Map Reduce by K-Nearest Neighbor Joins

Srikanth Bethu*[1], B Sankara Babu[2], S Govinda Rao[3], R Aruna Florence[4]
[1,2,3,4] Department of Computer Science and Engineering
[1,2,3,4] Gokaraju Rangaraju Institute of Engineering & Technology, JNTU Hyderabad, Telangana, India
*[1]srikanthbethu@gmail.com, [2] bsankarababu81@gmail.com, [3] govind.griet@gmail.com, [4] aruna1202@gmail.com

Abstract— **Knowledge discovery and Data mining plays a major role in computational intensive tasks with high range of applications. With the increase of volume and dimension of data, the distributed features perform operations in a reasonable period. MapReduce programming is suitable for distributed large scale data processing that provides different ways of solutions to the same problem, that (one) has particular constraints and properties. In this paper, we give comparative analysis and its approaches for computing KNN on MapReduce[1] theoretically and experimental evaluation. Load balancing, accuracy and complexity are analyzed on each step of data preprocessing, data partitioning and computation. The experiment results in this are produced by using variety of datasets. Time and Space complexity are analyzed periodically on each dataset and gives new advantages and short comings that are discussed for each algorithm. Finally this paper can be used as a reference material to handle KNN [2] based problems in the idea of Mapreducing in Big Data.**

Keywords— Big Data, Data Mining, KNN, Map Reducing.

I. INTRODUCTION

The k-Nearest Neighbor algorithm (k-NN) is a non-parametric method used for classification and regression [1]. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression: In k-NN classification [3][4], the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as the overlap metric (or Hamming distance [5][6]). The concise and abstract distributed parallel computing model MapReduce was first presented by Google in 2004 and then being used to simplify large-scale data processing, and it has been widely used in this field today, for example, the widely used open source implementation Hadoop. MapReduce runs on the basis of Hadoop Distributed File System (HDFS) [7][8][9]. The motive of MapReduce, which has two user defined functions, the Map function and the Reduce function, is to deal with files from HDFS and express them in key/value pairs, then these key/value pairs will be input to Map functions and output key/value pairs after a series of processing as intermediate results. These intermediate results from Map functions will be send to Reduce functions after the shuffle operation executed by MapReduce [10][11][12], final results will be output in key/value pairs after a series of operations executed by user defined Reduce functions. This makes only take Map functions and Reduce functions into consideration by programmers when dealing with big data in parallel into reality. Benefit from the thinking way of "dividing and dealing" that divide one big file into several small files, the complexity of data processing is reduced significantly and the processing efficiency is improved substantially. So it will be a very good substitute for the single process or single machine platform of the KNN algorithm in big data [13][14]time.

A. KNN algorithm in MapReducing

The situation of big data today generates demands of higher efficiency and better performance of the popular KNN algorithm. The irregular partitioning method partitions the whole text area into several small text areas and number them respectively. It fits very well with the operational mechanism that HDFS divide one big area into several small areas first and then pass them to the most popular parallel computing platform MapReduce. So run the irregular partitioning method based KNN algorithm [15] on MapReduce can significantly improve the comprehensive performance of KNN algorithm when dealing with big data. The basic thought of running the irregular partitioning method based KNN algorithm on MapReduce is: users upload files to HDFS [16] and a Map function downloads one text area each time, then the modified Map function divides it into several regions. Assume one small region in a mapper is Dq which number is q, and the input key/value pair can be (q, Dq) which key is q and value is Dq; this region will be transformed into a specific format file and the similarity value sq between this text and the given text t will be calculated out, using the former formula, after traversing the whole text. Set a collection S with t and sq, which form is S(t,sq), then, take q as the key and S(t,sq) as the value, output the intermediate key/value pair (q, S(t,sq)), which is the result of Map function as well as the input of Reduce function. After reading in, the Reduce function sorts sq with the same key in ascending order and select out the first

978-0-7695-6584-2/18/$31.00 ©2018 IEEE
DOI 10.1109/CyberC.2018.00050

A Survey Report on Data Analytics as a Tool in PoliticalCampaign

Srikanth Bethu[1], K.Madhavi[2], B.Rupa[3], A.Sai Hanuman[4], R Soujanya[5], B.Sankara Babu[6]
[1,3,5]Assistant Professor, Department of Computer Science & Engineering
[2,4]Professor, Department of Computer Science & Engineering
[1,2,3,4,5,6]GokarajuRangaraju Institute of Engineering & Technology, JNTU Hyderabad, India
[1]srikanthbethu@gmail.com, [2]bmadhaviranjan@yahoo.com,[3]rupa.bogolu@gmail.com,
[4]a_saihanuman@hotmail.com, [5]soujanya96@gmail.com, [6]bsankarababu81@gmail.com

ABSTRACT
Current Campaigns create databases of point by point data about natives to advise constituent procedure and to direct strategic endeavors. Campaign information investigators create models utilizing this data to deliver singular dimension forecasts about subjects' probabilities of playing out certain political practices, of supporting competitors and issues. As of late as twenty years back, a "numbers driven crusade" suggested that applicants and their guides gave careful consideration to survey numbers and balanced strategies in light of surveys.[1]Presidential Campaigns focused on states dependent on recorded ideas of which states were "swing" (i.e., could go in any case) and spending substances. Interestingly, contemporary political crusades gather colossal databases on individual natives and contract Campaign information examiners to make models anticipating nationals' practices, demeanors, and reactions to crusade contact.. This new type of information driven crusading gives hopefuls and their counselors incredible assets for plotting constituent system. Information examinations have been the piece of a Government Political Campaign since 2012. Present day human advancement's one of the key player is online life. Internet based life is changing existing data conduct by giving clients access to ongoing on the web Information channels without the limitations of existence. This produces immense unstructured information for information mining. This gives researcher a colossal degree for information investigation challenge. Propelled information mining and Machine learning [1] systems are as of late encouraged by our capacity to gather more fine-grained information and have been used for encouraging the tasks of Political parties [2]. Here we are utilizing Machine Learning and Data mining gives expansive comprehension of how the ideological groups are using Big Data has been featured.

In this paper investigates the job of Big Data examination in races over the world. Decisions can change the predeterminations of countries and the precise bearing in which nations are going in. It's imperative that both the applicants and the voters alike comprehend the effect of prescient examination and Big Data can have on decision results.

ICEIT 2019, March 2–4, 2019, Cambridge, United Kingdom

ACM ISBN 978-1-4503-6267-2/19/03…$15.00
http://doi.org/10.1145/3318396.3318452

CCS Concepts
• Computing Methodologies→Machine Learning • Computing methodologies→Big Data and Data Analytics.

Keywords
Big Data, Data and Analytics, Data Mining, Machine Learning.

1. INTRODUCTION
Utilization of investigation isn't new. US Presidential crusades have for since quite a while ago utilized huge information investigation to miniaturized scale focus on certain voter fragments, accumulate data on socioeconomics of states, create constituent maps, break down voter designs in past decisions, stall the board and comprehend issues most applicable to the voters and afterward making a message which will speak to the expansive electorate.

This was seen best amid the US Presidential decisions of 2012. The compelling utilization of enormous information examination by the Barack Obama Campaign is refered to by political experts as one of te significant purposes behind his triumph over Mitt Romney, his Republican challenger. The Obama Campaign had a staff of 100 individuals explicitly to deal with information investigation. The crusade made utilization of the HP Vertica MPP investigative database alongside R and Stata to advance beyond the Romney Campaign. The Campaign additionally utilized procedures like Aiwolf and media streamlining agent. With Air wolf, way to-entryway campaigners could take the reactions of voters and feed it to the HP database investigation motor. Media enhancer took into consideration better focusing of voters through Ad purchases by playing out a full investigation on the voter database accessible with the Democratic Party. Romney too utilized information examination, for example, through the Project Orca application, yet the numerous wasteful aspects in the framework at last prompted poor outcomes. These elements at last prompted Obama's triumph in the decision.

Someone else separated from Obama who left with every one of the honors following the decision was Nate Silver. Silver figured out how to anticipate the consequences all things considered, by and by the utilization of measurements and information investigation. What the Obama Campaign and Nate Silver did in 2012, Modi did in 2014, however on a littler scale. With the assistance of Prashant Kishor, Modi conveyed an expert touch to his Campaign, making substantial utilization of information

Feature Selection Based Supervised Learning Method for Network Intrusion Detection

Ch. Mallikarjuna Rao, G. Ramesh, D. V. Lalitha Parameswari
Karanam Madhavi, K. Sudheer Babu

Abstract*: Supervised learning is one of the data mining phenomena where a knowledge model is built for artificial intelligence. Learning from training samples has its advantages in predictive solutions. Such solution is essential for network intrusion detection problems. Networks of all kinds do have problem of intrusions as they are exposed to public communications in one way or other. Intrusions over a network are in the form of network flows that need to be analyzed. Manual observation of the flows and detecting intrusions is very time taking. Therefore it is essential to have an automated system for quickly detection of intrusions to safeguard network systems. There are many intrusion detection systems found in the literature. However, there is need for faster algorithm that makes sense in helping network administrators with accurate knowledge presented. Towards this end we proposed a framework with a feature subset selection mechanism to speed up detection process and improve accuracy of the same. The feature subset selection algorithm and Support Vector Machine (SVM) work together in order to have a faster detection system. Benchmark datasets like KDD and NSL-KDD are used for experiments. The empirical results showed that the proposed SVM-FSS framework shows better performance over the state of the art framework.

***Index Terms*: Data mining, feature selection, intrusion detection, Support Vector Machine, machine learning**

I. INTRODUCTION

To Data mining is widely used in real world applications. It is the discipline where historical data is analyzed to obtain hidden information. In other words, it is the process of extracting or discovering latent trends or patterns that are not known earlier. These trends or patterns uncovered from the databases are used to take expert decisions. The process of mining is essential for any enterprise in different domains. Knowledge discovery helps domain experts to have interpretation of knowledge and take decisions. Models are built in order to have solutions to different problems. The general steps involved in knowledge discovery from databases (KDD) are visualized in Figure 1.

There are many steps in KDD. First of all a problem is defined. Then data is gathered in order to solve the problem. Then data mining algorithms are used to built a model and evaluate it. This gives rise to knowledge needed. This knowledge is used to make expert decisions that result in business growth and profits. There are many algorithms related to data mining. They include association rule mining, decision trees, clustering and classification. These algorithms take time and resources to complete mining process. When high dimensional data is taken, these algorithms take long time to execute and consume more resources. To overcome this problem, it is important to reduce dimensions.

Many existing data mining based intrusion methods do not use feature selection method. For instance neural networks and SVM based approach [7], ANN and fuzzy clustering [10], SVM based approach [12], and fuzzy logic based approach [17] and Hidden Naive Bayes method [18]. There are some methods found with feature selection. They include Naive Bayes based method [15], Mutual information based intrusion detection [22] and [24] where many feature selection algorithms are reviewed. However, it is understood that feature selection is still an optimization problem which leads to further enhancement in accuracy and performance of data mining techniques for intrusion detection. Our contributions are as follows.

1. We proposed a framework named SVM-FSS for feature selection based intrusion detection that enhances the capability of SVM.
2. We proposed an algorithm named FSS for effective feature selection prior to employing classification technique on intrusion datasets like KDD and NSL-KDD.
3. We built an application to show the effectiveness of the framework and evaluated with the two datasets.

The remainder of the paper is structured as follows. Section 2 provides literature review on data mining techniques that are used for detecting network intrusions. Section 3 covers the proposed methodology for intrusion detection. Section 4 presents experimental results and evaluation while Section 5 provides conclusions and gives possible scope for future work.

Revised Manuscript Received on May 31, 2019.

Dr. Ch. Mallikarjuna Rao, Professor, Department of CSE, GRIET, Hyderabad, Telangana, India.
Dr. G. Ramesh, Associate Professor, Department of CSE, GRIET, Hyderabad, Telangana, India.
Dr. D. V. Lalitha Parameswari, Sr.Asst. Professor, Department of CSE, GNITS, Hyderabad, Telangana, India.
Dr. Karanam Madhavi, Professor, Department of CSE, GRIET, Hyderabad, Telangana, India.
Mr. K. Sudheer Babu, Assistant Professor, Department of CSE, GRIET, Hyderabad, Telangana, India.

FLAT Vs Hierarchical Routing Protocols in Wireless Sensor Networks: An In-depth Analysis

Shanthi.S, Padmalaya Nayak, A. Sai Hanuman

Abstract*: *The emerging applications of IoT require that Wireless Sensor Network should be energy proficient. To build the Wireless Sensor Network more energy proficient, many challenging issues like routing, localization and sensor fusion must be properly addressed. Although many routing protocols are in existence, there is a lack of research papers that contain an in depth analysis which can give an overview to the current researchers. In order to provide a big picture outlook, we have put an effort to analyze the comparative performance of various leading routing protocols available in WSN. Although many routing protocols are available in the literature under flat routing, SPIN is selected under flat routing protocol as it's a leading protocol. Similarly, LEACH, LEACH-C and PEGASIS is considered under hierarchical routing protocol. Simulations have been carried out by using the NS-2 simulator. The Performance metrics like energy utilization, delay, throughput and network lifetime are some of them which has been explored

Index Terms*: *LEACH, LEACH-C, NS-2, PEGASIS, SPIN

I. INTRODUCTION

A wireless sensor network is composed of a compilation of sensor nodes and they have been deployed in a field in order to examine the specific environment and to gather the data about the environment. Sensor nodes are usually small in size, resource constrained, less memory, limited battery power etc [1]. In spite of the above-mentioned drawbacks, sensors are capable of providing a real picture of the environment which is being sensed. Due to various resource constraints, WSN need to face many challenges in routing, communication, topology, efficient hardware components and algorithms etc [2]. Routing protocol takes part in packet delivery which includes routing of packets between various networks. The major goal is to deliver the data efficiently to the destination. Routing is a big difficult task in wireless sensor networks and has to to be focused more because of the densely populated sensor nodes and they have very minimal energy resource and a small memory. Generally, the routing protocols are classified into two major groups namely based on network architecture and application. On the basis of
network architecture, it is further classified into three types namely location ,flat and hierarchical based routing. The routing protocols can also be further divided on the basis of establishment of path, operations of the protocols and initiator of operation.

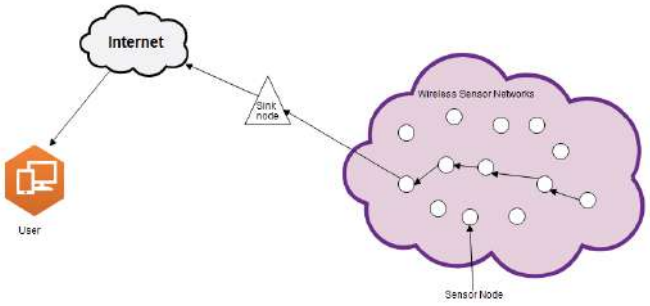

Figure 1: General Architecture of WSN

We are considering various routing protocol in this paper. Even though the concept is not new, but the performance parameters are presented in the literature is isolation to each other. The intention of this research paper is to make available the performance parameters in a single domain and give a sharp vision of most important protocols under flat and hierarchical routing protocol. Out of that, we have considered SPIN under flat routing protocol and LEACH, LEACH-C, PEGASIS under hierarchical routing protocol to verify the simulation results. The paper is structured as follows. Section 2 describes the existing works available in the literature and in Section 3, we have discussed the simulation results obtained from NS-2 Simulator. Section 4 provides the conclusion of the paper.

II. RELATED WORK

In this section, some of the famous protocols available in the current literature are discussed in detail. These protocols are designed with the intention of improving some factors like utilization of energy, network lifetime. WSN is basically classified into three types hierarchical, location and flat based routing [3, 14].Various techniques have been put-forward for the improvement of routing protocols. Categorization of routing protocols is depicted in Figure 2.

A. Flat Based Routing

In this type of routing, all the nodes are having the identical

Revised Manuscript Received on May 28, 2019.

First Author name, His Department Name, University/ College/ Organization Name, City Name, Country Name.
Second Author name, His Department Name, University/ College/ Organization Name, City Name, Country Name.
Third Author name, His Department Name, University/ College/ Organization Name, City Name, Country Name.

MapReduce Accelerated Signature-Based Intrusion Detection Mechanism (IDM) with Pattern Matching Mechanism

Chinta Someswara Rao and K. Butchi Raju

Abstract Network protection was rendered an ultimatum due to expeditious growth in the pace of wired Internet and also due to the constantly increasing number of attacks. Discovery of dubious activities and prevention of their malicious impact can be effectively done by intrusion detection systems (IDS). Due to the pattern matching operation, existing signature-based IDS will bear more searching time and memory. A diligent system to truncate the overhead is the need of the hour. Matching the pattern through parallelizing a matching algorithm on a GPU using MapReduce is the objective of this research. This paper attempts to parallelize a pattern matching technique under the MapReduce framework. A speedup of four times is achieved using the GPU implementations in juxtaposing to the CPU under MapReduce framework.

Keywords Information security · Intrusion detection systems
Pattern matching · MapReduce · GPU · CPU

1 Introduction

Human life has become much affable due to the advent of digital devices and services. A series of obvious hindrances and skeptical behaviors enhances as the connectivity between various services increases. For different services, various strong attack detection processes are required [1].

IDS is the most prominent technique that will be used to inspect network payloads of packets for malicious threats. Intrusion signatures are required for IDS. Usage of software or hardware is the most customary proposal of several

C. S. Rao (✉)
Department of CSE, SRKR Engineering College, Bhimavaram, AP, India
e-mail: chinta.someswararao@gmail.com

K. Butchi Raju
Department of CSE, GRIET, Hyderabad, TS, India
e-mail: butchiraju.katari@gmail.com

J. Nayak et al. (eds.), *Soft Computing in Data Analytics*,
Advances in Intelligent Systems and Computing 758,
https://doi.org/10.1007/978-981-13-0514-6_16

A Review on Different Defect Detection Models in Software Systems

B. Dhanalaxmi, Associate Professor, Department of Information Technology, Institute of Aeronautical Engineering, Hyderabad, Telangana, India, dinnu18@gmail.com

Dr.G.Appa Rao Naidu, Department of CSE, JBIET, Moinabad, Hyderabad, Telangana, India, apparaonaidug@gmail.com

Dr.K. Anuradha, Department of CSE, GRIET, Hyderabad, Telangana, India, kodali.anuradha@yahoo.com

Abstract:

Detecting defects in software product development requires serious effort. It's important to use the most efficient and effective methods. Many empirical studies have investigated defect detection techniques, inspections, and testing in isolation. In an effort to improve software quality, various project management systems have been developed. Although these project management systems improve the chances that projects will be completed in a timely manner, managers continue to find it difficult to predict the number of software defects for upcoming software releases. If the number of software defects could be reliably predicted, then managers would be able to commit the necessary resources to more accurately deal with problems that arise.

In software engineering, this area of software defect prediction has been the subject of considerable research. There are complex, quantitative methods that focus on the relationship between the number of defects and software complexity. Typically, these models make numerous, unrealistic assumptions. Still other models focus on the quality of the development process as the best predictor of a product's quality. Unfortunately, none of these approaches have yielded accurate results. Accordingly, it would be desirable and highly advantageous to provide improved and simplified techniques for predicting software defects.

Key words: Defect detection, Defect Prediction, In-appendage,Reliability.

I. Introduction:

Software failures are costly. Reports regarding software problems are published regularly, ranging from minor issues to the year 2000 problem[1]. Thus, a better understanding of software defects, their causes and possible improvements in the area are essential. Software failures are related to the reliability of the software. Reliability is an external attribute [3]. A program has certain reliability from the perspective of the user. The internal attributes related to reliability are defects or faults [2].

Many different techniques, tools and automation strategies have been developed to make testing more efficient[5]. Despite the wide variety of proposed solutions, the fundamental challenge of software testing revealing new defects in freshly developed software or after major modifications—is in practice still largely dependent on the performance of human testers doing manual testing.While test automation is becoming increasingly popular due to, e.g., approaches like Test-Driven Development and eXtreme Programming [1, 2, 3], empirical research shows that companies typically perform very little automated testing [4] and most new defects are found by manual testing. The role of automation is emphasized in regression testing and it is best viewed as a way of removing the enactment of simple and repetitive tasks from human testers in order to free up time for creative manual testing [4, 5, 6]. Interestingly, manual testing and especially test execution practices have been fairly little studied in the software engineering community. Testing research has focused on techniques for test case design, selection and prioritization, as well as on optimizing automated testing. However, we do not know, i.e., what factors affect the efficiency of manual testing, and how, or what practices industrial testers find useful. Previous research shows that aspects such as testers' skills and the type of the software have as strong an effect on test execution results as the test case design techniques.

II. Background Study:

Aybüke Aurum, Håkan Petersson, and Claes Wohlin summarize 25 years of empirical research on software inspections, including more than 30 studies that investigated different reading techniques, team sizes, meeting gains, and so on [7]. Similarly, Natalia Juristo, Ana Moreno, and Sira Vegas summarize 25 years of empirical research on software testing based on more than 20 studies [8]. They compare testing within and across so-called "families" of techniques. Despite the large number

ISSN 1943-023X
Received: 22 Apr 2018/Accepted: 12 Jun 2018

Optimal Feature Selection for Multivalued Attributes Using Transaction Weights as Utility Scale

K. Lnc Prakash and K. Anuradha

Abstract Attribute selection procedure is a key step in the process of Knowledge Discovery in Database (KDD). Majority of the earlier contributions of selection methods can handle easier attribute types. Such methods are not for multivalued attributes that comprise multiple values in simultaneously. Majority of the existing attribute selection methods can manage simple attribute types like the numerical and categorical. The methods cannot fit multivalued attributes, which are attributes that constitute multiple values simultaneously in the dataset for same instance. In this manuscript, a contemporary approach for selecting optimal values for features of multivalued attributes is proposed. In the proposed solution, the method is about adaptation of utility mining based pattern discovery approach. For evaluating the proposed approach, experiments are carried out with multivalued and multiclass datasets that are submitted to k-means clustering technique. The experiments show that the proposed method is optimal to assess the relevance of multivalued attributes toward mining models such as clustering.

Keywords Multivalued attributes · Utility scale · Transaction weight
Data mining · Feature selection by frequency

1 Introduction

Clustering, which is an unsupervised learning method, is a predominant area in which many researches are taking place. Based on the chosen optimal feature sets, the process of clustering occurs. Feature selection is one of the popular models that is adapted for improving the clustering process, as clustering plays a vital role in the

K. Lnc Prakash (✉)
Department of Computer Science and Engineering, AITS, Rajampet, Andhra Pradesh, India
e-mail: klnc.prakash@gmail.com

K. Anuradha
Department of Computer Science and Engineering, GRIET, Hyderabad, Telangana, India
e-mail: kodali.anuradha@yahoo.com

V. Bhateja et al. (eds.), *Proceedings of the Second International Conference on Computational Intelligence and Informatics*, Advances in Intelligent Systems and Computing 712, https://doi.org/10.1007/978-981-10-8228-3_49

Analysis of Early Detection of Emerging Patterns from Social Media Networks: A Data Mining Techniques Perspective

Yadala Sucharitha, Y. Vijayalata and V. Kamakshi Prasad

Contents

Abstract At present, social media networking sites like Twitter, Flickr, Facebook, YouTube, Instagram are offering a rich assistance for disparate information. Many people are used to extracting and penetrating information in Social Media Networks (SMNs). Detecting emerging patterns from the huge number of messages and tweets around the social networking blogs is crucial for information breeding and marking trends, especially early identification of the emerging patterns can intensively promote real-time intelligent systems. However, at present, we have many methods for discovering emerging patterns which are proposed by various researchers on long range, but they are not producing effective results. In this article, we provide a wide review of different approaches for discovering emerging trends (textual, audio, and video) in SMNs proposed by various researchers in data mining techniques perspective. In this paper, we also discuss the challenges and issues involved in discovering emerging patterns in social media blogs.

Y. Sucharitha (✉)
CMR Institute of Technology, Hyderabad, TS, India
e-mail: suchi.yadala@gmail.com

Y. Sucharitha
JNTUH, Hyderabad, TS, India

Y. Vijayalata
Gokaraju Rangaraju Institute of Engineering & Technology, Hyderabad, TS, India
e-mail: vijaya@griet.ac.in

V. Kamakshi Prasad
JNTUH College of Engineering, Hyderabad, TS, India
e-mail: kamakshiprasad@jntuh.ac.in

J. Wang et al. (eds.), *Soft Computing and Signal Processing*, Advances in Intelligent Systems and Computing 900, https://doi.org/10.1007/978-981-13-3600-3_2

A Secure Architecture of Design for Testability Structures

K. Swaraja, K. Meenakshi, Padmavathi Kora, Mamatha Samson, G. Karuna, A. Ushasree

Abstract*: *The structures of Scan-based Design for Testability are extremely susceptible towards unapproved access of the signals present inside the chip. This paper suggests a protected output based plan which averts the unapproved access without any compromise in the testability. A unique key for each test vector is provided in the proposed secure architecture. These inimitable keys are produced by a multi-polynomial linear feedback shift register (LFSR) in addition they are utilized as test vectors. The dimensions of the multi polynomial LFSR bit is saved bigger than the dimension of key so as to augment the level of security to the key. As the keys are concealed within the test vectors, there is reduction in area overhead. The amount of security is improved predominantly by changing the key for all test vectors, along with the location of the bit in the test vector by choosing a valid combination out of available test vector generated by multi polynomial LFSR.

Index Terms*: *Design for Test (DFT), Scan Chain, Multi polynomial LFSR, Testability, Security.

I. INTRODUCTION

The rising intricacy of the Integrated Circuits (IC) has proved testing very ambiguous. Thus, there is a need for minimal effort and high proficiency testing techniques [1] as there is a substantial rise in the proportion of testing results to the wide-ranging expense of an IC. This problem can be solved if the structures of test arrangement are concealed into the chips in the design cycle. The extensive and broadly explored area of present-day IC configuration is the emerging techniques that permit choosing and setting up the best test situation and device in the design of IC is Design for Testability (DFT) [2], [3]. It refers to those design techniques with the purpose of making test generation and test application cost-effective. DFT plays a vital role in chip manufacturing. Full scan design turns out to be utmost prevailing structured DFT approaches, because of its treatment to high faults besides decrease in overhead of hardware. In the testing of circuits a scan chain is utilized broadly which are sequential, as it resolves the challenges in controlling and perception of inner nodes of a circuit by providing entry to all components of storage in the design, so as to accomplish test spur and detect the responses to expand the coverage of faults. But security usually requires the opposite. The observability furnished by test structures can be utilized by a mugger to inspect the information being handled within the chip. Likewise, the test structures can uncover the confidential data concerning the design of chip. Scan chains open side channels for interlopers to look at the hidden information stored in cryptographic systems [1] and it has been verified as a safety hazard to the Intellectual Property (IP) of the chip. Thus concerning protection with testability, it is known that they were contradicts to each other. Though, testability itself is a crucial obligation of safety, as the scheme is not protected completely except it is fully testable. Enduring a counterbalance among the two is essential. In any case, DFT can't be kept away from insecure systems, in light of the fact that the IC requires elevated quality of testability necessities and the security might be undermined by some faults [4]. So as to fulfil both security and testability, additional equipment is combined into the ordinary scan chains so as to provide them safety with no arrangement in the testability of the target design. The remaining paper is structured as follows. Summary of the past methodologies is reported under Section II. Section III proposes a secure architecture for DFT structures, further the simulation outcomes are evaluated in section IV. At last, section V concludes the proposed work.

Revised Manuscript Received on July 9, 2019

K. Swaraja, ECE, GRIET, Hyderabad, India.
K. Meenakshi, ECE, GRIET, Hyderabad, India.
Padmavathi Kora, ECE, GRIET, Hyderabad, India.
Mamatha Samson, ECE, GRIET, Hyderabad, India.
G. Karuna, CSE, GRIET, Hyderabad, India.
A.Ushasree, ECE, GRIET, Hyderabad, India.

II. LITERATURE SURVEY

Improving both testability and security is a problematic task and usually a trade-off is upheld among the necessities for testability and security. By monitoring the circuit behavior in the scan mode the attacker can obtain the secret key. Yang et al. in [7] have determined that the conventional scan chains are prone to disclose significant details of advanced encryption standard (AES). In [8], a mirror key register method (MKR) has been presented in which a counterfeit secret key is stacked into the scan chain to shield MKR from unapproved access. To ensure crypto centers, counter to the scan based attacks can be characterized into two procedures of restricting the admission to avert muggers after perceiving the scan data and it is prone to a more overhead in timing and hiding the secret data while giving access to all clients. A power- reset is essential to shift the mode as of secure towards the test [9]. The stream of scan output is amended by [10] through accumulating gates of inverter haphazardly to the scan cells to control the yield data. Yet, the area of inverters can be resolved if appropriate data sources are connected to the scan chain. A minimal cost solution is proposed by adding sham flip-flops in the design of scan chain. If the correct key related to the area of these flip-flops is not entered, haphazard data will be displaced [11]. Configuration for Scan chain using scrambling is presented in [12], which separates the scan chain into littler sub-chains. At whatever point the test mode is secure it works in a predetermined order; else it reorders the sub-chains.

Measurement 178 (2021) 108974

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Routing in wireless sensor networks using machine learning techniques: Challenges and opportunities

Padmalaya Nayak *, G.K. Swetha, Surbhi Gupta, K. Madhavi

Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

ARTICLE INFO

Keywords:
WSNs
Artificial Intelligence
Machine Learning Techniques
Routing

ABSTRACT

Energy conservation is the primary task in Wireless Sensor Networks (WSNs) as these tiny sensor nodes are the backbone of today's Internet of Things (IoT) applications. These nodes rely exclusively on battery power to maneuver in hazardous environments. So, there is a requirement to study and design efficient, robust communication protocols to handle the challenges of the WSNs to make the network operational for a long time. Although traditional technologies solve many issues in WSNs, it may not derive an accurate mathematical model for predicting system behavior. So, some challenging tasks like routing, data fusion, localization, and object tracking are handled by low complexity mathematical models to define system behavior. In this paper, an effort has been made to provide a big outlook to the current "researchers" on machine learning techniques that have been employed to handle various issues in WSNs, and special attention has been given to routing problems.

1. Introduction

A WSN is a collection of a large number of sensor nodes, usually deployed in remote areas to monitor environmental parameters like temperature, humidity, moisture, etc. The sensor nodes are equipped with various types of sensors like acoustic, pressure, motion, image, chemical, weather, pressure, temperature, optical sensors, etc. Due to this diversity of sensor nodes, the applications of WSNs are huge in a range that starts with healthcare, military, defense, agriculture to our day to day life. Despite huge applications, WSN faces many typical challenges like limited energy sources, computational speed, memory, and limited communication bandwidth, making the sensor network degrade in performance and decreasing the network lifetime [1]. Developing different algorithms for different applications is quite a challenging task. In particular, the designer of WSNs must emphasize on various issues like data aggregation, clustering, routing, localization, fault detection, task scheduling, event tracking, etc. The various challenges and issues in WSNs are illustrated in Fig. 1. The complete description is given in section III. Among all the tasks, routing is one of the important tasks as major percentage of the energy consumption takes place while routing the data packet from the source node to the destination either through a single hop or multi-hop fashion. While routing the data, the sensor network designer must focus on all the sensor node's energy consumption issues to keep the network operating for a long time. Every routing protocol has its own characteristics and specifications based on network applications and structure.

Machine Learning (ML) is a part of Artificial Intelligence introduced in the late 1950s. Over the period, it evolved and moved towards algorithms that could computationally feasible and robust enough to handle different problems like classifications, clustering, regression, and optimization in the field of medical, engineering, and computing. ML is one of the most exciting and influential technologies in today's world. ML provides computer systems with the ability to learn automatically without human involvement and take action accordingly. It creates a model by analyzing complex data automatically, quickly, and accurately. ML has the ability to learn from the generalized structure to provide a general solution to improve system performance. Because of the diversified applications, it is applied in various scientific fields of medical, engineering, and computing like manual data entry, automatic detection of spam, medical diagnosis, image recognition, data cleansing, noise reduction [144,145], etc. Recent studies prove that ML has been applied to solve many issues in WSNs. Applying ML in WSNs not only improves the system performance but also reduces the complex tasks like reprogramming, accessing the large amount of data manually, and extracting useful information from the data. So, ML techniques are extremely helpful for fetching large amounts of data and extract useful information [2–4]. For more clarity, the requirements of Machine Learning Techniques in WSNs are briefly explained in the below

* Corresponding author.
E-mail addresses: padmalaya@griet.ac.in (P. Nayak), royal_surbhi@yahoo.com (S. Gupta).

https://doi.org/10.1016/j.measurement.2021.108974
Received 31 August 2020; Received in revised form 24 December 2020; Accepted 29 December 2020
Available online 31 January 2021