IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ENHANCED SELECTIVE ENCRYPTION METHOD FOR BIGDATA SENSING STREAM USING ONE WAY HASH CHAIN ALGORITHM

Ms.N.Suriyapriya[1], Mr.S.Nagaraj[2]. ME-CSE,
[1]PG Scholar, Department of Computer Science and Engineering,
[2]Assistant professor, Department of Computer Science and Engineering,
Selvam College of Technology, Pappinaickenpatti, Namakkal-637 001
Tamilnadu, India.

ABSTRACT— Resource constrained sensing devices are being used widely to build and deploy self-organizing wireless sensor networks for a variety of critical applications such as smart cities, smart health, precision agriculture and industrial control systems. A Data Stream Manager (DSM) at the server collects the data streams to perform real time analysis and decision-making for these critical applications. A malicious adversary may access or tamper with the data in transit. One of the challenging tasks in such applications is to assure the trustworthiness of the collected data so that any decisions are made on the processing of correct data. Assuring high data trustworthiness requires that the system satisfies two key security properties: confidentiality and integrity. To ensure the confidentiality of collected data, we need to prevent sensitive information from reaching the wrong people by ensuring that the right people are getting it. Sensed data are always associated with different sensitivity levels based on the sensitivity of emerging applications are sensed data types or the sensing devices. Providing multilevel data confidentiality along with data integrity for big sensing data streams in the context of near real time analytics is a challenging problem. We propose a Selective Encryption (SEEN) method to secure big sensing data streams that satisfies the desired multiple levels of confidentiality and data integrity. Our method is based on two key concepts: common shared keys that are initialized and updated by DSM without requiring retransmission and a seamless key refreshment process without interrupting the data stream encryption/decryption. A novel scheme to secure a multi-hop network programming protocol through the use of multiple one-way hash chains is proposed. The scheme is shown to be lower in computational, power consumption and communication costs yet still able to secure multi-hop propagation of program images.

Key words — *Big Data, Data Stream Manager, Selective Encryption, confidentiality, multi-hop*

1. INTRODUCTION:

Data mining or knowledge discovery is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides. For businesses, data mining is used to discover patterns and relationships in the data in order to help make better business decisions. Data mining can help spot sales trends, develop smarter marketing campaigns, and accurately predict customer loyalty. Big Data will help to create new growth opportunities and entirely new categories of companies, such as those that aggregate and analyze industry data. Many of these will be companies that sit in the middle of large information flows where data about products and services, buyers and suppliers, consumer preferences and intent can be captured and analyzed.

2 RELATED WORKS

Lei Tang and Huan et al [1] It demonstrates many advantages, especially suitable for large-scale networks, paving the way for the study of collective behavior in many real-world applications. Social media such as Facebook, MySpace, Twitter, BlogCatalog, Digg, YouTube and Flickr, facilitate people of all walks of life to express their thoughts, voice their opinions, and connect to each other anytime and anywhere. For instance, a popular content-sharing site like Delicious, Flickr, and YouTube allows users to upload, tag and comment different types of contents (e.g., bookmarks, photos, videos). Users registered at these sites can also become friends, a fan or follower of others. The prolific and expanded use of social media has turned online interactions into a vital part of human experience.The election of Barack Obama as the President of United States was partially attributed to his smart Internet strategy and access to millions of younger voters through the new social media, such as Facebook, a popular social networking site claiming to attract 400 million active users up to date. many challenges are still there and need further research. Below, they elaborated some interesting directions. Extraction of actor information: In social learning, the structural information of social networks alone is a weak indicator of user behavior. In all their experiment results, it is noticed that based on network information alone, the collective behavior prediction performance is far from satisfactory (with 20-30% F1-measure). Hybrid approach to social dimension extraction: In edge view methods, one fundamental assumption is that each edge belongs to only one affiliation.

M. E. J. Newman [2] Finding community structure in networks using the eigenvectors of matrices The result leads us to a number of possible algorithms for detecting community structure, as well as several other results, including a spectral measure of bipartite structure in networks and a new centrality measure that identifies those vertices that occupy central positions within the communities to which they belong. The algorithms and measures proposed are illustrated with applications to a variety of real-world complex networks. In social networks, for example, it has long been accepted that individuals who lie on the boundaries of communities, bridging gaps between otherwise unconnected people, enjoy an unusual level of influence as the gatekeepers of information flow between groups. . There is already a substantial body of theory supporting the view that community structure can be accurately quantified using the benefit function known as modularity and hence that communities can be detected by searching possible divisions of a network for ones that possess high modularity. They have demonstrated that the modularity can be succinctly expressed in terms eigenvectors of a matrix they call the modularity matrix which is a characteristic property of the network and is itself independent of any division of the network into communities. Using this expression they have derived a series.

Parag Singla, Matthew Richardson et al [3] Yes, There is a Correlation - From Social Networks to Personal Behavior on the Web They applied data mining techniques to study this relationship for a population of over 10 million people, by turning to online sources of data. The analysis reveals that people who chat with each other (using instant messaging) are more likely to share interests (their Web searches are the same or topically similar). The more time they spend talking, the stronger their relationship. People who chat with each other are also more likely to share other personal characteristics, such as their age and location and, they are likely to be of opposite gender. Similar findings hold for people who do not necessarily talk to each other but do have a friend in common. Their analysis is based on a well-defined mathematical formulation of the problem, and is the largest such study they were aware of.

Miller Mcpherson et al [4] BIRDS OF A FEATHER: Homophily in Social Networks"Similarity breeds connection". This principle the homophily principle-structures network ties of every type, including marriage, friendship, work, advice, support, information transfer, exchange, co-membership, and other types of relationship. The result is that people's personal networks are homogeneous with regard to many socio demographic, behavioral, and intrapersonal characteristics. Homophily limits people's social world in a way that has powerful implications for the information they receive, the attitudes, and the interactions they experience. Homophily in race and ethnicity creates the strongest divides in the personal environments, with age, religion, education, occupation, and gender following in roughly that order. Geographic propinquity, families, organizations, and isomorphic positions in social systems all create contexts in which homophilous relations form. Ties between non-similar individuals also dissolve at a higher rate, which sets the stage for the formation of niches (localized positions) within social space. they focused on the many types of network relationships that researchers have found to be homophilous, and on the wide range of dimensions on which similarity induces homophily. The sources of homophily, focusing on the social structures that induce propinquity among similar others and the cognitive processes that make communication between similar others more likely. Finally, they end with implications for future research.

Jan Zahálka et al [5] proposed City Melange, an interactive and multimodal content-based venue explorer. Our framework matches the interacting user to the users of social media platforms exhibiting similar taste. The data collection integrates location-based social networks such as Foursquare with general multimedia sharing platforms such as Flickr or Picasa. In City Melange, the user interacts with a set of images and thus implicitly with the underlying semantics. The semantic information is captured through convolutional deep net features in the visual domain and latent topics extracted using Latent Dirichlet allocation in the text domain. A linear SVM model learns the interacting user's preferences and determines similar users. The experiments show that our content-based approach outperforms the user-activity-based and popular vote baselines even from the early phases of interaction, while also being able to

recommend mainstream venues to mainstream users and off-the-beaten-track venues to afficionados. City Melange is shown to be a well-performing venue exploration approach. The classic recommender paradigm has a number of innate shortcomings with respect to exploration. Typically, classic recommenders depend either on a user-item matrix that matches users to venues, or on the overall popularity of individual venues. Both variants suffer from the cold start problem: they require an active user base before the recommendations start to make sense. Also, they often tend to gravitate towards mainstream venues visited by many users. This is due to the recommendations being based on topical relevance ranking: mainstream topics will rank higher for mainstream users, specialized topics will rank higher for afficionados, as long as the actor provides consistent feedback.

3 PROPOSED WORK

The system is an authentication scheme to secure multihop network programming with multiple one-way hash chains. Instead of the expensive asymmetric cryptographic primitives used in much prior work, the scheme employs only symmetric cryptographic primitives, in a circular geographic node deployment model. In the system discussed the possible attacks an adversary could mount on the scheme and provided simple and effective counter measures against them. Finally, it provided a comprehensive performance evaluation of the scheme in terms of end-to-end latency and power consumption, which it believes is the first power consumption evaluation of a security scheme for network programming protocols. In future work will include the design of security models for network using the scheme in hierarchical topologies for improved scalability, secure update of multiple one-way key chains when all the elements of the key chains have been consumed, and also the design of countermeasures against Denial of Service attack.

- Data set collection
- Network configuration
- selective encryption model
- one way hash chain model
- Hybrid hash chain model
- Top K-Query Recommendation
- Report

4 MODULE DESCRIPTIONS

4.1 Administrator login

Using the given username and password, the user logins to the application. A list of username and password are stored in the database table. First, the main form starts running. In that form's load event, the form is set visible to false and then login is displayed. After successful login, the main form will display; otherwise the application ends.

4.2 Data set collection

The length of the Hash Chain (In computer security, a **hash chain** is a method to produce many one-time keys from a single key or password) is denoted by L. The Hop group count is denoted by S. The L and S value are keyed in to the database table 'Parameters'. Text box controls are placed to key in the L and S values. The committed values i.e., final key in the hash chains are up to S count. Each Hop group nodes are given with a committed value so that they can decrypt the data. The data to be sent from base station to all nodes will split into L packets and so the hash chain will have L+1 key.

4.3 Network configuration

The Hop index is selected and the Node Id is keyed in to the database table 'Nodes'. The node id for first group is denoted as A1, A2, A3, etc, The node id for second group is denoted as B1, B2, B3, etc, and goes on up to 'S' number of Hops. All the nodes in various hop groups need to transmit data after processing packet information and cut out data of their corresponding part in the packet. For that, the nodes use the committed values already pre-distributed by the base station.

4.4 Selective encryption model

The hash chain is built with giving a prime number in $K_{L,1}$ level and creating keys of upto 'L' count. The hash chain is built for 'S' count. The committed value (i.e., the last key) produced are given to all nodes. The key details are stored in 'KeyValues' table. During packet preprocessing, the key values are fetched from 'KeyValues' and Packet Data is concatenated with key values and are broadcasted to all the nodes.

4.5 One Way Hash Chain Model

During packet receiving, the raw packet data is extracted from the received packet based on the given algorithm. The node receives the data and cut out the data of their corresponding part in the packet. For that, the nodes use the committed values already pre-distributed by the base station. Only the nodes already becomes the part of the network and having the hash chain's committed value can process the data. The node without having committed value, if tries to process the packet, then the packet data cannot be parsed out by the node and the network treats that nodes as suspicious, i.e., attacker node.

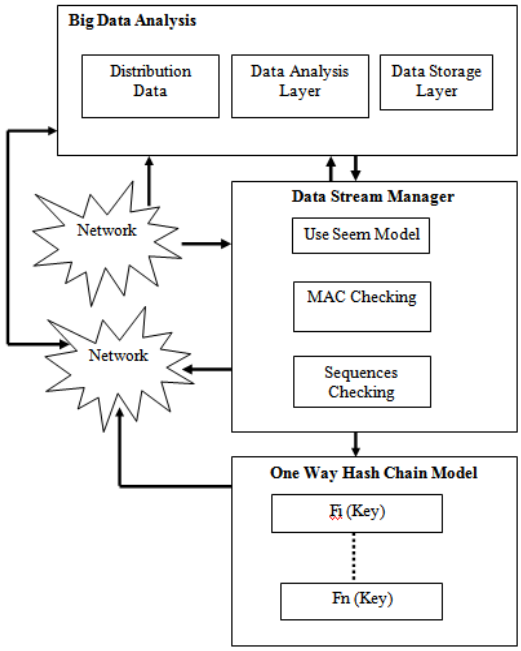

Fig 4.1 SYSTEM ARCHITECTURE

4.6 Top K-Query Recommendation

The ensemble of models is invoked both in the outlier detection and novel class detection modules. The **outlier detection process utilizes the decision boundary** of the ensemble of models to decide whether or not an instance is outlier. This decision boundary is built during training Top K-query. The novel class **detection process computes the cohesion among the outliers** in the buffer and separation of the outliers from the existing classes to decide whether a novel class has arrived. The proposed system enhances the existing novel class detection technique in three ways, which are

- outlier detection using adaptive threshold,
- novel class detection and
- Simultaneous multiple novel class detection

4.7 Report

4.7.1 Nodes List

The Hop index and Node id list are displayed using this report. The data grid view is filled with a data table which contains ‘nodes table’s record. The nodes started with A are treated as first hop nodes, B are treated as second hop nodes and goes on.

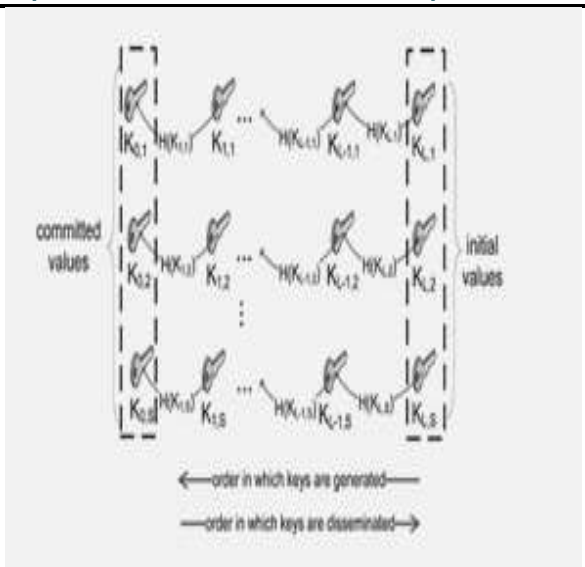

Fig 4.7.1 Nodes List

4.7.2 Key Values List and Broadcast Information

The Key values prepared for various hash chains at different hash levels are displayed using this report. Using data grid view control, the records are displayed. The data is taken from ‘keyvalues’ table. The From Node Id, To Node Id, packet index, packet data and key section values are displayed using this report.

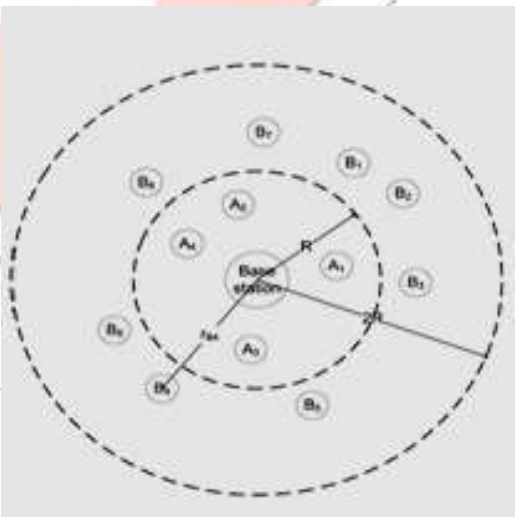

Fig 4.7.2 Key Values List and Broadcast Information

5 PERFORMANCE ANALYSES

First design a simplified mechanism to determine the number of neighboring nodes for any given node. Within time Tv, the given node crosses through an area and meets a number of neighbors N. Since mobile nodes are assumed uniformly distributed in the network, we may approximate N by Where r denotes the transmission range of nodes, *v* is the velocity, and *p* is the density of nodes in the network. Based on the obtained

number of neighboring nodes N, we can on firm the value of threshold K.

$$N = (\pi r^2 + 2rvT_v)\rho,$$

The following **Table 6.1** describes experimental result for existing system secure transmission node analysis. The table contains number of time slot interval and given time interval to calculate average numbers of send transmission node details are shown.

S.NO	NUMBER OF TIME SLOT (M)	RATIO OF SECURE TRANSMISSION NODE
1	10	0.43
2	20	0.52
3	40	0.61
4	60	0.69
5	80	0.74
6	100	0.80
7	120	0.86
8	140	0.90
9	150	0.93
10	160	0.97

Table 5.1 SEEN -Secure Transmission

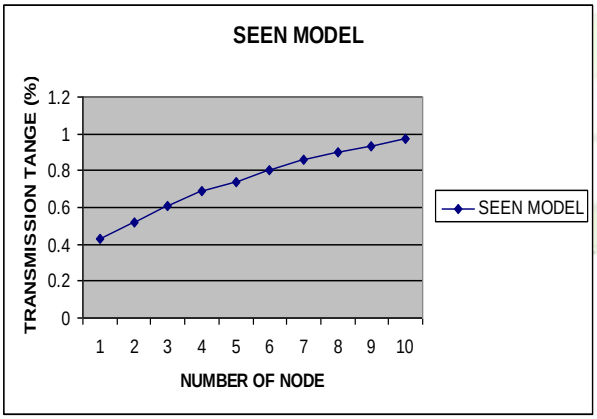

Fig 5.1 SEEN -Secure Transmission

S.NO	NUMBER OF TIME SLOT (M)	RATIO OF SECURE TRANSMISSION NODE
1	10	0.48
2	20	0.57
3	40	0.66
4	60	0.72
5	80	0.77
6	100	0.83
7	120	0.89
8	140	0.92
9	150	0.95
10	160	0.98

Table 5.2 HASH Secure Transmission

The following **Figure 6.2** describes experimental result for proposed system secure transmission node analysis. The table contains number of time slot interval and given time interval to calculate average numbers of send transmission node details are shown

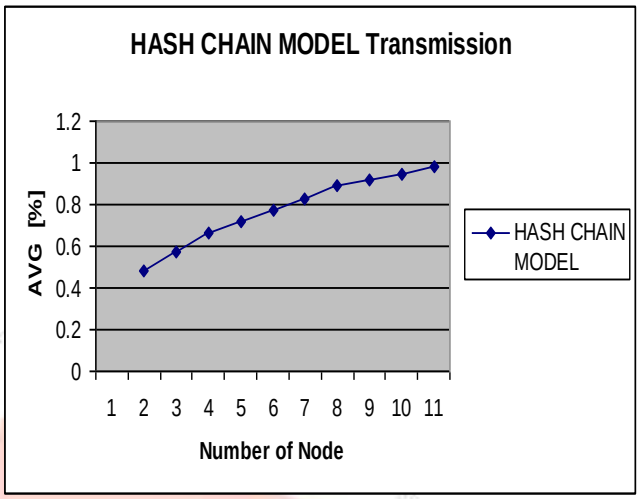

Fig 5.2 HASH CHAIN Secure Transmission

S.NO	NUMBER OF TIME SLOT (M)	RATIO OF SECURE TRANSMISSION NODE	
		SEEN	HASH
1	10	0.43	0.48
2	20	0.52	0.57
3	40	0.61	0.66
4	60	0.69	0.72
5	80	0.74	0.77
6	100	0.80	0.83
7	120	0.86	0.89
8	140	0.90	0.92
9	150	0.93	0.95
10	160	0.97	0.98

Table 6.3 Comparisons for SEEN and HASH Secure Transmission

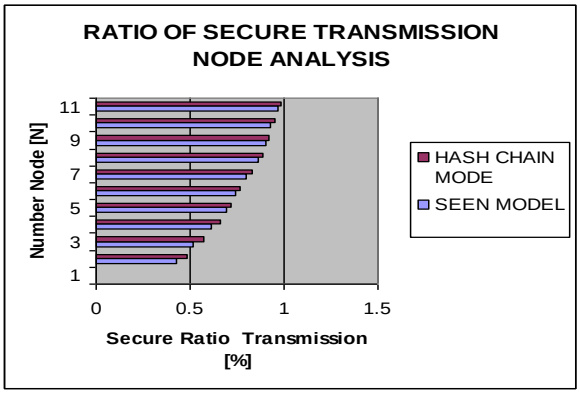

Fig 6.3 Comparisons for SEEN and HASH Secure Transmission

7 FUTURE ENHANCEMENTS

The future work will include the design of security models for network programming. The application if developed platform independent it can be used in may operating systems. The more hops the scheme supports for network programming, the more memory overhead there will be, as more key updates and packet authentications for downstream nodes are required. Most practical wireless networks will be restricted to fewer hops than four since the throughput significantly degrades for larger hop counts. In addition, large deployments will be likely to form some kind of hierarchy. Hence, the utilization of memory can be measured in future. Currently the scheme has a slightly less memory overhead, while in the more complex applications; the scheme may utilize more memory. The future study can be in the area of more significant memory savings. In this scheme for multi-hop program updates, as number of nodes increases, the keys used by the scheme increases due to the number of cryptographic operations in the scheme. The algorithms can be improved to reduce the key count and length further.

8 CONCLUSIONS

An authentication scheme is proposed to secure multihop network programming with multiple one-way hash chains. Instead of the expensive asymmetric cryptographic primitives used in much prior work, the scheme employs only symmetric cryptographic primitives, in a circular geographic node deployment model. Possible attacks an adversary could mount on the scheme is discussed and provided simple and effective counter measures against them. Finally, it provides a comprehensive performance evaluation of the scheme in terms of end-to-end latency and power consumption, which is said to be believed, is the first power consumption evaluation of a security scheme for network programming protocols. The application works well for given tasks in network environment. Any node with .Net framework installed can execute the application. The underlying mechanism can be extended to any or all kind of platform like Linux, Solaris and more. The system eliminates the difficulties in the existing system. It is developed in a user-friendly manner. The system is very fast and any transaction can be viewed or retaken at any level.

Error messages are given at each level of input of individual stages. This software is very particular in securing the multi-hop network programming.

REFERENCES

[1] L. Tang and H. Liu,"Toward predicting collective behavior via social dimension extraction," IEEE Intelligent Systems, vol. 25, pp. 19–25, 2010.

[2] P. Singla and M. Richardson, "Yes, there is a correlation: - from social networks to personal behavior on the web," in WWW '08: Proceeding of the 17th

[3] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," Annual Review of Sociology, vol. 27, pp. 415–444, 2001.

[4] H. W. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, "Homophily in the digital world: A LiveJournal case study," IEEE Internet Computing, vol. 14, pp. 15–23, 2010.

[5] M. Granovetter. Threshold models of collective behavior. American journal of sociology , 83(6):1420, 1978.

[6] T. C. Schelling. Dynamic models of segregation. Journal of Mathematical Sociology , 1:143186, 1971.

[7] M. Newman, "Finding community structure in networks using the eigenvectors of matrices," Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), vol.74,no.3,2006.[Online].Available: http://dx.doi.org/10.1103/PhysRevE.74.036104

[8] M. E. J. Newman, The structure and function of complex networks. SIAM

[9] M. Girvan and M. E. J. Newman, Community structure in social and biological networks. Proc. Natl . Acad. Sci USA 99, 7821–7826 (2002).

[10] R. Guimer`a and L. A. N. Amaral, Functional cartography of complex metabolic networks. Nature 433, 895–900 (2005).

[11] G. W. Flake, S. R. Lawrence, C. L. Giles, and F. M. Coetzee, Self-organization and identification of Web communities. IEEE Computer 35, 66–71 (2002).

[12] S. Gupta, R. M. Anderson, and R. M. May, Networks of sexual contacts: Implications for the pattern of spread of HIV. AIDS 3, 807–817 (1989).international conference on World Wide Web. New York, NY, USA: ACM, 2008, pp. 655–664. Review 45, 167–256 (2003).