Semantic Based Security for health data over Sensor Network

Muthukumar B[1], Sathiyamurthy K[2] , Jagan A [3]
[1] Associate Professor, Dept of CSE, Chennai Institute of Technology, Chennai, India,
anbmuthusbe@gmail.com
[2] Associate Professor, Dept of CSE, Pondicherry Engineering College, Puducherry India
[3] Associate Professor, Dept of CSE, Suriya Group of Institution, Tamil Nadu, India

Abstract: *An acquiescent juxtaposition of wireless technologies, micro-electromechanical systems (MEMS), micro-services and the internet paved an ecosystem named Internet of Things (IoT), which ascertains link between physical objects that are reachable through the internet. The embedded technology in those objects succors them to interact with internal states or the external milieu, which in turn influences on decisions. This new connectivity, bridges the gap between physical objects and digital world to improve the quality and productivity of life, has become common and going beyond laptops and smartphones, in applications like cars, smart homes, smart cities, healthcare, retails, energy management agriculture, wearables etc. IoT connects smart objects together (through internet and intelligent sensors) using internet protocol, and make them to be read, controlled, and managed at any time at anywhere. Since this communication is in the public environment, these devices are vulnerable to attacks, and hence the security and privacy are vitiated. Detection of abnormality in propositional information must be followed by recovery action to ensure the correct semantics of the frame network. This paper focuses on building a semantic based security platform to analyze the data received from sensors using Hidden Markov Model (HMM), semantic sensor network ontology, and temporal ontology to detect the malicious attack data. The HMM is used for reasoning purpose and the label for visible states are created. The Stream Annotation Ontology is used to represent the quality of the data over the Semantic Sensor Network Ontology.*

Keywords: Abnormal data, Hidden Markov Model, Internet of Things (IoT), k- means clustering algorithm,security breaches, semantic, Stream Annotation Ontology, Semantic Sensor Network (SSN), Quality Ontology.

1. Introduction

The internet of things (IoT) is the nascent technology which has the potential to change the environment into smart by incorporating various type of sensors and actuators like vehicle sensor, medical sensor, camera surveillance and other home appliances together. It has the savvy and vision to make machines smart enough to wane human labour to almost nil [1]. The sensors and actuators are configured and can be controlled remotely through the internet. The smart devices connected to internet by using network protocol which led to the advent of many applications such as smart city, home automation, smart grid, traffic management, smart parking, smart waste management, smart health etc. According to Gartner report, by 2020 connected devices across all technologies will reach to 20.6 billion [2]-[4].

IoT is changing the lane of medical environment by attaching tiny smart devices and sensors in the patient body. The construction of smart healthcare system needs to connect to the internet directly or indirectly always, which allows the physician to monitor the arrhythmia events and abnormal ECG signals for medical diagnosis and correct treatment [12]. The IoT objects has the wherewithal to change the state of environment like increasing the room temperature or the flow of fluid

to the patient in a hospital environment automatically. IoT is interconnection of highly heterogeneous entities and the communication patterns could be human to human, human to things, and also things to things. This results in colossal amount of data from sensors and actuators which have to be stored and processed. IoT data can describe our environment status of our health, home and cities which are often personnel. The communication between the sensors is established through the internet in public environment; this raises several security issues such as attack against the IoT devices, attack while communicating, and attacks in the master devices. Thus attention has to be paid on the security issues of IoT, due to the palpable fact that from a security perception, this IoT revolution represents a potential disaster.

Due to their limited assets and constricted interaction with the environment, sensor nodes can report corrupt readings due to environmental disturbance, accidental faults in the sensor hardware or software, and malicious activities, such as an adversary capturing and altering a number of sensor nodes. The corrupt reading may be caused by failure and the error may originate in degraded sensor devices which directly deals with the environment that are exposed to a variety of forces like physical or chemical activities. It is worthwhile to note that the data in the IoT environment are significant in making the decisions in the system.

The data from the sensors are automatically transferred over internet through the networking appliances; the sensors use its unique identifier to transfer the data to other entities. IoT products are mostly using password for authenticating each other. The usage of default/weak password often ease the embezzling of the intruders and the wake of this impediment, anomaly behaviours occur in sensor networks. Protection of data over the internet are achieved by encryption and hashing algorithms; whereas in the IoT entities due to insufficient storage and computational capabilities, sensor are not support by high level computational cryptographic algorithms. So the anomaly detection algorithms needs to be applied to detect the malicious activities.

2. Literature Survey

2.1 Security Issues In The Internet Of Things

The security issues of the internet of things are directly related to the wide application of its system. The architecture and features of IoT security are depicted in the work by Matsemela et.al[14].

The structure of internet of things [17] has been divided into three layers, perception layer, network layer and application layer. The perception layer includes main equipment's namely RFID (radio frequency identification), ZigBee and all kinds of sensors. After data collection the way of information transmission through IOT can be basically the wireless mode. The signals are transmitted in the public place. If the data trnasmitted lacks effective protection measures, the signals will be monitored, intercepted, and disturbed easily by the hackers. The most of the sensing devices are deployed in the automation monitoring sites. The attackers can easily gain access to the equipment, control or physically damage them [19]. Several kinds of attacks can perform in the perception layer [5] namely node capture attack (controlling the sensor node may leads to information leakage), fake node or malicious node attack (inserting new malicious node into the sensor network may leads to misuse of energy and resources of the sensor network) likewise the denial of service attack, timing attack, routing attack and replay attack can perform in the perception layer. Some security measures to overcome this attacks are cryptography technology scheme, key management, Security routing protocol, authentication and access control, physical security design, and intrusion detection technologies. The network layer [13] includes communication, the problems or attack for example if the large number of malicious nodes send data at the same time it will lead to DoS attack [8]. The problem arise in confidentiality and integrity of the data. The security measures to overcome this attack are by using Public Key Infrastructure (PKI). In application layer the security issues are in data protection, data access and software vulnerabilities, few counter measures to overcome this issues are authentication management and access control of data.

2.2 IEEE 802.15.4 (ZigBee)

ZigBee uses the IEEE 802.15.4 protocol as a base [15]. The IEEE 802.15.4 protocol is developed for low rate wireless private areas networks (LR-WPAN), due to low consumption, low cost, high throughput, low rate of this protocol it is used as communication protocols for IoT. This protocols also ensure security, reliability with both authentication and encryption process and can handle up to 65000 nodes. To transfer the data from one to another the IEEE 802.15.4 protocol works well in three frequencies using DSSS (direct-

sequence spread spectrum) method. With a DSSS method, capable of transferring data at 250 kbps at 2.4 GHz [15]. (Carrier Sense Multiple Access/Collision Avoidance) protocol to avoid collisions and to control the flow through delivering acknowledged frame, validating frames. Full Function Device (FFD), and Reduce Function Device (RFD) nodes are supported by this IEEE 802.15.4 standard. Creating the networks, controlling and maintain the network can be done by Full Function Device (FFD) nodes. Reduce Function Device (RFD), with low resources for communicate with coordinator in star topology.

2.3 Internet of Things Data Integrity

Physical objects in IoT networks such as devices, vehicles, buildings and machines lodged with electronic circuits, software, sensors and network connectivity are used efficiently in modern technology. Data can be collected, communicate and exchanged with internal states and external environment using these physical objects in Internet of Things network, while transferring and communicating the data with external environment these data are often violated by intelligent thieves and hence we have to ensure the data integrity using encryption and decryption algorithm such as RSA, AES, and Triple DES

RSA: It stands for Ron Rivest, Adi Shamir and Leonard Adleman who as created this algorithm. This algorithm uses public key cryptosystem Key size of RSA algorithm is 1024 to 3072 bits [14]. Two different keys are used by the sender and receiver. The keys are public key and private key in which public keys are known by everyone whereas private key are known by receivers. Data encrypted by public key can only be decrypted by a specific private key

Advanced Encryption Standard (AES): This AES is a symmetric cryptosystem. In Rijndael Cipher, AES is a subgroup and this subgroup can process a block cipher of 128 bits. AES algorithms allow cipher keys of 128 (AES-128), 192 (AES-192) and 256 bits (AES-256) [14]. Rounds are performed to process the encryption and decryption process, Number of round in AES are dependent on the key size.

Triple DES: Triple DES was created to overcome the issues in DES algorithm (Data Encryption Standard). Triple Data Encryption Algorithm is a symmetric key cryptosystem. Sender and receiver having the same key to perform the encryption and decryption process. In each and every blocks, three cipher blocks and three keys are provided to perform the rounds.

2.4 Data Representation

The ontology-based framework for Intelligent Data Analysis [6] of sensed data is based on a knowledge model composed by two existing ontologies such as Semantic Sensor Network ontology (SSN), SWRL Temporal Ontology (SWRLTO) .The data received from the sensor are real world data are converted into ontology for data representation. The ontology-based framework for Intelligent Data Analysis [6] of sensed data is based on a knowledge model composed by three ontologies [11] such as Semantic Sensor Network ontology (SSN), SWRL Temporal Ontology (SWRLTO) and Temporal Abstractions Ontology (TAO). In each observation, a sensor measures a system property and provides an estimated value, a time stamp and some contextual data such as a measurement quality estimation. These observation records are validated and stored in large repositories which usually implemented as time series databases.

Distributed sensor networks are vulnerable to accidental error and malicious activities in the data received from the sensor [7]. Distinguishing faults and attacks is essential to ensure the correct semantic of the network. The approach Hidden Markov Model (HMM) is used here to differentiate the attack or error data. HMM captures a hidden stochastic process that is inferred through a sequence of observations, which are stochastically related to the state of the hidden process. From the observations collected in each time window and a set of potential states of the environment generate: a sequence of the observable states of the sensed environment (derived using all the collected data regardless of their correctness), a sequence of the hidden states of the environment (i.e., actual, unknown states traversed by the environment), and a sequence of the erroneous states traversed by the observations that are (potentially) corrupted by accidental errors or malicious attacks[10]. The Stream Annotation Ontology is used to semantically represent features of the stream data, it is an important requirement in semantic stream data applications and knowledge based environment like smart applications. The SAO uses the definition of the Stream Event concepts in order to express the artificial classification of a time region, along with the particular stream data and its features. SAO is also used to extend the sensor observations described in SSN ontology through a concept, Stream data

The ontology-based framework for Intelligent Data Analysis [6] of sensed data is based on a

knowledge model composed by two existing ontologies such as Semantic Sensor Network ontology (SSN), SWRL Temporal Ontology (SWRLTO) .The data received from the sensor are real world data are converted into ontology for data representation. The ontology-based framework for Intelligent Data Analysis [6] of sensed data is based on a knowledge model composed by three ontologies [11] such as Semantic Sensor Network ontology (SSN), SWRL Temporal Ontology (SWRLTO) and Temporal Abstractions Ontology (TAO). In each observation, a sensor measures a system property and provides an estimated value, a time stamp and some contextual data such as a measurement quality estimation. These observation records are validated and stored in large repositories which usually implemented as time series databases.

Distributed sensor networks are vulnerable to accidental error and malicious activities in the data received from the sensor [7]. Distinguishing faults and attacks is essential to ensure the correct semantic of the network. The approach Hidden Markov Model (HMM) is used here to differentiate the attack or error data. HMM captures a hidden stochastic process that is inferred through a sequence of observations, which are stochastically related to the state of the hidden process. From the observations collected in each time window and a set of potential states of the environment generate: a sequence of the observable states of the sensed environment (derived using all the collected data regardless of their correctness), a sequence of the hidden states of the environment (i.e., actual, unknown states traversed by the environment), and a sequence of the erroneous states traversed by the observations that are (potentially) corrupted by accidental errors or malicious attacks[10].

Eduardo Xamena et al [23] has proposed a method of detecting and classifying web application attacks based on efficient semantic rules. Eduardo Xamena et al [24] have investigated the presence of various features that characterize small-world networks. Emna Amdouni et al [25] have proposed the state-of-the-art methods for a semantic representation of the imaging biomarker. Prodromal Kolyvakis et al [26] have derived a novel phrase retrofitting strategy for pre-trained word vectors. The Stream Annotation Ontology is used to semantically represent features of the stream data, it is an important requirement in semantic stream data applications and knowledge based environment like smart applications. The SAO uses the definition of the Stream Event concepts in order to express the artificial classification of a time region, along with the particular stream data and its features. SAO is also used to extend the sensor observations described in SSN ontology through a concept, Stream data.

Neelima p et all [27] have proposed hybrid model combines gravitational search algorithm (GSA) for scheduling the task in the application by using the benefits of both SLGSA algorithm and firefly algorithm This method is used to design the hybridization process and suitable fitness function of the corresponding task.

Qahtan M. Yas et all [28] have conducted a comprehensive survey using the keywords “skin cancer,” “apps,” and “Smartphone” or “m-Health” in different variations to find all the relevant articles in three major databases: Web of Science, Science Direct, and IEEE Xplore to develop and improve skin cancer apps in several ways since 2011.

Jing Hua, et all [29] have proposed a cardiac arrhythmia classification scheme that performs classification task directly in the compressed domain, skipping the reconstruction stage. It first employs the Pan–Tompkins algorithm to preprocess the ECG signals, including denoising and QRS detection, and then compresses the ECG signals by CS to obtain the compressive measurements. Amarjit Roy et all [30] all have proposed k-means clustering has been incorporated with fuzzy-support vector machine (FSVM) classifier for classification of noisy and non-noisy pixels in removal of impulse noise from gray images. S.P Raja et all [31] have discussed the various challenges, issues and applications confronting the Internet of Things.

2.5 Svm Based Anomaly Detection

Anomaly detection in the sensor network is an important challenge for tasks such as fault diagnosis and intrusion detection. A key problem is how to minimize the communication overhead in the network while performing in-network computation when detecting anomalies. The approach to this problem is based on a formulation that uses distributed, one-class quarter-sphere support vector machines [12] to identify anomalous measurements in the data. The detection algorithm [20] is proposed which finds a hyper sphere that captures the majority of the data vectors in the feature space at every sensor node. The data vectors that fall outside the hypersphere are identified as anomalies. The radius of the hypersphere is communicated among the nodes to compute a global radius. Each individual node uses

the global radius to classify its data vectors as globally normal or anomalous. This method incurs little communication overhead as it only communicates radius values. It is suitable for sensor networks deployed in a homogeneous environment, where the data distribution at each node is the same but unknown.

The choice of anomaly detection technique in a given [18] context depends on factors such as the computation and communication resources in the sensors, prior knowledge of the underlying data distribution, and the stationary of the environment. The challenges exists in the techniques used in sensor networks are providing higher accuracy in detecting anomaly detection in the heterogeneous sensor network. Moreover, sensor networks may consist of heterogeneous nodes with varying resources and capabilities [19]. An open issue is allocating and dynamically scheduling various anomaly detection processes among these heterogeneous nodes to achieve the accuracy and energy targets collectively. Finally, once anomalies are detected and the necessary actions can be taken to mitigate the damage caused by faults or attacks in the system [21].

Table 1- Analysis of evaluation metrics for the existing system (SVM)

Size of data	Recall	Precision	f-score
10k	0.8334	0.8695	0.8509
20k	0.825	0.7857	0.8048
30k	0.810	0.7916	0.8197
40k	0.806	0.7596	0.7710

Evaluation metrics for the existing system (SVM)

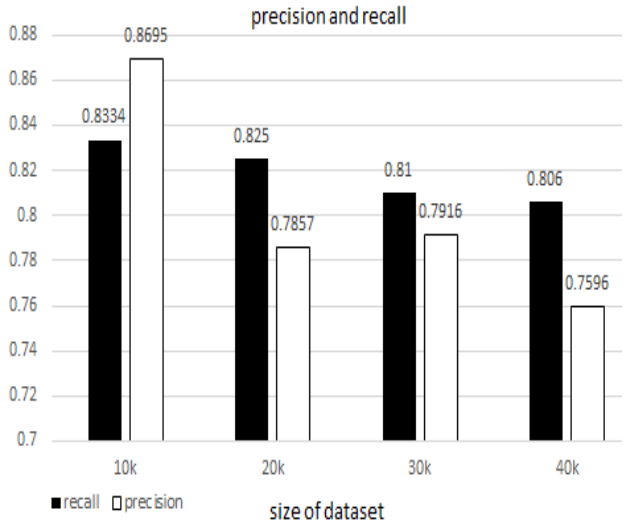

Figure 1 Evaluation of Precision and Recall

The precision and recall values are evaluated for the Mhealth dataset as shown in figure 1, the changes in the value occur based on the correct predictions and wrong predictions of anomaly data generated by SVM classifier. The reduction in value is due to homogeneous sensor value. These methods is used when the less computational time is required for anomaly detection, whereas the prior knowledge is required to analyse the data and these methods will work constrained with homogeneous sensors data and it is also trouble with the large dataset. The proposed methodology is used to overcome the difficulties with the heterogeneous data received from the sensor.

3 Proposed Architectural Diagram

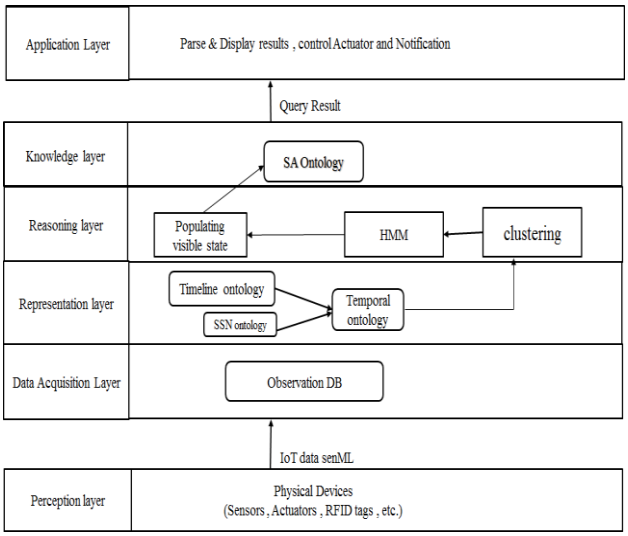

Figure 2 Proposed architecture

The figure 2 shows the proposed architecture of the system. The process starts with the sensing tasks

over the dynamic system. It is usually implemented by programmed sensor devices, but it can also be achieved manually. In each observation, a sensor measures a system stuff and provides an estimated value, a time stamp and some contextual data such as a measurement quality estimation. In the acquisition layer these observation records are validated and stored in large repositories. Data acquisition and validation tasks are highly oriented on domain and therefore they must be specifically designed for each application environment. In the representation layer an ontology of measurements, incorporated by semantically annotating the observation records. Since the data came from heterogeneous sources, the entities of the KB must be integrated in a consistence model expressed in a machine-interpretable language. The two main ontologies has to be populated in the representation layer. The ontologies are Semantic Sensor Network (SSN) ontology and temporal ontology. In the reasoning layer the hidden states are populated from the clustering algorithm using the k-means clustering. The main role of the Hidden Markov Model (HMM) is extracting the temporal patterns over the observations using HMM. Using, the concepts of ontology labelling from the observed states from the HMM can be constructed. Finally, by making use of these temporal relations and the labels, the higher level abstraction Stream Annotation Ontology (SAO) has been constructed. The temporal relations and the labels infers the quality of the data like whether the data is normal or abnormal this ensures the trustworthiness of the stream data. The SAO is used to represent the quality of the data stream over the SSN ontology. As a final point in the application layer the output can be a notification, display message or a control to actuator.

3.1. Modules in Proposed System

There are three modules in the system namely Data Assortment, Data Depiction and Abnormal Identification. The figure 3 shows the module description for the proposed system.

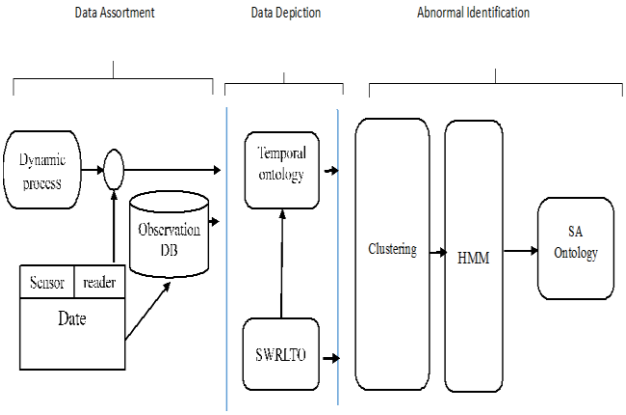

Figure 3 Module Description

3.1.1. Data Assortment

The torrent of data along with time stamp is collected from the sensors which are attached for observation. The temporal factors are needed to be considered since sensors are event based in nature. In each observation, a sensor measures a system property and provides an estimated value with their time cast and some contextual data such as a measurement quality estimation. To properly evaluate the stored measurements, the consistent sensors metadata must be attached with the observation. This information should include sensors precision, operation range, Eng. Units. In this work, the health care related raw sensor reading has to be processed in order to extract the semantic knowledge among the sensor data. The datasets are collected from the Mhealth environment [34]. Mhealth sensors are ECG leads, chest sensor, anklet sensor and arm sensor. The use of multiple sensors permits us to measure the motion experienced by various body parts namely, the acceleration, the rate of turn and the magnetic field orientation, thus better capturing the physique movements. The sensor positioned on the chest provides 2-lead ECG measurements. This information can be used for basic heart monitoring, checking for various arrhythmias or looking at the effects of exercise on the ECG. The real-time implementation requires Intel Galileo Gen 2. The sensors has to be connected to the Intel Galileo, Arduino or any other development board and their values has to be sent to the cloud for every time, t intervals

3.1.2. Data Depiction

The data received from the sensor network system is a stream of time-stamped values usually produced by one or more sensors attached to observe a significant process variable. Sensor data are event based in nature and the time-based and spatial

dimensions that need to be considered. There has been a rising interest in ontologies and other semantic technologies to expand the semantic between sensor networks. The basic idea under these techniques is annotating sensor data with temporal that increases interoperability and as well as provide contextual information.

The Semantic Sensor Network (SSN) ontology is used to represent the sensor data in machine-processable format. The representation includes sensor capabilities, sensor properties, observation values, primitives and measurement units. The primitive property is used to infer the qualitative Representation. SSN allows the network, its sensors and the resulting data to be ordered, managed, queried, understood and controlled through high-level specifications. The SSN ontology has implicated a large conceptualization effort to merge sensor-centric and observation-centric approaches. Observation is defined as ‘‘a Situation in which a Sensing method has been used to estimate or calculate a value of a Property of a Feature Of Interest’’.

The class ssn: Observation provides the structure to represent a single observation, hence it is related to a single measurement (class ssn: Sensor Output) and attributed to a single property (i.e. classes ssn: Property and ssn: Feature Of Interest) and to a particular ssn: Sensor. A proper temporal model must be a principled model that enforces a consistent representation of temporal information in the system. The SSN ontology which provides high-energy consuming services such as being alerted when a specific event occurs or asking for more detailed measurements.

SWRLTO is an open source OWL ontology that can be layered on existing ontologies without requiring them to be significantly rewritten. In fact, it has been successfully applied to several works on medicine. The temporal references are known as valid time. The SWRL temporal ontology consist of start time and end time for the valid episode of data received from the data. The figure 4.2 shows the arrangement of SWRLTO

SWRLTO is based on the valid-time temporal model. In this model, every temporal fact can be associated with an instant or an interval denoting the Fact’s Valid-Time. These temporal references are known as the Valid-Time as the fact is held to be true. No conclusions can be made about the fact for time regions outside of this.

The class swrlto:ValidTime has two subclasses: swrlto:ValidInstant and swrlto:ValidPeriod. A Valid Instant denotes a point on a time-line. A Valid Period models the time between two instants. These are specified by the swrlto:hasStartTime and swrlto:hasFinishTime date-time properties. The swrlto:Granularity class represent the unit of measure for temporal datum. In the ontology, this concept is modeled by the class tao:Episode.

The temporal extent is given by a swrlto:ValidTime while the qualitative context is given by a sao:Primitive here the primitive refers to the nature of the data such as normal or abnormal. Like ssn: Observation, tao:Episode is associated with the property of the tao:feature of interest that is abstracted by the episode.

3.1.3. Abnormal Identification

This module is used to identify the abnormal data in the data stream. This layer plays a key role in extracting the temporal patterns over the observations using Hidden Markov Model (HMM). The anomalies are identified using the observation states from Hidden Markov Model and finally the temporal relations and labels are used for constructing the semantic ontology. The hidden states for the markov model are the derived results from the clustering algorithm. The data are grouped into five different clusters with five different centroids.

Clustering Algorithm:

k- Means clustering algorithm is used to cluster the data into finite number of groups. The sensor data are given as an input to the clustering algorithm and the algorithm calculates the centroids for the data, based on the centroids the data are grouped into definite groups related to the number of centroids [36].

$$\mu_j = \frac{\sum_{i=1}^{m} 1\{c^i=j\}x^i}{\sum_{i=1}^{m} 1\{c^i=j\}}$$

Hidden Markov Model

The above algorithm is used to group the sensor data into five clusters namely very low, low, medium, high and very high. The Euclidean distance is used to calculate the centroids for the above clustering groups. The clustered groups are given as a hidden states to the Hidden Markov Model [33].

The definition of a HMM is as follows

$$H = \langle \omega, V, \{\pi_i\}, \{a_{ij}\}, \{b_{jk}\}, t \rangle \text{ where}$$

ω is the number of hidden states, and V is the number of visible state

$$\omega = (\omega_1, \omega_2, \dots\dots, \omega_n)$$

$$V = (v_1, v_2, \dots\dots, v_n)$$

We define π is the initial state which has to be set to the initial probability

$$\pi_i = \{\pi_1, \pi_2, \dots \pi_N\}$$

a_{ij} is a transition probability array, storing the probability of state j following state i, the transition probability is independent of time.

$$a_{ij} = P(\omega_i | \omega_i) \text{ or } P(\omega_i | \omega_{i+1})$$

b_{jk}, is a emission probability array, loading the probability of observation state or visible state for each hidden states and the visible state is dependent of time t.

$$b_{jk} = P(\omega_i | v_k)$$

Hidden = { sensor reading}, States ω visible
States V = { Normal, Borderline, Abnoral}
Algorithm 1:
Step: 1 Initialize t←0, a_{ij}, b_{jk}, V^T, $\alpha_j(0)$
Step: 2 For t ←t+1
Step: 3 $\alpha_t(i) = \max_j(\alpha_{t-1(j)}\, a_{ji}\, b_{ik})$
Step: 4 until t=T
Step: 5 return $P(V^T/\theta)$ ← $\alpha_0(T)$
Step: 6 T_p for final state
Step: 7 End

The probability of occurrence of abnormal or normal data are calculated by using the above algorithm formula. Here αt refers to possibility of moving to the next state, αt-1(j) denotes to pervious state, aji refers to the transition probability and bik signifies emission probability.

The figure 4 shows the trellis diagram for the attack identification. A trellis is a graph whose nodes are ordered into vertical slices (time), and with each node at each time connected to at least one node at an earlier and at least one node at a later time. Here the nodes refers to the attributes in the dataset values. With the pattern obtained from the above trellis diagram attack is sensed. The patterns are taken as a temporal feature for the abnormal identification in the data. For example the state moving from one state to another state with considerable probability is used to identify the abnormality in the data.

3.1.4. Stream Annotation Ontology (SAO)

As the Stream Annotation Ontology is used to semantically represent features of the stream data, it is an important requirement in semantic stream data applications and knowledge based environment like smart applications. The SAO uses the definition of the Stream Event concepts in order to express the artificial classification of a time region, along with the particular stream data and its features. SAO is also used to extend the sensor observations described in SSN ontology through a concept, Stream data.

The figure 5 represents the Quality Ontology in detail. It allows to annotate Stream Data of the SAO with quality. The quality ontology has five subcategories to describe the attributes of the annotated data stream regarding its quality. In

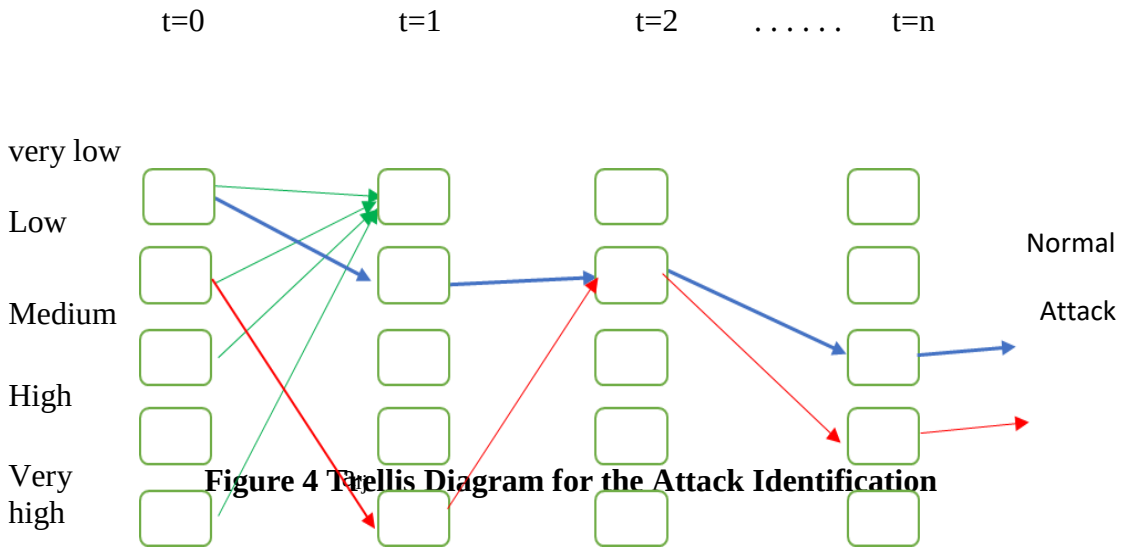

Figure 4 Trellis Diagram for the Attack Identification

addition it provides a concept of trust worthiness for data sources [35].

The results from the Hidden Markov Model is annotated as the quality of the data annotated in SSN ontology. The a tl: instant refers the each instant of the dataset, tl:at: Date and time shows the received time of the data from the sensor. Sao: value: sensor data which express the value of the data received. Sao: has unit of measurement used to denote the unit of the observation. The sample SA ontology with the quality of the data below

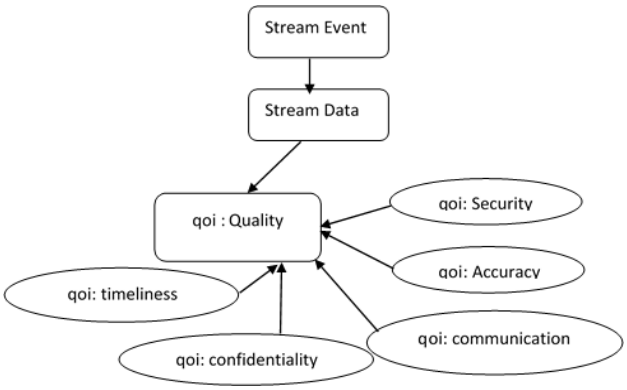

Figure 5 Quality Ontology

EXAMPLE: 1

@prefix- ssn: <http://purl.oclc.org/NET/ssnx/ssn #>.
@prefix prov: <http://purl.org/NET/provenance .owl#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix tl: <http://purl.org/NET/c4dm/timeline.owl#>.
@prefix to: <http://swrl.stanford.edu/ontologies/built-ins/3.3/temporal.owl#>.
@prefix xml: <http://www.w3.org/XML/1998/namespace>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix sao: <http://purl.oclc.org/NET/UNIS/sao/sao#>.

<http://www.archieve.ics.uni.edu/Mhealth/datasets#Chest>a sao:StreamEvent ;
sao:time [a tl:Interval ;
tl:beginsAtDateTime "2014-08-01T00:00:00"^^xsd:duration ;
tl:endsAtDateTime "2014-08-01T09:00:00"^^xsd:duration
] ;

<http://www.archieve.ics.uni.edu/Mhealth/dataset#Chest123 a sao:Point, ssn:Observation;
sao:time [a tl:Instant ;
tl:at "22014-08-04T10:50:00"^^xsd:dateTim;];
sao:hasUnitOfMeasurement <http://unit1:mV>;
sao:value "-0.85178"^^xsd:double ;
sao:hasQuality "Normal" .
<http://www.archieve.ics.uni.edu/Mhealth/dataset#Chest1112> a sao:Point , ssn:Observation ;
sao:time [a tl:Instant ;
tl:at "22014-09-01T04:20:00"^^xsd:dateTime ;
] ; sao:hasUnitOfMeasurement <http://unit1:mV>;
sao:value "-11.58"^^xsd:double ;
sao:hasQuality "Abnormal" .

The data stream from the sensors are analysed and the quality of the data is reported along with the values and time instant. Further the SA ontology is used to identify the cause of the abnormal data. The abnormal data can be caused by the malicious behaviour or damaged sensors.

4 Performance metrics

A performance metrics is a measure of an organization activities and performance. In projects, performance metrics are used to assess the efficiency of the project which is based on various criteria.

Accuracy

The accuracy (AC) is the proportion of the total number of the correct predictions to the actual data set size. It is determined using the equation [19]:

$$AC = \frac{TP+TN}{(TP+TN+FP+FN)}$$

Here the TP refers to True Positive, TN refers to True Negative, FP refers to Four Positive, and FN refers to False Negative. The four instances TP, TN, FP and FN are counted due to the relation between the predicted and actual classes [19].

The recall calculated by using the equation

$$R= \frac{TP}{(TP+FN)}$$

The precision is calculated by using the equation

$$P= \frac{TP}{(TP+FP)}$$

The F-score is calculated by using the equation

$$\text{F- score}=\frac{2*P*R}{(P+R)}$$

Table 2. Evaluation Metrics

Size of dataset	Existing System		SVM	Proposed System		HMM
	Precision	Recall	F – Score	Precision	Recall	F-score
10k	0.8695	0.8334	0.8509	0.615	0.6843	0.6478
20k	0.7857	0.825	0.8048	0.829	0.8012	0.8146
30k	0.7916	0.810	0.8197	0.866	0.8468	0.8559
40k	0.7596	0.806	0.7710	0.893	0.8755	0.8837
50k	0.7489	0.756	0.7456	0.899	0.8963	0.8976
60k	0.7384	0.742	0.7365	0.954	0.9413	0.9476

The above table shows the precision, recall and F- score of the proposed system and the existing system for the different data size from the sensor network with heterogeneous data.

The figure 6 and figure 7 shows the performance evaluation of both existing and proposed work. The graph 6 shows the degradation in performance when the size of the heterogeneous dataset increased as input to the existing algorithm Support Vector Machine, whereas in the proposed system the Hidden markov model is used to detect the anomaly in the data which leads to efficient prediction of anomalies in the sensor data as show in graph 7.

In the existing system the SVM used quadratic sphere to analyse the data to identify the anomalies in the data but if the data size increased with heterogeneity the SVM performance degraded. The proposed system uses Hidden Markov Model which is independent of temporal patterns and previous state. The dataset used in the proposed system is health data so the patient condition is totally not depend on the present so the Hidden Markov Model is used to provide efficient outcome.

The figure 8 shows the compared graph for the proposed and existing system

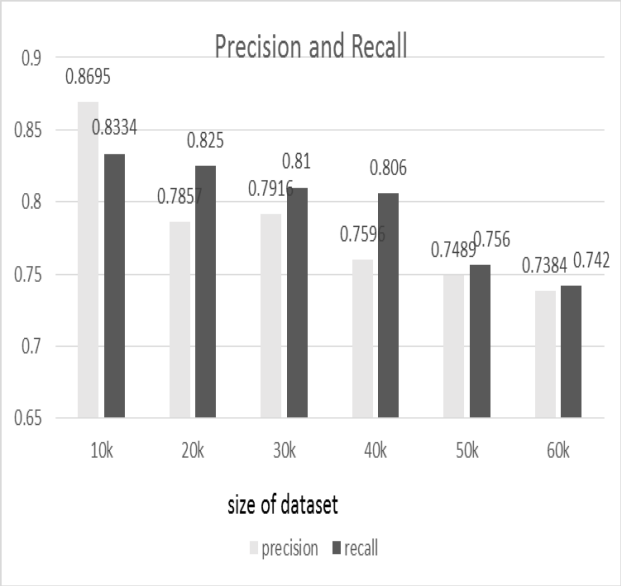

Figure 6 Precision and Recall for the Existing System

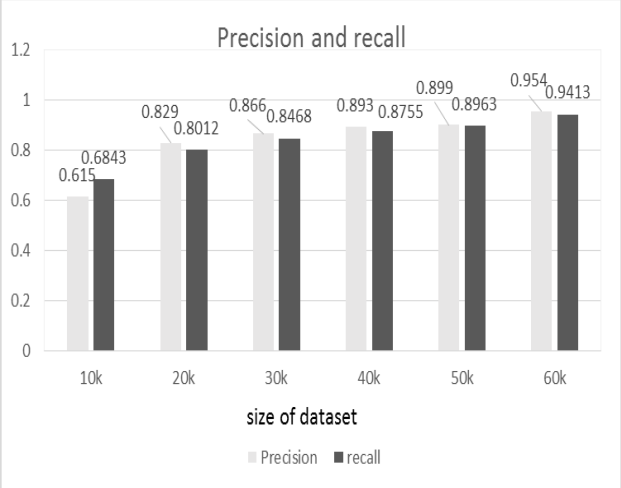

Figure 7: Precision and recall of the proposed system

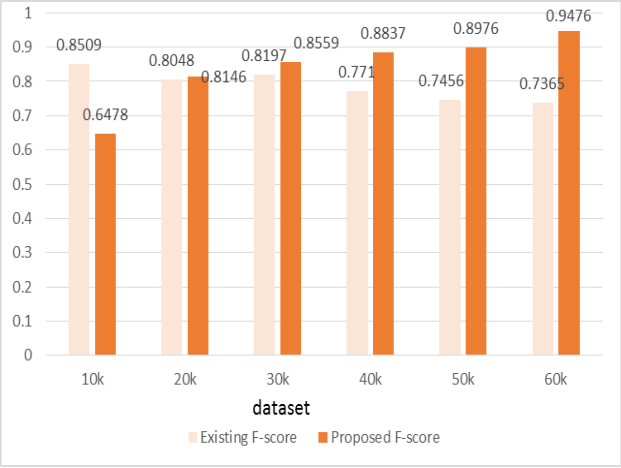

Figure 8 Compared Graph of proposed and existing system

Table 3 Accuracy for proposed and existing system

Size of dataset	Existing system accuracy	Proposed System accuracy
10k	0.8055	0.6905
20k	0.7681	0.7530
30k	0.7681	0.8428
40k	0.7476	0.9181
50k	0.7328	0.9437
60k	0.7107	0.9713

The Table 3 shows the accuracy of the proposed system and existing system for the different data size.

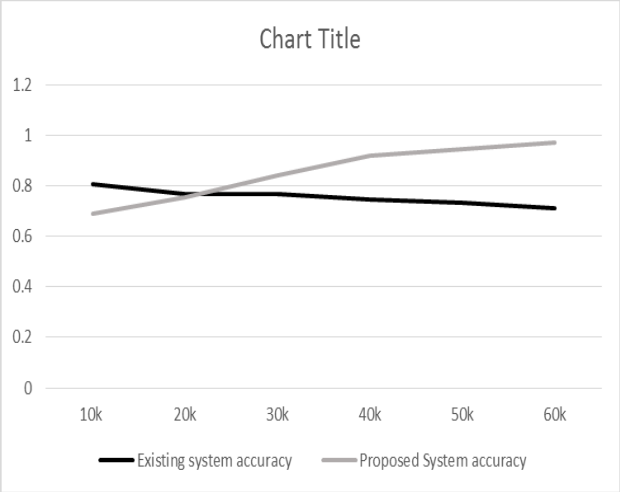

Figure 9 Comparison of accuracy

5 Conclusions

The explosive growth of Internet of Things which is vulnerable to security threats and attacks because of the communication between the sensors and base node are automated communication in public environment. In order to secure the IoT application the proposed work lay on to a semantic based platform. The primary objective of the system is to analyse the obtained sensor value has been opposed to any malicious attack. The data obtained from sensors used in IoT are possible to malicious activity and techniques for detecting anomaly activities are not consistent based on events. To overcome these issues a semantic based platform for analyzing the data received from the sensor is build and hidden markov model is used to address the temporal features. The proposed work was implemented and the data which has been obtained from sensor are converted to CSV format and the values are annotated to SSN and SWRL temporal ontologies. The SSN ontology provides the primitive property of the SSN ontology which helps to ease analyze of sensor data to identify the malicious behavior in the data with more detailed measurements. SWRLTO is based on the valid-time temporal model which helps to divide the data into episodes. The existing machine learning algorithm SVM are used for detecting anomaly behavior in data are implemented and limitations are recognized. The proposed Hidden Markov Model is implemented and the anomalies in data are identified. The Stream Annotation Ontology is annotated to represent the quality of the data over SSN ontology. The accuracy of the system is measured by the precision, recall and the f-score. This value helps to demonstrate the proposed system was attained the better performance compared to the SVM classifier.

7. FUTURE WORK

The future enhancement of this work includes identification of an attack type and distinguish between the error data and attack data by Stream Annotation Ontology and if the attack has been identified then the type of the attack should be determined. The attacks are labelled by using the security ontology, to help non security expert's software designers to be aware of security issues by notification.

References

[1] Giaffreda, Raffaele, Dagmar Cagáňová, Yong Li, Roberto Riggio, and Agnès Voisard, eds. *Internet of Things. IoT Infrastructures: First International Summit, IoT360 2014, Rome, Italy, October 27-28, 2014, Revised Selected Papers*. Vol. 151. Springer, 2015.

[2] Greengard, Samuel. *The internet of things*. Essential Knowledge Series, MIT Press, 2015.

[3] Coetzee, Louis, and Johan Eksteen. "The Internet of Things-promise for the future? An introduction.", IEEE In *IST-Africa Conference Proceedings, 2011*, pp. 1-9. 2011.

[4] De Boeck, Jo. "IoT: The impact of things." IEEE In *VLSI Technology (VLSI Technology), 2015 Symposium on*, pp. T82-T83. 2015.

[5] Gyrard, Amelie, Christian Bonnet, and Karima Boudaoud. "The stac (security toolbox: attacks & countermeasures) ontology." In *Proceedings of the 22nd International Conference on World Wide Web*, pp. 165-166. ACM, 2013.

[6] Gyrard, Amelie, Christian Bonnet, and Karima Boudaoud. "An ontology-based approach for helping to secure the ETSI machine-to-machine architecture." In *Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom)*, pp. 109-116. IEEE, 2014.

[7] da Silva, Ana Paula R., Marcelo HT Martins, Bruno PS Rocha, Antonio AF Loureiro, Linnyer B. Ruiz, and Hao Chi Wong. "Decentralized intrusion detection in wireless sensor networks." In *Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks*, pp. 16-23. ACM, 2005.

[8] Jun, Chen, and Chen Chi. "Design of complex event-processing ids in internet of things." In *Measuring Technology and Mechatronics Automation (ICMTMA), 2014 Sixth International Conference on*, pp. 226-229. IEEE, 2014.

[9] Eik Loo, Chong, Mun Yong Ng, Christopher Leckie, and Marimuthu Palaniswami. "Intrusion detection for routing attacks in sensor networks." *International Journal of Distributed Sensor Networks* 2, no. 4 (2006): 313-332.

[10] Basile, Claudio, Meeta Gupta, Zbigniew Kalbarczyk, and Ravi K. Iyer. "An approach for detecting and distinguishing errors versus attacks in sensor networks." In *Dependable Systems and Networks, 2006. DSN 2006. International Conference on*, pp. 473-484. IEEE, 2006.

[11] Roda, Fernando, and Estanislao Musulin. "An ontology-based framework to support intelligent data analysis of sensor measurements." *Expert Systems with Applications* 41, no. 17 (2014): 7914-7926.

[12] Abuaitah, Giovani Rimon, and Bin Wang. "Data-centric anomalies in sensor network deployments: Analysis and detection." In *Mobile Adhoc and Sensor Systems (MASS), 2012 IEEE 9th International Conference on*, pp. 1-6. IEEE, 2012.

[13] Shi, Elaine, and Adrian Perrig. "Designing secure sensor networks." *IEEE Wireless Communications* 11, no. 6 -2004.

[14] Matsemela, Gift, Suvendi Rimer, Khmaies Ouahada, Richard Ndjiongue, and Zinhle Mngomezulu. "Internet of things data integrity." In *IST-Africa Week Conference (IST-Africa), 2017*, pp. 1-9. IEEE, 2017.

[15] Glória, André, Francisco Cercas, and Nuno Souto. "Comparison of communication protocols for low cost Internet of Things devices." In *Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), 2017 South Eastern European*, pp. 1-6. IEEE, 2017.

[16] Zhao, Kai, and Lina Ge. "A survey on the internet of things security." In *Computational Intelligence and Security (CIS), 2013 9th International Conference on*, pp. 663-667. IEEE, 2013.

[17] Sutharshan rajasegarar, Christopher leckie, and Marimuthu Palaniswami," Anomaly Detection in Wireless Sensor Networks", IEEE Wireless Communications, 2008.

[18] Y.Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", J. comput. Syst . sci, vol.5, no,2009.

[19] S. Rajasegarar, Leckie C, Palaniswami M, Bezdek J.C "Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks", in IEEE international conference 07, 2007.

[20] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kaloguraki, D. Gunapulos, "Online Outlier Detection in Sensor Data Using Nonparametric Models," Proceedings of the 32nd international conference on Very large data bases, 2006.

[21] M. E. Elhamahmy, Hesham N. Elmahdy and Imane A.Saroit, "A New Approach for Evaluating Intrusion Detection System", in CiiT International Journal of Artificial Intelligent System and Machine Learning, 2010.

[22] Orestic Banos, Rafael Garcia, Juan A. Holgado – Terriza, Migiual Damas, Hector Promares, Ignacia Rojas, Alejandro Saez, Claudia Villalonga, "mHealthDroid: a novel framework for agile development of mobile health applications", in 6th International Work-Conference, 2014.

[23] Abdul Razzaq, Khalid LatifH, ,FarooqAhmad, AliHur, ZahidAnwar, Peter CharlesBloodsworth, "Semantic security against web application attacks" Information Sciences ,Vol 254, No 1, pp 19- 38, January 2014.

[24] EduardoXamena, Nélida BeatrizBrignole,Ana GabrielaMaguitman," A Structural Analysis of topic ontologies 'Information Sciences ,Vol 421, Pp 15- 29 , December 2017

[25] Emna Amdouni, Bernard Gibaud ," Imaging Biomarker Ontology (IBO): A Biomedical Ontology to Annotate and Share Imaging Biomarker Data " Journal on Data Semantics ,Vol 7, No 4, pp 223–236 , Dec 2018

[26] Prodromal Kolyvakis ,Alexandros Kalousis, Barry Smith, Dimitris Kiritsis ," Biomedical ontology alignment: an approach based on representation learning " Journal of Biomedical Semantics , Vol 9, No 21, PP 1-20, December 2018

[27] Neelima .P and Rama Mohan Reddy A,(2018)" An Efficient Hybridization Algorithm Based Task Scheduling in Cloud

Environment “ Journal of circuit , system and Computers, Vol .27, No .02,Pp. 1850018

[28] Qahtan M. Yas , A. A. Zaidan, B.B Zaidan, M. Hashim and C.K Lim ,” A Systematic Review on Smartphone Skin Cancer Apps: Coherent Taxonomy, Motivations, Open Challenges and Recommendations, and New Research Direction”, Journal of Circuits, Systems, and Computers Vol. 27, No. 5 (2017) 1830003

[29] Jing Hua, Hua Zhang, Jizhong Liu, Yiju Xu and Fumin Guo , “ Direct Arrhythmia Classification from Compressive ECG signals in wearable Health Monitoring System” Journal of Circuits, Systems, and Computers Vol. 27, No. 06(2018).

[30] Amarjit Roy , Joyeeta singha and Rabul Hussain Laskar , “ Removal of Impulse Noise from Gray Images Using Fuzzy SVM Based Histogram Fuzzy Filter” Journal of Circuits, Systems, and Computers 27(9) · October 2017 .

[31] S.P Raja, T. Dhiliphan Rajkumar and vivek pandiya Raj ,” Internet of things: Challenges,Issues and Applications “Journal of Circuits, Systems, and Computers 27(12):1830007 · February 2018

[32] Hidden Markov Model, http\\http://www.comp.leeds.ac.uk/roger/HiddenMarkovModels

[33] MHealth Dataset, http://archive.ics.uci.edu/ml/datasets/MHEALTH+Dataset

[34] Stream Annotation Ontology, https://mobcom.ecs.hs-osnabrueck.de/cp_quality/

[35] K-means algorithm, http://home.deib.polimi.it/matteucc/clustering/tutorial

[36] k-means http://stanford.edu/~cpiech/cs221/handouts/kmeans.html