriate obfuscation algorithm depending of the type of multimedia data.
  - Finally, each provider submits its concealed data to the IoMT- enabled data mashup service that in turn unites these results and performs further analysis on them.

The obtained information is delivered to environmental cognition service. The environmental deep learning service uses these datasets to accomplish its data analytics goals.

III. DATA MASHUPS IN IOT-ENABLED ENVIRONMENTAL MONITORING SCENARIO

We consider the scenario where the IoMT-enabled data mashup service (IoMT-enabled DMS) integrates various types of datasets from multiple IoMT context-based services for the IoT-enabled environmental monitoring;

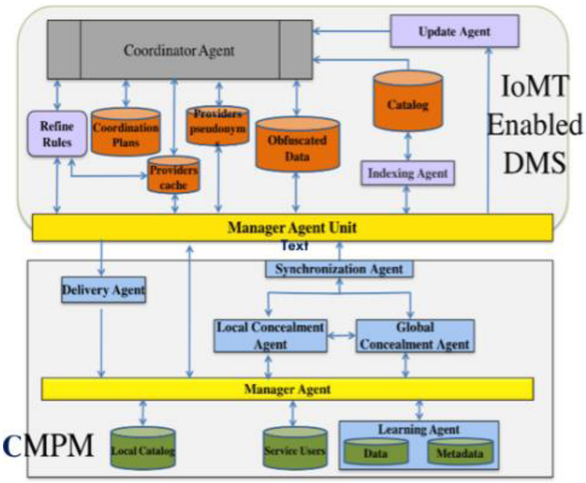

FIGURE 2. The Building Blocks of IoMT-enabled DMS and CMPM.

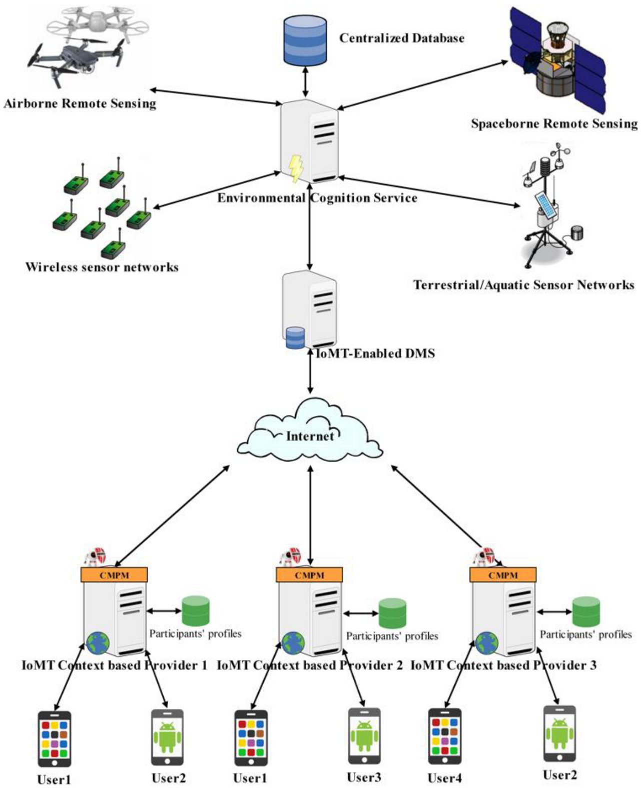

FIGURE 3. IoMT-enabled data mashup service with Third Party IoMT context-based services.

figures (2) and (3) illustrates the scenario used in this work. We assume all the involved parties follow the semi-honest model, which is a realistic assumption because each party needs to accomplish some business goals and increases its revenues. Also, we assume all parties involved in the data mashup have similar items set (activities' catalogue) but the users' sets are not identical. Each IoT context-based service has its own ETL (Extract, Transform, Load) service that has the ability to learn behavioural and neighbouring environment data of citizens. The data mashup process based on CMPM can be summarized as follows; the environmental deep learning service sends a query to the IoMT-enabled DMS to gather information related to behavioural and neighbouring environment data of citizens in a specific region to leverage its predictions and performance. The coordinator Agent in IoMT-enabled DMS lookup in its providers' cache to determine the providers could satisfy that query then it transforms query of the environmental deep learning service into appropriate sub-queries languages suitable for each provider's database. The manager agent unit sends each sub-query to the candidate providers to incite them about the data mashup process. The provider who decides to participate in that process, forwards the sub-query to its manager agent to refine it considering its privacy preferences. This step allows the manager agent to audit all issued sub-queries and prevent ones that can extract sensitive information. The resulting dataset sent to the local concealment agent (LOA) to hide real participants' data using the appropriate concealment algorithm. Then, every synchronization agents at each provider along with the coordinator agent engage in distributed joint process to identify frequent and partially frequent items in each dataset, then send the joined results to the coordinator. The coordinator agent builds a virtualized schema for the datasets and submits it to each provider involved in the mashup process. Based on this virtualized schema, the providers incite their global concealment agent (GOA) to start the appropriate concealment algorithm on the locally concealed datasets. Finally, the providers submit all the resulting datasets to IoMT-enabled DMS that in turn unites these results and delivers them to the environmental cognition service. The environmental deep learning service uses these datasets to accomplish the required data analytics goals. We use anonymous pseudonyms identities to alleviate providers' identity problems, as the database providers does not want to reveal their ownership of the data to competing providers moreover the IoMT-enabled DMS will keen to hide the identities of providers as a business asset.

The recommendation process based on the two stage obfuscation algorithms can be summarized as following:

The IoMT-enabled DMS acts as an integrator that collects data from lower level IoMT context provider, processing them, and delivering the result to its upper level environmental deep learning service. Each IoMT context provider is responsible for sensing and collecting various types of data from the physical world. This data can be textual data or multimedia data. This data represents some parameters, measurements and conditions in streets, specific regions and buildings, transportation and the air quality. That means IoMT devices can be utilized to monitor and collect data from everything. The process of merging the sensed data at each IoMT context provider can be summarized as follows. The IoMT context provider broadcast message to devices' owners in its network to incite them to submit their personal profiles and multimedia data in order to start notifying the users about the status of specific environmental hazard. Individual users who decide to respond to this request, specify their privacy preferences to their IoMT context provider then submit the requested data. More details about this process can be summerized as follows:

1. An IoMT context provider, broadcasts a message to other devices' owners in their network to indicate its intention to start notifying the users of specific region about the status of specific environmental hazard in their surroundings. The provider requests may require either of textual and/or multimedia data.
2. Individual users that decide to participate in that request, integrates all textual and/or multimedia data that they collected for a specific region. In addition, each participant specifies its privacy preferences regarding textual and/or multimedia data. Finally, they submit the collected data and preferences to the requester.
3. In order to hide the identities and personal information of the participants' group from The IoMT context provider, each participant masks the list of items provided by responding users using anonymous indexes which are linked to the actual items indexes through a secret map Ω known only by them as in table 1. One important issue to standardize this secret map is to use hashing functions using a group generated key to mask the list of regions and users from the IoMT context provider.

TABLE 1. Secret map Ω used by the participants.

Anonymous Index or Hash value	Region Index	Item Name	Data1	Data2	Image1
A1	R1	…	…	…	..
A2	R2	…	…	…	…
A3	R3	…	…	…	…

   Due to that, the IoMT context provider will not be able to deal directly with items names but their hash values or anonymous index. Additionally, the users' data are also anonymized.

4. Each participant submits the collected data together with pseudonyms of devices' owners who participated in collection process to the IoMT context provider.
5. The IoMT context provider inserts pseudonyms into user database and their data into its database. The IoMT context provider updates its model using received data, then produces a list $\mathrm{A_i} = \{\mathrm{A_1}, \mathrm{A_2}, \ldots \mathrm{A_n}\}$ of anonymous indexes that users in the same cluster have chosen in the past.
6. The participants then submit their secret maps, so they able to update their data. They can unmask the list $\mathrm{A_i}$ using the shared secret map Ω to get final list.

IV. PROPOSED COGNITIVE CONCEALMENT ALGORITHMS

In the next sub-sections, we introduce our proposed cognitive algorithms used to preserve the privacy of the resulting datasets with minimum loss of accuracy. A closer look at the attack model proposed in [43] reveals that, if a set of behavioural and neighbouring environment data of certain citizen is fully distinguishable from the data of other citizens in the dataset with respect to some features. This citizen can be identified if an attacker correlates the revealed data with data from other publicly accessible databases. Therefore, it is highly desirable that the dataset has at least a minimum number of items should have a similar feature vector to every real item released by each participant. A real item in the released dataset can be described by a certain number of features in a feature vector, such as place of activity, type of activity, duration, time, date and so on. Both implicit and explicit ways can be used to extract this information and to construct these feature vectors and to maintain them. Additionally, the data sparsity problem associated with ETL services can be used to formulate some attacks as also shown in [43]. Before starting, we introduce a couple of relevant definitions.

Definition 1 (Dissimilarity Measure): This metric measures the amount of divergence between two items with respect to their feature vector. We use the notation $\mathcal{D}_m(I_u, I_n)$ to denote the dissimilarity measure between items I_u and I_n based on the feature vector of each item. $\mathcal{D}_m\,(I_u, I_n) < \delta \Rightarrow I_u \sim I_n$ [I_u is similar to I_n], δ is a user defined threshold value.

Definition 2 (Affinity Group): The set of items that are similar to item I_u with respect to *pth* attribute A_p of the feature vector and it is called affinity group of I_u and denoted by $C_{A_p}\,(I_u)$.

$$\begin{aligned} C_{A_p}\,(I_u) &= \left\{I_n \in D_n |\, (I_u \sim I_n) \wedge \left(A = A_p\right)\right\} \\ &= \{I_n \in D_n | \mathrm{D}_m\,(I_u, I_n) < \delta\} \end{aligned}$$

Definition 3 (K-Similar Item Group): Let D_ϖ be the real items dataset and $\widetilde{D_\varpi}$ its locally concealed version. We say $\widetilde{D_\varpi}$ satisfies the property of k-Similar item group (where K is defined value) provided for every item $I_u \in D_\varpi$. There is at least k-1 other distinct fake items $I_{n_1}, \ldots I_{n_{(k-1)}} \in D_n$ forming affinity group such that:

$$FV\left(I_{n_i}\right) \sim FV\,(I_u)\,, \quad \forall 1 \leq i \leq k-1$$

A. LOCAL CONCEALMENT USING CLUSTERING BASED OBFUSCATION (CBO) ALGORITHM

Our motivation to propose CBO is the limitation of the current anonymity models. The current anonymity models proposed in the literature failed to provide an overall anonymity as they don't consider matching items based on their feature' vectors. CBO uses the feature vectors of the current real items to select fake items highly similar to real items to create homogeneous concealed dataset. Using fake transactions to maintain privacy was presented in [3], [44], and [45], the authors considered adding fake transactions to anonymise the original data transactions. This approach has several advantages over other schemes including that any off-the-shelf data analytics algorithms can be used for analysing the concealed data and the ability to provide a high theoretical privacy guarantee. The locally concealed dataset obtained using CBO should be indistinguishable from the original dataset in order to preserve privacy. The core idea for CBO is to split the dataset into two subsets, the first subset is modified to satisfy K-Similar item group definition, and the other subset is concealed by substituting real items with fake items based probabilistic approach. CBO creates a concealed dataset D_P as following:

1. The sensitive items are suppressed from the dataset based on provider preferences thereafter we will have the suppressed dataset D as the real dataset.
2. Selecting a ϖ percent of highest frequent items in dataset D to form a new subset D_{ϖ}. This step aims to reduce the substituted fake items inside the concealed dataset D_P. Moreover, it maintains data quality by preserving the aggregates of highly frequent preferences.
3. CBO builds affinity groups for each real item $\forall I_u \in D_{\varpi}$ through adding fake items to form *K-Similar items group*. We implemented this task as a text categorization problem based on the feature vectors of real items. We also implemented a bag-of-words naive Bayesian text classifier [46] that extended to handle a vector of bags of words. The task continues until all items in D_{ϖ} are belonging to different affinity groups, then we get a new dataset $\widetilde{D_{\varpi}}$.
4. For each $I_u \in D_u = D - D_{\varpi}$, CBO selects a real item $\{I_u\}$ from real item set D_u with probability α or selects a fake item $\{I_n\}$ from the candidate fake item set D_n with probability $1 - \alpha$. The selected item I_P is added as a record to the concealed dataset D_P. This method achieves the desired privacy guarantee because the type of selected item and α are unknowns to external parties. The process continues until all real items in D_u are selected.
5. Finally, the concealed dataset D_P is merged with the subset $\widetilde{D_{\varpi}}$, which obtained from step 3.

1) ANALYSIS OF LOCAL CONCEALMENT USING CBO

In terms of performance, CBO requires supplementary storage costs and computations costs. The supplementary storage costs can be reduced by clustering items in the resulting dataset into C clusters and use the feature' vectors of top N items with high rates in each cluster for CBO algorithm. Thus supplementary storage costs will be in order of $O(CN)$. The computation costs for CBO are divided between computational complexities required to create affinity groups and adding fake items. Obviously, the computation overhead in creating affinity groups dominates, and it can be reduced by selecting lower values for ϖ.

B. GLOBAL CONCEALMENT USING RANDOM RATINGS GENERATION (RRG) ALGORITHM

After executing CBO, the synchronization agents build a virtualized schema with the aid of the coordinator agent at IoMT-enabled DMS then the global concealment agent starts executing the RRG algorithm. The coordinator agent will not be able to know the real items in the merged datasets as they already concealed locally using CBO algorithm. The main aim for the RRG is to alleviate data sparsity problem by filling the empty cells in such a way to improve the accuracy of the predictions at the environmental monitoring side and increase the attained privacy for providers. The RRG algorithm consists of following steps:

1. The global concealment agent finds the number of majority frequent items I_r and partially frequent items by all users $I - I_r$, where I denotes the total number of items in merged datasets.
2. The global concealment agent randomly selects an integer ρ between 0 and 100, and then chooses a uniform random number ξ over the range $[0, \rho]$.
3. The global concealment agent decides ξ percent of the partially frequent items in merged datasets and uses the KNN to predicate the values of the empty cells for that percentage.
4. The remaining empty cells are filled by random values chosen using a distribution reflecting the frequent items in the merged datasets.

1) ANALYSIS OF GLOBAL PERTURBATION USING RRG

The privacy of the merged datasets is maintained because all the processing is done on the datasets that previously processed using CBO. The global concealment agent improves the overall privacy and accuracy by increasing the density of the merged datasets due to the filled cells. With increasing ρ values, the RRG reduces the randomness in the frequencies. That might increase the accuracy of the predictions while decreases the privacy level. So, RRG should select ρ in a way to achieve the required balance between privacy and accuracy.

V. PROPOSED ENVIRONMENTAL DEEP LEARNING SERVICE

In this section, we proposed a new service for anomaly detection with markov based segmentation approach to detect possible regions of interest, then fed the extracted features within each region into a discriminative classifier to extract various patterns.

Figure 4 depicts the basic flowchart of our approach, which consists of four modules. Firstly, the noise present in the captured images is eliminated with the help of a noise removal and background subtraction processes. The second module executes Markov based approach to segment the possible regions of interest. The third module extracts the viable features within each region using a multiresolution wavelet transform. Finally, the last module, is a discriminative classifier that learns effective features from each region and distinguish it into anomaly or normal region. In the next sub-sections, we introduce the various steps involved in our proposed anomaly detection service. Each step utilizes an effective technique, which plays an important role in the system. The building blocks of the proposed systems are

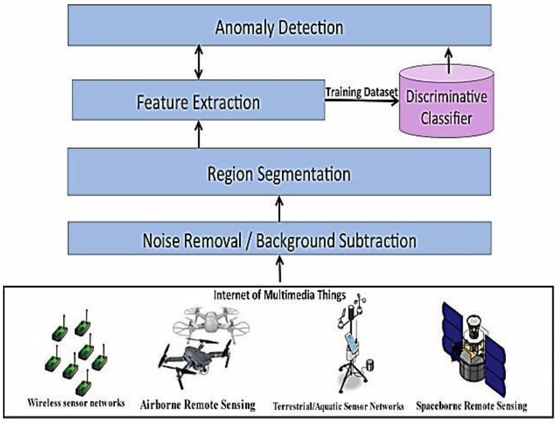
Anomaly Detection
Feature Extraction
Training Dataset
Discriminative Classifier
Region Segmentation
Noise Removal / Background Subtraction
Internet of Multimedia Things
Wireless sensor networks
Airborne Remote Sensing
Terrestrial/Aquatic Sensor Networks
Spaceborne Remote Sensing

FIGURE 4. **Building Blocks of the Anomaly Detection System.**

depicted in the figure (4). The proposed anomaly detection system consists of following steps:

Step 1 (Noise Removal and Background Subtraction Processes): The noise present in the captured images is eliminated with the help of anisotropic diffusion combined with non-local mean and Gaussian background process [47]. The method successfully analyses the images according to each and every pixel present without eliminating the important features such as line, interpretations, and edges. More over the anisotropic diffusion process can effectively analyse blurred images. This process is applied to images using following equation:

$$\frac{\partial I}{\partial t} = div\,(c(x, y, t)\nabla I) = \nabla c.\nabla I + c\,(x, y, t)\,\Delta I \tag{1}$$

Where $div\,(c(x, y, t)\nabla I)$ represents the divergence operator of the diffusion coefficient c(x,y,t) in relation with the image gradient operator ∇I. Based on (1) the anisotropic diffusion is applied to the image, if the pixel corrupted with the noise can be replaced using a non-local approach [43], where the similarity between pixels of the image is determined using the pixel intensity and is defined as follow:

$$v\,(i) = u\,(i) + n\,(i) \tag{2}$$

Where $v(i)$ is defined as the current value of pixel i in given image I, $u(i)$ is defined as the ''true'' value of pixel i and $n(i)$ is defined as the noise mixed with the value of pixel i. The noise exists in the image is analyzed according to the following assumption, that $n(i)$ is an independent value extracted from Gaussian distribution with a variance $\sigma 2$ and mean μ equal to 0. Based on that, the similarity between the neighboring pixels is defined depending on the weights $w\,(p, q_1)$ and $w\,(p, q_2)$. Then the non-local mean value of each pixel [43] is calculated as follows,

$$NL\,(V)\,(p) = \sum_{q\varepsilon V} w\,(p, q)\,V(q) \tag{3}$$

V is defined as the image with noise, and the weights $w(p, q)$ satisfy that $0 \leq w\,(p, q) \leq 1$ and $\sum_q w\,(p, q) = 1$ and is defined as follow:

$$w\,(p, q) = \frac{1}{Z(p)} e^{\frac{-max(d^2 - 2\sigma^2(p,q)}{h}} \tag{4}$$

σ is as defined as the standard deviation of the noise and $Z(p)$ is defined as a normalizing constant and is defined as follow:

$$Z\,(p) = \sum_q e \frac{-d(p, q)}{h} [1, 2] \tag{5}$$

h is defined as the weight-decay control parameter. After that the neighborhood similarity value is defined as using the weighted value of the pixel and is calculated as follow:

$$d\,(p, q) = \left\| V\left(N^p\right) - V(N^q) \right\|_{2,F}^2 [1, 2] \tag{6}$$

F is defined as the neighborhood filter employed on the neighborhood's squared difference R_{sim} and is defined as following:

$$F = \frac{1}{R_{sim}} \sum\nolimits_{i=m}^{R_{sim}} 1/(2 \neq i|1)^2 \tag{7}$$

m is the distance between the weight and the center of the neighborhood filter. F provides higher values if the pixels near the neighborhood center, and provide lower values if the pixels near the neighborhood edge. Finally, these values are used to generate the final image.

The background subtraction was performed [48] using the Gaussian model. The background model has been constructed using the selective average method for eliminating the unwanted background pixel information as follow:

$$BM_N\,(x, y) = \frac{\sum_{m=1}^{N} I_m(x, y)}{N} \tag{8}$$

Where $BM_N\,(x, y)$ is defined as the intensity of pixel (x, y) of the background model, $I_m(x, y)$ is defined as the intensity of pixel (x, y) of the m^{th} frame of the captured video, and N is defined as the number of video frames utilized to construct the background model. The background model is defined using a Gaussian mixture model as follow,

$$p\,(x\,|\,\lambda) = \sum_{i=1}^{M} w_i g(X|\mu_i, \Sigma_i) \quad \forall i = 1, \ldots, M \tag{9}$$

Where x is defined as continuous-valued data vector, w_i are defined as the mixture weights, and $g(X|\mu_i)$ are defined as the component of gaussian density functions [49], After that the probability value of each pixel is calculated,

$$\sum\nolimits_{i=1}^{k} w_i N(\mu_t \Sigma_t, Z) \tag{10}$$

N is the probability density function that has a mean vector μ and covariance Σ. w_i is defined as the weight of the i^{th} Gaussian. The new pixel value Z_t is compared to each Gaussian, if the Gaussian weight is matched

$\|Z-\mu_h\| < d\sigma_h$, then the Gaussian parameters are updated in accordance with:

$$\begin{aligned} w_{i,t} &= (1-a) * w_{i,t-1} + a * M_{i,t} \\ \mu_t &= (1-\rho) * \mu_{t-1} + \rho * Z_t \\ \sigma_t^2 &= (1-\rho) * \mu_{t-1} + \rho * (Z_t - \mu_t)^T * (Z_t - \mu_t) \\ \rho &= a * N(\mu_t) \end{aligned}$$

a is the learning rate for the Gaussian weight. Additionally, the unmatched pixel are eliminated using $w_{i,t} = (1 = a) * w_{i,t} - 1$. If none of the pixel matches the Gaussians weight, lowest weight pixel is replaced with Z_t. When the Gaussians values are stored in a corresponding index with a descending order, the initial values of this index will probably represent the background. After eliminating the background pixels and noise, the images are fed into the next step.

Step 2 (Regions Segmentation): The segmentation approach is done using markov random field [10] to ensure the effective extraction of meaningful regions. It uses the local image feature value, prior probability, and marginal distribution value of the image. At the start, Markov random neighboring value must be defined from the image in terms of both first and second order neighboring values. Then the initial probability value for each feature value is set as 0 or 1. After that the mean and variance value of each pixel value is computed and labelled in the image. From the computed values, marginal distribution value is calculated according to the Bayes theorem. Finally, the probability value must be calculated and the pixels with similar values are grouped into the particular cluster or region. This process is repeated until the prior probability value reaches to a maximum value other than the defined one. The extracted regions are fed into the next step.

Step 3 (Features Extraction): The multiresolution wavelet transform was employed for feature extraction. At first, the segmented regions are divided into sub-regions [50] in all the directions and then the key elements of the scale descriptors are selected. This step starts with applying a Gaussian filter on the image to detect the key elements. The maximum and minimum values of the edges are determined using the following equation $D(x, y, \sigma) = L(x, y, K_i\sigma) - L(x, y, K_j\sigma)$, where $D(x, y, \sigma)$ is the difference in the Gaussian image, $L(x, y, K\sigma)$ is the convolution value of the image $L(x, y, K\sigma) = G(x, y, k\sigma) * I(x, y)$, and $I(x, y)$ is the Gaussian blur value. Detecting key elements is accomplished using Taylor series, which is calculated as:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \tag{11}$$

From the detected key elements and their locations, each key elements is assigned magnitude $m(x, y)$ and orientation $\theta(x, y)$ in every direction as shown at the bottom of this page.

Based on the extracted key elements, different features can be calculated such as mean, standard deviation, entropy and variance. The extracted features are fed into the next step.

Step 4 (Anomaly Detection): In the last step, the extracted features are used to train support vector machine classifier to detect anomalies from the captured videos, the training stage reduces misclassification error and increases the recognition rate. Each feature in the training dataset is represented as $D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{-1, 1\}\}$. The output value of this stage is defined as {1,-1}, in which 1 is represented as the normal feature and -1 denoted as the anomaly feature. Then the feature belongs to the class is defined by applying the hyper plane which is calculated as $w \cdot x - b = 0$, where x is represented as the features exists in the training set, The normal hyper plane vector is w and hyper plane offset is b. The extreme learning neural networks [51] were utilized to reduce the maximum margin classification, which in turn improves the anomaly detection process. At the testing stage, the extracted features are matched with the training features to successfully detect the anomaly features. The accuracy of the proposed system was examined using the experimental results.

VI. EXPERIMENTAL RESULTS

The proposed algorithms are implemented in C++, we used message-passing interface (MPI) for a distributed memory implementation of RRG algorithm to mimic a distributed network of nodes. Since, there is no publicly available datasets for environmental hazards on the internet repositories. Therefore, we constructed our own datasets that utilizes the video footages of Forest fires dataset, which was provided by the National Protection and Rescue Directorate of Croatia, and other fire videos from an online social service such as YouTube. This dataset consists of 1020 fire and non-fire video clips. There are 130 forest fire video clips, 260 indoor fire video clips, 320 outdoor fire video clips and 310 non-fire video clips among the collected dataset. The resolutions of video clips were 480 × 360 pixels and each video clips consists of 200 ∼ 300 frames. Almost 1/4 of the video clips were used for testing while the remaining were used for training. The testing set contains 40 forest fire video clips, 60 indoor fire video clips, 80 out fire video clips and 80 non-fire video clips. For negative video clips, collection of videos contains some kind of flame, such as ambulance light, Flame Effect Light, and so on. Table 2 shows that the proposed techniques achieved the real-time performance for this resolution. The most time-consuming part was related to the calculation of pixel intensity. To check the effect of the resolution of videos on the processing time, another video sequence, which had 1280 × 720 pixels resolution, was tested. Tests have shown

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) &= atan2\,(L(x, y+1) - L(x, y-1)), (L(x+1, y) - L(x-1, y)) \end{aligned}$$

TABLE 2. Performance of proposed anomaly detection system.

Precision	Recall	True Negative Rate	Accuracy	F-Measure	Processing Time (MS)	
					480x360	1280x720
0.8998	0.89	0.82	0.9543	0.9128	25.1	55.2

that the complexity was increased by more than 2 times for 5.33 larger frame size. The Precision of this solution applied on these videos is about 94%.

In order to evaluate the effect of our proposed concealment algorithms on mashuped datasets used in problem solving. A dataset pulled from the SportyPal® network that was linked to another dataset containing behavioural and neighbouring environment data of 8000 students in the University of Zagreb in Croatia in the period of 2006 to 2008. For the purpose of this work, we intended to measure two aspects in this dataset, which are privacy breach levels and accuracy of results. We divide the dataset into a training set and testing set. The training set is concealed then used as a database for the monitoring service. To evaluate the accuracy of the generated predictions, we used the mean average error (MAE) metric proposed in [52]. To measure the privacy breach levels, we used mutual information as a measure for the notion of privacy breach of D_u through D_P.

In the first experiment, we want to measure the relation between the quantity of real items in the concealed dataset and privacy breach, we select α in a range from 1.0 to 5.5, and we increased the number of real items from 100 to 1000. We select fake items set using uniform distribution as a baseline. As shown in figure (5), our generated fake set reduces the privacy breach and performs much better than uniform fake set. As the number of real items increase the uniform fake set get worse as more information is leaked while our optimal fake set does not affect with that attitude.

In the second experiment, we measured the relation between the quantity of fake items in the subset D_ϖ and the accuracy of the classification results. We select a set of real items from our dataset, then we split it into two subsets D_ϖ and D_u. We concealed subset D_u with fixed value for α to obtain the subset D_p. We append the subset D_ϖ with either items from optimal fake set or uniform fake set. Thereafter, we gradually increased the percentage of real items in D_ϖ that are selected from our dataset from 0.1 to 0.9. Figure (6) shows MAE values as a function of the concealment rate for the whole concealed dataset D_p. The IoMT context-based service can select a concealment rate based on its privacy preferences. Hence, with a higher value for the concealment rate, higher accurate predictions can be attained by the monitoring service. Adding items from the optimal fake set have a minor impact on MAE of the results without having to select a higher value for the concealment rate.

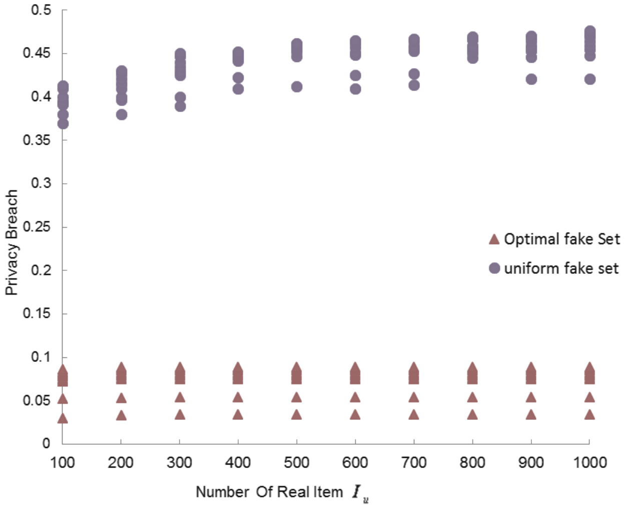

FIGURE 5. Privacy breach for optimal and uniform fake sets.

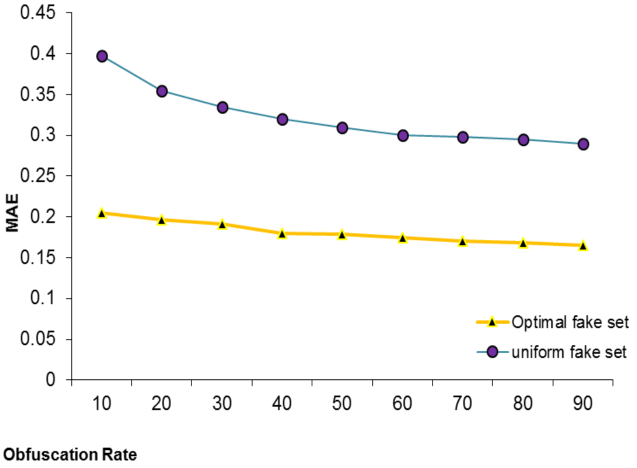

FIGURE 6. MAE of the generated predictions vs. concealment rate.

Finally, Due to different levels of privacy concerns between IoT context based providers, they might select various values for the ρ parameter that might affect the accuracy and privacy of the overall predictions. This probably influences on their revenues, since IoT-enabled CMS pays for the usage of their databases to achieve a certain prediction quality. To evaluate how various privacy levels, affect the accuracy of predictions, we performed two experiments using our dataset. We varied the value of ρ from 0 to 100 to show how the different values of ξ affect the accuracy and privacy of the results Note that when the value of ρ is 0, this means select all the partially frequent and infrequent items then fill the selected items with random values chosen using a distribution reflecting the data in the merged datasets. Once we set the value of ρ, we can randomly select the value of ξ over the range $[0, \rho]$, after calculating the values of MAE and privacy breach of

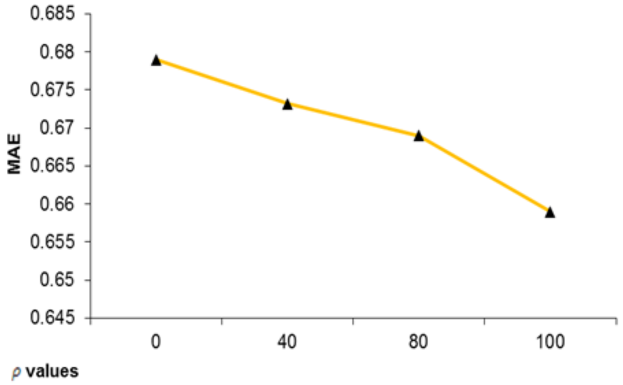

FIGURE 7. MAE of the generated predictions for different ρ values.

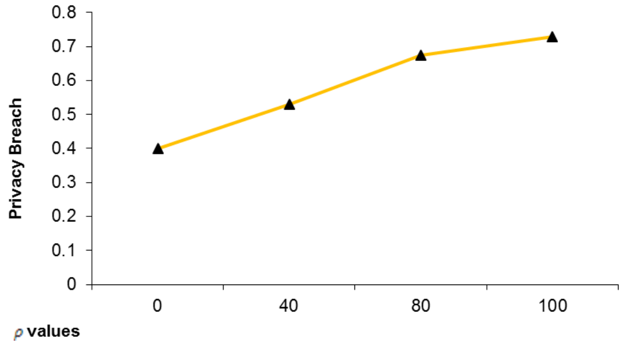

FIGURE 8. Privacy breach of the generated predictions for different ρ values.

the results, figures (7) and (8) depict the results. As seen from figure (5) accuracy becomes better with augmented ρ values, as the size of selected portion filled using *KNN* is increased and the size of randomized portion is decreased. Although, augmenting the values of ρ attains lower values for MAE. However, we still have a decent accuracy level for the predictions. Accuracy losses result from an error in the predictions such that the predicted items might not represent true frequent items in regard to those infrequent items. Also there is an error yield from using *KNN* predictions with different values for the K parameter. Using these errors; we guarantee in the merged datasets lower values for the privacy breach metric as shown in figure (7). This can contribute to overcoming some privacy breaches that might happen due to the mashup process of various datasets from independent IoT context based services [53]. We can conclude that accuracy losses due to privacy concerns are small and our proposed algorithms make it possible to offer accurate predictions.

VII. CONCLUSIONS

In this work, we presented our ongoing work on building a cognitive -based middleware for private data mashup (CMPM) to serve centralized IoT-enabled environmental monitoring service. We gave a brief overview over the mashup process and two concealment mechanisms. A Novel anomaly detection solution was also presented in detail, which achieves promising results in terms of performance. The experiments were conducted on a real dataset and it shows that the accuracy of our solution is more than 94%. However, the ability to detect environmental hazards such as fires and reduce false positives depends mainly on the image quality. Additionally, the experiments show our approach reduces privacy breaches and attains accurate results. We realized many challenges in building an IoMT-enabled data mashup service. As a result, we focused on environmental monitoring service scenario. This allows us to move forward in building an integrated system while studying issues such as a dynamic data release at a later stage and deferring certain issues such as virtualized schema and auditing to future research agenda.

REFERENCES

[1] T. Catarci *et al.*, "Pervasive software environments for supporting disaster responses," *IEEE Internet Comput.*, vol. 12, no. 1, pp. 26–37, Jan. 2008.

[2] J. San-Miguel-Ayanz *et al.*, "Comprehensive monitoring of wildfires in Europe: the European forest fire information system (EFFIS)," in *Approaches to Managing Disaster-Assessing Hazards, Emergencies and Disaster Impacts*. Rijeka, Croatia: InTech, 2012.

[3] T. Trojer, B. C. M. Fung, and P. C. K. Hung, "Service-oriented architecture for privacy-preserving data mashup," in *Proc. IEEE Int. Conf. Web Services*, 2009, pp. 767–774.

[4] R. D. Hof, "Mix, match, and mutate," Business Week, Jul. 2005.

[5] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, Mar. 2007, Art. no. 3.

[6] L. F. Cranor, "'I didn't buy it for myself': Privacy and ecommerce personalization," in *Proc. ACM Workshop Privacy Electron. Soc.*, Washington, DC, USA, 2003, pp. 57–73.

[7] *Cyber Dialogue Survey Data Reveals Lost Revenue for Retailers Due to Widespread Consumer Privacy Concerns*, Cyber Dialogue, New York, NY, USA, Nov. 2001.

[8] J. S. Olson, J. Grudin, and E. Horvitz, "A study of preferences for sharing and privacy," in *Proc. CHI Extended Abstracts Hum. Factors Comput. Syst.*, Portland, OR, USA, 2005, pp. 1985–1988.

[9] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song, "On the feasibility of side-channel attacks with brain-computer interfaces," in *Proc. 21st USENIX Conf. Secur. Symp.*, Bellevue, WA, USA, 2012, p. 34.

[10] D. Storm, "MEDJACK: Hackers hijacking medical devices to create backdoors in hospital networks," Computerworld, Jun. 2015.

[11] A. M. Elmisery and D. Botvich, "Enhanced middleware for collaborative privacy in IPTV recommender services," *J. Converg.*, vol. 2, no. 2, p. 10, 2011.

[12] A. M. Elmisery and D. Botvich, "Agent based middleware for private data mashup in IPTV recommender services," in *Proc. CAMAD*, 2011, pp. 107–111, doi: 10.1109/CAMAD.2011.5941096.

[13] A. M. Elmisery and D. Botvich, "Multi-agent based middleware for protecting privacy in IPTV content recommender services," *Multimedia Tools Appl.*, vol. 64, no. 2, pp. 249–275, 2012, doi: 10.1007/s11042-012-1067-3.

[14] A. M. Elmisery, "Private personalized social recommendations in an IPTV system," *New Rev. Hypermedia Multimedia*, vol. 20, no. 2, pp. 145–167, 2014, doi: 10.1080/13614568.2014.889222.

[15] A. M. Elmisery, S. Rho, and D. Botvich, "A fog based middleware for automated compliance with OECD privacy principles in Internet of healthcare things," *IEEE Access*, vol. 4, pp. 8418–8441, 2016, doi: 10.1109/ACCESS.2016.2631546.

[16] A. M. Elmisery, S. Rho, and D. Botvich, "A distributed collaborative platform for personal health profiles in patient-driven health social network," *Int. J. Distrib. Sensor Netw.*, vol. 2015, p. 11, Jan. 2015, doi: 10.1155/2015/406940.

[17] A. M. Elmisery, S. Rho, M. Sertovic, K. Boudaoud, and S. Seo, "Privacy aware group based recommender system in multimedia services," *Multimedia Tools Appl.*, vol. 76, no. 24, pp. 26103–26127, Dec. 2017, doi: 10.1007/s11042-017-4950-0.

[18] A. Esma, G. Brassard, J. M. Fernandez, F. S. M. Onana, and Z. Rakowski, ''Experimental demonstration of a hybrid privacy-preserving recommender system,'' in *Proc. ARES*, 2008, pp. 161–170.

[19] H. Polat and W. Du, ''Privacy-preserving collaborative filtering using randomized perturbation techniques,'' in *Proc. 3rd IEEE Int. Conf. Data Mining*, Nov. 2003, pp. 625–628.

[20] H. Polat and W. Du, ''SVD-based collaborative filtering with privacy,'' in *Proc. ACM Symp. Appl. Comput.*, Santa Fe, NM, USA, 2005, pp. 791–795.

[21] Z. Huang, W. Du, and B. Chen, ''Deriving private information from randomized data,'' in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Baltimore, MD, USA, 2005, pp. 37–48.

[22] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, ''On the privacy preserving properties of random data perturbation techniques,'' in *Proc. 3rd IEEE Int. Conf. Data Mining*, 2003, pp. 99–106.

[23] J. R. Martinez-de Dios, B. C. Arrue, A. Ollero, L. Merino, and F. Gómez-Rodríguez, ''Computer vision techniques for forest fire perception,'' *Image Vis. Comput.*, vol. 26, no. 4, pp. 550–562, 2008.

[24] J. R. Martínez-de Dios, L. Merino, F. Caballero, and A. Ollero, ''Automatic forest-fire measuring using ground stations and unmanned aerial systems,'' *Sensors*, vol. 11, no. 6, pp. 6328–6353, 2011.

[25] L. Rossi, T. Molinier, A. Pieri, M. Akhloufi, Y. Tison, and F. Bosseur, ''Measurement of the geometric characteristics of a fire front by stereovision techniques on field experiments,'' *Meas. Sci. Technol.*, vol. 22, no. 12, p. 125504, 2011.

[26] L. Rossi, T. Molinier, M. Akhloufi, Y. Tison, and A. Pieri, ''Estimating the surface and volume of laboratory-scale wildfire fuel using computer vision,'' *IET Image Process.*, vol. 6, no. 8, pp. 1031–1040, 2012.

[27] L. Rossi, T. Toulouse, M. Akhloufi, A. Pieri, and Y. Tison, ''Estimation of spreading fire geometrical characteristics using near infrared stereovision,'' *Proc. SPIE*, vol. 8650, p. 86500A, Mar. 2013.

[28] S. Verstockt *et al.*, ''FireCube: A multi-view localization framework for 3D fire analysis,'' *Fire Safety J.*, vol. 46, no. 5, pp. 262–275, 2011.

[29] S. Verstockt, ''Multi-modal video analysis for early fire detection,'' Ph.D. dissertation, Faculty Eng. Archit., Ghent University, Ghent, Belgium, 2011.

[30] W. P. Iii, M. Shah, and N. da Vitoria Lobo, ''Flame recognition in video,'' *Pattern Recognit. Lett.*, vol. 23, no. 1, pp. 319–327, 2002.

[31] R. Lucile, A. Moulay, and T. Yves, ''Dynamic fire 3D modeling using a real-time stereovision system,'' *J. Commun. Comput.*, vol. 6, no. 10, pp. 54–61, 2009.

[32] J.-F. Collumeau, H. Laurent, A. Hafiane, and K. Chetehouna, ''Fire scene segmentations for forest fire characterization: A comparative study,'' in *Proc. ICIP*, 2011, pp. 2973–2976.

[33] T. Toulouse, L. Rossi, T. Celik, and M. Akhloufi, ''Automatic fire pixel detection using image processing: A comparative analysis of rule-based and machine learning-based methods,'' *Signal, Image Video Process.*, vol. 10, no. 4, pp. 647–654, 2016.

[34] C. Yuan, Y. Zhang, and Z. Liu, ''A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques,'' *Can. J. Forest Res.*, vol. 45, no. 7, pp. 783–792, 2015.

[35] C. C. Wilson and J. B. Davis, ''Forest fire laboratory at Riverside and fire research in California: Past, present, and future,'' Pacific Southwest Res. Station, Berkeley, CA, USA, Tech. Rep. PSW-105, 1988.

[36] V. G. Ambrosia and T. Zajkowski, ''Selection of appropriate class UAS/sensors to support fire monitoring: Experiences in the United States,'' in *Handbook of Unmanned Aerial Vehicles*, K. Valavanis and G. Vachtsevanos, Eds. Dordrecht, The Netherlands: Springer, 2015.

[37] V. G. Ambrosia, ''Remotely piloted vehicles as fire imaging platforms: The future is here,'' *Wildfire Mag.*, vol. 11, no. 3, pp. 9–16, 2002.

[38] R. Charvat, R. Ozburn, S. Bushong, K. Cohen, and M. Kumar, ''SIERRA team flight of zephyr UAS at West Virginia Wild Land Fire Burn,'' *Infotech Aerosp.*, vol. 2012, p. 2544, Jun. 2012.

[39] A. Ollero *et al.*, ''Multiple eyes in the skies: Architecture and perception issues in the COMETS unmanned air vehicles project,'' *IEEE Robot. Autom. Mag.*, vol. 12, no. 2, pp. 46–57, Jun. 2005.

[40] J. R. Martínez-de-Dios, L. Merino, A. Ollero, L. M. Ribeiro, X. Viegas, ''Multi-UAV experiments: Application to forest fires,'' in *Multiple Heterogeneous Unmanned Aerial Vehicles* (Springer Tracts in Advanced Robotics), vol 37, A. Ollero and I. Maza, Eds. Berlin, Germany: Springer, 2007.

[41] L. Merino, F. Caballero, J. R. Martínez-de-Dios, I. Maza, and A. Ollero, ''An unmanned aircraft system for automatic forest fire monitoring and measurement,'' *J. Intell. Robot. Syst.*, vol. 65, no. 1, pp. 533–548, 2012.

[42] L. Merino, J. R. Martínez-de Dios, and A. Ollero, ''Cooperative unmanned aerial systems for fire detection, monitoring, and extinguishing,'' in *Handbook of Unmanned Aerial Vehicles*, K. Valavanis and G. Vachtsevanos, Eds. Dordrecht, The Netherlands: Springer, 2015.

[43] A. Narayanan and V. Shmatikov, ''Robust de-anonymization of large sparse datasets,'' in *Proc. IEEE Symp. Secur. Privacy*, May 2008, pp. 111–125.

[44] J.-L. Lin and J. Y.-C. Liu, ''Privacy preserving itemset mining through fake transactions,'' in *Proc. ACM Symp. Appl. Comput.*, Seoul, South Korea, 2007, pp. 375–379.

[45] J.-L. Lin and Y.-W. Cheng, ''Privacy preserving itemset mining through noisy items,'' *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5711–5717, 2009.

[46] D. D. Lewis, ''Naive (Bayes) at forty: The independence assumption in information retrieval,'' in *Proc. 10th Eur. Conf. Mach. Learn.*, 1998, pp. 4–15.

[47] D. Tschumperlé and L. Brun, ''Non-local image smoothing by applying anisotropic diffusion PDE's in the space of patches,'' in *Proc. ICIP*, 2009, pp. 2957–2960.

[48] M. Piccardi, ''Background subtraction techniques: A review,'' in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2004, pp. 3099–3104.

[49] A. Elgammal, D. Harwood, and L. Davis, ''Non-parametric model for background subtraction,'' in *Proc. Comput. Vis.-ECCV*, 2000, pp. 751–767.

[50] T. Ojala, M. Pietikäinen, and T. Mäenpää, ''Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[51] G. Feng, G.-B. Huang, Q. Lin, and R. Gay, ''Error minimized extreme learning machine with growth of hidden nodes and incremental learning,'' *IEEE Trans. Neural Netw.*, vol. 20, no. 8, pp. 1352–1357, Aug. 2009.

[52] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, ''Evaluating collaborative filtering recommender systems,'' *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, 2004.

[53] P. Golle, F. McSherry, and I. Mironov, ''Data collection with self-enforcing privacy,'' in *Proc. 13th ACM Conf. Comput. Commun. Secur.*, Alexandria, VA, USA, 2006, pp. 69–78.

AHMED M. ELMISERY received the B.S. degree in computer science from the Faculty of Computer Science, Mansoura University, Egypt, in 2001, the M.S. degree in computer science from the Arab Academy for Science, Technology and Maritime Transport, Egypt, in 2007, and the Ph.D. degree in computer science from the Waterford Institute of Technology, Ireland, in 2014. He was a Researcher in computer security at Telecommunications Software and Systems Group, Department of Computing, Mathematics and Physics, Waterford Institute of Technology, Ireland. He visited the Center for Advanced Technology in Telecommunications and Secure Systems, Monash University, Australia, from 2014 to 2015. He is currently working as an Assistant Professor with the Electronic Engineering Department, Federico Santa María Technical University, Chile. He is also a Research Fellow with the Internet of Things and People Research Center, Malmö University, Sweden, and an Adjunct Assistant Professor with the Computer Science Department, Technical College, Egypt. He has published over 36 research papers in national and international conferences. His research interests include security, cryptography, and machine learning. He is conducting research on privacy and security for future telecommunication services. His research has been grounded to develop privacy-enhanced algorithms for outsourced data in healthcare systems and recommender systems scenarios.

MIRELA SERTOVIC received the B.S. degree in education from the University of Tuzla, Bosnia and Herzegovina, in 2004. She is currently pursuing the Ph.D. degree with University in Zagreb, Croatia. Her research interests include social-humanistic informatics, social networks, blended learning, technology integration in teacher education, application of technology in language learning and teaching, and technology-assisted language learning.

BRIJ B. GUPTA (M'09) received the Ph.D. degree from IIT Roorkee, India. He was a Post-Doctoral Research Fellow in UNB, Canada. He is currently working as an Assistant Professor with the Department of Computer Engineering, National Institute of Technology Kurukshetra, India. He spent over six months with the University of Saskatchewan, Canada, to complete a portion of his research. He has visited several countries to present his research. His biography is selected to publish in the 30th Edition of prestigious Marquis *Who's Who in the World* (2012). He has published over 45 research papers in international journals and conferences of high repute. His research interest includes information security, cyber security, cloud computing, Web security, intrusion detection, computer networks, and phishing. He is member of ACM, SIGCOMM, The Society of Digital Information and Wireless Communications (SDIWC), Internet Society, and the Institute of Nanotechnology, and a Life Member of the International Association of Engineers and the International Association of Computer Science and Information Technology. He has also served as a technical program committee member of over 20 international conferences worldwide. In 2009, he was selected for Canadian Commonwealth Scholarship and awarded by the Government of Canada Award ($10 000). He is an Editor of various international journals and magazines.