RESEARCH ARTICLE OPEN ACCESS

An Effective Technique for PSO Based Clustering and Polynomial Regression in Wireless Sensor Network

M.Arthi [1], M.Jayashree [2]
Assistant Professor [1], Research Scholar [2]
Department of Computer Scienc
,Tiruppur Kumaran College for Women
Tirupur,Tamilnadu
India

ABSTRACT

Wireless sensor networks (WSNs) consist of sensor nodes. These networks have huge application in habitat monitoring, disaster management, security and military, etc. Wireless sensor nodes are very small in size and have limited processing capability very low battery power. This restriction of low battery power makes the sensor network prone to failure. Data aggregation is very crucial technique in wireless sensor networks. With the help of data aggregation we reduce the energy consumption by eliminating redundancy. In this paper we discuss about data aggregation and its various energy-efficient technique used for data aggregation in WSN. Data aggregation is a process of aggregating the sensor data using aggregation approaches. Particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. The cluster head choice is predicated on the space, residual energy, position, velocity parameters. The optimized cluster head selected by exploitation PSO algorithm. Particle swarm optimization (PSO) could be a population-based random search method, shapely once the social behavior of a bird flock. The algorithmic rule maintains a population of particles, wh ere every particle represents a possible answer to an optimisation drawback.

Keywords:- Wireless sensor networks, PSOC (Particle swarm optimization clustering), Data aggregation, PRDA(Polynomial regression based secure aggregation) and Security.

I. INTRODUCTION

The data aggregation is a technique used to solve the implosion and overlap problems in data centric routing. Data coming from multiple sensor nodes are aggregated as if they are about the same attribute of the phenomenon when they reach the same routing node on the way back to the sink. Data aggregation is a widely used technique in wireless sensor networks. The security issues, data confidentiality and integrity, in data aggregation become vital when the sensor network is deployed in a hostile environment. Data aggregation is a process of aggregating the sensor data using aggregation approaches. The general data aggregation algorithm works that data aggregation is the process of aggregating the sensor data using aggregation approaches. Then the algorithm uses the sensor data from the sensor nodes and then aggregates the data by using some aggregation algorithms such as centralized approach, LEACH(Low Energy Adaptive Clustering Hierarchy), TAG(Tiny Aggregation) etc. This aggregated data is transfer to the sink node by selecting the efficient path. This paper proposes a Polynomial Regression based secure Data aggregation protocol, called PRDA, to preserve the privacy of the data being aggregated. PRDA is an additive data aggregation protocol and achieves data privacy by employing polynomial regression on sensor data series. The novel idea behind PRDA protocol is to perform data aggregation using polynomial coefficients that represent sensor data.

Security Issues in Data Aggregation Data aggregation in Wireless sensor Network refers to exploit the sensed data from the sensors to the gateway node. Data aggregation plays a significant role in Wireless sensor Networks since the aggregation schemes followed here involve in reducing the amount of power consumed throughout data transmission between the sensor nodes. Within the data aggregation of WSN, security requirements, confidentiality and integrity, ought to be consummated. Specifically, the fundamental security issue is data confidentiality, that protects the sensitive transmitted data from passive attacks, such as eavesdropping. Data confidentiality is especially very important in a hostile environment, where the wireless channel is at risk of eavesdropping. Though there are many methods provided by cryptography, the difficult encryption and decryption operations, like modular multiplications of large numbers in public key primarily based cryptosystems, will assign the sensor's power quickly. The other security issue is data integrity, that prevents the compromised source nodes or aggregator nodes from considerably altering the final aggregation value. sensor nodes are easy to be compromised because they lack expensive tampering-resistant hardware, and even that tampering-resistant hardware may not continually be reliable. A compromised node will modify, forge or discard messages. The data's should be transmitted from member node to cluster head and from cluster head to either cluster head or base station inside a given time. If a time exceeds or any modifications wiped out the information then the certificate authority checks the threshold value of that node. If the threshold value is in vary then the node it trustworthy node and data aggregation is finished through this node. If the threshold value is in out of vary then the node is marked as malicious node. once marking the malicious node the information is not transferred at the actual node. so the information is transmitted solely the trustworthy node and it is collective additional securely and with efficiency. Provides safer for all the nodes due to exploitation the certificate authority. It will increase the packet delivery ratio and additionally improves the performance of non-stochastic elements errors like node fault etc.

II. RELATED WORKS

Ozdemir S. and Xiao Y. (2009), ' Mentioned the Secure data aggregation in wireless sensor networks: The algorithm uses the sensor data from the sensor node and then aggregates the data by using some aggregation algorithms such as centralized approach.

Centralized Approach: This is an address centric approach where each node sends data to a central node via the shortest possible route using a multihop wireless protocol. The sensor nodes simply send the data packets to a leader, which is the powerful node. The leader aggregates the data which can be queried. Each intermediate node has to send the data packets addressed to leader from the child nodes. So a large number of messages have to be transmitted for a query in the best case equal to the sum of external path lengths for each node. In-Network Aggregation.

In-network aggregation is the global process of gathering and routing information through a multi-hop network, processing data at intermediate nodes with the objective of reducing resource consumption (in particular energy), thereby increasing network lifetime. There are two approaches for in-network aggregation: with size reduction and without size reduction. In-network aggregation with size reduction refers to the process of combining & compressing the data packets received by a node from its neighbours in order to reduce the packet length to be transmitted or forwarded towards sink. In-network aggregation without size reduction refers to the process merging data packets received from different neighbours in to a single data packet but without processing the value of data.

Tree-Based Approach: In the tree-based approach perform aggregation by constructing an aggregation tree, which could be a minimum spanning tree, rooted at sink and source nodes are considered as leaves. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and therein the aggregation done by parent nodes.

Cluster-Based Approach[6]: In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to sink.

Data aggregation is the process of collecting and aggregating the useful data. Data aggregation is considered as one of the fundamental processing procedures for saving the energy. In WSN data aggregation is an effective way to save the limited resources. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced.. Wireless sensor networks have limited computational power and limited memory and battery power, this leads to increased complexity for application developers and often results in applications that are closely coupled with network protocols. In this paper, a data aggregation framework on wireless sensor networks is presented and a survey on various energy-efficient algorithm for data aggregation. The framework works as a middleware for aggregating data measured by number of nodes within a network.

The data aggregation is a technique used to solve the implosion and overlap problems in data centric routing. Data coming from multiple sensor nodes are aggregated as if they are about the same attribute of the phenomenon when they reach the same routing node on the way back to the sink. Data aggregation is a widely used technique in wireless sensor networks. The security issues, data confidentiality and integrity, in data aggregation become vital when the sensor network is deployed in a hostile environment. Data aggregation is a process of aggregating the sensor data using aggregation approaches. The general data aggregation algorithm works as shown in the below figure1. Figure 1 illustrates that data aggregation is the process of aggregating the sensor data using aggregation approaches. Then the algorithm uses the sensor data from the sensor nodes and then aggregates the data by using some aggregation algorithms such as centralized approach, LEACH(Low Energy Adaptive Clustering Hierarchy), TAG(Tiny Aggregation) etc. This aggregated data is transfer to the sink node by selecting the efficient path.

J Kennedy, RC Eberhart, Mentioned the "Particle Swarm Optimization", Proceedings of the IEEE International Joint Conference on Neural Networks The most widely used transmission support is radio waves. Wireless transmissions utilize the microwave spectre: the available frequencies are situated around the 2.4 GHz ISM (Industrial, Scientific and Medical) band for a bandwidth of about 83 MHz, and around the 5 GHz U-NII (Unlicensed-National Information Infrastructure) band for a bandwidth of about 300 MHz divided into two parts. The exact frequency allocations are set by laws in the different countries; the same laws also regulate the maximum allotted transmission power and location (indoor, outdoor). Such a wireless radio network has a range of about 10–100 meters to 10 Km per machine, depending on the emission power, the data rate, the frequency, and the type of antenna used. Many different models of antenna can be employed: omnis (omnidirectional antennas), sector antennas (directional antennas), yagis, parabolic dishes, or waveguides (cantennas).

The crucial point in channel access techniques for wireless networks is that it is not possible to transmit and to sense the carrier for packet collisions at the same time. Therefore there is no way to implement a CSMA/CD (Carrier Sense Multiple Access / Collision Detection) protocol such as in the wired Ethernet.

IEEE 802.11 uses a channel access technique of type CSMA/CA, which is meant to perform Collision Avoidance (or at least to try to). The CSMA/CA protocol states that a node, upon sensing that the channel is busy, must wait for an interframe spacing before attempting to transmit, then choose a random delay depending on the Contention Window.

The reception of a packet is acknowledged by the receiver to the sender. If the sender does not receive the acknowledgement packet, it waits for a delay according to the binary exponential backoff algorithm, which states that the Contention Window size is doubled at each failed try.

Unicast data packets are sent using a more reliable mechanism. The source transmits a RTS (Request To Send) packet for the destination, which replies with a CTS (Clear To Send) packet upon reception. If the source correctly receives the CTS, it sends the data packet.

Wireless networks offer the following productivity, convenience, and cost advantages over traditional wired networks.

de Kerchove and P. Van Dooren Mentioned the "PRDA(polynomial regression based secure data aggregation)," polynomial regression based secure data aggregation protocol in which sensor nodes represent their sensed data as polynomial functions. Instead of their original data, sensor nodes secretly send coefficients of these polynomial functions to data aggregators. Data aggregation is performed based on these coefficients and the base station is able to extract a good approximation of the network data from the aggregation result. The security analysis and simulation results show that the proposed scheme is able to reduce the amount of data transmission in the network without compromising data confidentiality.

III. DATA AGGREGATION AND CLUSTERING

Flat Networks: In flat networks, each sensor node plays the same role and is equipped with approximately the same battery power. In such networks, data aggregation is accomplished by data centric routing where the sink usually transmits a query message to the sensors, for example, via flooding and sensors which have data matching the query send response messages back to the sink. The choice of a particular communication protocol depends on the specific application at hand.

3.1.1 Diffusion: Directed diffusion (DD) may be a popular information aggregation paradigm for wireless device networks. it's a data-centric and application aware paradigm, within the sense that every one information generated by sensor nodes is called by attribute-value pairs. Such a scheme combines the information coming back from totally different sources en-route to the sink by eliminating redundancy and minimizing the amount of transmissions. during this means, it saves the energy consumption and will increase the network lifespan of WSNs. during this theme usually base station broadcast the message to the interested supply node. subsequently every node receives interest. These interests outline the attribute worth like name of object. every node get the interest will cache it for later use. because the interest is broadcasted by the network hop by hop, gradient square measure setups to draw information satisfying the query toward the requesting node. A gradient may be a reply link to the nearer from that the interest was received

3.1.2 SPIN: The sensor protocol for data via negotiation The staring node that has new data advertises the data to the close nodes within the network using the meta data. A close node that is interested in this type of information sends asking to the leader node for data. The leader node responds and send data to the sinks every node has a resource managing capability to keeps track of its energy usage within the sensing element network. every node polls its resources like battery power before data transmission. SPIN is also well-suited for environments with mobile sensors, since the forwarding decisions are based on native neighbourhood data.

3.2 Hierarichical Networks: In the hierarchical network, In which data aggregation data has to be done at special nodes, with the help of these special node we can reduce the number of number of data packet transmitted to the sink. So with this network improves the energy efficiency of the whole network. Various type hierarchical data-aggregation protocols as follows

3.2.1 Cluster-Based Networks for data aggregation: These Wireless sensor network is resource constraint that’s why sensor cannot directly transmit data to the base station. In which all regular sensors can send data packet to a cluster head (local aggregator) which aggregates data packet from all the regular sensors in its cluster and sends the concise digest to the base station. With the help of the scheme we save the energy of the sensors. LEACH: Low energy adaptive clustering has been proposed to organise a sensor network into a set of clusters so that the energy consumption can be event distributed among all the sensor nodes.

3.2.2 Chain –Based Networks for Data Aggregation In which each sensor sends data to the closer neighbour. Power- Efficient DataGathering Protocol for Sensor Information Systems (PEGASIS) is type of chain based data aggregation. In PEGASIS, all sensors are structured into a linear chain for data aggregation. The nodes can form a chain by employing a greedy algorithm or the sink can decide the chain in a centralized manner. In the Greedy chain formation assumes that all sensors have inclusive knowledge of the network. The farthest node from the sink initiates chain formation and, at each step, the closest neighbour of a node is selected as its successor in the chain. In each data-gathering round, a node receives data packet from one of its neighbours, aggregates the data with its own, and sends the aggregates data packet to its other neighbour along the chain. Eventually, the leader node in the are similar to cluster head sends the aggregated data to the base station. Figure below shows the chain based data-aggregation procedure in PEGASIS. 3.2.3 Tree

Based Networks for Data Aggregation In which all node are organized in form of tree means hierarchical, with then help of intermediate node we can perform data aggregation process and data transmit leaf node root node. Tree based data aggregation is suitable for applications which involve innetwork data aggregation. An example application is radiation-level monitoring in a nuclear plant where the maximum value provides the most useful information for the safety of the plant. One of the main aspects of tree-based networks is the construction of an energy efficient data-aggregation tree.

To overcome the matter occurred within the iterative filtering algorithm new technique referred to as Certificate Authority (CA) is introduced in every cluster. knowledge Aggregation is employed to mixture data's by the cluster head finally transmit it to the base station. the base station collects all the data's from cluster head and mixture for secure data transmission. To perform the aggregation safer the CA is employed to ascertain every node condition whether or not a node is trust node or malicious node. By exploitation the CA the node method are monitored.

Working Principle of Data Aggregation The working of WSN proposed architecture model illustrated in Figure 2 below that starts working by choosing selecting of nodes and divided into clusters. These clusters can satisfy the intended parameter requirements and conditions. The parameters like RSSI, TTL, MRIC, bandwidth, battery consumption are accustomed verify the amount of nodes that will be considered in a cluster. thereafter a cluster head (CH) is selected among nodes lies within the each cluster. CH are going to be responsible for administration of all different nodes inside several cluster and collecting the data} from the nodes within the cluster and transferring the information to the neighbouring cluster head for more information exchange and updation . The newly arrived nodes will be assigned as cluster head if the global cost of arrived node is minimum , otherwise other cluster nodes are going to be given opportunity to participate and global cost is once more recalculated. thereafter the data aggregation approach is presumed as the collection of data and numerous queries from the user end are checked and transformed into low level schemes by a query processor. All data collected and aggregated is stored at a storage location in database server. Finally at last the data is aggregated by data cube approach and every one the aggregated data are going to be transfer to the base station for further use.

Different Energy-efficient Techniques in Data Aggregation

Grid-Based Architecture It is energy–efficient data storage scheme in which Snake-like Energy Efficient Scheduling is given in the network is divided into 2 dimensional logical grids where the number of sensors in a grid is N. This works on Active and sleep mode procedure if one sensor is active at one time slot then the other is in sleep mode at that time of slot. The time slots are assigned in a snake-like direction into T with 4*3. When the active sensor changes from one row to another row in figure 4, i.e., from row A to row B, there exists one column sensor in active mode with two sequential time slots, i.e., column G. Hence, when two active sensors receive the query packet, exactly one of two sensors can continue to stay in active mode in the next time slot. This mechanism can guarantee that the query packet is preserved in one node of the grid to continue performing the query task.Temporal Correlation Based Data Aggregation Scheme In this scheme the author uses the ARIMA model also called Box-Jenkins model is a widely used forecast model for univariate time series. Data aggregation in this scheme the ordinary sensor node collects sensed value from environment. If the periodical update time is up, it will save the sensed value into the historical data queue and send the sensed value to the aggregator. Otherwise, it will calculate the forecast value using ARIMA model and compare the sensed value with the forecast value. If the difference between them is less than the predefined error threshold, the sensor will store the forecast value into the historical data queue. Otherwise, it will store the sensed value into the historical data queue and send the sensed value to the aggregator at the same time. The periodical update time is a preset and tunable scheme parameter which is used to periodically collect real sensed value and avoid cumulative error in continuous forecasts. The aggregator listens on the wireless channel to retrieve sensed values from ordinary sensor node and store them into the historical data queue. If the aggregator does not receive any data from sensor node after a predefined periodical data collect time, it means the difference between the sensed value and forecast value is within an acceptable range. Then the aggregator will calculate the forecast value using ARIMA model with historical data. The periodical data collect time should be selected carefully to ensure it is enough to deliver the message from sensor to the aggregator.

Steiner Tree

In this technique the author describe the wireless sensor network communication model as an undirected graph G , where the node set V contains all nodes that have been aware within the region, the distance between nodes in the graph are the weights E associate with the vertices, the communication distance of the node is R, introduces a Steiner tree a weighted undirected graph G is given, and the demand Steiner tree T, and find the shortest path from the root of T to other nodes

A basic variant of the PSO algorithm works by having a population (called a swarm) of candidate solutions (called particles). These particles are moved around in the search-space according to a few simple formulae. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known position. When improved positions are being discovered these will then come to guide the movements of the swarm. The process is repeated and by doing so it is hoped, but not guaranteed, that a satisfactory solution will eventually be discovered.

Formally, let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be the cost function which must be minimized. The function takes a candidate solution as argument in the form of a vector of real numbers and produces a real number as output which indicates the objective function value of the given candidate solution. The gradient of f is not known. The goal is to find a solution $\mathbf{a}$ for which $f(\mathbf{a}) \leq f(\mathbf{b})$ for all $\mathbf{b}$ in the search-space, which would mean $\mathbf{a}$ is the global minimum. Maximization can be performed by considering the function $h = -f$ instead.

Let S be the number of particles in the swarm, each having a position $\mathbf{x}_i \in \mathbb{R}^n$ in the search-space and a velocity $\mathbf{v}_i \in \mathbb{R}^n$. Let $\mathbf{p}_i$ be the best known position of particle i and let $\mathbf{g}$ be the best known position of the entire swarm. A basic PSO algorithm is then:

- For each particle $i = 1, ..., S$ do:
- Initialize the particle's position with a uniformly distributed random vector: $\mathbf{x}_i \sim U(\mathbf{b}_{\mathbf{lo}}, \mathbf{b}_{\mathbf{up}})$, where $\mathbf{b}_{\mathbf{lo}}$ and $\mathbf{b}_{\mathbf{up}}$ are the lower and upper boundaries of the search-space.
- Initialize the particle's best known position to its initial position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
- If $(f(\mathbf{p}_i) < f(\mathbf{g}))$ update the swarm's best known position: $\mathbf{g} \leftarrow \mathbf{p}_i$
- Initialize the particle's velocity: $\mathbf{v}_i \sim U(-|\mathbf{b}_{\mathbf{up}}-\mathbf{b}_{\mathbf{lo}}|, |\mathbf{b}_{\mathbf{up}}-\mathbf{b}_{\mathbf{lo}}|)$
- Until a termination criterion is met (e.g. number of iterations performed, or a solution with adequate objective function value is found), repeat:
- For each particle $i = 1, ..., S$ do:
- For each dimension $d = 1, ..., n$ do:
- Pick random numbers: $r_p, r_g \sim U(0,1)$
- Update the particle's velocity: $\mathbf{v}_{i,d} \leftarrow \omega\, \mathbf{v}_{i,d} + \varphi_p\, r_p\, (\mathbf{p}_{i,d}-\mathbf{x}_{i,d}) + \varphi_g\, r_g\, (\mathbf{g}_d-\mathbf{x}_{i,d})$
- Update the particle's position: $\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{v}_i$
- If $(f(\mathbf{x}_i) < f(\mathbf{p}_i))$ do:
- Update the particle's best known position: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
- If $(f(\mathbf{p}_i) < f(\mathbf{g}))$ update the swarm's best known position: $\mathbf{g} \leftarrow \mathbf{p}_i$
- Now $\mathbf{g}$ holds the best found solution.

- The parameters ω, φ_p, and φ_g are selected by the practitioner and control the behaviour and efficacy of the PSO method, see below.

Parameter Selection

The choice of PSO parameters can have a large impact on optimization performance. Selecting PSO parameters that yield good performance has therefore been the subject of much research.

The PSO parameters can also be tuned by using another overlaying optimizer, a concept known as meta-optimization. Parameters have also been tuned for various optimization scenarios.

Neighborhoods and Topologies

The basic PSO is easily trapped into a local minimum. This premature convergence can be avoided by not using the entire swarm's best known position g but just the best known position l of a sub-swarm "around" the particle that is moved. Such a sub-swarm can be a geometrical one - for example "the m nearest particles" - or, more often, a social one, i.e. a set of particles that is not depending on any distance. In such a case, the PSO variant is said to be local best (vs global best for the basic PSO).

If we suppose there is an information link between each particle and its neighbours, the set of these links builds a graph, a communication network, that is called the topology of the PSO variant. A commonly used social topology is the ring, in which each particle has just two neighbours, but there are many others. The topology is not necessarily fixed, and can be adaptive (SPSO, stochastic star, TRIBES, Cyber Swarm, C-PSO)

Inner workings

There are several schools of thought as to why and how the PSO algorithm can perform optimization.

A common belief amongst researchers is that the swarm behaviour varies between exploratory behaviour, that is, searching a broader region of the search-space, and exploitative behaviour, that is, a locally oriented search so as to get closer to a (possibly local) optimum. This school of thought has been prevalent since the inception of PSO. This school of thought contends that the PSO algorithm and its parameters must be chosen so as to properly balance between exploration and exploitation to avoid premature convergence to a local optimum yet still ensure a good rate of convergence to the optimum. This belief is the precursor of many PSO variants, see below.

Another school of thought is that the behaviour of a PSO swarm is not well understood in terms of how it affects actual optimization performance, especially for higher-dimensional search-spaces and optimization problems that may be discontinuous, noisy, and time-varying. This school of thought merely tries to find PSO algorithms and parameters that cause good performance regardless of how the swarm behaviour can be interpreted in relation to e.g. exploration and exploitation. Such studies have led to the simplification of the PSO algorithm,

Convergence

In relation to PSO the word convergence typically refers to two different definitions:

Convergence of the sequence of solutions (aka, stability analysis, converging) in which all particles have converged to a point in the search-space, which may or may not be the optimum,

Convergence to a local optimum where all personal bests p or, alternatively, the swarm's best known position g, approaches a local optimum of the problem, regardless of how the swarm behaves.

Convergence of the sequence of solutions has been investigated for PSO. These analyses have resulted in guidelines for selecting PSO parameters that are believed to cause convergence to a point and prevent divergence of the swarm's particles (particles do not move unboundedly and will converge to somewhere). However, the analyses were criticized by Pedersen for being oversimplified as they assume the swarm has only one particle, that it does not use stochastic variables and that the points of attraction, that is, the particle's best known position p and the swarm's best known position g, remain constant throughout the optimization process. However, it was shown that these simplifications do not affect the boundaries found by these studies for parameter where the swarm is convergent.

Convergence to a local optimum has been analyzed for PSO in and. It has been proven that PSO need some modification to guarantee to find a local optimum.

This means that determining convergence capabilities of different PSO algorithms and parameters therefore still depends on empirical results. One attempt at addressing this issue is the development of an "orthogonal learning" strategy for an improved use of the information already existing in the relationship between p and g, so as to form a leading converging exemplar and to be effective with any PSO topology. The aims are to improve the performance of PSO overall, including faster global convergence, higher solution quality, and stronger robustness. However, such studies do not provide theoretical evidence to actually prove their claims.

Biases

As the basic PSO works dimension by dimension, the solution point is easier found when it lies on an axis of the search space, on a diagonal, and even easier if it is right on the centre.

One approach is to modify the algorithm so that it is not any more sensitive to the system of coordinates. Note that some of these methods have a higher computational complexity (are in O(n^2) where n is the number of dimensions) that make the algorithm very slow for large scale optimization.

The only currently existing PSO variant that is not sensitive to the rotation of the coordinates while is locally convergent has been proposed at 2014. The method has shown a very good performance on many benchmark problems while its rotation invariance and local convergence have been mathematically proven.

Variants

Numerous variants of even a basic PSO algorithm are possible. For example, there are different ways to initialize the particles and velocities (e.g. start with zero velocities instead), how to dampen the velocity, only update pi and g after the entire swarm has been updated, etc. Some of these choices and their possible performance impact have been discussed in the literature.

A series of standard implementations have been created by leading researchers, "intended for use both as a baseline for performance testing of improvements to the technique, as well as to represent PSO to the wider optimization community. Having a well-known, strictly-defined standard algorithm provides a valuable point of comparison which can be used throughout the field of research to better test new advances." The latest is Standard PSO 2011 (SPSO-2011).

Hybridization

New and more sophisticated PSO variants are also continually being introduced in an attempt to improve optimization performance. There are certain trends in that research; one is to make a hybrid optimization method using PSO combined with other optimizers, e.g., combined PSO with biogeography-based optimization,[43] and the incorporation of an effective learning method.

Alleviate Premature

Another research trend is to try and alleviate premature convergence (that is, optimization stagnation), e.g. by reversing or perturbing the movement of the PSO particles, another approach to deal with premature convergence is the use of multiple swarms (multi-swarm optimization). The multi-swarm approach can also be used to implement multi-objective optimization. Finally, there are developments in adapting the behavioural parameters of PSO during optimization.

Simplifications

Another school of thought is that PSO should be simplified as much as possible without impairing its performance; a general concept often referred to as Occam's razor. Simplifying PSO was originally suggested by Kennedy and has been studied more extensively, where it appeared that optimization performance was improved, and the parameters were easier to tune and they performed more consistently across different optimization problems.

Another argument in favour of simplifying PSO is that metaheuristics can only have their efficacy demonstrated empirically by doing computational experiments on a finite number of optimization problems. This means a metaheuristic such as PSO cannot be proven correct and this increases the risk of making errors in its description and implementation. A good example of this presented a promising variant of a genetic algorithm (another popular metaheuristic) but it was later found to be defective as it was strongly biased in its optimization search towards similar values for different dimensions in the search space, which happened to be the optimum of the benchmark problems considered. This bias was because of a programming error, and has now been fixed.

Initialization of velocities may require extra inputs. A simpler variant is the accelerated particle swarm optimization (APSO), which does not need to use velocity at all and can speed up the convergence in many applications. A simple demo code of APSO is available.

Multi-objective optimization

PSO has also been applied to multi-objective problems,[55][56] in which the objective function comparison takes pare to dominance into account when moving the PSO particles and non-dominated solutions are stored so as to approximate the pare to front.

Binary, Discrete, and Combinatorial PSO

As the PSO equations given above work on real numbers, a commonly used method to solve discrete problems is to map the discrete search space to a continuous domain, to apply a classical PSO, and then to demap the result. Such a mapping can be very simple (for example by just using rounded values) or more sophisticated.[57]

However, it can be noted that the equations of movement make use of operators that perform four actions:

computing the difference of two positions. The result is a velocity (more precisely a displacement)

multiplying a velocity by a numerical coefficient

adding two velocities

applying a velocity to a position

Usually a position and a velocity are represented by n real numbers, and these operators are simply -, *, +, and again +. But all these mathematical objects can be defined in a completely different way, in order to cope with binary problems (or more generally discrete ones), or even combinatorial ones. One approach is to redefine the operators based on sets.

IV. RESULTS AND DISCUSSIONS

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It is often necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Besides the term clustering, there are a number of terms with similar meanings, including automatic classification, numerical taxonomy and typological analysis. The subtle differences are often in the usage of the results: while in data mining, the resulting groups are the matter of interest, in automatic classification the resulting discriminative power is of interest. This often leads to misunderstandings between researchers coming from the fields of data mining and machine learning[citation needed], since they use the same terms and often the same algorithms, but have different goals.

Notations of basic Clustering

The notion of a "cluster" cannot be precisely defined, which is one of the reasons why there are so many clustering algorithms. There is a common denominator: a group of data objects. However, different researchers employ different cluster models, and for each of these cluster models again different algorithms can be given. The notion of a cluster, as found by different algorithms, varies significantly in its properties. Understanding these "cluster models" is key to understanding the differences between the various algorithms. Typical cluster models include:

Connectivity models: for example, hierarchical clustering builds models based on distance connectivity.

Centroid models: for example, the k-means algorithm represents each cluster by a single mean vector.

Distribution models: clusters are modeled using statistical distributions, such as multivariate normal distributions used by the Expectation-maximization algorithm.

Density models: for example, DBSCAN and OPTICS defines clusters as connected dense regions in the data space.

Subspace models: in Bi-clustering (also known as Co-clustering or two-mode-clustering), clusters are modeled with both cluster members and relevant attributes.

Group models: some algorithms do not provide a refined model for their results and just provide the grouping information.

Graph-based models: a clique, that is, a subset of nodes in a graph such that every two nodes in the subset are connected by an edge can be considered as a prototypical form of cluster. Relaxations of the complete connectivity requirement (a fraction of the edges can be missing) are known as quasi-cliques, as in the HCS clustering algorithm.

A "clustering" is essentially a set of such clusters, usually containing all objects in the data set. Additionally, it may specify the relationship of the clusters to each other, for example, a hierarchy of clusters embedded in each other. Clustering can be roughly distinguished as:

hard clustering: each object belongs to a cluster or not

Soft clustering (also: fuzzy clustering): each object belongs to each cluster to a certain degree (for example, a likelihood of belonging to the cluster)

There are also finer distinctions possible, for example:

strict partitioning clustering: here each object belongs to exactly one cluster

Strict partitioning clustering with outliers: objects can also belong to no cluster, and are considered outliers.

Overlapping clustering (also: alternative clustering, multi-view clustering): while usually a hard clustering, objects may belong to more than one cluster.

Hierarchical clustering: objects that belong to a child cluster also belong to the parent cluster

Subspace Clustering: while an overlapping clustering, within a uniquely defined subspace, clusters are not expected to overlap.

Clustering using PSO

PSO clustering is a fundamental operation used in unsupervised document organization, automatic topic extraction, and information retrieval. Clustering involves dividing a set of objects into a specified number of clusters. The motivation behind clustering a set of data is to find inherent structure in the data and to expose this structure as a set of groups. The data objects within each group should exhibit a large degree of similarity while the similarity among different clusters should be minimized. There are two major clustering techniques: “Partitioning” and “Hierarchical”. Most document clustering algorithms can be classified into these two groups. The hierarchical techniques produce a nested Sequence of partition, with a single, all-inclusive cluster at the top and single clusters of individual points at the bottom. The partitioning clustering method seeks to partition a collection of sets into a set of non-overlapping groups, so as to maximize the evaluation value of clustering. Although the hierarchical clustering technique is often portrayed as a better quality clustering approach, this technique does not contain any provision for the reallocation of entities, which may have been poorly classified in the early stages of the text analysis. Moreover, the time complexity of this approach is quadratic. In recent years, it has been recognized that the partitional clustering technique is well suited for clustering a large document dataset due to their relatively low computational requirements. The time complexity of the partitioning technique is almost linear, which makes it widely used. The best known partitioning clustering algorithm is the K-means algorithm and its variants. This algorithm is simple, straightforward and is based on the firm foundation of analysis of variances. In addition to the K-means algorithm, several algorithms, such as Genetic Algorithm (GA) and Self-Organizing Maps (SOM), have been used for document clustering. Particle Swarm Optimization (PSO) is another computational intelligence method that has already been applied to image clustering and other low dimensional datasets. However, to the best of the author's knowledge, PSO has not been used to cluster text documents. In this study, a document clustering algorithm based on PSO is proposed. The remainder of this paper is organized as follows: provides the methods of representing documents in clustering algorithms and of computing the similarity between documents. Section 3 provides a general overview of the K-means and PSO optimal algorithm. The PSO clustering algorithms are described in Section 4. Section 5 provides the detailed experimental setup and results for comparing the performance of the PSO algorithm with the K-means approaches. The discussion of the experiment's results is also presented.

Problem Definition

The system performs knowledge aggregation with security and attack handling mechanism. Repetitive filtering techniques with initial approximation model square measure accustomed secure knowledge aggregation method. Owing to restricted procedure power and energy resources, aggregation of data's from multiple device nodes done at the aggregating node. Such aggregation is understood to be extremely liable to node compromising attacks. Repetitive filtering algorithms hold nice promise for such a purpose. Such algorithms at the same time mixture knowledge from multiple sources and supply trust assessment of those sources, typically in an exceedingly sort of corresponding weight factors allotted to knowledge provided by every supply. The present paper demonstrate that many existing repetitive filtering algorithms, whereas considerably a lot of sturdy against collusion attacks than the easy averaging strategies, are even so susceptive to a unique subtle collusion attack. To handle this security issue, this paper proposes Associate in Nursing improvement for repetitive filtering techniques by providing an initial approximation for such algorithms that makes them not solely collusion sturdy, however additionally a lot of correct and quicker convergence. This algorithm doesn't handle packet drop attack and not economical for centralized approach.

To overcome the matter occurred within the iterative filtering algorithm new technique referred to as Certificate Authority (CA) is introduced in every cluster. knowledge Aggregation is employed to mixture data's by the cluster head finally transmit it to the base station. the base station collects all the data's from cluster head and mixture for secure data transmission. To perform the aggregation safer the CA is employed to ascertain every node condition whether or not a node is trust node or malicious node. By exploitation the CA the node method are monitored.

V. CONCLUSION AND FUTURE WORK

In introduction, the scope and objective of the proposed method has been described. In this final chapter, conclusion is made by describing the progress made towards this goal in terms of the development of proposed framework and also suggested some future research directions that could provide the next steps along the path to a practical and widely applicable for laser speckle authentication.

Trust management systems for WSN is constrained by the nature and features of these type of networks (computational power, energy constraint) and also depending on the underlying problem that the trust management aims to solve. Thus, a system designed for detecting misbehaving nodes could be different than another one designed, for instance, for routing. Special attention should be paid to the way of gathering

information and what sort of information is relevant to be gathered. Thus, causes of mistrust could be dropping packets or, appearing or disappearing from the network without an apparent reason. The information is gathered the underlying mathematical model used for computing the trust or reputation values of the no des is also different from one model to the other. Even if in some cases simple averages or linear functions like a pro duct are used, the values obtained in these cases might not be very significant

REFERENCES

[1] Ozdemir S. and Xiao Y. (2009), 'Secure data aggregation in wireless sensor networks: A comprehensive overview', Comput. Netw., vol. 53, no. 12, pp. 2022–2037.

[2] Ayday E., Lee H., and Fekri F. (2009), 'An iterative algorithm for trust and reputation

management', Proc. IEEE Int. Conf. Symp. Inf.Theory, vol. 3, pp. 2051–2055

[3] de Kerchove and P. Van Dooren, "Iterative filtering in reputation systems," SIAM J. Matrix Anal. Appl., vol. 31, no. 4, pp. 1812–1834, Mar. 2010.

[4] Hoffman K., Zage D., and Nita-Rotaru C. (2009), 'A survey of attack and defense techniques for reputation systems', ACM Comput. Surveys, vol. 42, no. 1, pp. 1:1–1:31.

[5] Yang Y., Wang X., Zhu S., and S. Cao S. (2006), 'SDAP: A secure hop-by hop data aggregation protocol for sensor networks', in Proc. 7th ACM Int. Symp. Mobile Ad Hoc Netw.Comput., pp. 356–367.

[6] Chan H., Perrig A., and Song D. (2006), 'Secure hierarchical in-network aggregation in sensor networks', in Proc. 13th ACM Conf. Comput. Commun. Security, pp. 278–287

[7] Ho J.-W., Wright M., and Das S. (2012), 'Zone Trust: Fast zone-based node compromise detection and revocation in wireless sensor networks using sequential hypothesis testing', IEEE Trans. Dependable Secure Comput ., vol. 9, no. 4, pp. 494–511.

[8] Lim H.-S., Ghinita G., Bertino E., and Kantarcioglu M. (2012), 'A game-theoretic approach for high-assurance of data trustworthiness in sensor networks', in Proc. IEEE 28th Int. Conf. Data Eng., pp. 1192–1203.

[9] Roy S., Conti M., Setia S., and Jajodia S. (2012), 'Secure data aggregation in wireless sensor networks', IEEE Trans. Inf. Forensics Security, vol. 7, no. 3, pp. 1040–1052.

[10] Tang L.-A., Yu X., Kim S., Han J., Hung C.-C., and Peng W.-C. (2010), 'Tru-Alarm: Trustworthiness analysis of sensor networks in cyber-physical systems', in Proc. IEEE Int. Conf. Data Mining, pp. 1079–1084.

[11] J Kennedy, RC Eberhart, "Particle Swarm Optimization", Proceedings of the IEEE International Joint Conference on Neural Networks, Vol. 4, pp 1942–1948, 1995.

[12] J Kennedy, RC Eberhart, Y Shi, "Swarm Intelligence", Morgan Kaufmann, 2002.